# CRITERION VALIDITY OF THE MMPI-2 IN A STATE HOSPITAL SETTING

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Richard Connell, B.A.

Denton, Texas

August, 1996

# CRITERION VALIDITY OF THE MMPI-2 IN A STATE HOSPITAL SETTING

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Richard Connell, B.A.

Denton, Texas

August, 1996

Connell, Richard, Criterion validity of the MMPI-2 in a state hospital setting. Doctor of Philosophy (Clinical Psychology), August, 1996, 85 pp., 10 tables, references, 68 titles.

The current study investigated the criterion validity of the Minnesota Multiphasic Personality Inventory - 2 (MMPI-2) by comparing participants' profiles with other variables, including diagnosis, length of hospitalization, and chronicity. The specific diagnostic groups investigated were depressed (major depressive disorder; dysthymic disorder; and bipolar disorder, depressed), schizophrenic (schizophrenia, schizophreniform disorder, and schizoaffective disorder), and borderline personality disorder (BPD). Statistical analyses included use of univariate analyses of variance (ANOVAs), multivariate analyses of variance (MANOVAs), regression analyses, and measures of sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). MANOVA results indicated significant differences between diagnostic groups on Scales F, 2, 3, 4, 7, ANX, FRS, DEP, BIZ, ANG, LSE, and FAM. There were considerable differences between males and females when separate MANOVAs were performed for gender groups. Cutoff scores for classification by diagnosis resulted in significant specificity rates and negative predictive power, but sensitivity rates and positive predictive power were not significant.

TABLE OF CONTENTS

LIST OF TABLES

# CHAPTER I

# INTRODUCTION

The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1942) was the most widely used objective test of psychopathology (Helmes & Reddon, 1993). The 1989 revision of the MMPI, resulting in the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) sparked considerable interest and debate regarding the comparability of MMPI and MMPI-2 profiles (Adler, 1990; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1990; Dahlstrom, 1992; Edwards, Morrison, & Weissman, 1993; Edwards & Weissman, 1990; Elwood, 1993; Greene, 1991; Morrison, Edwards, & Weissman, 1994). The present study will provide an overview of the development of the MMPI, followed by rationale and procedures used in the revision of this test, and concerns regarding interpretations of the MMPI-2 based on validity studies of the MMPI. Next, the need for validity studies of the MMPI-2 will be discussed, as well as types of statistical analyses that would be most appropriate. Finally, a study will be presented that addresses some of this need for appropriate validity investigation of the MMPI-2.

Cohen, Montague, Nathanson, and Swerdlik (1988) defined criterion-related validity as: "... a judgement regarding how adequately a test score can be used to infer an individual's most probable standing on some measure of interest--

the measure of interest being the criterion (p. 128). They also described the two types of validity evidence that are subsumed under the broader heading of criterion-related validity. These are predictive and concurrent validity. Predictive validity refers to a measure of the degree to which a test score predicts some criterion measure to be collected at a later time. Concurrent validity refers to a measure of the degree to which a test score relates to a criterion measure obtained at the same time, which is the case in the present investigation (Cohen et al., 1988). As detailed in later sections, the present study investigated the criterion validity of the MMPI-2 by comparing individual's test scores with concurrent diagnostic information.

Development of the MMPI

The original MMPI was published in 1942 (Hathaway & McKinley, 1942) as a paper-and-pencil personality test for the assessment of psychopathology. It was derived empirically by administering 504 true-false statements to a group of "normal" participants and to several clinical groups comprised of discreet diagnoses. These groups of clinical participants were labeled hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, psychasthenia, schizophrenia, and hypomania. These groups represented the major psychiatric categories being utilized clinically at the time of construction of the test (Graham, 1990).

An item analysis was then performed for each of the clinical groups to identify items from the pool of 504 that significantly differentiated the clinical groups from one another and from the control group. Thus, the clinical scales

were formed by grouping together items which appeared to differentiate one clinical group from the other groups (Graham. 1990). After about a decade of use and continued research, it became evident that the clinical scales were not pure and accurate measures of the syndromes suggested by the scale names. This led not to dismissal of the MMPI as useless, but to a new approach in the interpretation of MMPI profiles. Because reliable differences in scales were found among individuals known to differ in other important ways, it was concluded the scales must be measuring something other than error variance. In order to attempt to understand what these scales might be measuring, the construct underlying each MMPI scale was treated by researchers as an "unknown" until empirical research and clinical experience more clearly identified the correlates of each scale. In this vein, more than 10,000 studies have been published on the MMPI (Graham, 1990). See Appendices A and B for names and descriptions of the various scales of the MMPI-2.

Interpretation of MMPI and MMPI-2 profiles usually involves examination of high-point, two-point, and/or three-point codes. A high-point code is generally considered to be the highest T score elevation among the clinical scales, excluding Scales 5 and 0, because these scales do not reflect clinical constructs. Similarly, a two-point code is the highest two T score elevations among the clinical scales, excluding Scales 5 and 0; and a three-point code is the highest three clinical scales, excluding Scales 5 and 0 (Butcher & Williams, 1992; Graham, 1987; Graham, 1990; Greene, 1980; Greene, 1991).

The cumulative information from many of the MMPI studies has been examined in several meta-analytic studies (Gartner, Hurt, & Gartner, 1989; Zalewski & Archer, 1991; Zalewski & Gottesman, 1991). The reviews have allowed some generalities to be inferred regarding many diagnostic constructs currently in use. The present paper will focus on three of the most common of such constructs in hospital settings: depression, schizophrenia, and borderline personality disorder.

MMPI scales which have been found to correspond with a diagnosis of depression include clinical Scales 2 and 8. Vincent et al. (1983) found that 66% of participants who had a high-point on Scale 2 were diagnosed with an Affective Disorder (Bipolar Disorder: Depressed; Major Depression; Dysthymic Disorder) or Adjustment Disorder (with depressed mood; with mixed emotional features). In the same investigation, 80% of participants who had a two-point code of 2/8 or 8/2 were diagnosed with an Affective Disorder (Major Depression; Dysthymic Disorder) or Adjustment Disorder (with depressed mood).

MMPI scales found to correspond with a diagnosis of schizophrenia include clinical Scales 2, 4, 6, and 8. Because Scales 2 and 8 also associate strongly with a diagnosis of depression, differential diagnosis of these groups has proved difficult based on MMPI scores. In fact, Schizophrenic participants have been shown to score higher on Scale 2 (originally labeled the Depression Scale) than depressed participants (Vincent et al., 1983; Zalewski & Gottesman, 1991).

MMPI scales found to correspond with Borderline Personality Disorder (BPD) are more varied. Participants with this diagnosis tend to score higher across all clinical scales than do other groups. However, their scores on Scales 2, 3, 4, 7, and 8 have been shown to be more highly elevated than other scales (Evans, Ruff, Braff, & Ainsworth, 1984; Gartner, Hurt, & Gartner, 1989; Vincent et al., 1983; Zalewski & Archer, 1991).

Development of the MMPI-2

The MMPI underwent significant revision which yielded the current form of the test, the MMPI-2 (Butcher et al., 1989). Reasons for this revision included concerns that the original standardization sample was not adequate and was not representative of the general population. Also, there were concerns that the average United States citizen had changed since the normative data were collected in the late 1930s. Other concerns included the use of language and references in some of the items which had become archaic or obsolete. Some items also included sexist language or statements concerning sexual behavior and bowel and bladder functions. These items were judged to be inappropriate or irrelevant for use in psychological tests. For these reasons, the task of revising and renorming the MMPI was undertaken in 1982 and completed in 1989 (Graham, 1990).

Seventy-one items in the MMPI-2 are at least slightly revised from the original form (Ben-Porath & Butcher, 1989; Levitt, 1990). All of the three validity and ten clinical scales contain some revised items, ranging from a low of one new

item on the K Scale to a high of 14 on F and Scale 8 (Butcher et al., 1989). Initially it was suggested that the profiles obtained via administration of the MMPI-2 were comparable to those of the original MMPI (Butcher, 1990; Graham, Timbrook, Ben-Porath, & Butcher, 1991). This assumption, however, has been called into question by more current investigations (Dahlstrom, 1992; Edwards et al., 1993; Elwood, 1993; Greene, 1991; Morrison, Edwards, & Weissman, 1994).

Questions of Comparability

One basis for caution in extrapolating MMPI diagnostic indicators to the MMPI-2 is that current diagnostic criteria are quite different from those used in early MMPI studies (Marks, Seeman, & Haller, 1974; Loranger, 1990). Also, the MMPI-2 utilizes uniform T scores as well as linear T scores, whereas the MMPI used only linear T scores. This change was made in an attempt to improve percentile comparability of T scores across the various validity and clinical scales (Tellegen & Ben-Porath, 1992).

The MMPI used linear T scores which were derived from raw scores using the following formula:

$$T\ score = 50 + [10\ (X - M)] / SD,$$

where X is the raw score for a particular scale, and M and SD are the mean and standard deviation of the raw scores for that scale in the normative sample. This formula yields a T score with a mean of 50 and a standard deviation of 10. Once the formula for transforming raw scores to T scores is established for each scale

using normative data, that formula can then be applied to raw scores obtained in other populations, such as in a clinical or research setting. An important disadvantage of using linear T scores is that when these scores are assembled into a profile their relative height or strength can be misleading. This is because the same T score can signify different percentile standings for different scales due to varying degrees of skewness of scores in the normative distributions. The MMPI-2 uniform T scores are essentially linear T scores which have been adjusted so that there is percentile comparability between T scores. It should be noted that only 8 of the 10 main scales of the MMPI-2 use uniform T scores. Scales 5 and 0 do not represent clinical constructs and, therefore, were not subjected to uniform T score transformations. Scales 5 and 0, then, continue to use linear T scores (Tellegen & Ben-Porath, 1992).

This change to linear T scores may not address the most significant potential problem in comparing scales within a profile. Helmes and Reddon (1993) pointed out that for these comparisons to be meaningful, it is important to consider the standard errors of measurement of the scales being compared. This procedure has long been appreciated in the fields of intellectual and achievement testing, but has not become common practice in the literature on MMPI interpretation.

The impact of changing to the use of uniform and linear T scores was investigated by Greene (1991) by simulating MMPI-2 profiles from a database of over 7,700 MMPI protocols. This approach did not take into account item

revisions, but did estimate the revised scale scores using the new uniform and linear T scores. He found only a 62.7% concordance rate of high-point and two-point codes with the revised T scores (Greene, 1991).

As stated earlier, norms for the MMPI-2 were derived from a new standardization sample. A normative group of 2,600 participants from seven U.S. states was selected, approximating major aspects of the 1980 census (Helmes & Reddon, 1993). This was a vast improvement over the original normative group of 724 visitors to the University of Minnesota Hospitals. However, the new normative group was administered the MMPI-AX, which was a 704 item form developed for the purpose of the MMPI restandardization project (Butcher et al., 1989) and not the actual 567 item MMPI-2. This is because the same group was used for development of the MMPI-2, as well as for obtaining norm values. Helmes and Reddon (1993) pointed out that because of this, the new norms are somewhat inappropriate for the MMPI-2, because item order and the length of form MMPI-AX may have affected participants' responses. Although norms based on the administration of the actual MMPI-2 may have been very similar to the current norms, no studies have been published at this time to allow such a comparison.

Munley (1991) investigated changes in MMPI-2 T scores as compared to MMPI T scores which would result from use of the new normative population for the MMPI-2. He did this by utilizing Table K-1 of the MMPI-2 manual (Butcher et al., 1989) which provides estimates of MMPI T scores based on MMPI-2 T

scores. Munley (1991) found that, "for both men and women, Scales 1, 2, 3, 6, and 0 appear to retain more relative elevation in comparison to original MMPI norms" (p. 89). He then described varying degrees of attenuation of the remaining scales as compared to those derived from MMPI norms. This study had limited clinical relevance because it provided no information regarding the possible impact of the new norms on profile configuration, high-point codes, and two- or three-point code patterns. Furthermore, this study did not attempt to investigate the impact of item revisions and uniform T scores on MMPI-2 profiles.

Ben-Porath and Butcher (1989) investigated the comparability of the MMPI and MMPI-2 among 403 college students' profiles. Of the 403 participants, 377 provided valid profiles. Half of these participants completed both the MMPI and the MMPI-2 in the form of the MMPI-AX. The other half of the participants were administered the original MMPI twice. The differences between the test-retest scores in this group were then compared to the differences between the scores obtained on the two different versions of the test by the experimental group. The original to original comparison (O-O) yielded a 53.5% agreement of high-point for males, and the original to revised comparison (O-R) yielded a 58.7% agreement for males. Results were comparable for female high-point codes, for male and female two-point codes, and for within-normal-limits profiles. Unfortunately, Ben-Porath and Butcher (1989) did not report agreement rates for any specific scales. These results appeared to provide evidence in favor of

generalizing data on MMPI interpretation to the interpretation of MMPI-2 scores. This study did, however, have a number of significant limitations.

Because college students were used as participants in the Ben-Porath and Butcher (1989) study, the results lack obvious generalizability to clinical populations. The authors did not provide information regarding the number of within-normal-limits profiles or the relative elevation of scores across profiles; however, it would be assumed that these participants produced considerably less elevated profiles than would be obtained in a clinical population. The use of the experimental form MMPI-AX also presented a potentially confounding variable. Because the items from the MMPI-2 did not appear in the same order on the MMPI-AX as they did in the final version, participants may have responded somewhat differently to the items on form MMPI-AX as compared to how they would have responded to the MMPI-2.

The results of the O-O comparison in the Ben-Porath and Butcher (1989) study were consistent with other investigations of temporal stability of MMPI high-point and two-point codes. Chojnacki and Walsh (1992) obtained 62.8% agreement for high-point codes among males, 44.2% agreement for two-point codes among males, and 37.2% agreement for three-point codes among males when all 10 clinical scales were compared. Agreement rates were consistently lower among females, with 55.3% agreement for high-point codes, 31.9% agreement for two-point codes, and 31.9% agreement for three-point codes. When they examined agreement rates between the MMPI and MMPI-2,

Chojnacki found three comparisons that were significantly lower than the corresponding 0-0 comparison. High-point code agreement rate for males was only 43.5% ($p < .01$), two-point code agreement for males was 28.3% ($p < .05$), and two-point code agreement for females was only 18.8% ($p < .05$). This study also relied on college student participants, resulting in questionable clinical relevance of the results.

These test-retest agreement rates, which appear to be quite low, raise questions regarding the reliability of the clinical scales. This issue was addressed in the MMPI-2 manual by providing reliability data, including standard error of measurement ($SE_{meas}$) for each of the validity, clinical and content scales. For the validity and clinical scales, these values ranged from a low of 1.00 raw score point on the L Scale for males, to a high of 2.36 raw score points on Scale 0 for females. The $SE_{meas}$ values were based on 82 male, and 111 female community adults who were retested at an average interval of 8.58 days (median interval of 7 days). These data are provided in Appendix D of the MMPI-2 manual (Butcher et al., 1989). Although the T score variance that would result from adding or subtracting 3 raw score points varies from scale to scale, the variability of 1 $SE_{meas}$ translates to approximately 5 T score points on any of the validity or clinical scales. Therefore, variability of less than 5 T score points on any scale should not be considered clinically or statistically significant. In terms of the impact of $SE_{meas}$ on high-point, two-point, or three-point codes, that would depend on the relative elevation of the next highest scales. If these differences

were less that 5 T score points, then changes in high-point, two-point or three-point codes could be attributable to limitations in the reliability of the test. For example, assume a person obtained a two-point code of 4/7, with T scores of 87 and 73 respectively. However, the next highest clinical scale on their profile was Scale 9, with a T score of 71. To Label their two-point code as 4/7 might be inaccurate or irrelevant, since their "true" two-point code could just as easily be 4/9, given the impact of $SE_{meas}$ on Scales 7, and 9.

Morrison, Edwards, and Weissman (1994) compared the MMPI and MMPI-2 as predictors of psychiatric diagnosis in 200 outpatients. The 200 participants (100 male, 100 female) completed test items from both forms of the test. This was done by administering the complete MMPI, along with the revised or added items from the MMPI-2, in four counterbalanced combinations. Diagnoses, which were rendered by referring clinicians, were assigned to one of five categories in a system devised by Lachar, Dahlstrom, and Moreland (1986). The five possible categories were normal, neurotic, character disordered, psychotic, and other. Next, MMPI and MMPI-2 responses were scored and assigned high-point or two-point codes based on the highest score(s) in the clinical range (i.e., $T \geq 70$ on MMPI, $T \geq 65$ on MMPI-2). Each profile was then assigned to one of five categories, again based on the system devised by Lachar et al. (1986). Results of their analyses indicated 75% agreement between category placement based on MMPI scores, and category placement based on MMPI-2 scores. Both the MMPI and the MMPI-2 had a 39% agreement rate between

category placement based on test scores, and category placement based on clinician's diagnoses. These results would appear to be quite discouraging to clinicians wishing to utilize the MMPI or MMPI-2 in the diagnosis of mental illness. However, the clinical relevance of the classification system employed in this study was questionable. For example, the psychotic classification included participants with a diagnosis of bipolar disorder, regardless of whether it was a primarily depressed or manic presentation. This inclusion of bipolar disorder with psychotic participants made little clinical sense.

The same outpatient population utilized in the above study was also examined by Edwards et al. (1993). This study investigated agreement between MMPI and MMPI-2 code types. They found a 58% agreement rate for elevated codes (i.e., $T \geq 70$ on MMPI, $T \geq 65$ on MMPI-2). They also found a 58% agreement rate for two-point codes independent of elevation. The two-point code concordance rates were 50% for males and 66% for females when gender groups were analyzed separately.

Edwards et al. (1993) also examined the effect of including only well defined codes in their comparisons. The use of well defined versus nonrestrictive codes has been an arena of great debate among MMPI and MMPI-2 researchers. Well defined codes are generally defined as those in which the scale(s) included in the code are at least 5 T score points higher than the next highest scale(s) in a given profile. Proponents of the use of well defined codes (Edwards et al., 1993; Graham et al., 1991; Tellegen & Ben-Porath, 1993) have argued that classification

of protocols utilizing criteria for well defined codes results in more homogeneous groupings and, therefore, more meaningful interpretations. As a proponent of the use of nonrestrictive codes, Dahlstrom (1992) argued that interpreting only well defined codes seriously handicapped MMPI users by limiting the number of interpretable protocols. Dahlstrom also pointed out that the vast majority of studies of MMPI code patterns, which formed the foundation for MMPI and MMPI-2 interpretation, were carried out without restriction to well defined codes. He further stated that the use of well defined codes, "makes little sense either psychometrically or clinically" (Dahlstrom, 1992, p. 161).

Edwards et al. (1993) found that when only elevated and well defined profiles were included, there was a 72% agreement rate. However, fewer than 40% of the participants met these criteria, supporting the arguments of Dahlstrom (1992).

Even those involved in developing the revised version of the test have not been consistent in suggesting that MMPI-2 interpretations can rely on the empirical foundation of MMPI research and clinical usage. Dahlstrom (1992) commented, "It is a matter of some urgency that research be carried out to establish the external correlates of code patterns on modern norms provided on the MMPI-2" (p. 163). This was based on his investigation of 2600 records used in the MMPI-2 restandardization sample (Butcher et al., 1989). This study also utilized form MMPI-AX. Results of a comparison of two-point codes between each participant's MMPI and MMPI-2 profile indicated only 40.8% agreement for

males and 43.8% agreement for females. This was considerably lower than agreement ratios reported by Butcher et al. (1989) which were approximately two-thirds. It should also be noted that these percentages are considerably lower than those reported by Graham et al. (1991) who examined the same data set (i.e. the 2600 records from the restandardization sample). They reported 64.5% agreement of two-point codes for males and 62.3% for females. This discrepancy is apparently due to the fact that Dahlstrom included Scales 5 and 0 as possible two-point code components; these scales were excluded by Graham and colleagues (Tellegen & Ben-Porath, 1993).

What can be surmised from the preceding studies is that two-point code agreement between MMPI and MMPI-2 profiles likely ranges between 40% and 65%. The implications of these findings, as stated by Dahlstrom (1992), are that they "... highlight the need for new empirical data on the correlates of coding patterns based on these [MMPI-2] norms" (p. 153).

MMPI-2 Content Scales

One change resulting from the revision of the MMPI was the inclusion of fifteen new scales, the MMPI-2 content scales (see Appendix B). Most of these were based on the Wiggins (1966) content scales for the original MMPI (Helmes & Reddon, 1993). A few studies have investigated the clinical utility of these new scales in the diagnosis of mental illness.

Ben-Porath, Butcher, and Graham (1991) administered the MMPI-2 to 423 psychiatric inpatients in an investigation of the contribution of the MMPI-2

content scales to the differential diagnosis of schizophrenia and major depression. Of these, 76 carried a diagnosis of schizophrenia and 84 were diagnosed with major depression. Diagnoses were rendered by staff psychiatrists or psychologists at the time of hospitalization (prior to administration of the MMPI-2). Results of multivariate analyses of variance (MANOVAs) indicated that two content scales, Depression (DEP) and Bizarre Mentation (BIZ), contributed incrementally to the differential diagnosis of schizophrenia and major depression. The addition of these two scales raised the total amount of variance in diagnosis accounted for by the MMPI-2 from 22% to 37%. The authors concluded that the content scales were a viable source of information in the differential diagnosis of schizophrenia and major depression. They also acknowledged that this study was limited by its narrow focus on two diagnoses and its small sample size.

Ben-Porath, McCully, and Almagor (1993) investigated the validity of the MMPI-2 content scales in the assessment of psychopathology by administering the MMPI-2 and other psychological tests to 596 college students (339 females, 257 males). They calculated zero-order correlations between the 10 MMPI-2 clinical and 15 MMPI-2 content scales versus 8 extratest measures from four other psychological tests. These other tests were: the Beck Depression Inventory (BDI; Beck, 1987), the Symptom Checklist 90-Revised (SCL-90-R; Derogatis, 1983), the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1979), and the State-Trait Anger Expression Inventory (Spielberger, 1987). Next, the researchers performed combined hierarchical stepwise regression analyses on those clinical

and content scales which were found to correlate significantly with various scales in the extratest measures. These regression analyses indicated that the content scales contributed between 2% and 24% of additional explained variance over the clinical scales for males, and between 7% and 33% of additional explained variance for females. They concluded that the content scales did add incrementally to the prediction of variance on these self-report measures. They also acknowledged that this study was limited by the use of college students for MMPI-2 data and by the use of self-report inventories as measures of psychopathology.

Boone (1994) studied the validity of the MMPI-2 Depression content scale (DEP) in identifying depressed and/or suicidal psychiatric inpatients. He administered the MMPI-2 to 62 psychiatric inpatients (40 males, 22 females), along with the BDI, the Beck Hopelessness Scale (Beck, 1988), and the Suicide Probability Scale (Cull & Gill, 1983). Participants were classified as depressed or non-depressed based on psychiatric diagnoses rendered by a staff psychiatrist. Analyses indicated that both the DEP Scale and clinical Scale 2 correlated significantly with the other measures of depression, with all correlations being stronger for the DEP Scale. Also, mean DEP Scale scores were higher for depressed participants than non-depressed participants (71.4 vs. 62.8, respectively), and were higher for participants assessed to be at risk for suicide as compared to those considered not at risk (83.5 vs. 59.3, respectively) as assessed by the Suicide Probability Scale. Both of these differences were reported to be statistically

significant. Limitations of this study included the small sample size and failure of the researcher to report mean scores on Scale 2 for depressed versus non-depressed and suicidal versus non-suicidal participants. This investigation did, however, provide evidence that the DEP Scale accounts for significant variance in measures of depression and suicidal ideation.

Statistical Issues

The statistical analyses employed to validate the diagnostic utility of tests such as the MMPI have typically been ANOVA and MANOVA (Elwood, 1993). Validation studies of the MMPI-2 have followed in this tradition of comparing group mean test scores of target diagnostic groups with reference groups (Ben-Porath et al., 1991; Weed, Butcher, McKenna, & Ben Porath, 1992). These comparisons, however, may have little relevance to clinical utility because group means are not necessarily representative of the individual profiles that make up those means (Butcher & Tellegen, 1978; Brems, 1991; Elwood, 1993). Furthermore, clinicians interpret individual profiles, not group means.

In an attempt to overcome the limitation of group mean comparisons, some researchers have used measures of sensitivity and specificity. Sensitivity refers to the true positive rate and is expressed as the proportion of subjects with a target disorder who are identified by positive test scores. Specificity refers to the true negative rate and is expressed as the proportion of subjects without the disorder who are identified by normal scores. In Table 1, cell a represents true positives; cell b represents false positives; cell c represents false negatives; and

cell d represents true negatives. Table 2 shows that sensitivity would be calculated a/(a+c), and specificity would be calculated d/(b+d).

Although these ratios are more useful and relevant than comparisons of group means, they also have significant limitations. The task faced by clinicians is exactly the opposite of that expressed by sensitivity and specificity. The clinician must make a decision on diagnosis based upon positive or negative test results. Even if a test has an established sensitivity rate of 1.0, it does not follow that all positive test scores are indicative of that disorder. Similarly, even if a test has an established specificity rate of 1.0, it does not follow that all negative test scores are indicative of the absence of that disorder. These shortcomings are due in part to the fact that sensitivity and specificity are not affected by variations in the base rate of the target disorder. The base rate, or prevalence rate, of a disorder defines the probability of obtaining a positive test result by chance (Meehl & Rosen, 1955). Meehl and Rosen (1955) provided evidence that the accuracy of diagnostic decisions based on test scores while ignoring prevalence rates of the target disorder could actually increase diagnostic errors. For these reasons, it has been recommended that prevalence rates be considered in discriminant validity studies of the MMPI-2 (Butcher, 1990; Gottesman & Prescott, 1989; Weed et al., 1992). However, this advice has been followed by only a handful of researchers (Elwood, 1993; Gerardi, Keane, & Penk, 1989. Moldin, Gottesman, Rice, & Erlenmeyer-Kimling, 1991).

Table 1

Distribution of Participants by Diagnoses and Test Results

| | Diagnosis | | |
|---|---|---|---|
| Test Result | Positive | Negative | Totals |
| Positive | a | b | a+b |
| Negative | c | d | c+d |
| Totals | a+c | b+d | N |

Table 2

Definitions and Computations of Sensitivity, Specificity, PPP, NPP & Prevalence

| | Definition[a] | Computation |
|---|---|---|
| Sensitivity | True + / all Dx | a/(a+c) |
| Specificity | True - / all no Dx | d/(b+d) |
| PPP | True + / all + | a/(a+b) |
| NPP | True - / all - | d/(c+d) |
| Prevalence | All Dx / all participants | (a+c)/N |

Note. Pluses indicate positive results; minuses, negative results; Dx, diagnosis present; and no Dx, diagnosis absent. Tables adapted from Baldessarini, Finklestein, & Arana (1983, p. 570). [a]Test results were considered to dichotomize into "positive" or "negative" outcomes, even though graduations or ambiguous results are common in reality.

Baldessarini et al. (1983) described a method for computing the predictive power of diagnostic tests by taking into account the base rate of a given diagnosis. In their system, positive predictive power (PPP) refers to the ratio of true test positives to all test positives, and would be calculated $\underline{a/(a+b)}$ using Tables 1 and 2. PPP, then, expresses the probability that a participant has a target disorder given a significant elevation on a corresponding test scale. Conversely, negative predictive power (NPP) refers to the ratio of true test negatives to all test negatives, and would be calculated $\underline{d/(c+d)}$ using Tables 1 and 2. NPP, then, expresses the probability that a participant does not have a target disorder given a subclinical score on a corresponding test scale. This procedure has been applied to psychological test validation (Keane, Caddell, & Taylor, 1988; Olin, Schneider, Eaton, Zemansky, & Pollock, 1992; Rapp, Parisi, Walsh, & Wallace, 1988) and to diagnostic classification (Elwood, 1993; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984). Only one study has been published which utilized PPP and NPP in validating the MMPI-2 (Elwood, 1993).

It should be noted that these terms are not universally recognized, and some authors who refer to sensitivity and specificity may, in fact, be describing statistical procedures which are defined here as PPP and NPP, respectively (Bagby, Rogers, & Buis, 1994; Rogers, Bagby, & Chakraborty, 1993). This confusion is compounded by the fact that most published articles using these terms do not provide operational definitions. Glaros and Kline (1988) provide an

excellent description of these statistical procedures and make recommendations for their use in evaluating diagnostic instruments which utilize cutting scores.

Elwood (1993) investigated the clinical utility of the MMPI-2 in the diagnosis of unipolar depression among male alcoholics. One hundred six male inpatients admitted to an alcohol treatment unit were administered the MMPI-2. Of these 106 men, 87 provided valid profiles, all of whom were diagnosed with alcohol dependence. Of these 87 men, 17 (20%) were depressed within the preceding month, 13 with unipolar depression and 4 with alcohol-induced depression. Diagnoses were rendered by an experienced clinical psychologist using the Structured Clinical Interview for DSM-III-R-Patient Version (SCID-P; Spitzer, Williams, Gibbon, & First, 1988).

Elwood (1993) found that by using only Scale 2 with a T score cutting score of 65, 42% of the depressed participants were detected (sensitivity rate = .42). However, when prevalence rates were taken into consideration, the test failed to predict the presence of depressive disorder significantly (PPP = .31), although the test did predict the absence of depressive disorder significantly (NPP = .88). When test positives were limited to two-point codes of 2-3/3-2, 2-7/7/2, and 2-4/4-2, the PPP rose to 1.00, however sensitivity fell to the level of chance. The failure of this study to achieve significant results was attributed by the author to its limited composition and focus on Scale 2 elevations. The use of only male veterans in an alcohol-treatment program severely limited the generalizability of this investigation. Also, the MMPI-2 content scales were not

evaluated because half of the participants completed a short version of the MMPI-2 that did not allow for scoring the content scales.

To summarize, there are several reasons to caution against interpreting the MMPI-2 by relying on the vast literature base of the original MMPI. These limitations include changes in diagnostic criteria over the past fifty years, the new normative group utilized in validating the MMPI-2, and the change from the use of linear T scores to the use of uniform and linear T scores. Other concerns have been raised by studies comparing code-types obtained from the MMPI and MMPI-2. These studies have revealed agreement rates ranging from 40% to 65%. Furthermore, there have been very few studies of the clinical utility of the new content scales, as these were not a formal part of the original MMPI. Limitations of many of the past studies of the MMPI and MMPI-2 include the use of statistical procedures that fail to reflect clinically relevant information, the lack of comparisons between gender and ethnicity groups, and the use of non-clinical populations. For these reasons, it appears that validity studies of the MMPI-2, using sound methodology and clinically relevant populations and statistical procedures, are needed in order to understand the strengths and limitations of this diagnostic tool.

## The Present Study

The present study investigated the diagnostic utility of the MMPI-2 in classifying participants according to three diagnostic categories. These categories were depressed (major depressive disorder; dysthymic disorder; and bipolar

disorder, depressed), schizophrenic (schizophrenia, schizophreniform disorder, and schizoaffective disorder), and borderline personality disorder (BPD). Concurrent validity was examined by comparing MMPI-2 scores associated with these disorders with DSM-III-R (American Psychiatric Association, 1987) admission and discharge diagnoses rendered by psychiatrists in the course of hospitalization at a state hospital.

The particular questions which the current study attempted to address were: (1) Will the MMPI-2 profiles of participants vary significantly based on diagnostic classification? (2) Will the MMPI-2 profiles vary in a manner consistent with expectations based on MMPI research? (3) Will diagnostic classifications based on MMPI-2 scores stand up to measures of sensitivity, specificity, PPP, and NPP? (4) Will MMPI-2 scores account for significant variance in DSM-III-R diagnoses on Axis IV and Axis V? (5) Will MMPI-2 scores account for significant variance in length and frequency of hospitalization?

Hypotheses

Hypotheses were as follows:

(1) Multivariate analysis of variance (MANOVA) will yield significant differences in MMPI-2 profiles between diagnostic groups.

(2) Univariate analyses of variance (ANOVAs) will yield significant differences between diagnostic groups on various clinical and content scales, including the following:

(a) Clinical Scale 2 and the DEP content scale will be higher for the depressed category than for the schizophrenic or BPD categories. Also, Scales 3 and 7 will be higher for the depressed category than for the schizophrenic category (Boone, 1994; Ben-Porath et al., 1991; Elwood, 1993; Nichols, 1988; Vincent et al., 1983).

(b) Clinical Scales 4, 6, and 8, and the BIZ content scale will be higher for the schizophrenic category than for the depressed category. Scale 6 will be higher for the schizophrenic category than for both the depressed and BPD categories (Ben-Porath et al., 1991; Vincent et al., 1983; Walters, 1984; Walters & Greene, 1988).

(c) Clinical Scales 3, 4, 7, and 8, will be higher for the BPD category than for the depressed or schizophrenic categories (Evans, Ruff, Braff, & Ainsworth, 1984; Gartner et al., 1989; Vincent et al., 1983; Zalewski & Archer, 1991).

(3) For each diagnostic group, a set of clinical and content scales will correctly classify a significant proportion of participants according to diagnostic group as evidenced by measures of sensitivity, specificity, PPP, and NPP. These measures will be computed twice for each diagnostic group, once using well defined codes, and then using non-restrictive codes. The hypothesized predictors are as follows.

(a) Test scores that will predict classification in the depressed category will be Scale 2 high-point (defined as Scale 2 being the highest

elevation of the clinical scales and having an elevation of T ≥ 65), and/or elevation of the depression content scale (DEP T ≥ 65), and/or two-point codes of 2-3/3-2, 2-7/7-2, and 2-4/4-2.

(b) Test scores that will predict classification in the Schizophrenic category will be elevation of the BIZ content Scale, two-point codes of 6-8/8-6 and/or three-point codes of 2-4-8 (with at least one of these three T scores ranging from 65 to 75).

(c) Test scores that will predict classification in the BPD category will be Scale 4 high-point (as defined above) and/or Scale 9 high-point and/or two-point codes of 3-4/4-3, and 7-8/8-7 and/or three-point codes of 2-4-8 (with T ≥ 75 on all three scales).

(4) Axis IV (severity of psychosocial stressors) rating is hypothesized to have a significant positive correlation with total MMPI-2 elevation; and Axis V (GAF) rating is hypothesized to have a significant negative correlation with total MMPI-2 elevation.

(5) Length of hospitalization is hypothesized to have a significant positive correlation with the negative treatment indicators content scale (TRT).

# CHAPTER II

# METHOD

## Participants/Data Source

Data were collected at Big Spring State Hospital (BSSH) in Big Spring, Texas. Data collection was archival. A total of 477 case files and corresponding MMPI-2 tests were reviewed for this study. The average age was 34.9 years, with 388 (81.3%) participants being White, 47 (9.9%) Hispanic, 38 (8.0%) Black, 2 (.4%) American Indian, and 2 (.4%) coded as other. Of the 477 participants, 116 had a diagnosis on either admission or discharge that qualified them for inclusion in the depressed category, 82 for the schizophrenic category, and 54 for the BPD category.

## Procedures

The MMPI-2 raw data were computer scored, and had been stored in a computer file as part of another investigation (Greene, 1991). The MMPI-2 tests had been given as part of the routine diagnostic procedure following admission to BSSH. MMPI-2 tests were administered between November, 1989 and May, 1995. Diagnostic information was obtained from the Client Assignment and Registration (CARE) system, a computer data-base which tracks individuals receiving services from any agency of Texas Department of Mental Health and Mental Retardation. The accuracy of this information was checked by randomly selecting 25 case files

and checking diagnoses against those obtained from the CARE system. No inconsistencies were found.

MMPI-2 profiles were considered invalid if the VRIN scale raw score exceeded 13, or if the F Scale T score exceeded 99 or if the ? ("cannot say") count exceeded 10. Determination of cutting scores for validity purposes always requires a balance between an ideal confidence level (i.e. 99% certainty that included tests are valid) and loss of statistical power through attrition. Furthermore, cutting scores vary considerably across investigations. For example, Bagby et al. (1994) used a cutting score on the F Scale of T score >103 for males, and > 112 for females. Graham et al. (1991) chose a cutting score on the F Scale of T score > 118 for males, and > 119 for females. In studies reviewed by Berry et al. (1991), optimal hit rates for F Scale T scores ranged from 88 to 92. In the present investigation, the cutting scores selected are consistent with acceptable values established by these and other investigators for use with clinical populations (Bagby et al., 1994; Berrey et al., 1991; Graham et al., 1991; Rogers et al., 1993; Rogers, Sewell, & Salekin, 1994; Wetter, Baer, Berry, Smith, & Larsen, 1992).

Using these validity criteria, a total of 345 participants were included (166 male, 179 female), 162 of whom had at least one of the target diagnoses. Of the 345 participants with valid profiles, 286 (82.9%) were White, 29 (8.4%) were Hispanic, 28 (8.1%) were Black, 1 (.3%) was American Indian and 1 (.3%) was classified as other. There were 70 participants included in the depressed category

(29 male, 41 female), 62 in the schizophrenic category (37 male, 25 female) and 30 in the BPD category (7 male, 23 female). Median length of hospital stay for all participants was 28 days.

Base rates for each of the three diagnostic categories were established by calculating the ratio of participants who met diagnostic criteria for each category, either at admission or at discharge, to total number of participants. Using this method, the base rate for depression was .24; for schizophrenia, .17; and for BPD, .11.

A high-point code was defined as the highest clinical scale elevation with $T \geq 65$, not including Scale 5 or Scale 0. Well defined high-point codes were statistically analyzed separately as a sub-group. A well defined high-point code was operationally defined as one in which the most elevated scale exceeded the next highest scale by at least five points. This definition of well defined high-point codes has been utilized by other researchers (Graham et al., 1991; Tellegen & Ben-Porath, 1993).

A non-restrictive two-point code was defined as the highest two scale elevations among clinical scales with $T \geq 65$, not including Scale 5 or Scale 0. Well defined two-point codes were also analyzed separately. A well defined two-point code was operationally defined as one in which the lower of the two most elevated scales exceeds the third highest scale by at least five points. This definition of well defined two-point codes has also been utilized by other researchers (Graham et al., 1991; Tellegen & Ben-Porath, 1993).

Three-point codes were defined as the highest three scale elevations among clinical scales with T ≥ 65, not including Scale 5 or Scale 0. In case of tied two-point or three-point codes (i.e. if more than one two-point or three-point code is possible due to equal T scores) the customary rule for ranking tied scales was used, which involved choosing the scale with the lowest number first (Edwards et al., 1993; Morrison et al., 1994; Vincent et al., 1983). Other variables included participants' Global Assessment of Functioning (GAF) score from Axis V of the DSM-III-R (American Psychiatric Association, 1987), severity of psychosocial stressors as documented on Axis IV of each participant's admitting diagnosis and length of hospitalization (expressed in number of days of hospitalization in which the MMPI-2 was administered).

Analyses

Statistical analyses included use of multivariate analyses of variance (MANOVAs), univariate analyses of variance (ANOVAs), correlations, and measures of sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). The first step in statistical analysis was to perform MANOVAs for the five validity scales, ten clinical scales, and fifteen content scales of each of the three target groups (depressed, schizophrenic, and BPD), comparing the scores of each group.

The second step was to perform ANOVAs to determine which specific scales were contributing to the overall differences between groups. Because three groups were being compared in this analysis, Duncan's multiple ranges tests were

used to examine how the groups differed, with alpha set at $p < .05$. In an effort to control for between group differences on potentially confounding variables such as gender and ethnicity, separate MANOVAs and ANOVAs were performed to examine the degree to which these variables contributed to test score variance. Only participants whose primary discharge diagnosis qualified them for placement in one of the three diagnostic groups were selected for purposes of MANOVAs and ANOVAs. This insured that the three groups were independent, with no participants scores being entered into two groups.

Next, for each MMPI-2 predictor (as specified in the third Hypothesis), calculations were performed to determine rates of sensitivity, specificity, PPP, and NPP. A given PPP value was considered significant if it exceeded 1 minus the identified base rate for that diagnostic group. This test for significance was recommended by Meehl and Rosen (1955) for populations in which the base rate of a target disorder is less than 50%. Similarly, an NPP value was considered significant if it exceeded the identified base rate for that diagnostic group. A given sensitivity rate was considered significant if it exceeded 1 minus the base rate for that diagnostic group. Similarly, a specificity rate was considered significant if it exceeded the identified base rate for that diagnostic group.

Finally, correlations were computed. The first correlation compared Axis IV (severity of psychosocial stressors) ratings and sum of all T scores (K corrected, excluding Scales 5 and 0). The second correlation compared Axis V (GAF) ratings and sum of all T scores (K corrected, excluding Scales 5 and 0).

The third and final correlation compared leng:h of hospitalization (during which the MMPI-2 was administered) and scores on the negative treatment indicators content scale (TRT).

# CHAPTER III

# RESULTS

## MANOVAs and ANOVAs

It was hypothesized that MMPI-2 scores would be significantly different between diagnostic groups. This hypothesis was supported by results of MANOVA, which indicated significant differences between diagnostic groups, Wilk's lambda = .406, $F(2, 159) = 2.47$, $p < .001$.

It was further hypothesized that univariate analyses of variance (ANOVAs) would yield significant differences between diagnostic groups on various clinical and content scales, including the following:

> Clinical Scale 2 and the DEP content scale will be higher for the depressed category than for the schizophrenic or BPD categories. Also, Scales 3 and 7 will be higher for the depressed category than for the schizophrenic category.

This hypothesis was partially supported. As summarized in Table 3, significant differences were found across diagnostic groups. Specifically, depressed participants scored significantly higher than schizophrenics on Scale 2, but did not score higher than BPDs on this scale. Similarly, depressed participants scored significantly higher than schizophrenics on the DEP Scale, but BPDs scored even higher. Depressed participants did not score higher than schizophrenics on Scales

Table 3

Mean (M) T Scores, Standard Deviations, Univariate F Ratios, and Significance by Diagnostic Category

| | Diagnostic Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SCH | | DPR | | BPD | | | |
| Scale | M | SD | M | SD | M | SD | F (2, 159) | p |
| L | 58.98[a] | 11.88 | 56.53[a] | 11.64 | 53.03[a] | 13.08 | 2.52 | .08 |
| F | 69.02[a] | 16.73 | 67.94[a] | 16.03 | 77.03[b] | 15.13 | 3.53 | .03 |
| K | 45.90[a] | 9.37 | 43.50[a] | 9.22 | 43.87[a] | 10.27 | 1.14 | .32 |
| 1 | 59.00[a] | 13.10 | 62.01[a] | 14.33 | 59.83[a] | 14.13 | 0.81 | .44 |
| 2 | 62.08[a] | 11.42 | 71.83[b] | 16.60 | 71.17[b] | 14.13 | 8.49 | < .01 |
| 3 | 53.29[a] | 13.54 | 61.21[a,b] | 16.59 | 59.67[b] | 15.56 | 4.66 | .01 |
| 4 | 62.23[a] | 14.70 | 69.70[b] | 13.82 | 76.47[c] | 15.75 | 10.47 | < .01 |
| 5 | 52.94[a] | 10.39 | 52.74[a] | 10.03 | 52.23[a] | 11.62 | 0.05 | .96 |
| 6 | 67.92[a] | 18.14 | 65.63[a] | 15.79 | 69.50[a] | 14.31 | 0.67 | .51 |
| 7 | 62.69[a] | 16.56 | 67.37[a,b] | 13.72 | 71.57[b] | 14.53 | 3.81 | .02 |
| 8 | 70.32[a] | 16.38 | 67.84[a] | 13.84 | 73.70[a] | 15.41 | 1.61 | .20 |
| 9 | 59.42[a] | 12.62 | 55.76[a] | 12.38 | 58.37[a] | 11.96 | 1.51 | .22 |
| 0 | 57.82[a] | 11.09 | 59.26[a] | 12.46 | 60.43[a] | 12.36 | 0.53 | .59 |
| VRIN | 60.38[a] | 11.53 | 56.66[a] | 11.00 | 59.47[a] | 11.42 | 1.90 | .15 |

(table continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TRIN | 64.21[a] | 12.25 | 62.17[a] | 11.22 | 63.03[a] | 11.33 | 0.50 | .60 |
| ANX | 58.29[a] | 13.86 | 66.40[b] | 12.43 | 67.57[b] | 13.01 | 8.06 | < .01 |
| FRS | 60.89[a] | 13.70 | 53.14[b] | 11.88 | 54.67[b] | 12.08 | 6.51 | < .01 |
| OBS | 57.84[a] | 13.50 | 59.34[a] | 12.66 | 59.60[a] | 13.15 | 0.28 | .75 |
| DEP | 62.00[a] | 13.04 | 72.39[b] | 12.78 | 75.83[b] | 12.46 | 15.94 | < .01 |
| HEA | 61.21[a] | 14.23 | 62.70[a] | 12.36 | 61.67[a] | 12.71 | 0.21 | .80 |
| BIZ | 67.10[a] | 14.20 | 56.71[b] | 11.87 | 60.07[b] | 10.15 | 11.44 | < .01 |
| ANG | 52.05[a] | 12.44 | 58.37[b] | 10.81 | 62.87[b] | 14.90 | 8.89 | < .01 |
| CYN | 60.53[a] | 9.77 | 61.01[a] | 12.17 | 57.70[a] | 12.83 | 0.92 | .40 |
| ASP | 55.84[a] | 8.60 | 57.01[a] | 11.61 | 61.43[a] | 15.04 | 2.54 | .08 |
| TPA | 53.13[a] | 10.82 | 53.06[a] | 10.50 | 56.50[a] | 12.80 | 1.16 | .32 |
| LSE | 59.92[a] | 13.84 | 65.37[b] | 14.91 | 68.33[b] | 14.82 | 4.10 | .02 |
| SOD | 55.63[a] | 10.62 | 56.94[a] | 13.18 | 59.20[a] | 15.18 | 0.80 | .45 |
| FAM | 60.34[a] | 13.50 | 64.77[a,b] | 13.81 | 69.60[b] | 12.97 | 4.95 | <.01 |
| WRK | 60.97[a] | 13.48 | 66.01[a] | 13.70 | 65.40[a] | 15.21 | 2.36 | .10 |
| TRT | 62.94[a] | 13.94 | 66.80[a] | 15.29 | 68.37[a] | 15.75 | 1.74 | .18 |

Note. In each row, means with the same superscript do not differ at the .05 level or better using Duncan's multiple-range tests. SCH = schizophrenic category (n = 65), DPR = depressed category (n = 70), and BPD = borderline personality disorder category (n = 30). See Appendices A and B for an explanation of abbreviations.

3 or 7. Additionally, depressed participants scored higher than schizophrenics on Scales 4, ANX, ANG, and LSE.

Clinical Scales 4, 6, and 8, and the BIZ content scale were hypothesized to be higher for the schizophrenic category than for the depressed category. Also, Scale 6 was hypothesized to be higher for the schizophrenic category than for both the depressed and BPD categories. This hypothesis was also partially supported. Schizophrenics did score significantly higher than depressed and BPD participants on the BIZ Scale, but did not score significantly higher on Scales 4, 6, or 8. Schizophrenics also scored significantly higher on the FRS Scale than either of the other two groups.

Clinical Scales 3, 4, 7, and 8, were hypothesized to be higher for the BPD category than for the depressed or schizophrenic categories. This hypothesis was partially supported as well. Scales 3 and 7 were significantly higher for the BPD participants than for the schizophrenic participants, and Scale 4 was higher for the BPD participants than either of the other two groups. Scale 8, however, was not significantly different for any group. Additionally, the BPD group scored significantly higher than schizophrenic participants on Scales 2, ANX, DEP, ANG, LSE, and FAM. Also, BPD scores were higher than either of the other two groups on the F Scale. ANOVA results are summarized in Table 3.

Because of concerns regarding heterogeneity of the depressed group, ANOVAs and MANOVAs were performed for two sub-groups of depressed participants. These groups were labeled depressed[2] (DEP[2]), which included

participants with major depression and dysthymia; and bipolar (BIPLR), which included only participants with bipolar disorder (depressed). The results of these analyses were highly consistent with results for the two groups combined. The results of these ANOVAs and MANOVAs are presented in Appendix C.

Ethnicity and Gender

MANOVAs and ANOVAs were computed separately for White Females, White Males, All Females, and All Males. Table 4 lists the number of participants in each group, by diagnosis. The number of non-white participants was considered too small for meaningful interpretation of MANOVAs and ANOVAs.

Because of the low number of non-white participants in each diagnostic group, statistical power was inadequate to reveal statistically significant differences between diagnostic groups. Thus, ethnic comparability cannot be established in the present study. It should, however, be noted that the validity attrition rate due to invalid profiles was appreciably different for Hispanics, as compared to Blacks and Whites (38% for Hispanics, 26% for Blacks, 26% for Whites, and 1 out of 2 for American Indians and for others). When all female participants and all male participants were examined separately, considerable differences emerged. Fewer significant results were found for males as compared to females, particularly for the clinical scales. These results are summarized in Table 5, and are presented in detail in Appendices D,E,F and G.

Table 4

Frequency Data: Ethnicity & Gender

| | Diagnostic Classification | | | | |
|---|---|---|---|---|---|
| Ethnicity & Gender | SCH | DEP[1] | BPD | DEP[2] | BIPLR |
| Non-white Females | 4 | 5 | 2 | 1 | 4 |
| Non-white Males | 12 | 9 | 2 | 4 | 5 |
| White Females | 21 | 36 | 21 | 9 | 27 |
| White Males | 25 | 20 | 5 | 6 | 14 |
| All Females | (25) | (41) | (23) | (10) | (31) |
| All Males | (37) | (29) | (7) | (10) | (19) |
| Totals | 62 | 70 | 30 | 20 | 50 |

Note. SCH = schizophrenia category; DEP[1] = depressed category, including bipolar disorder (depressed); BPD = borderline personality disorder; DEP[2] = depressed category, excluding bipolar disorder; BIPLR = bipolar disorder (depressed). Values listed in ( ) are not included in totals.

As can be seen by comparing Table 3 and Appendix F, a few comparisons were significant for all females that were not significant for all participants combined. The L Scale was significantly higher for schizophrenics than for BPDs. Also, Scales 3, 7, FAM, and WRK, were significantly higher for depressed participants than for schizophrenics. As can be seen by comparing Table 3 and

Table 5

Scale Differences between Diagnoses: P values by Ethnicity and Gender

| Scale Abbreviation | All Valid Profiles | White Females | White Males | All Females | All Males |
|---|---|---|---|---|---|
| L | .08 | .03* | .96 | .06 | .57 |
| F | .03* | .20 | .14 | .27 | .27 |
| K | .32 | .30 | .90 | .29 | .93 |
| 1 | .44 | .60 | .85 | .49 | .95 |
| 2 | < .01* | .01* | .03* | < .01* | .14 |
| 3 | .01* | .18 | .07 | .10 | .14 |
| 4 | < .01* | < .01* | .01* | < .01* | .04* |
| 5 | .96 | .62 | .87 | .49 | 1.00 |
| 6 | .51 | .51 | .51 | .66 | .74 |
| 7 | .02* | .02* | .21 | .01* | .49 |
| 8 | .20 | .26 | .51 | .32 | .46 |
| 9 | .22 | .14 | .43 | .23 | .70 |
| 0 | .59 | .25 | .66 | .31 | .57 |
| VRIN | .15 | .03* | .86 | .06 | .66 |
| TRIN | .60 | .96 | .65 | .92 | .52 |
| ANX | < .01* | < .01* | .14 | < .01* | .26 |
| FRS | < .01* | .63 | .04* | .46 | < .01* |

(table continued)

| | | | | | |
|---|---|---|---|---|---|
| OBS | .75 | .35 | .93 | .31 | .95 |
| DEP | < .01* | < .01* | < .01* | < .01* | < .01* |
| HEA | .80 | .70 | .87 | .51 | .78 |
| BIZ | < .01* | < .01* | < .01* | < .01* | < .01* |
| ANG | < .01* | < .01* | .24 | < .01* | .07 |
| CYN | .40 | .40 | .61 | .35 | .91 |
| ASP | .08 | .54 | .19 | .70 | .03* |
| TPA | .62 | .59 | .90 | .62 | .66 |
| LSE | .02* | .02* | .34 | .01* | .75 |
| SOD | .45 | .79 | .69 | .55 | .90 |
| FAM | < .01* | < .01* | .92 | < .01* | .74 |
| WRK | .98 | .04* | .41 | .04* | .79 |
| TRT | .18 | .41 | .77 | .37 | .69 |

*Denotes significant Manova results ($p < .05$).

Appendix G, only one scale, ASP, had significant differences for males that were non-significant with all participants combined. This scale was significantly higher for BPDs than for depressed and schizophrenic participants. It should be noted that there were only 7 males with a diagnosis of BPD.

Sensitivity, Specificity, PPP, and NPP

Hypothesis (3) stated: For each diagnostic group, a set of clinical and content scales will correctly classify a significant proportion of participants according to diagnostic group as evidenced by measures of sensitivity, specificity,

PPP, and NPP. It is further hypothesized that these relations will appear less robust than is suggested by results of ANOVAs. The hypothesized predictors are as follows: Test scores that will predict classification in the depressed category will be Scale 2 high-point (defined as Scale 2 being the highest elevation of the clinical scales and having an elevation of T ≥ 65), and/or elevation of the depression content scale (DEP T ≥ 65), and/or two-point codes of 2-3/3-2, 2-7/7-2, and 2-4/4-2.

These criteria yielded a specificity rate of .43 and a NPP rate of .88, both of which were significant. The sensitivity rate of .80 and PPP rate of .28 were not significant. In order to be considered significant, the sensitivity rate and PPP needed to exceed .83, and specificity rate and NPP needed to exceed .17. The results of Hypothesis (3) are presented in Table 6.

Table 6

Sensitivity, Specificity, Positive Predictive Power, & Negative Predictive Power

| | Diagnostic Group | | |
|---|---|---|---|
| Measure | Depressed | Schizophrenic | BPD |
| Sensitivity | .80 | .74 | .40 |
| Specificity | .43 * | .39 * | .77 * |
| PPP | .28 | .22 | .15 |
| NPP | .88 * | .86 * | .93 * |

* denotes significant values.

Sensitivity, specificity, PPP and NPP rates were also calculated separately for the depressed[2] group (excluding bipolar disorder, depressed) and for the bipolar disorder (depressed) group. These values were comparable to those for the two groups combined (see Table 7).

Table 7

Sensitivity, Specificity, Positive Predictive Power, & Negative Predictive Power, Depressed[2] and Bipolar categories.

| | Diagnostic Group | |
|---|---|---|
| Measure | Depressed[2] | Bipolar |
| Sensitivity | .82 | .79 |
| Specificity | .41 * | .43 * |
| PPP | .09 | .20 |
| NPP | .97 * | .92 * |

Note: Depressed[2] = depressed participants, excluding bipolar, depressed ($n$ = 20); Bipolar = Bipolar, depressed only, ($n$ = 50).

Hypothesis (3b) stated: Test scores that will predict classification in the Schizophrenic category will be elevation of the BIZ content scale, two-point codes of 6-8/8-6, and/or three-point codes of 2-4-8 (with at least one of these three T scores ranging from 65 to 75). These criteria yielded a specificity rate of .39 and a NPP rate of .86, both of which were significant. The sensitivity rate of .74 and the PPP rate of .22 were not significant. In order to be considered significant, the

sensitivity rate and PPP needed to exceed .76, and specificity rate and NPP needed to exceed .24.

Hypothesis (3c) stated: Test scores that will predict classification in the BPD category will be Scale 4 high-point (as defined above) and/or Scale 9 high-point and/or two-point codes of 3-4/4-3 and 7-8/8-7, and/or three-point codes of 2-4-8 (with T ≥ 75 on all three scales). These criteria yielded a specificity rate of .77 and a NPP rate of .93, both of which were significant. The sensitivity rate of .40 and the PPP rate of .15 were not significant. In order to be considered significant, the sensitivity rate and PPP needed to exceed .89, and specificity rate and NPP needed to exceed .11.

Table 8

Well Defined Scores: Sensitivity, Specificity, Positive Predictive Power, & Negative Predictive Power

| | Diagnostic Group | | |
|---|---|---|---|
| Measure | Depressed | Schizophrenic | BPD |
| Sensitivity | .80 | .80 * | .30 |
| Specificity | .45 * | .32 * | .82 * |
| PPP | .29 | .22 | .13 |
| NPP | .89 * | .88 * | .92 * |

* denotes significant values.

Table 9

Females: Sensitivity, Specificity, Positive Predictive Power, & Negative Predictive Power

| | Diagnostic Group | | |
|---|---|---|---|
| Measure | Depressed | Schizophrenic | BPD |
| Sensitivity | .83 | .85 * | .35 |
| Specificity | .45 * | .31 * | .85 * |
| PPP | .34 | .18 | .25 |
| NPP | .88 * | .92 * | .90 * |

Table 10

Males: Sensitivity, Specificity, Positive Predictive Power, & Negative Predictive Power

| | Diagnostic Group | | |
|---|---|---|---|
| Measure | Depressed | Schizophrenic | BPD |
| Sensitivity | .77 | .76 | .14 |
| Specificity | .46 * | .34 * | .79 * |
| PPP | .24 | .26 | .03 |
| NPP | .90 * | .83 * | .95 * |

* denotes significant values.

As can be seen by comparing Tables 6 and 8, sensitivity, specificity, PPP and NPP values were comparable when well-defined scores were analyzed separately. Only one cell, the sensitivity rate for schizophrenics, improved to a significant level as compared to the use of non well-defined scores.

Gender Differences

When females and males were analyzed separately, values remained fairly consistent. Only one cell, the sensitivity rate for schizophrenia among females, improved to a significant level as compared to males (see Tables 9 and 10).

Correlations

Hypothesis (4) stated: Axis IV (severity of psychosocial stressors) rating is hypothesized to have a significant positive correlation with total MMPI-2 elevation; and Axis V (GAF) rating is hypothesized to have a significant negative correlation with total MMPI-2 elevation. The first comparison resulted in a correlation coefficient of .009. The second comparison resulted in a correlation coefficient of -.002. Neither coefficient approached statistical significance.

Hypothesis (5) stated: Length of hospitalization is hypothesized to have a significant positive correlation with the negative treatment indicators content scale (TRT). This comparison resulted in a correlation coefficient of .072 (not significant).

# CHAPTER IV

# DISCUSSION

The MMPI-2 performed modestly in differentiating participants with diagnoses of depression, schizophrenia, and borderline personality disorder (BPD) from each other and from other psychiatric inpatients. MANOVA and ANOVA results indicated that the groups differed significantly on 12 of the validity, clinical, and content scales. The ANOVA results reported here are consistent with some previous investigations of the MMPI-2. Ben-Porath et al. (1991) reported significantly higher elevations for depressed males on the DEP Scale as compared with schizophrenic males, and higher scores on the BIZ Scale for schizophrenic males than depressed males. They also found higher elevations on Scales 7, ANX, and DEP for depressed females as compared to schizophrenic females. Boone (1994) also reported significantly higher scores on the DEP Scale for depressed participants than for non-depressed participants (males and females combined).

There were considerable differences, however, when males and females were analyzed separately, with the test performing much better for females than for males. This may be due in part to changes made from the original MMPI to the MMPI-2. Early validation research on the MMPI-2 indicated generally lower T scores as compared to the MMPI. Therefore, the point at which scores were

considered elevated was lowered from 70 to 65 (Butcher et al., 1989). More recent MMPI-2 research has indicated that male profiles show more pronounced reduction in elevation than do female profiles, particularly for Scales 2, 4, 5, 8, and 9 (Edwards et al., 1993; Munley, 1991; Munley & Zarantonello, 1990). This may also account for findings of lower agreement rates between MMPI and MMPI-2 profiles for males (40.8%) than for females (43.8%) according to Dahlstrom (1992). Edwards et al. (1993) found a 50% concordance rate for males and 66% for females when gender groups were analyzed separately. This trend was also apparent in data collected on the restandardization sample, with males (psychiatric sample) obtaining a congruence rate of 80.1% and females 91.9% (Graham et al., 1991). This comparison was of MMPI items scored using the original MMPI norms versus MMPI-2 uniform T scores, and highlights the possible impact of the new normative population and the shift to using linear and uniform T scores. Chojnacki and Walsh (1992), however, reported higher agreement rates for males (43.1) than for females (36.4). Thus, given the present results for males versus females and past findings, it is clear that further research is needed to clarify the impact of these changes on MMPI-2 use with populations including both males and females.

In the current investigation, Scales 2, ANX, and LSE differentiated schizophrenic participants from depressed and BPD participants among females, but not among males. Also, Scale 4 differentiated BPD participants from depressed and schizophrenic participants for females, but not among males. In

comparing schizophrenic and BPD participants, Scales 7 and FAM were significantly different for females, but not for males. The ANG Scale differed significantly for schizophrenic versus depressed females, but not for males. Two scales performed better for males than for females. The FRS Scale significantly differentiated schizophrenic participants from depressed and BPD participants among males, but not among females. The BIZ Scale was significantly different for schizophrenic versus BPD males, but not for females. The clinical implications of these gender differences include better overall performance of the MMPI-2 in differentiating schizophrenia, depression, and BPD among females as compared to males. These data also suggest that the adjustments built into the MMPI-2 to accommodate gender differences may not be sufficient to allow for comparisons between genders.

The scales that were predicted to correctly classify a significant proportion of participants according to diagnostic category were moderately successful. Using the criteria listed in Hypothesis 3a, 80% of depressed participants were correctly classified (sensitivity rate = .80). However, the PPP was only .28, meaning that of those participants who were classified by the test as depressed, only 28% had a diagnosis of depression. This means that a clinician relying solely on these criteria would misdiagnose 72% of all persons appearing to be depressed on the test. When looking at negative diagnoses for depression, 43% of participants who were not depressed were correctly classified by the identified criteria (specificity rate = .43). On the other hand, given a negative indication of depression on the

MMPI-2, a clinician using the identified criteria would be correct 88% of the time (NPP = .88). When the depressed group was separated, with bipolar disorder (depressed) analyzed separately, and other depressed participants also analyzed separately, results were comparable to those reported above for all depressed participants combined. Thus, in the present investigation the MMPI-2 performed just as well (or as poorly) for classification of participants with major depression and dysthymia as compared to participants with bipolar disorder (depressed).

It can be concluded that the criteria identified for diagnosis of depression are quite useful for ruling out the disorder, but are less effective at ruling in depression. These results were fairly consistent with those reported by Elwood (1993) who obtained a PPP of .31 and NPP of .88 when using Scale 2 elevations to predict diagnosis of depression. However, the sensitivity rate of .42 and specificity rate of .70 reported by Elwood (1993) were almost the reverse of the current findings. Possible reasons for this include a lower base rate for depression (20%) in the Elwood (1993) investigation, and his inclusion of all invalid profiles submitted by depressed individuals as false-negatives.

In the current investigation, classification rates for schizophrenic participants were similar to those of depressed participants. The identified criteria correctly classified 74% of schizophrenic participants (sensitivity rate = .74). However, only 22% of the participants identified as schizophrenic by the test had a diagnosis of schizophrenia (PPP = .22). This mean that 78% of those people who were diagnosed by the identified criteria as schizophrenic were

incorrectly classified. Looking at negative diagnoses, 39% of participants without schizophrenia were correctly classified by the identified criteria (specificity rate = .39); and a negative test result was correct 86% of the time (NPP = .86). There were no previous investigations of MMPI-2 diagnosis of schizophrenia utilizing sensitivity, specificity, PPP and NPP.

The identified criteria performed even more poorly in classifying BPD participants. Only 40% of BPD participants were correctly classified by the identified criteria (sensitivity rate = .40), and only 15% of the participants identified as BPD had that diagnosis (PPP = .15). Negative diagnosis, however, was much better with 77% of participants without BPD being correctly classified by the identified criteria (specificity rate = .77), and 93% of negative test results being correct (NPP = .93). There were no previous investigations of MMPI-2 diagnosis of BPD utilizing sensitivity, specificity, PPP and NPP. The current results indicate that the identified criteria were too stringent, especially considering the low base rate of BPD in the population under investigation. This illustrates the need for researchers and clinicians to take base rates into account when setting cutoff levels and interpreting test results. This point has been made by other investigators as well (Baldessarini et al., 1983; Gerardi et al., 1989; Glaros & Kline, 1988; Meehl & Rosen, 1955; Moldin et al., 1991; Wideger et al., 1984).

The gender differences that were quite apparent in MANOVA and ANOVA results virtually disappeared under examination of sensitivity, specificity,

PPP and NPP. The MMPI-2 performed slightly better in classifying schizophrenic females as compared to males, as evidenced by a significant sensitivity rate of .85 for females versus a non-significant rate of .76 for males. These findings suggest that gender differences will have little impact on interpretation of the MMPI-2. Clearly of more importance is the generally poor performance of the MMPI-2 in correctly classifying individuals using the criteria identified here. The PPP rates obtained, ranging from .15 for BPD participants to .28 for Depressed participants are very discouraging to anyone who wishes to utilize the MMPI-2 to diagnose these disorders among clinical populations. These rates did not improve with the use of well defined scores.

Recommendations for Clinical Practice

Clinicians should use great caution when classifying people as depressed, schizophrenic or BPD based on MMPI-2 scores, especially when the base rate of the target diagnosis is low for the population being examined. Of course, clinicians are advised against ever rendering diagnoses based on data from only one diagnostic instrument. The current study shows greater promise for the MMPI-2 as a "rule-out" instrument, rather than a tool for assigning a diagnosis. This is based on statically significant specificity and NPP rates across all three diagnostic groups, as compared to non-significant sensitivity and PPP rates across all three groups. Therefore, clinicians may wish to rely on other diagnostic tests, or perhaps a structured interview format, for generating diagnostic hypotheses.

The MMPI-2 might then be relied upon to assist in ruling out erroneous diagnoses.

Based on the current study, there is little evidence to support the use of well defined MMPI-2 scores. Although some sensitivity, specificity, PPP and NPP rates were slightly higher when only well defined scores were included, the differences were generally neither clinically nor statistically significant. Only one cell, the sensitivity rate for schizophrenics, improved to a significant level as compared to the use of non well-defined scores.

The poor performance of the MMPI-2 in the current investigation to correctly classify individuals by diagnosis, as evidences by low sensitivity and PPP rates, does not mean that the MMPI-2 lacks clinical utility. Gynther, Altman and Sletten (1973) described a useful set of behavioral and clinical correlates for 20 commonly occurring two-point codes for the MMPI. These correlates were found to hold up well in replication studies, and provided useful clinical information. Similar studies investigating reliable behavioral correlates for the MMPI-2 could prove quite useful to clinicians who wish to utilize this instrument.

Limitations and Recommendations for Future Research

The current investigation had several limitations. The low number of non-white participants made comparisons of various ethnic groups impractical. Very few investigations of the effects of ethnic group membership on MMPI performance have been published, and no studies were found that investigated these effects on validity of MMPI-2 scores. Greene (1987) reviewed 53 studies of

the effects of ethnic group membership on MMPI performance. Of these, only 3 included empirical correlates of test scores. Greene (1987) found no simple relation between ethnic group membership and MMPI performance. He did state that the more rigorously moderator variables and profile validity were controlled, the less likely it was that differences between ethnic groups would be found. Thus, a further limitation of the current study is that the potential impact of moderator variables such as education, level of intelligence, and socio-economic status were not investigated.

Another limitation involved the fact that the MMPI-2 was given as part of the routine assessment procedure at a state hospital and, therefore, may have had some bearing on the diagnosis that was rendered. This may have resulted in artificially high agreement rates between MMPI-2 scores and diagnosis. Given the relatively poor agreement rates, as evidenced by sensitivity and PPP values, this limitation does not appear to warrant great concern. The poor performance of the MMPI-2 in this investigation may have been due, in part, to the nature of the population examined. Given the high rate of co-morbidity between and among Axis I and Axis II diagnoses in state hospital settings, detection of any specific construct of pathology might be more difficult than, for example, in an outpatient sample.

A third limitation concerned the number of participants in the BPD category. Having only 30 participants in this diagnostic group undoubtedly hampered statistical power and contributed to poor performance of the MMPI-2

in differentiating BPDs from other participants. Also, having only 9 males in the BPD group made comparisons of males and females within this group, as well as analysis of male BPDs alone, virtually pointless. Future MMPI-2 research with larger BPD samples is needed to validate its efficacy in classifying this disorder.

A fourth limitation of the current investigation involved relying solely on diagnoses rendered by staff psychiatrists as the criterion measure. Because no effort was made to establish the reliability and/or validity of these diagnoses, it is impossible to know whether the poor performance of the MMPI-2 in classifying individuals was due to limitations of the instrument or inaccurate diagnoses. Future studies should rely on a structured clinical interview format for at least part of the participants in order to determine the accuracy of diagnoses.

A final limitation involved investigating only three diagnostic groups among a relatively homogenous psychiatric population. Future research should utilize larger and more varied samples, and investigate various diagnostic groups. The current investigation does, however, provide evidence warranting re-evaluation of the criterion validity of the MMPI-2 rather than relying on the vast research base of MMPI literature.

Important areas for future MMPI-2 research include investigation of differences between males and females in regards to MMPI-2 criterion validity. Also, ethnic group membership should be more systematically evaluated as a moderator variable. Ethnicity studies should control for variables that might cloud the effects of ethnic group membership, including level of education,

intelligence, socio-economic status, and the degree to which participants identify with their ethnic group. These issues are discussed at length by Greene (1987).

Other important areas of MMPI-2 research include discriminant validity studies which examine several diagnostic constructs simultaneously. This type of investigation, employing the use of sensitivity, specificity, PPP and NPP, would closely approximate the task faced by clinicians. Therefore, results of such an investigation would have high clinical relevance.

The ideal follow-up study to the current investigation would employ a psychiatric inpatient population large enough to accomplish several goals. The first goal would be to obtain valid MMPI-2 profiles from a minimum of 200 participants (100 male, 100 female) for each major ethnic group (Asian, Black, Hispanic, Native American, and White) in 10 diagnostic groups (major depression, dysthymia, bipolar disorder, generalized anxiety disorder, schizophrenia, BPD, antisocial personality disorder, alcohol abuse disorder, poly-substance abuse, and post-traumatic stress disorder). These diagnoses would be rendered by a seasoned clinical psychologist and by an experienced psychiatrist, with only those participants who received matching diagnoses being included. MMPI-2 results would not be known prior to rendering diagnoses. The MMPI-2 scores would then be analyzed using the statistics employed in the current investigation. Level of education, intelligence and socio-economic status would be coded and entered prior to computation of multivariate statistics in order to control for these moderator variables. Thus, the utility of the MMPI-2 in correctly classifying

participants of both sexes and various ethnic group membership could be carefully examined in a highly systematic manner. Of course, this sort of study would require a minimum of 10,000 participants to be included, which would mean that a total N of 30,000 to 50,000 would be necessary. Because it is unlikely that such an enormous study will be conducted, validation of the MMPI-2 will probably continue in much the same manner as did investigation of the original MMPI. That is to say that numerous smaller, methodologically flawed, but still-meaningful studies will slowly accumulate and further clarify the relative strengths and limitations of this diagnostic instrument.

A reasonable follow-up of the current study would examine the same three diagnostic constructs in a larger and more heterogenous sample. Participants would be recruited equally from inpatient and outpatient populations. Diagnoses would be rendered by a licensed psychologist utilizing a structured clinical interview such as the SCID-P (Spitzer et al., 1988). A minimum of 30 male and 30 female Black, Hispanic and White participants in each diagnostic group would be included. The validity exclusion criteria and statistical procedures of the current study would be replicated. This would allow for a reasonable comparison of the diagnostic utility of the MMPI-2 among the various combinations of gender and ethnicity. Also, by including both inpatient and outpatient participants, the results would warrant greater generalization than do those of the current investigation.

# APPENDIX A

# VALIDITY AND CLINICAL SCALES OF THE MMPI-2

Validity and Clinical Scales of the MMPI-2

| Scale | Original Scale Name | Number of Items | Description |
|---|---|---|---|
| L | Lie | 15 | Measures deliberate attempts by the subject to present himself in a good light. |
| F | Fake bad | 60 | These items are rarely answered in the scored directions by normals (< 10%). Indicates random responding or deliberate attempts by the subject to present himself in a bad light. |
| K | Fake good | 30 | Indicates a general test-taking attitude of defensiveness about psychological weaknesses. The K-score is used as a correction to certain clinical scales (1,4,7,8,9) to improve their ability to discriminate normal from abnormal profiles. |
| 1 | Hypochondriasis | 32 | Reflects abnormal concern over bodily functions and preoccupation with physical complaints. |
| 2 | Depression | 57 | Reflects a pessimistic world view, feelings of hopelessness, and self-depreciation, possible considerations of suicide. |
| 3 | Hysteria | 60 | Measures the tendency to use physical or mental symptoms to avoid stressful conflicts. Often accompanied by an unwillingness to accept adult responsibilities. |

(table continued)

| Scale | Original Scale Name | Number of Items | Description |
|---|---|---|---|
| 4 | Psychopathic Deviate | 50 | Measures the tendency toward conflicts with authority figures, disregard of social conventions and laws, inability to learn from experience, and shallowness in personal attachments: the most frequently elevated scale among juvenile delinquent and criminal populations. |
| 5 | Masculinity-Femininity | 56 | Differentiates tendency toward traditional masculine or feminine interests, attitudes, and forms of self-expression. |
| 6 | Paranoia | 40 | Reflects abnormal suspiciousness and sensitivity, possible delusions of persecution or grandeur. |
| 7 | Psychasthenia | 48 | Measures the tendency toward obsessive ruminations, guilty feelings, anxiety, indecision and worrying, and compulsive ritualistic behavior. |
| 8 | Schizophrenia | 78 | Reflects bizarre or unusual thinking and behavior, interpersonal withdrawal and alienation, inappropriate affect, possible hallucinations or delusions. |
| 9 | Hypomania | 46 | Reflects high activity level often without productivity, emotional agitation, possible euphoria and flight of ideas. |
| 0 | Social Introversion | 69 | Reflects shyness, social withdrawal and insecurity, and disinterest in others. |

Note: Table adapted from Ben-Porath et al. (1991, pp. 639-640) and Megargee and Bohn (1979, pp. 77-78).

# APPENDIX B

## CONTENT SCALES OF THE MMPI-2

Content Scales of the MMPI-2

| Scale | Abbreviation | Number of Items | Description |
|---|---|---|---|
| Anxiety | ANX | 23 | Various psychological and somatic manifestations of anxiety |
| Fears | FRS | 23 | General tendency toward fearfulness and reporting of various specified fears |
| Obsessiveness | OBS | 16 | Obsessive-compulsive behaviors and cognitions |
| Depression | DEP | 33 | Various cognitive and behavioral symptoms of depression |
| Health Concerns | HEA | 36 | General health concerns and specific somatic complaints |
| Bizarre Mentation | BIZ | 23 | Psychotic symptomatology |
| Anger | ANG | 16 | Irritability and explosive behavior |
| Cynicism | CYN | 23 | Interpersonal distrust and misanthropic beliefs |
| Antisocial Practices | ASP | 22 | Various antisocial behaviors and attitudes |
| Type A | TPA | 19 | Type A personality behaviors and attitudes |
| Low Self-Esteem | LSE | 24 | Negative and demeaning self-view |
| Social Discomfort | SOD | 24 | Shyness and introversion |
| Family Problems | FAM | 25 | Tension and conflict in the family |
| Work Interference | WRK | 33 | Psychological problems that might interfere with adequate job performance |
| Negative Treatment Indicators | TRT | 26 | Beliefs, personality characteristics, and symptoms that may interfere with a client's treatment for psychological problems |

Note: Table adapted from Ben-Porath et al. (1991, pp. 639-640).

# APPENDIX C

## MEAN T SCORES, STANDARD DEVIATIONS, UNIVARIATE F RATIOS, AND SIGNIFICANCE FOR SCHIZOPHRENIA, DEPRESSED (EXCLUDING BIPOLAR), BPD, AND BIPOLAR (DEPRESSED) CATEGORIES

Mean T Scores, Standard Deviations, Univariate F Ratios, and Significance for Schizophrenia, Depressed (excluding Bipolar), BPD, and Bipolar (depressed) Categories

| | Diagnostic Category | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | SCH | | DPR[2] | | BPD | | BIPLR | | | |
| Scale | M | SD | M | SD | M | SD | M | SD | F (2, 157) | P |
| L | 58.98[a] | 11.88 | 58.00[a] | 15.16 | 53.03[a] | 13.08 | 55.85[a] | 10.96 | 11.24 | < .30 |
| F | 69.02[a] | 16.73 | 67.00[a,b] | 14.92 | 77.03[b] | 15.13 | 69.36[a,b] | 16.65 | 1.87 | .14 |
| K | 45.90[a] | 9.37 | 43.10[a] | 8.61 | 43.87[a] | 10.27 | 43.94[a] | 9.77 | 0.59 | .62 |
| 1 | 59.00[a] | 13.10 | 65.75[a] | 12.64 | 59.83[a] | 14.13 | 60.81[a] | 14.47 | 1.39 | .25 |
| 2 | 62.08[a] | 11.42 | 71.83[b] | 16.60 | 71.17[b] | 14.13 | 71.77[b] | 16.03 | 6.41 | < .01 |
| 3 | 53.29[a] | 13.54 | 61.21[b] | 16.59 | 59.67[b] | 15.56 | 60.26[b] | 16.33 | 4.00 | < .01 |
| 4 | 62.23[a] | 14.70 | 69.70[b,c] | 13.82 | 76.47[b] | 15.75 | 68.74[c] | 13.66 | 8.00 | < .01 |
| 5 | 52.94[a] | 10.39 | 53.55[a] | 11.06 | 52.23[a] | 11.62 | 51.81[a] | 8.73 | 0.19 | .90 |
| 6 | 67.92[a] | 18.14 | 62.80[a] | 17.74 | 69.50[a] | 14.31 | 66.49[a] | 15.21 | 0.90 | .44 |
| 7 | 62.69[a] | 16.56 | 66.95[a,b] | 12.27 | 71.57[b] | 14.53 | 68.26[a] | 14.50 | 2.43 | .07 |
| 8 | 70.32[a] | 16.38 | 67.45[a] | 12.87 | 73.70[a] | 15.41 | 68.98[a] | 14.50 | 0.58 | .63 |
| 9 | 59.42[a] | 12.62 | 53.95[a] | 12.78 | 58.37[a] | 11.96 | 56.21[a] | 11.83 | 1.28 | .28 |
| 0 | 57.82[a] | 11.09 | 60.65[a] | 13.42 | 60.43[a] | 12.36 | 59.96[a] | 12.90 | 0.35 | .79 |
| VRIN | 60.38[a] | 11.53 | 54.15[a] | 10.73 | 59.47[a] | 11.42 | 57.76[a] | 11.14 | 1.58 | .20 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| TRIN | 64.21[a] | 12.25 | 58.00[a] | 8.63 | 63.03[a] | 11.33 | 63.23[a] | 111.65 | 1.49 | .22 |
| ANX | 58.29[a] | 13.86 | 67.45[b] | 13.49 | 67.57[b] | 13.01 | 66.23[b] | 12.05 | 5.51 | < .01 |
| FRS | 60.89[a] | 13.70 | 53.55[b] | 12.15 | 54.67[b] | 12.08 | 53.43[b] | 11.70 | 3.65 | .01 |
| OBS | 57.84[a] | 13.50 | 58.10[a] | 13.30 | 59.60[a] | 13.15 | 59.70[a] | 12.39 | 0.24 | .87 |
| DEP | 62.00[a] | 13.04 | 74.30[b] | 14.24 | 75.83[b] | 12.46 | 72.30[b] | 12.64 | 10.27 | < .01 |
| HEA | 61.21[a] | 14.23 | 65.10[a] | 11.93 | 61.67[a] | 12.71 | 61.60[a] | 12.22 | 0.46 | .71 |
| BIZ | 67.10[a] | 14.20 | 55.60[b] | 11.74 | 60.07[b] | 10.15 | 57.62[b] | 11.97 | 6.66 | < .01 |
| ANG | 52.05[a] | 12.44 | 57.75[a,b] | 12.79 | 62.87[b] | 14.90 | 58.40[b] | 10.44 | 6.34 | < .01 |
| CYN | 60.53[a] | 9.77 | 61.65[a] | 11.06 | 57.70[a] | 12.83 | 60.17[a] | 12.79 | 0.37 | .77 |
| ASP | 55.84[a] | 8.60 | 57.00[a] | 11.63 | 61.43[a] | 15.04 | 56.81[a] | 11.79 | 2.17 | .09 |
| TPA | 53.13[a] | 10.82 | 52.05[a] | 10.89 | 56.50[a] | 12.80 | 53.42[a] | 10.45 | 1.00 | .40 |
| LSE | 59.92[a] | 13.84 | 65.10[a,b] | 15.29 | 68.33[b] | 14.82 | 66.02[b] | 15.41 | 2.40 | .07 |
| SOD | 55.63[a] | 10.62 | 58.60[a] | 13.29 | 59.20[a] | 15.18 | 58.87[a] | 13.29 | 0.49 | .69 |
| FAM | 60.34[a] | 13.50 | 63.75[a,b] | 14.61 | 69.60[b] | 12.97 | 65.94[b] | 13.64 | 3.09 | .03 |
| WRK | 60.97[a] | 13.48 | 67.10[a] | 13.56 | 65.40[a] | 15.21 | 66.00[a] | 13.75 | 1.80 | .15 |
| TRT | 62.94[a] | 13.94 | 67.45[a] | 15.66 | 68.37[a] | 15.75 | 67.38[a] | 15.49 | 1.24 | .30 |

Note. In each row, means with the same superscript do not differ at the .05 level or better using Duncan's multiple-range tests. SCH = schizophrenic category (n = 61); DPR[2] = depressed category, excluding bipolar, depressed (n = 20); BPD = borderline personality disorder category (n = 30); and BIPLR = bipolar, depressed category (n = 50).

# APPENDIX D

## FOR WHITE FEMALES: MEAN T SCORES, STANDARD DEVIATIONS, UNIVARIATE F RATIOS, AND SIGNIFICANCE BY DIAGNOSIS

For White Females: Mean T Scores, Standard Deviations, Univariate F Ratios, and Significance by Diagnosis

| | Diagnostic Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SCH | | DPR | | BPD | | | |
| Scale | M | SD | M | SD | M | SD | F (2, 75) | p |
| L | 63.71[a] | 13.83 | 56.75[a] | 10.61 | 53.67[a] | 13.53 | 3.72 | .03 |
| F | 73.00[a] | 17.75 | 70.53[a] | 15.70 | 78.52[a] | 14.92 | 1.65 | .20 |
| K | 46.95[a] | 10.79 | 42.72[a] | 9.26 | 44.10[a] | 10.02 | 1.21 | .30 |
| 1 | 61.24[a] | 13.84 | 64.53[a] | 14.97 | 61.00[a] | 15.70 | 0.51 | .60 |
| 2 | 61.91[a] | 13.58 | 74.97[b] | 17.25 | 71.67[b] | 14.95 | 4.64 | .01 |
| 3 | 54.38[a] | 14.93 | 62.75[a] | 17.07 | 60.05[a] | 16.21 | 1.75 | .18 |
| 4 | 60.52[a] | 14.46 | 68.89[b] | 15.47 | 77.81[c] | 15.66 | 6.74 | < .01 |
| 5 | 55.91[a] | 9.94 | 55.33[a] | 10.09 | 52.91[a] | 12.24 | 0.49 | .62 |
| 6 | 71.33[a] | 20.60 | 66.36[a] | 15.87 | 70.24[a] | 14.86 | 0.68 | .51 |
| 7 | 60.57[a] | 13.93 | 70.19[b] | 11.89 | 71.29[b] | 14.75 | 4.43 | .02 |
| 8 | 70.38[a] | 14.40 | 69.92[a] | 12.42 | 76.00[a] | 16.03 | 1.38 | .26 |
| 9 | 59.86[a] | 10.35 | 54.00[a] | 11.91 | 58.29[a] | 11.52 | 2.02 | .14 |
| 0 | 56.43[a] | 12.19 | 62.14[a] | 12.60 | 61.67[a] | 13.92 | 1.43 | .25 |
| VRIN | 62.76[a] | 11.65 | 54.81[a] | 10.74 | 59.48[a] | 11.36 | 3.56 | .03 |
| TRIN | 63.33[a] | 13.07 | 62.42[a] | 12.45 | 62.62[a] | 11.58 | 0.03 | .96 |
| ANX | 55.81[a] | 12.21 | 67.72[b] | 10.99 | 67.91[b] | 13.70 | 7.54 | < .01 |

(table continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| FRS | 57.43[a] | 11.38 | 54.56[a] | 11.06 | 56.29[a] | 11.23 | 0.46 | .63 |
| OBS | 55.52[a] | 11.28 | 60.28[a] | 12.47 | 60.10[a] | 13.77 | 1.08 | .35 |
| DEP | 60.52[a] | 11.95 | 73.00[b] | 12.71 | 75.52[b] | 12.79 | 9.05 | < .01 |
| HEA | 62.67[a] | 14.53 | 65.56[a] | 13.09 | 63.33[a] | 13.51 | 0.36 | .70 |
| BIZ | 65.49[a] | 13.43 | 55.17[b] | 10.31 | 62.19[a] | 10.24 | 6.24 | < .01 |
| ANG | 51.29[a] | 10.26 | 59.25[b] | 10.13 | 62.52[b] | 15.21 | 5.20 | < .01 |
| CYN | 61.57[a] | 10.71 | 60.53[a] | 11.85 | 56.95[a] | 12.56 | 0.92 | .40 |
| ASP | 57.00[a] | 8.97 | 56.81[a] | 11.40 | 60.05[a] | 12.49 | 0.63 | .54 |
| TPA | 54.19[a] | 9.75 | 54.36[a] | 9.75 | 57.14[a] | 13.44 | 0.53 | .59 |
| LSE | 58.24[a] | 13.02 | 68.94[b] | 14.61 | 69.29[b] | 15.57 | 4.30 | .02 |
| SOD | 56.81[a] | 11.81 | 58.83[a] | 13.51 | 59.67[a] | 16.48 | 0.24 | .79 |
| FAM | 56.52[a] | 11.00 | 65.50[b] | 12.66 | 71.71[b] | 13.39 | 7.94 | < .01 |
| WRK | 58.33[a] | 12.84 | 68.31[a] | 13.59 | 65.38[a] | 15.83 | 3.37 | .04 |
| TRT | 63.19[a] | 13.47 | 68.39[a] | 15.47 | 68.86[a] | 18.08 | 0.90 | .41 |

Note. In each row, means with the same superscript do not differ at the .05 level or better using Duncan's multiple-range tests. SCH = schizophrenic category (n = 21), DPR = depressed category (n = 36), and BPD = borderline personality disorder category (n = 21).

# APPENDIX E

## FOR WHITE MALES: MEAN T SCORES, STANDARD DEVIATIONS, UNIVARIATE F RATIOS, AND SIGNIFICANCE BY DIAGNOSIS

For White Males: Mean T Scores, Standard Deviations, Univariate F Ratios, and Significance by Diagnosis

| | Diagnostic Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SCH | | DPR | | BPD | | | |
| Scale | M | SD | M | SD | M | SD | F (2, 47) | p |
| L | 55.40[a] | 8.62 | 54.60[a] | 9.13 | 54.60[a] | 15.60 | 0.04 | .96 |
| F | 66.04[a] | 17.67 | 63.70[a] | 15.65 | 80.60[a] | 16.07 | 2.06 | .14 |
| K | 45.92[a] | 8.70 | 45.30[a] | 9.06 | 47.40[a] | 13.61 | 0.10 | .90 |
| 1 | 57.20[a] | 13.61 | 58.80[a] | 14.23 | 60.60[a] | 7.20 | 0.17 | .85 |
| 2 | 61.20[a] | 11.03 | 69.60[b] | 16.08 | 75.00[b] | 6.86 | 3.64 | .03 |
| 3 | 54.32[a] | 13.68 | 61.45[a,b] | 16.43 | 69.40[b] | 5.77 | 8.90 | .07 |
| 4 | 63.04[a] | 15.72 | 71.70[b] | 11.77 | 82.60[b] | 12.70 | 4.96 | .01 |
| 5 | 50.64[a] | 10.59 | 51.80[a] | 9.47 | 52.80[a] | 9.96 | 0.13 | .88 |
| 6 | 65.12[a] | 16.57 | 62.90[a] | 14.43 | 71.80[a] | 10.92 | 0.68 | .51 |
| 7 | 63.88[a] | 19.22 | 64.60[a] | 16.19 | 78.80[a] | 7.76 | 1.61 | .21 |
| 8 | 70.16[a] | 19.53 | 65.75[a] | 16.48 | 75.00[a] | 7.84 | 0.69 | .51 |
| 9 | 61.00[a] | 14.58 | 55.65[a] | 13.33 | 61.20[a] | 16.80 | 0.85 | .43 |
| 0 | 57.92[a] | 10.74 | 55.15[a] | 11.30 | 55.40[a] | 3.65 | 0.42 | .66 |
| VRIN | 57.12[a] | 12.37 | 58.35[a] | 12.16 | 55.20[a] | 10.73 | 0.15 | .86 |
| TRIN | 64.80[a] | 12.74 | 61.35[a] | 11.29 | 63.00[a] | 13.42 | 0.44 | .65 |
| ANX | 58.92[a] | 15.42 | 64.75[a] | 15.10 | 72.60[a] | 11.91 | 2.06 | .14 |

(table continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| FRS | 59.68[a] | 16.16 | 50.00[b] | 11.90 | 46.40[a,b] | 13.67 | 3.40 | .04 |
| OBS | 59.68[a] | 15.24 | 59.55[a] | 11.96 | 57.20[a] | 13.85 | 0.07 | .93 |
| DEP | 63.04[a] | 14.90 | 73.50[b] | 12.72 | 82.20[b] | 11.39 | 5.70 | < .01 |
| HEA | 60.00[a] | 14.62 | 58.05[a] | 12.29 | 57.80[a] | 13.42 | 0.14 | .87 |
| BIZ | 67.84[a] | 15.18 | 52.75[b] | 9.56 | 53.60[b] | 11.06 | 8.44 | < .01 |
| ANG | 52.56[a] | 13.69 | 58.85[a] | 11.68 | 60.00[a] | 18.45 | 1.49 | .24 |
| CYN | 59.12[a] | 9.81 | 57.05[a] | 11.45 | 54.20[a] | 11.79 | 0.50 | .61 |
| ASP | 54.24[a] | 7.62 | 55.15[a] | 11.59 | 64.80[a] | 25.15 | 1.72 | .19 |
| TPA | 50.72[a] | 10.54 | 51.90[a] | 11.97 | 52.80[a] | 11.10 | 0.11 | .90 |
| LSE | 62.40[a] | 16.57 | 62.80[a] | 14.07 | 73.20[a] | 9.60 | 1.12 | .34 |
| SOD | 54.76[a] | 10.45 | 52.45[a] | 12.00 | 51.00[a] | 10.08 | 0.38 | .69 |
| FAM | 62.44[a] | 14.66 | 63.55[a] | 14.58 | 65.20[a] | 14.50 | 0.09 | .92 |
| WRK | 61.60[a] | 15.36 | 65.10[a] | 13.51 | 70.60[a] | 14.72 | 0.91 | .41 |
| TRT | 62.64[a] | 16.16 | 65.25[a] | 13.98 | 66.80[a] | 11.39 | 0.26 | .77 |

Note. In each row, means with the same superscript do not differ at the .05 level or better using Duncan's multiple-range tests. SCH = schizophrenic category (n = 25), DPR = depressed category (n = 20), and BPD = borderline personality disorder category (n = 5).

# APPENDIX F

## FOR ALL FEMALES: MEAN T SCORES, STANDARD DEVIATIONS, UNIVARIATE F RATIOS, AND SIGNIFICANCE BY DIAGNOSIS

For All Females: Mean T Scores, Standard Deviations, Univariate F Ratios, and Significance by Diagnosis

| | Diagnostic Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SCH | | DPR | | BPD | | | |
| Scale | M | SD | M | SD | M | SD | F (2, 86) | p |
| L | 61.68[a] | 13.80 | 56.32[a,b] | 10.30 | 53.35[b] | 13.11 | 2.98 | .06 |
| F | 73.36[a] | 16.38 | 71.15[a] | 15.61 | 77.78[a] | 14.88 | 1.33 | .27 |
| K | 46.28[a] | 10.38 | 42.44[a] | 8.84 | 43.74[a] | 9.67 | 1.27 | .29 |
| 1 | 60.52[a] | 13.68 | 64.24[a] | 14.43 | 60.57[a] | 15.54 | 0.72 | .49 |
| 2 | 62.00[a] | 13.01 | 75.00[b] | 16.62 | 71.35[b] | 14.99 | 5.69 | < .01 |
| 3 | 53.72[a] | 14.59 | 62.51[b] | 16.65 | 59.04[a,b] | 16.25 | 2.35 | .10 |
| 4 | 61.40[a] | 13.97 | 68.81[b] | 14.69 | 77.22[c] | 15.25 | 7.00 | < .01 |
| 5 | 56.48[a] | 10.69 | 54.29[a] | 10.17 | 52.74[a] | 12.15 | 0.73 | .49 |
| 6 | 71.08[a] | 20.17 | 67.66[a] | 15.89 | 70.74[a] | 14.28 | 0.42 | .66 |
| 7 | 61.52[a] | 13.31 | 70.02[b] | 11.28 | 71.49[b] | 14.57 | 4.59 | .01 |
| 8 | 70.68[a] | 13.38 | 69.95[a] | 12.08 | 75.17[a] | 16.13 | 1.16 | .32 |
| 9 | 59.48[a] | 10.41 | 54.59[a] | 11.89 | 57.30[a] | 11.47 | 1.49 | .23 |
| 0 | 57.44[a] | 11.80 | 61.76[a] | 12.34 | 62.26[a] | 13.49 | 1.18 | .31 |
| VRIN | 62.88[a] | 11.30 | 56.22[a] | 11.11 | 60.91[a] | 11.84 | 2.99 | .06 |
| TRIN | 63.84[a] | 13.20 | 62.66[a] | 11.85 | 63.57[a] | 11.52 | 0.09 | .92 |
| ANX | 55.96[a] | 12.07 | 68.29[b] | 10.57 | 67.17[b] | 13.43 | 9.29 | < .01 |

(table continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| FRS | 58.24[a] | 11.27 | 54.66[a] | 11.24 | 56.52[a] | 11.77 | 0.79 | .46 |
| OBS | 55.56[a] | 11.68 | 60.10[a] | 11.97 | 59.96[a] | 13.39 | 1.20 | .31 |
| DEP | 61.20[a] | 11.85 | 73.17[b] | 12.16 | 75.30[b] | 12.59 | 10.05 | < .01 |
| HEA | 61.72[a] | 14.33 | 65.15[a] | 12.49 | 62.91[a] | 13.15 | 0.57 | .51 |
| BIZ | 61.01[a] | 13.92 | 56.17[b] | 10.89 | 61.48[a,b] | 10.06 | 5.75 | < .01 |
| ANG | 52.08[a] | 11.10 | 59.22[b] | 9.77 | 62.65[b] | 14.56 | 5.39 | < .01 |
| CYN | 61.48[a] | 10.41 | 61.02[a] | 11.60 | 57.17[a] | 12.00 | 1.08 | .35 |
| ASP | 57.32[a] | 9.79 | 56.93[a] | 10.87 | 59.30[a] | 12.24 | 0.48 | .62 |
| TPA | 54.44[a] | 11.10 | 53.80[a] | 9.42 | 56.57[a] | 13.06 | 0.48 | .62 |
| LSE | 57.92[a] | 12.28 | 67.68[b] | 14.34 | 69.04[b] | 14.99 | 4.91 | .01 |
| SOD | 56.72[a] | 11.30 | 59.17[a] | 13.05 | 61.00[a] | 16.32 | 0.61 | .55 |
| FAM | 58.64[a] | 11.75 | 65.27[b] | 12.46 | 71.13[b] | 13.14 | 6.06 | < .01 |
| WRK | 58.80[a] | 12.57 | 67.61[b] | 13.20 | 65.13[a,b] | 15.44 | 3.27 | .04 |
| TRT | 63.32[a] | 13.59 | 68.34[a] | 15.02 | 68.65[a] | 17.40 | 1.02 | .37 |

Note. In each row, means with the same superscript do not differ at the .05 level or better using Duncan's multiple-range tests. SCH = schizophrenic category (n = 25), DPR = depressed category (n = 41) and BPD = borderline personality disorder category (n = 23).

# APPENDIX G

## FOR ALL MALES: MEAN T SCORES, STANDARD DEVIATIONS, UNIVARIATE F RATIOS, AND SIGNIFICANCE BY DIAGNOSIS

For All Males: Mean T Scores, Standard Deviations, Univariate F Ratios, and Significance by Diagnosis

| | Diagnostic Category | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SCH | | DPR | | BPD | | | |
| Scale | M | SD | M | SD | M | SD | F (2, 70) | p |
| L | 57.16[a] | 10.19 | 56.83[a] | 13.49 | 52.00[a] | 14.00 | 0.56 | .57 |
| F | 66.08[a] | 16.54 | 63.41[a] | 15.77 | 74.57[a] | 16.88 | 1.34 | .27 |
| K | 45.65[a] | 8.76 | 45.00[a] | 9.68 | 44.29[a] | 12.88 | 0.08 | .93 |
| 1 | 57.97[a] | 12.78 | 58.86[a] | 13.82 | 57.43[a] | 8.40 | 0.06 | .95 |
| 2 | 62.14[a] | 10.39 | 67.35[a] | 15.78 | 70.57[a] | 11.86 | 2.04 | .14 |
| 3 | 53.00[a] | 12.99 | 59.38[a] | 16.62 | 61.71[a] | 13.95 | 2.06 | .14 |
| 4 | 62.78[a] | 15.33 | 70.97[b] | 12.64 | 74.00[a,b] | 18.37 | 3.41 | .04 |
| 5 | 50.54[a] | 9.60 | 50.55[a] | 9.58 | 50.57[a] | 10.37 | 0.00 | 1.00 |
| 6 | 65.78[a] | 16.57 | 62.76[a] | 15.45 | 65.43[a] | 14.72 | 0.30 | .74 |
| 7 | 63.49[a] | 18.57 | 63.62[a] | 16.03 | 71.86[a] | 15.54 | 0.73 | .49 |
| 8 | 70.08[a] | 18.32 | 64.86[a] | 15.75 | 68.86[a] | 12.56 | 0.79 | .46 |
| 9 | 59.38[a] | 14.05 | 57.41[a] | 12.74 | 61.86[a] | 13.80 | 0.37 | .70 |
| 0 | 58.08[a] | 10.74 | 55.72[a] | 11.95 | 54.43[a] | 3.82 | 0.57 | .57 |
| VRIN | 58.70[a] | 11.53 | 57.28[a] | 11.02 | 54.71[a] | 9.03 | 0.42 | .66 |
| TRIN | 64.46[a] | 11.75 | 61.48[a] | 10.43 | 61.29[a] | 11.34 | 0.66 | .52 |
| ANX | 59.87[a] | 14.91 | 63.72[a] | 14.43 | 68.86[a] | 12.42 | 1.37 | .26 |

(table continued)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| FRS | 62.68[a] | 15.00 | 51.00[b] | 12.61 | 48.57[b] | 11.89 | 7.12 | < .01 |
| OBS | 59.38[a] | 14.56 | 58.28[a] | 13.72 | 58.43[a] | 13.28 | 0.05 | .95 |
| DEP | 62.54[a] | 13.92 | 71.28[b] | 13.74 | 77.57[b] | 12.86 | 5.37 | < .01 |
| HEA | 60.87[a] | 14.35 | 59.24[a] | 11.50 | 57.57[a] | 10.98 | 0.25 | .78 |
| BIZ | 67.81[a] | 14.54 | 57.48[b] | 13.30 | 55.43[b] | 9.71 | 5.66 | < .01 |
| ANG | 52.03[a] | 13.42 | 57.17[a,b] | 12.22 | 63.57[b] | 17.16 | 2.75 | .07 |
| CYN | 59.89[a] | 9.40 | 61.00[a] | 13.14 | 59.43[a] | 16.20 | 0.09 | .91 |
| ASP | 54.84[a] | 7.67 | 57.14[b] | 12.79 | 68.43[b] | 21.69 | 4.00 | .02 |
| TPA | 52.24[a] | 10.68 | 52.00[a] | 11.95 | 56.29[a] | 12.87 | 0.42 | .66 |
| LSE | 61.27[a] | 14.81 | 62.10[a] | 15.34 | 66.00[a] | 15.13 | 0.29 | .75 |
| SOD | 54.89[a] | 10.23 | 53.79[a] | 12.94 | 53.29[a] | 9.12 | 0.11 | .90 |
| FAM | 61.49[a] | 14.61 | 64.07[a] | 15.71 | 64.57[a] | 11.89 | 0.30 | .74 |
| WRK | 62.43[a] | 14.04 | 63.76[a] | 14.31 | 66.29[a] | 15.59 | 0.24 | .79 |
| TRT | 62.68[a] | 14.35 | 64.62[a] | 15.66 | 67.43[a] | 9.36 | 0.31 | .69 |

Note. In each row, means with the same superscript do not differ at the .05 level or better using Duncan's multiple-range tests. SCH = schizophrenic category (n = 37), DPR = depressed category (n = 29), and BPD = borderline personality disorder category (n = 7).

## REFERENCES


Adler, T. (1990, April). Does the 'new' MMPI beat the 'classic'? APA Monitor, pp.18-19.

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

Bagby, R.M., Rogers, R., & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. Journal of Personality Assessment, 62, 191-203.

Baldessarini, R.J., Finklestein, S., & Arana, G.W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. Archives of General Psychiatry, 40, 569-573.

Beck, A. (1987). Beck Depression Inventory. San Antonio, TX: Psychological Corp.

Beck, A. (1988). Beck Hopelessness Scale. San Antonio, TX: Psychological Corp.

Ben-Porath, Y.S., & Butcher, J.N. (1989). Psychometric stability of rewritten MMPI items. Journal of Personality Assessment, 53, 645-653.

Ben-Porath, Y.S., Butcher, J.N., & Graham, J.R. (1991). Contribution of the MMPI-2 content scales to the differential diagnosis of schizophrenia and

major depression. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3, 634-640.

Ben-Porath, Y.S., McCully, E., & Almagor, M. (1993). Incremental validity of the MMPI-2 content scales in the assessment of personality and psychopathology by self-report. Journal of Personality Assessment, 61, 557-575.

Berry, D.T., Wetter, M.W., Baer, R.A., Widiger, T.A., Sumpter, J.C., Reynolds, S.K., & Hallam, R.A. Detection of random responding on the MMPI-2: Utility of F, Back F, and VRIN Scales. Psychological Assessment, 3, 418-423.

Boone, D.E. (1994). Validity of the MMPI-2 depression content scale with psychiatric inpatients. Psychological Reports, 74, 159-162.

Brems, C. (1991). Depression and personality disorder: Differential diagnosis with the MMPI. Journal of Clinical Psychology, 47, 669-675.

Butcher, J.N. (1990). MMPI-2 in psychological treatment. New York: Oxford University Press.

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

Butcher, J.M., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1990, July). Tangible improvements [Letter to the editor]. APA Monitor, p.2.

Butcher, J.N., & Tellegen, A. (1978). Common methodological problems in MMPI research. Journal of Consulting and Clinical Psychology, 46, 620-628.

Butcher, J.N., & Williams, C.L. (1992). Essentials of MMPI-2 and MMPI-A interpretation. Minneapolis: University of Minnesota Press.

Chojnacki, J.T., & Walsh, W.B. (1992). The consistency of scores and configural patterns between the MMPI and MMPI-2. Journal of Personality Assessment, 59,267-289.

Cohen, R.J., Montague, P., Nathanson, L.S., & Swerdlik, M.E. (1988). Psychological Testing: An Introduction to Tests & Measurement. Mountain View, CA: Mayfield Publishing Company.

Cull, J.G., & Gill, W.S. (1983). Suicide Probability Scale. Los Angeles, CA: Western Psychological Services.

Dahlstrom, W.G. (1992). Comparability of two-point high-point code patterns from original MMPI norms to MMPI-2 norms for the restandardization sample. Journal of Personality Assessment, 59, 153-164.

Derogatis, L.R. (1983). SCL-90-R: Administration, scoring, and procedures manual-II. Townson, MD: Clinical Psychometric Research.

Edwards, D.W., Morrison, T.L., & Weissman, H.N. (1993). The MMPI and MMPI-2 in an outpatient sample: Comparisons of code types, validity scales, and clinical scales. Journal of Personality Assessment, 61, 1-18.

Edwards, D.W., & Weissman, H.N. (1990, July). Shared concerns [Letter to the editor]. APA Monitor, p.2.

Elwood, R.W. (1993). The Clinical utility of the MMPI-2 in diagnosing unipolar depression among male alcoholics. Journal of Personality Assessment, 60, 511-521.

Evans, R.W., Ruff, M.R., Braff, D.L., & Ainsworth, T.L. (1984). MMPI characteristics of borderline personality inpatients. Journal of Nervous and Mental Disease, 172, 742-748.

Gartner, J., Hurt, S.W., & Gartner, A. (1989). Psychological test signs of borderline personality disorder: A review of the empirical literature. Journal of Personality Assessment, 53, 423-441.

Gerardi, R., Keane, T.M., & Penk, W. (1989). Sensitivity and specificity in developing tests of combat-related posttraumatic stress disorder (PTSD). Journal of Clinical Psychology, 45, 691-703.

Glaros, A.G., & Kline, R.B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. Journal of Clinical Psychology, 44, 1013-1023.

Gottesman, I.I., & Prescott, C.A. (1989). Abusers of the MacAndrew MMPI Alcoholism Scale: A critical review. Clinical Psychology Review, 9, 661-669.

Graham, J.R. (1987). The MMPI: A practical guide (2nd. ed.). New York: Oxford University Press.

Graham, J.R. (1990). MMPI-2: Assessing personality and psychopathology. New York: Oxford University Press.

Graham, J.R., Timbrook, R.E., Ben-Porath, T.S., & Butcher, J.N. (1991). Code-type congruence between MMPI and MMPI-2: Separating fact from artifact. Journal of Personality Assessment, 57, 205-215.

Greene, R.L. (1980). The MMPI: An interpretive manual. New York: Grune & Stratton.Greene, R.L. (1987). Ethnicity and MMPI performance: A review. Journal of Consulting and Clinical Psychology, 55, 497-512.

Greene, R.L. (1991). The MMPI-2/MMPI: An interpretive manual. Boston, MA: Allyn & Bacon.

Gynther, M.D., Altman, H., & Sletten, I.F. (1973). Replicated correlates of MMPI two-point code types: The Missouri Actuarial System. Journal of Clinical Psychology, 29, 263-289.

Hathaway, S.R., & McKinley, J.C. (1942). The Minnesota Multiphasic Personality Inventory. Minneapolis, MN: University of Minnesota Press.

Helmes, E., & Reddon, J.R. (1993). A perspective on developments in assessing psychopathology: A critical review of the MMPI and MMPI-2. Psychological Bulletin, 113, 453-471.

Keane, T.M., Caddell, J.M., & Taylor, K.L. (1989). Mississippi scale for combat-related Posttraumatic Stress Disorder: Three studies in reliability and validity. Journal of Consulting and Clinical Psychology, 56, 85-90.

Lachar, D., Dahlstrom, W.G., & Moreland, K.L. (1986). Relationship of ethnic background and other demographic characteristics to MMPI patterns in psychiatric samples. In W.G Dahlstrom, D. Lachar, & L.E. Dahlstrom, MMPI

patterns in American minorities (pp. 139-157). Minneapolis: University of Minnesota Press.

Levitt, E.E. (1990). A structural analysis of the impact of MMPI-2 on the MMPI. Journal of Personality Assessment, 55, 562-577.

Loranger, A.W. (1990). The impact of DSM-III on diagnostic practice in a university hospital. Archives of General Psychiatry, 47, 672-675.

Marks, P.A., Seeman, W., & Haller, D.L. (1974). The actuarial use of the MMPI with adolescents and adults. New York: Oxford University Press.

Meehl, P.E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 52, 194-216.

Megargee, E., & Bohn, M., Jr. (1979). Classifying criminal offenders. Beverly Hills: Sage

Moldin, S.O., Gottesman, I.I., Rice, J.P., & Erlenmeyer-Kimling, L. (1991). Replicated psychometric correlates of schizophrenia. American Journal of Psychiatry, 148, 762-767.

Morrison, T.L., Edwards, D.W., & Weissman, H.N. (1994). The MMPI and MMPI-2 as predictors of psychiatric diagnosis in an outpatient sample. Journal of Personality Assessment, 62, 17-30.

Munley, P.H. (1991). A comparison of MMPI-2 and MMPI T-scores for men and women. Journal of Clinical Psychology, 47, 87-91.

Munley, P.H., & Zarantonello, M.M. (1990). A comparison of MMPI profile types with corresponding estimated MMPI-2 profiles. Journal of Clinical Psychology, 46, 803-811.

Nichols, D.S. (1988). Mood disorders. In R.L. Greene (Ed.), The MMPI: Use with specific populations (pp. 75-109). Philadelphia, PA: Grune & Stratton.

Olin, J.T., Schneider, L.S., Eaton, E.M., Zemansky, J.F., & Pollack, V.E. (1992). The Geriatric Depression scale and the Beck Depression Inventory as screening instruments in an older adult outpatient population. Psychological Assessment, 4, 190-192.

Rapp, S.R., Parisi, S.A., Walsh, D.A., & Wallace, C.E. (1988). Detecting depression in elderly medical inpatients. Journal of Consulting and Clinical Psychology, 56, 509-513.

Rogers, R., Bagby, R.M., & Chakraborty, D. (1993). Faking schizophrenic disorders on the MMPI-2: Detection of coached simulators. Journal of Personality Assessment, 60, 215-226.

Rogers, R., Sewell, K.W., & Salekin, R.T. (1994). A meta-analysis of malingering on the MMPI-2. Assessment, 1, 227-237.

Spielberger, C.D. (1987). State Trait Anger Expression Inventory Manual. Odessa, FL: Psychological Assessment Resources.

Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. (1979). Manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologist's Press.

Spitzer, R.L., Williams, J.B.W., Gibbon, M., & First, M.B. (1988). Structured Clinical Interview for DSM-III-R-Patient Version (SCID-P). New York: Biometrics Research, New York State Psychiatric Institute.

Tellegen, A., & Ben-Porath, Y.S. (1992). The new uniform T scores for the MMPI-2: Rationale, derivation, and appraisal. Psychological Assessment, 4, 145-155.

Tellegen, A., & Ben-Porath, Y.S. (1993). Code-type comparability of the MMPI and MMPI-2: Analysis of recent findings and criticisms. Journal of Personality Assessment, 61, 489-500.

Vincent, K.R., Castillo, I., Hauser, R.I., Stuart. H.J., Zapata, J.A., Cohn, C.K., & O'Shanick, G.J. (1983). MMPI code types and DSM-III diagnosis. Journal of Clinical Psychology, 83, 829-842.

Walters, G.D. (1984). Identifying schizophrenia by means of Scale 8 Sc of the MMPI. Journal of Personality Assessment, 48, 390-391.

Walters, G.D., & Greene, R.L. (1988). Differentiating between schizophrenic and manic inpatients by means of the MMPI. Journal of Personality Assessment, 52, 91-95.

Weed, N.C., Butcher, J.N., McKenna, T., & Ben-Porath Y.S. (1992). New measures for assessing alcohol and drug abuse with the MMPI-2: The APS and AAS. Journal of Personality Assessment, 58, 389-404.

Wetter, M.W., Baer, R.A., Berry, D.T.R., Smith, G.T., & Larsen, L.H. (1992). Sensitivity of MMPI-2 validity scales to random responding and malingering. Psychological Assessment, 4, 369-374.

Widiger, T.A., Hurt, S.W., Frances, A., Clarkin, J.F., & Gilmore, M. (1984). Diagnostic efficiency and DSM-III. Archives of General Psychiatry, 41, 1005-1012.

Wiggins, J.S. (1966). Substantive dimensions of self-report in the MMPI item pool. Psychological Monographs, 80, (22, Whole No. 630).

Zalewski, C.E., & Archer, R.P. (1991). Assessment of borderline personality disorder: A review of MMPI and rorschach findings. The Journal of Nervous and Mental Disease, 179, 338-345.

Zalewski, C.E., & Gottsman, I.I. (1991). (Hu)man versus mean revisited: MMPI group data and psychiatric diagnosis. Journal of Abnormal Psychology, 100, 562-568.