

DIAGNOSTIC VALIDITY OF MMPI-2 COMPUTER BASED TEST INTERPRETATIONS

Hina Pant

A dissertation submitted in partial fulfillment of  
the requirements for the degree of  
Doctor of Philosophy

Department of Psychology

Central Michigan University  
Mount Pleasant, Michigan  
December, 2011

Accepted by the Faculty of the College of Graduate Studies,  
Central Michigan University, in partial fulfillment of  
the requirements for the doctoral degree

Dissertation Committee:

Nathan Weed, Ph.D.	Committee Chair
Kyunghee Han, Ph.D.	Faculty Member
George Ronan, Ph.D.	Faculty Member
Mark Deskovitz, Ph.D.	Faculty Member
December 15, 2011	Date of Defense
Roger Coles, Ed.D.	Dean College of Graduate Studies
March 19, 2012	Approved by the College of Graduate Studies

Copyright by  
Hina Pant  
2011

## ACKNOWLEDGMENTS

I thank Dr. Nathan Weed for his assistance during planning and implementation of this project. In his role as a graduate advisor, Dr. Weed was instrumental in fostering my growth as a clinical psychologist, balancing giving advice and suggestions with playing Socrates to enhance my critical thinking skills. His humor often made lab meetings on even the driest statistical procedures seem like a Saturday Night Live show and I am grateful for the respite it provided from the rigors of graduate school training.

I am grateful to Dr. Mark Deskovitz for his help with procuring the archival data used in this dissertation. I also thank members of the dissertation committee, Dr. Kyunghye Han and Dr. George Ronan for their advice and suggestions. I am grateful to the University of Minnesota Press for providing Pearson CBTI interpretive reports for this study. I am also thankful to Dr. Roger Greene and Psychological Assessment Resources for providing the PAR CBTI program used to generate interpretive reports for this study.

Finally, I am grateful to my parents for giving me both the genes *and* the environment needed to achieve my goals.

## ABSTRACT

### DIAGNOSTIC VALIDITY OF MMPI-2 COMPUTER BASED TEST INTERPRETATIONS

by Hina Pant

Diagnostic classification is a common part of psychological assessment. The Minnesota Multiphasic Personality Inventory (MMPI) was initially developed as a diagnostic instrument to aid in assigning psychiatric diagnoses to individuals. However, due to large item overlap between different MMPI scales, it was not found to be optimum for that purpose (Graham, 2006). However, the MMPI and now its revision, the MMPI-2, continue to be used to guide diagnostic formulations by examining patterns of scale scores or codetypes. Most commercially available computer based test interpretations (CBTIs) for the MMPI-2 also offer diagnostic suggestions. However, data on the validity of these diagnostic suggestions is lacking.

This study aimed to answer three fundamental questions with respect to the diagnostic validity of CBTIs. Firstly, this study questioned whether the diagnostic suggestions offered by the CBTIs are valid with respect to descriptive information provided by treating clinicians. Archival MMPI-2 scores were obtained for 65 clients at an outpatient mental health center in the Midwestern United States. Diagnostic Q-sorts completed by using information provided by treating clinicians and information provided by CBTI reports were compared to examine the level of similarity between diagnostic judgments derived from these two sources. Results indicated that the overall mean Q-correlation between criterion diagnostic sorts based on therapist ratings and those based on CBTI reports was .29.

Secondly, this study questioned whether specific diagnostic categories were associated with higher diagnostic validity coefficients, indicating greater CBTI agreement with information provided by treating clinicians. This was examined by comparing mean diagnostic validity

coefficients across different psychodiagnostic categories. Results indicated that profiles receiving highest-ranked diagnoses of Histrionic Personality Disorder and Generalized Anxiety Disorder had a higher representation among a subset of profiles, for which membership was based on high criterion sort reliability values (greater than .69).

Thirdly, this study examined how CBTIs compare with each other with respect to diagnostic validity. The diagnostic validity of these CBTIs was compared by examining validity coefficients obtained for each CBTI after being averaged across all the cases. The CBTI by PAR had the lowest overall correlation (.23) with the criterion sorts and the CBTI by PS had the highest overall correlation (.28) with the criterion sorts. The overall diagnostic reliability across CBTI programs was found to be high ( $\alpha = .89$ ), indicating that, in general, CBTI programs provide similar diagnostic information.

Results suggest that, at a minimum, CBTIs provide moderately valid information with respect to diagnostic suggestions, as compared to diagnostic information based on descriptive statements provided by treating clinicians. It is likely that validity values obtained in this study were attenuated due to lower reliability of diagnostic information used as the criterion.

Limitations to this study along with future directions are discussed.

## TABLE OF CONTENTS

LIST OF TABLES.....		viii
CHAPTER		
I.	INTRODUCTION.....	1
	Diagnosis and the MMPI.....	1
	Uses and Advantages of CBTIs.....	4
	<i>Cautions Regarding the Use of CBTIs</i> .....	5
	Research on CBTIs.....	7
	Research on CBTIs for the MMPI-2.....	8
	<i>Reliability Research</i> .....	8
	<i>Validity Research</i> .....	11
	<i>Customer Satisfaction Studies</i> .....	11
	<i>External Criterion Studies</i> .....	15
	<i>Review of CBTIs</i> .....	19
	The Q-sort Method.....	20
	Present Study.....	22
II.	METHOD.....	24
	Participants.....	24
	Raters.....	25
	Clinicians.....	25
	Instruments.....	26
	<i>MMPI-2</i> .....	26
	<i>CBTIs</i> .....	26
	<i>Diagnostic Q-sort</i> .....	26
	<i>Midwestern Q-sort</i> .....	27
	Procedures.....	30
	Analysis.....	32
III.	RESULTS.....	33
	Data Screening.....	33
	Diagnostic validity.....	40
IV.	DISCUSSION.....	51
	Reliability of criterion Q-sorts.....	51
	Diagnostic validity.....	52
	Limitations.....	59
	Future Directions.....	60
	REFERENCES.....	77

LIST OF TABLES

TABLE	PAGE
1. Mean K-Corrected MMPI-2 Profile.....	24
2. Diagnostic Q-set Items.....	27
3. Example of Clinician-Rated Statements Provided to Raters.....	28
4. Reliability of Criterion Sorts Based on Clinician Information .....	34
5. Inter-CBTI Reliability.....	36
6. Inter-interpreter Reliability for CBTI-based Diagnostic Q-sorts.....	38
7. Diagnostic Validity of CBTIs .....	41
8. Most Highly Rated Diagnostic Category for Each Profile-CBTI Combination .....	43
9. Diagnostic Validity of Profiles with High Criterion Reliability Values .....	47
10. Characteristics of WNL Profiles. ....	49



# CHAPTER I

## INTRODUCTION

According to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (American Psychiatric Association, 2000), the first attempt to gather diagnostic data on mental disorders in the United States was made in 1840 when census data was collected for individuals falling into the idiocy/insanity category. Since then, diagnostic classification of mental disorders has gone through multitudinous iterations to become the currently used multiaxial system. Providing diagnostic classification using the five-axis system of the DSM-IV TR is ubiquitous in the practice of clinical psychology. This is especially evident in psychological assessment, where results from psychological tests are interpreted to provide diagnoses and recommendations for treatment.

### Diagnosis and the MMPI

The Minnesota Multiphasic Personality Inventory (MMPI) was developed by Hathaway and McKinley in 1943 to assist in making psychodiagnostic judgments about individuals. The MMPI was developed using an empirical keying approach; items were added to a scale contingent on their success in differentiating people belonging to a particular diagnostic group from people who did not have that exact diagnosis but may have had some symptoms in common. However, research on the Clinical Scales of the MMPI indicated that these scales were not optimal for assigning specific diagnoses as they were correlated with each other due to having many common items (Graham, 2006). Despite these findings, scores obtained from the MMPI, and later the MMPI-2, are used to guide diagnostic judgment. Formulating diagnostic

hypotheses from MMPI-2 results is commonly done by examining patterns of scale scores or codetypes.

There is a substantial amount of research on the MMPI (and the MMPI-2) and psychiatric diagnosis. Moreland (1983) examined the diagnostic validity of the MMPI and two short versions of the MMPI, the Faschingbauer Abbreviated MMPI (FAM) and MMPI-168. Scores on the MMPI were obtained from a sample of 514 inpatients diagnosed as psychotic, neurotic or having a personality disorder. This sample was divided by gender (males = 193 and females = 321) and each group was randomly divided in half. Scale scores obtained from the full MMPI were used to develop multiple discriminant functions for the first half of male and female samples by a computerized statistical program. Using the other half of male and female samples, these discriminant functions were cross validated with both short versions and the full MMPI as predictors. Results indicated that each of the three forms correctly classified 46% of the participants.

Morrison, Edwards and Weissman (1994) evaluated both the MMPI and MMPI-2 with respect to predicting psychiatric diagnosis in a sample of 200 male and female outpatients. Referring clinicians provided DSM III-R Axis I and Axis II diagnoses, along with Global Assessment of Functioning (GAF) scores. Using Lachar's classification system (Lachar, Dahlstrom, & Moreland, 1986), each MMPI and MMPI-2 profile was classified as either neurotic, psychotic, character disorder, normal, or other. Two psychologists used decision rules to classify the diagnoses provided by referring clinicians into one of Lachar's diagnostic categories. For cases where comorbid Axis I diagnoses fell into different Lachar categories, the diagnosis with the most pronounced symptoms was selected. Cases with comorbid Axis II disorder were placed in the character disorder category. To check the diagnostic accuracy of the

referring clinicians, an independent psychologist reviewed diagnoses for 12 randomly selected cases. It was found that there was 75% agreement with respect to assigned diagnoses. Results of this study indicated that both the MMPI and MMPI-2 had an agreement rate of 39% with clinician assigned diagnostic categories based on Lachar's classification system and the best agreement was for the neurosis category. There were no significant differences between results based on the MMPI and the MMPI-2.

Pancoast, Archer, and Gordon (1988) examined consensus between psychiatrist assigned diagnoses and diagnoses derived by applying four diagnostic classification systems to MMPI data. Clinical diagnoses for 150 adult inpatients were provided by psychiatrists who were blind to MMPI information for each subject. These diagnoses were grouped as psychotic, neurotic, personality disorders, and indeterminate. Four diagnostic classification systems including Goldberg's equations, Lachar's system, Heinrich's revision of the Meehl-Dahlstrom rules, and assigning diagnoses based on the scale that had the highest score were applied to K-corrected MMPI scores for each subject. Pancoast et al. examined the level of similarity between discharge diagnoses and diagnostic judgments obtained by using the MMPI diagnostic systems by using the kappa statistic. Overall, the authors obtained similar results across the four systems. Modest hit rates, ranging from 26% to 34%, were obtained across systems. The results also indicated that diagnostic judgments based on the highest scale elevation were not significantly different from other more elaborate systems with respect to level of agreement with clinical diagnoses. Based on these findings related to modest level of diagnostic agreement, the authors cautioned against using MMPI information in isolation with other data for the purpose of making diagnostic judgments.

Despite the fact that the MMPI/MMPI-2 is not recommended as a stand-alone measure of diagnostic classification, its use has always included drawing inferences based on scale scores and configural patterns to guide diagnostic formulations (Bagby, Marshall, Basso, Nicholson, Bacchioni, & Miller, 2005; Ben-Porath, Butcher & Graham, 1991; Morey, Waugh & Blashfield, 1985; Walters, 1985). Commensurate to this application of the MMPI-2, Computer Based Test Interpretations (CBTIs) are also commonly used to make diagnostic inferences.

#### Uses and Advantages of CBTIs

McMinn, Ellens, & Soref (1999) surveyed 364 members of the Society for Personality Assessment regarding ways in which they used CBTIs and ethical perspectives associated with using CBTIs. Results from the survey indicated that the MMPI-2 was the most frequently used test, with a mean of 4.8 administrations per month and a range of 0-250 administrations per month. Across psychological tests, CBTIs for the MMPI-2 were most frequently used with a mean of 3.3 uses per month and a range of 0-250 uses per month. Survey responses also indicated that CBTI software was most frequently used for scoring psychological tests. 87% of respondents indicated that they never used CBTI reports as the main source of information to guide case conceptualization. 28% of respondents indicated that they often used CBTI reports as a source of *additional* information to guide case conceptualization whereas 26% of respondents indicated that they very often used CBTI reports *in tandem* with other sources of information to guide case conceptualization. McMinn et al. (1999) define support functions of CBTIs as those involving the use of CBTIs to supplement information obtained from using other assessment measures such as interviews. Results of this survey indicate that most CBTI users are using them for support functions. However, even seemingly innocuous practices such as using CBTIs to

supplement one's case conceptualization (based on assessment scores) are fraught with the pitfalls of clinical judgment. For example, it is unclear how clinical psychologists decide whether their own formulation offers a superior or inferior assessment of the client, as compared to information in the CBTI report.

Butcher (1995) outlines several advantages of using computer-based reports. He writes that using computer based test reports ensures that interpretive information from all relevant scale scores is being considered, whereas human beings often tend to neglect information when they have to attend to large amounts of data. Additionally, using CBTI reports minimizes subjectivity in interpreting scale scores as compared to interpreting data gathered by diagnostic interviewing. Other advantages offered by computerized scoring and interpretation include efficiency, accuracy and reliability in scoring, the possibility of adaptive testing (Graham, 2006), and the possibility of accessing stored information including normative data and decision rules based on prior research for making unbiased interpretations (Jackson, 1985).

#### *Cautions Regarding the Use of CBTIs*

Butcher (1995) also provides cautionary guidelines for using CBTIs. He writes that using computerized interpretive reports may facilitate an excessively passive approach by blindly using information from the report without integrating it with other sources of data (such as the clinical interview). He adds that for some people, using CBTIs can "introduce a mystical aura" (Butcher, 1995; p. 91) and they may be too enamored by the efficiency of computers to be able to evaluate sections of the interpretive report critically. Other cautions include the susceptibility of CBTIs to the Barnum effect and inapplicability of some of the prototypical information present in these reports to the specific individual for which the report has been generated. The Barnum effect

(Meehl, 1956) refers to vague personality descriptions that seem to apply to everyone due to their ambiguity. According to Butcher, CBTI reports may contain generic statements that seem to apply to the test-taker even though they may not be valid.

Matarazzo (1986) critiqued CBTI reports as providing generic information for the *typical* person, given specific scale scores, and lacking validation for the *individual* who is being assessed. By stating that CBTIs are characteristically “all mean and no sigma” (Matarazzo, 1986; p.14) he strongly cautioned against using CBTIs as ends to psychological assessment, as opposed to the means of assessment. Fowler and Butcher (1986) responded to Matarazzo’s critique and pointed out that Matarazzo failed to account for the fact that the large amount of validity data available for assessment instruments such as the MMPI and MMPI-2 positively contributed to the validity of a carefully formulated report. They argued that in the hands of inexperienced or poorly trained psychologists, CBTIs were as susceptible to faulty interpretations as paper and pencil tests.

In light of some of these cautions, Moreland (1987) underscores the importance of examining validity research on CBTIs before using them in one’s clinical practice. He provides fourteen guidelines to consumers of CBTIs for examining validity research on CBTIs. He also adds the caveat that the validity study being evaluated does not have to meet all the guidelines outlined by him. Moreland’s guidelines can be broadly grouped to include the following:

- 1) Raters participating in the study should have previous experience with the CBTI program being evaluated. Raters should also have prior experience with the rating system being used to evaluate the CBTI program.
- 2) A representative and unbiased sample of raters and test respondents (or codetypes) should be used in CBTI validity studies.

- 3) Discriminant validity of CBTIs should be assessed by using both authentic and bogus reports.
- 4) Both inter-rater reliability and intra-rater reliability should be examined.
- 5) Raters should be asked to outline aspects of interpretive reports that were useless and contradictory.
- 6) Raters should also be asked to indicate whether significant information was missing in the interpretive report.
- 7) Accuracy ratings should be obtained at the rule-level of the interpretive algorithm.

#### Research on CBTIs

Jackson (1985, p. 260) noted that “the greatest challenge facing both those who offer and those who employ computer-assisted testing is in appraising the accuracy and validity of interpretations derived from test scores.” He writes that one of the hurdles in obtaining information regarding interpretive algorithms used by CBTIs is that test publishers often do not want to disclose this information. Jackson suggests that in order to circumvent this problem, test publishers should at least select a qualified group of reviewers to examine the decision rules on which the program is based, without having to disclose the actual computer algorithms.

Jemelka, Wiegand, Walker and Trupin (1992) conducted a validation study on a CBTI for assessment of criminal offenders in correctional settings. Using different assessment devices, the authors developed algorithms to provide data on risk assessment and diagnostic impressions. Variables were included in the algorithm based on data from previous research identified by the authors. A total of 100 male felony offenders in prison were administered a battery of tests including a brief background history questionnaire, the MMPI, the Revised Beta IQ

Examination, the Suicide Probability Scale, the Buss-Durkee Hostility Inventory, the Monroe Dyscontrol Scale, and the Veterans Alcohol Screening Test within two weeks of prison admission. Inmates were also interviewed by two clinicians who provided risk ratings (on a scale of 1 to 5) for violence, suicide, victimization and substance abuse potential. Intra-class correlations were used to assess similarity between clinician-assigned risk ratings and those obtained from the CBTI developed by the authors. The highest correlations were found for substance abuse potential (.83) and the lowest were found for victimization (.49). Regression analyses were conducted to identify variables that significantly predicted clinicians' risk ratings. These included age at first offense, recidivism, and violence history. In the second phase of this study, inmates were administered the Diagnostic Interview Schedule (Robins & Helzer, 1985) by trained clinicians to ascertain DSM-III R diagnoses. Results indicated high agreement between clinician-assigned and CBTI-assigned diagnoses. The authors then used discriminant function analyses to identify variables that could improve diagnostic classification by the CBTI. This study illustrates some of the steps taken by CBTI developers in validating the algorithms that constitute computerized interpretive programs.

### Research on CBTIs for the MMPI-2

#### *Reliability Research*

Deskovitz (2003) examined the interpretive reliability of two commonly occurring MMPI-2 codetypes, the 27 and 68 codetypes. Participants included twenty master's and doctoral level clinicians who were randomly assigned to interpret either the 27 or the 68 codetype. Participants interpreted the selected MMPI-2 profile by using a web-based Q-sorting program that involved sorting a set of 100 descriptive statements into seven categories containing 5, 10,



20, 30, 20, 10, and 5 items respectively. Participants were given feedback on the correlation between their Q-sort and a Q-sort completed by an experienced MMPI-2 interpreter. Results indicated that the mean pairwise interpretive reliability (extent to which participants interpreted the MMPI-2 profile similarly) for the 27 and 68 codetypes was .41 and .38, respectively. There was no statistically significant difference between educational degrees with respect to interpretive reliability. Additionally, Deskovitz examined specific items that received highest and lowest mean ratings for each codetype interpreted. In discussing possible reasons behind low interpretive reliability values obtained in this study, Deskovitz lists issues including misunderstanding the Q-sort procedure and lack of motivation as there was no compensation offered in exchange for participation.

Later, Deskovitz (2005) examined the interpretive reliability of the six commercially available CBTIs. Deskovitz defined inter-program reliability as the level of consensus between the six CBTIs and inter-interpreter reliability as the extent to which CBTI reports were understood consistently by different users of these reports. A total of 20 MMPI-2 profiles were interpreted by each of the six CBTIs, giving a total of 120 reports. Q-sorts based on these reports were completed by four doctoral students enrolled in the clinical psychology program at a Midwestern university. The students used a Windows based computer program called the Midwestern Q-sort (Williams & Weed, 2002) to complete sorts on the CBTI reports. Inter-interpreter reliability was calculated as the intercorrelation between sorts of all four raters for one particular MMPI-2 profile and a particular CBTI report. The overall reliability coefficient was found to be .60 and no significant differences were found between different CBTI programs. Considering profile-level results, Psych Screen (PS) 1/3, Psychological Assessment Resources (PAR) 7/8 and Automated Assessment Associates (AAA) Within Normal Limits (WNL) profiles

had relatively low inter-rater reliability coefficients. Inter-program reliability was calculated as the inter-correlation between all six CBTIs for each of the 20 MMPI-2 profiles. The overall mean reliability coefficient was found to be .83 and there was no significant overall difference between the six CBTIs. With respect to particular profiles, the 2/4 and 7/8 codetype produced lower mean reliability coefficients whereas the 4/9 and 1/2/3 codetypes produced higher mean reliability coefficients. The author attributed high reliability coefficients obtained in this study to aggregation of data by using mean values of sorts from all four raters. Additionally, no particular CBTI was found to have extremely high or extremely low reliability.

Pant and Weed (2008) evaluated the diagnostic reliability of six commercially available CBTIs by using a Q-sort procedure to operationalize diagnostic agreement. Four raters from the doctoral program in clinical psychology at a Midwestern university provided diagnostic classifications based on CBTI reports for twenty MMPI-2 profiles using the Diagnostic Q-sort, which involved ranking a set of twenty DSM-IV TR (American Psychiatric Association, 2000) diagnoses into five categories (ranging from most characteristic to least characteristic), with each category having four diagnoses each. It was found that CBTIs showed high inter-interpreter diagnostic reliability ( $\alpha = .81$ ), indicating that diagnostic suggestions offered by them are understood consensually by users. Notable exceptions were the 7/8 profile interpreted by the Psychological Assessment Resources program ( $\alpha = .29$ ), the over five scales elevated profile interpreted by the Psych Screen program ( $\alpha = .41$ ), the 4/8 profile interpreted by the Psych Screen program ( $\alpha = .50$ ) and the 6/8 profile interpreted by the Psych Screen program ( $\alpha = .51$ ). It was also found that there was high inter-CBTI diagnostic reliability ( $\alpha = .94$ ), which implies that they provide similar diagnostic suggestions. A notable exception was the profile with over five scale elevations which had the lowest inter-CBTI reliability ( $\alpha = .86$ ). The program by

Automated Assessment Associates was found to be most consensual with all other programs whereas the program by Psych Screen was found to be least consensual with all other programs.

### *Validity Research*

Moreland (1985) evaluated two sets of studies pertaining to the validity of CBTIs which he terms “customer satisfaction studies” (involving consumers rating the degree to which CBTIs provide accurate information) and “external criterion studies” (involving comparisons of ratings on specific variables based on CBTI reports and ratings of the same variables based on other information). Moreland (1985) states that some of the early studies focusing on customer satisfaction evaluated the CBTI program developed for the MMPI by Fowler and involved using Likert-type rating scales to evaluate the accuracy and utility of CBTIs. However, the author adds that these studies only evaluated content areas of the report as opposed to a more detailed (sentence-by-sentence or paragraph-by-paragraph) comparison and it was difficult to identify specific flaws in a particular CBTI program. In addition to the Fowler program, other early studies evaluated the accuracy of the CBTI program for the MMPI developed by Gilberstadt. The third CBTI program that was evaluated in the early years of CBTI development was Lachar’s system which was first sold by Automated Psychological Assessment and then by Western Psychological Services.

### *Customer Satisfaction Studies*

“Customer satisfaction” studies involve asking consumers of CBTIs to rate the degree to which CBTI reports provide accurate information. Webb, Miller and Fowler (1970) conducted a customer satisfaction study evaluating the CBTI for MMPI which was developed by Fowler and sold by Roche Services. Eighteen participants (including psychiatrists, clinical psychologists and

psychiatric social workers) used a five point rating scale to rate a total of 86 CBTI reports on their clients. Ratings were provided on global judgments of narrative accuracy, usefulness and clarity. Overall results were in favor of the Roche CBTI report with 60% of raters indicating that the reports were not missing significant information. Moreland (1985) points out that it is difficult to identify specific weaknesses in these reports as raters provided global ratings as opposed to evaluating the reports at the statement or paragraph level.

Eyde, Kowal and Fishburne (1991) conducted a comprehensive “customer satisfaction” study on seven CBTIs for the MMPI, including Behaviordyne, Caldwell Report, NCS Minnesota Report, WPS, Morris-Tomlinson Report, NCS FASTTEST, and Applied Innovations. MMPI-2 data on both inpatients and outpatients at the Walter Reed Army Medical Center was used to generate the CBTI reports. Common cases rated by all the participants (a total of 12 Army mental health professionals such as clinical psychologists) included a pair of the 27/72 codetype matched for race (Caucasian and African American), a pair of subjects (matched for race) with subclinical scores who participated in an MMPI normative study, and a pair of subjects (matched for race) with a spike or two-point MMPI codetype. Thus, each rater evaluated six MMPI CBTI reports generated by seven different CBTI systems (a total of 42 reports per rater). The authors standardized each CBTI report and numbered each statement in the report, except for those occurring in footnotes. Raters were instructed to rate each numbered statement in the CBTI narrative report according to a prearranged order. Results indicated that the Minnesota Report was rated as being most accurate and the Behaviordyne report was rated as being least accurate. With respect to ratings on overall helpfulness of the reports (including diagnostic evaluation and disposition planning), the Minnesota Report received the highest median rank whereas

Behaviordyne received the lowest median rank. This study is unique in that it asked raters to rate *each statement* of the CBTI reports and thus, provides more detailed accuracy data.

Green (1992) examined adequacy of CBTI reports for the MMPI generated by NCS and Roche Systems, and compared those with CBTI reports for the Millon Clinical Multiaxial Inventory or the MCMI (Millon, 1977). Participants (or raters) in this study were twenty three clinicians (including psychiatrists, social workers and clinical psychologists). MMPI and MCMI reports were generated for a total of 100 randomly selected patients from both inpatient and outpatient settings and these patients had received a minimum for four contact hours with the participating clinicians who submitted the MMPI and MCMI test scores for their patients to the investigators. Each clinician submitted scores on a range of two to nine patients and then rated the interpretive reports. Ratings were obtained for three broad areas. 'Information Accuracy' included items such as confirmation of knowledge, clarification of case, exclusion of important information etc. 'Accuracy of Descriptions' contained items such as interpersonal attitudes and relationships, affective tone and mood, self-image, primary symptoms, coping styles, etc. 'Format and Utility' included items such as internal consistency, organization, intelligibility and clarity and helpful in treatment. Overall results indicated that the Roche reports and MCMI reports were judged to be better than the NCS report. Additionally, the MCMI reports were judged to be more useful for information on coping style and interpersonal relationships. Even though this study did not control for Barnum effects, the authors described conducting a small pilot study that used graduate students to rank reports based on simulated data. This study has some drawbacks that limit generalization. For example, the possibility of rater bias (ratings of reports on the MMPI and MCMI being influenced by clinicians preference for one of these tests) and the possibility of Barnum effects.

Williams and Weed (2004) conducted a “customer satisfaction” study to evaluate the relative utility of eight commercially available CBTIs while controlling for the Barnum effect. Forty-one practitioners from fields such as clinical psychology, neuropsychology, counseling psychology and social work were asked to send deidentified MMPI-2 answer sheets completed by their clients. Each MMPI-2 answer sheet that had been sent by practitioners was randomly assigned to one of two conditions – being interpreted by a CBTI (the ‘authentic’ condition) or being substituted by a report based on modal MMPI-2 profiles obtained in inpatient, outpatient, corrective, and college counseling settings (the ‘modal’ condition). All the participants were notified that they would receive either an authentic interpretive report or a modal report. So, each MMPI-2 protocol received was randomly assigned to be interpreted by one of the eight CBTI programs and either the authentic or modal report method. Practitioners downloaded an electronic version of the report from an Internet website (only one report per MMPI-2 profile submitted) and rated the report on ten aspects by using a seven-point Likert-type scale. These aspects included conciseness, confirmation of therapist’s impressions of the client, usefulness for diagnosis and/or treatment, accuracy, provision of new information, presence of contradictory information, organization and clarity, presence of useless information, omission of important information and appropriateness of diagnostic considerations.

To deal with the Barnum effect, the ratings of genuine CBTI interpretations were compared to those of modal interpretive reports by examining Cohen’s *d* values for the difference in ratings between the two reports for each CBTI program. The authors found that, in general, authentic reports received higher ratings than modal reports (as evidenced by mostly positive effect sizes). The Automated Assessment Associates (AAA) report received most favorable overall ratings in this study whereas the report by Behaviordata received the lowest

overall ratings. It is interesting to note that the category ‘Appropriate Diagnosis’ on the rating scale received some of the lowest ratings across CBTIs for the authentic reports. Given values ranging from -3 (strongly disagree) to +3 (strongly agree) on the rating scale, the AAA report received the highest mean rating of 1.31 whereas the other programs received generally lower ratings with a mean rating of -.20 for the Behaviordata (BD) report, .33 for the Caldwell Report (CR), .31 for the report by Pearson Assessments (PA), .38 for the Psychological Assessment Resources (PAR) report, .60 for the report by Psychometric Software Inc. (PSI), .75 for the report by Psych Screen Inc. (PS), and .47 for the report by Western Psychological Services (WPS).

#### *External Criteria Studies*

“External criteria” studies involve a comparison of CBTI reports with data from rating scales or Q-sorts based on interaction with the patient completing the MMPI. Anderson (1985) examined the Roche interpretive system for the MMPI by using ratings by the therapists of participants (28 inpatients) as the external criteria. The participants were rated on twelve variables by their therapists and the Roche CBTI report was rated on the same twelve variables by experts in MMPI interpretation. Results indicated that the average correlation between CBTI ratings and ratings made by the therapists was .20.

Studies involving a comparison of the validity of different CBTI programs for the MMPI were pioneered by Chase (1974). In her study, she compared CBTI reports obtained from three different program providers – Roche, Behaviordyne, and Caldwell Report – and asked clinicians familiar with the patients who had completed the MMPI to rate the accuracy of each report on a five-point scale. Results from her study indicated that Roche and Caldwell reports were superior

to Behaviordyne reports. However, when external criteria were used to judge accuracy of reports, Behaviordyne was judged to be more accurate as compared to the other two systems. In using external criteria, Chase used six different methods to interpret patients' MMPI profiles. Two MMPI experts provided interpretive reports and subsequently rated patients' MMPI profiles by using a 59-item subset of the Minnesota-Ford phenotypic item pool. An actuarial atlas developed by Marks and Seeman (1963) was used to write reports and the three aforementioned CBTI programs were used to generate reports. All these reports were rated using items from the Minnesota-Ford pool by clinical psychologists, psychiatrists, social workers and psychiatric nurses. Clinicians who had either worked with the patients or studied their histories provided criterion ratings.

Crumpton (1974) was the first to use Q-sort methodology to evaluate the accuracy of computer generated reports for the MMPI. In her doctoral dissertation, she used MMPI reports generated by the Roche system, Clinical Psychological Services, and the Institute for Clinical Analysis for nine randomly selected patients receiving psychotherapy and compared them with therapists' ratings (provided after four sessions of psychotherapy) on the Marks Q-sort. This Q-sort consists of 108 phenotypic and genotypic statements that are placed into nine categories, with each category having twelve items. In addition to therapist ratings, each of the CBTI reports were rated by two clinical psychologists and one advanced clinical psychology graduate student (a total of three judges) by using the Marks Q-sort. Crumpton found the mean inter-rater reliability coefficient for the judges' ratings to be .62 with a standard deviation of .12. Results also indicated that the average correlation between the judges' Q-sort ratings, based on the computer report and the therapists' Q-sort ratings for all nine patients was .36 with a standard deviation of .24. Post-hoc analyses indicated that the computer reports were less accurate for



patients that were reporting severe psychopathology (operationalized as having T-scores above 70 on one or more of the clinical scales). Results obtained in this study highlight the low level of agreement between information contained in computer based test interpretations and information provided by therapists, which raises concerns regarding use of computer reports as an important source of MMPI interpretive information.

Moreland (1983) used a sample of 70 MMPI profiles representative of normal, psychotic, neurotic, characterological, and indeterminate types (based on Lachar's typology). Experienced clinicians who were not familiar with the MMPI and who reported not having previously used CBTIs by Roche and Western Psychological Services were asked to rate the 70 reports. Additionally, reports on the same patient were not provided to the same rater in order to prevent recognition problems. Criterion ratings were obtained at the first MMPI administration to the patient. Moreland obtained low inter-rater reliabilities (in the .50s) and low validity coefficients (.36) for both programs.

Harrington (2000) evaluated the reliability and relative validity of computer based interpretations provided by six interpretive programs - Automated Assessment Associates (AAA), Behaviordata (BD), Caldwell Report (CR), National Computer Systems (NCS), Psychological Assessment Resources (PAR) and Psychometric Software (PS). The MMPI-2 was administered to 26 subjects from a university counseling center. By using the Q-sort method, two sorts on each subject were completed by their therapist and friend/family member. Additionally, two advanced graduate students from the clinical psychology program at the same university Q-sorted the narrative reports obtained by the six CBTIs for each subject. Harrington found considerable inter-rater reliability, calculated as the Q-correlation between sorts generated by the two student raters for each CBTI, with a mean value of .64 (averaged across the six CBTIs). The

author concluded that these high reliability estimates indicate that the CBTIs present similar information to users. The median value of correlations between therapist and friend/family member Q-sorts was .49, indicating modest agreement. Validity estimates were calculated by using three criteria – therapist Q-sorts, friend/family member Q-sorts, and the average value of these two sorts. Validity estimates based on correlation with therapist sorts ranged from .07 to .22, those based on correlations with friend/family sorts ranged from -.08 to .16, whereas estimates based on the average of therapist and friend/family sorts ranged from -.07 to .14. With respect to relative validity, the AAA report had the highest validity coefficient whereas PS had the lowest. According to the author, one possible explanation behind these low validity estimates may be selection bias as the requirement of having a friend/relative complete a Q-sort on the client may have hindered participation of clients with serious psychopathology.

McNeal (1999) used Q-sort methodology to examine the accuracy of CBTI reports created by using the Minnesota Report (sold by Pearson Associates) by using therapist Q-sorts as the criterion. In this study, five different Q-sorts were completed for each of the 20 participants going through individual psychotherapy. The therapist provided a Q-sort on his/her client without access to any data from the MMPI-2, two investigators each completed a Q-sort on the client based solely on scores on MMPI-2 validity, clinical, content and supplementary scales, and the same two investigators each completed Q-sorts on the client based only on narrative sections of the CBTI report (including sections like profile validity, symptomatic patterns, interpersonal relations, diagnostic considerations etc.). Mean inter-rater agreement (between sorts completed by the two investigators) for Q-sorts based on MMPI-2 scores was found to .66 and mean inter-rater agreement for sorts using the narrative report was found to be .74. Mean correlation between therapist Q-sorts and the mean (averaged) Q-sorts of both investigators was found to be

.25 for the narrative report and .27 for the sorts based on MMPI-2 scores. Even though validity coefficients were modest for Q-sorts based on both narrative reports and MMPI-2 scale scores, it is interesting to note that, compared to information based only on MMPI-2 scale scores, the narrative report (without specific scale scores) provides an almost equally valid estimate of the client.

In summary, results from the “external-criterion” studies outlined in this section have found that the correlation between subjects’ CBTI reports based on MMPI/MMPI-2 scale scores and their therapists’ Q-sort ratings ranges from .07 (Harrington, 2000) to .36 (Crumpton, 1974; Moreland, 1983). However, higher correlation coefficients, ranging from .60 (Deskovitz, 2005) to .64 (Harrington, 2000), have been obtained for inter-interpreter reliability of CBTIs. Notably, consumers have provided consistently low ratings on appropriateness of diagnostic suggestions, across available CBTI programs for the MMPI-2 (Williams & Weed, 2004). Results regarding diagnostic reliability of CBTIs for the MMPI-2 indicate high overall reliability values across commonly obtained codetypes (Pant & Weed, 2008). However, reliability does not guarantee validity. Even if all CBTIs are consensual in providing diagnostic suggestions, these suggestions could still have weak correlations with diagnoses provided by treating clinicians. This issue regarding the validity of diagnostic suggestions offered by CBTIs for the MMPI-2 merits further empirical investigation.

### *Review of CBTIs*

Williams and Weed (2004) provided a review of six commercially available CBTIs for the MMPI-2. The authors compared the six programs based on software features such as type of operating system required for running the program, type of media used by the program (compact

disk, floppy disk, etc.), requirement of a dongle, price of the program, the number of times it can be used (single use, unlimited use, etc.), verification of input data and options regarding saving or exporting reports, searching saved reports and sorting saved reports. Additionally, reports from all six programs were obtained for a commonly observed MMPI-2 codetype (738 codetype) and the similarities as well as differences across the six reports were examined. It was found that all six reports included cautions against using the report as the only basis for making therapeutic decisions and most reports also indicated that the statements contained in the report were probabilistic. There were also similarities regarding likelihood of certain symptoms such as anxiety, guilt, and depression. However, the authors reported that differences between the six reports were “striking” (Williams & Weed, p. 80). For example, there were differences regarding presence or absence of certain symptoms, and treatment recommendations, and “dramatic differences” between diagnostic suggestions provided by each report for the same case. These substantial differences suggest that there may be significant disparities in the overarching interpretive methods used by the developers of these reports.

### The Q-sort Method

Examining diagnostic judgment with respect to various MMPI-2 CBTI reports is problematic because absolute diagnoses cannot be generated easily by relying solely on scores obtained on the MMPI-2. The Q-sort method circumvents this problem. Originally proposed by Stephenson in 1953, this method involves sorting a set of given statements into different categories having fixed frequencies, subject to their level of importance or relevance. For example, a set of 100 statements can be sorted by raters into seven different categories ranging from most characteristic of a specific property to least characteristic, with each category

containing a total of 5, 10, 20, 30, 20, 10 and 5 statements respectively. In this way, the ratings represent a forced normal distribution and facilitate both inter-rater and intra-rater comparison by using quantitative techniques. In this way, the Q-sort technique can enable raters to assign diagnostic labels on a continuum ranging from most characteristic of a particular MMPI-2 profile to least characteristic of it. Additionally, it can facilitate diagnostic comparison across CBTI programs by operationalizing possible diagnoses given by each CBTI as participants' ratings on a diagnostic Q-sort.

Ozer (1993) outlines several advantages of the Q-sort method. One of the advantages is that raters/sorters don't require rigorous training in order to complete the sorting task and the level of training required largely depends on the nature of items in the Q-set. A second advantage is the "transportability" of this method as it allows for the object to be rated or judged by multiple sources of information. In this regard, the Q-sort method not only facilitates judgments by multiple raters but also enables the examination of multiple data sources ranging from biographies of past presidents of the United States to information gleaned from MMPI profiles or reports. Another major advantage is that it controls for rater error by utilizing a forced-choice method. Having a fixed number of categories reduces the possibility of extreme ratings and socially desirable responding as the number of items that can be placed in any category are less than the total number of items in the Q-set. Ozer also discusses the effects of the shape of Q-sort distributions, for example, using a normal distribution versus using a flat/rectangular one. He shows that the number of discriminations increases by a very small number if a rectangular distribution (in which all categories have equal number of items) is used instead of a forced normal distribution.

## Present Study

Hitherto, no study has examined diagnostic validity of CBTIs for the MMPI-2. However, as reported in the survey conducted by McMinn et al. (1999), psychologists use CBTIs as a source of secondary information in clinical case formulation. Therefore, it is important to subject them to empirical study with respect to the validity of their diagnostic suggestions. This study aimed to examine the diagnostic validity of four commercially available CBTIs for the MMPI-2 and questioned whether the diagnostic suggestions offered by the CBTIs are valid with respect to descriptive information provided by treating clinicians. The answer to this question was found with the help of Q-sort methodology by correlating mean diagnostic Q-sorts completed by raters based on descriptive information by treating clinicians with mean diagnostic Q-sorts completed by raters using CBTI reports generated for MMPI-2 results (by each of four CBTI programs), for the same client. The correlation coefficients provided an index of similarity between the diagnostic suggestions offered by the CBTIs and diagnostic judgments made by treatment providers. It was hypothesized that the diagnostic validity coefficients obtained in this study will be in the low to moderate range.

Ancillary to the question on diagnostic validity, this study explored whether specific diagnostic categories were associated with higher diagnostic validity coefficients, indicating greater CBTI agreement with information provided by treating clinicians. This was examined by comparing mean diagnostic validity coefficients across the different psychodiagnostic categories included in the Diagnostic Q-sort.

Finally, this study examined how CBTIs compare with each other with respect to diagnostic validity. The diagnostic validity of these CBTIs was compared by examining validity coefficients obtained for each CBTI after being averaged across all the cases. Based on results

from previous research regarding diagnostic reliability (Pant & Weed, 2008), it was hypothesized that CBTIs will not show significant differences in diagnostic validity with respect to each other.

## CHAPTER II

### METHOD

The aim of this study was to examine the diagnostic validity of four commercially available MMPI-2 Computer Based Test Interpretation (CBTI) programs by using diagnostic information given by treatment providers as the criterion. The instrument used to facilitate diagnostic comparison between the CBTIs and information provided by clinicians is the Diagnostic Q-sort, which comprises 20 diagnoses from both Axis I and II of DSM IV-TR (American Psychiatric Association, 2000) judged to be most germane to the MMPI-2.

#### Participants

This study used archival data from 65 clients at a Midwestern mental health center, including 42 males and 23 females with a mean age of 35.6 years. Information on these clients was provided in terms of ranked descriptive statements by their treating clinicians. Table 1 shows the mean K-corrected MMPI-2 profile for the 65 cases, along with the range, minimum and maximum values, as well as standard deviations for scale scores.

Table 1. Mean K-corrected MMPI-2 Profile

Scale	Range	Minimum	Maximum	Mean	Std. Deviation
L	40	38	78	54.6	10.2
F	84	36	120	55.1	14.0
K	41	35	76	53.5	10.1
S	42	34	76	54.2	10.9
Fp	79	41	120	56.0	11.9
TRIN	21	50	71	57.2	5.3
VRIN	42	31	73	51.9	10.8
Hs	51	35	86	55.8	11.0
D	62	32	94	59.0	14.4
Hy	62	34	96	59.1	12.7
Pd	60	35	95	60.0	12.7



Scale	Range	Minimum	Maximum	Mean	Std. Deviation
Mf	42	32	74	50.8	9.4
Pa	57	37	94	56.6	13.4
Pt	69	33	102	58.1	14.4
Sc	84	36	120	56.9	13.1
Ma	34	31	65	48.4	7.9
Si	50	34	84	51.2	11.7
RCd	47	36	83	55.4	13.8
RC1	58	36	94	52.9	12.1
RC2	66	33	99	55.1	15.0
RC3	45	33	78	47.0	8.2
RC4	44	33	77	50.3	11.2
RC6	52	41	93	51.8	11.8
RC7	60	32	92	50.7	13.0
RC8	49	38	87	48.3	10.8
RC9	34	30	64	44.7	7.7

As seen in Table 1, this sample evidenced comparatively higher mean scores for Psychopathic Deviate (T = 60), Depression (T = 59), and Hysteria (T = 59.1) scales. However, none of the validity, clinical, and Restructured Clinical (RC) scales had clinically elevated scores (T > 64).

#### Raters

Raters in this study included six advanced clinical psychology doctoral students and recent clinical psychology doctoral graduates having more than 2 years of experience with using the MMPI-2. One set, including three raters, completed diagnostic Q-sorts on descriptive information on clients provided by treating clinicians. A second set of three raters completed diagnostic Q-sorts based on interpretive reports generated by CBTI programs.

#### Clinicians

Four doctoral-level psychologists working at a Midwestern outpatient mental health center completed Q-sort ratings on 65 clients seen by them for psychotherapy services. The clients were administered the MMPI-2 at the third session and the clinicians completed Q-sorts (without knowledge of MMPI-2 results) after the fourth session. The clients received feedback on their MMPI-2 results by another doctoral-level psychologist working at the same facility.

## Instruments

*MMPI-2.* The Minnesota Multiphasic Personality Inventory – Second Edition (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 2001) is a self-report measure of personality and consists of 567 items that can be given a true or false response. The CBTIs were based on scores obtained for the 65 archival MMPI-2 protocols. Participants were provided the entire report generated by a CBTI, including sections on diagnostic considerations.

*CBTIs.* The four CBTI programs used in this study were Automated Assessment Associates [AAA] Strassberg & Cooper, 2005; Pearson Assessments, formerly known as National Computer Systems: The Minnesota Report [PA] Butcher, 2005; Psychological Assessment Resources [PAR] Green, Brown, & Kovan, 2011; and Psych Screen, Inc. [PS] Minnesota Multiphasic Personality Inventory- 2 Interpreter, 1998. These four CBTIs were selected because, at present, they are the only commercially available CBTI programs for the MMPI-2 which provide instant scoring and/or interpretation services after being purchased.

*Diagnostic Q-sort.* The Diagnostic Q-Sort consists of 20 diagnoses from both Axis I and II of the DSM IV-TR (American Psychiatric Association, 2000). The diagnostic categories were initially selected based on a comprehensive review of the book MMPI-2: Assessing Personality and Psychopathology – Fourth Edition (Graham, 2006), which is currently one of the most

widely used textbooks on the MMPI-2. The Q-sort distribution was chosen to be flat, that is, there were equal number of items in each category. Thus, the 20 items are sorted in five categories, ranging from most characteristic to least characteristic, with four items each. Table 2 provides a list of these 20 diagnostic categories. For the purpose of this study, the Diagnostic Q-sort was adapted as a computer application to be accessed by the raters by using their personal computers to go to a website. The Diagnostic Q-sort was used to operationalize diagnostic decisions for Q-sorts derived from descriptive information by treating clinicians, as well as for diagnostic decisions for Q-sorts based on CBTI interpretive reports.

Table 2. Diagnostic Q-set Items

No.	Diagnosis
1.	Somatization Disorder
2.	Conversion Disorder
3.	Generalized Anxiety Disorder
4.	Obsessive Compulsive Disorder
5.	Major Depressive Disorder
6.	Bipolar Disorder
7.	Schizophrenia – Paranoid Type
8.	Schizophrenia
9.	Schizoaffective Disorder
10.	Anorexia Nervosa
11.	Factitious Disorder
12.	Substance Abuse
13.	Paranoid Personality Disorder
14.	Schizoid Personality Disorder
15.	Schizotypal Personality Disorder
16.	Anti-social Personality Disorder
17.	Histrionic Personality Disorder
18.	Narcissistic Personality Disorder
19.	Passive Aggressive Personality Disorder
20.	No Diagnosis

*Midwestern Q-sort.* Descriptive information on the 65 cases included in this study was obtained by treating clinicians in the form of completed Midwestern Q-sorts on each case. The Midwestern Q-sort was developed by Noland & Weed (1995) as a Windows-based computer

program consisting of a set of 100 descriptive statements which are ranked into seven categories ranging from most descriptive to least descriptive of a person. The Midwestern Q-sort distribution was designed to be normal, with 5, 10, 20, 30, 20, 10, and 5 statements in each of the seven categories, respectively. Table 3 provides an example of descriptive information provided to raters based on ranked statements imported from clinician-completed Midwestern Q-sorts.

Table 3. Example of Clinician-Rated Statements Provided to Raters

---

Most Characteristic:

- acts anxious
- acts coarse and tough
- feels like she/he does not belong
- is a productive worker/student
- worries about own sexual feelings

Somewhat Characteristic:

- acts confused
- adapts easily to new situations
- avoids responsibility because of health problems
- has mood swings
- is argumentative
- is comfortable with her/his sexual orientation
- is willing to discuss most personal problems
- manipulates others
- sees or hears things that do not exist
- values intellectual activities

Characteristic

- accepts advice and suggestions
- acts on impulse
- believes things that obviously are not true
- complains about aches and pains
- creates excitement to overcome boredom
- demands sympathy from others
- feels guilty
- has close friendships
- has little energy
- is involved in a variety of activities
- is irritable
- is self-confident

---

---

is suspicious  
makes decisions with ease  
participates in many social activities  
reasons logically  
resists change  
sees projects through to completion  
trembles, sweats, or shows other physical signs of anxiety  
uses alcohol or drugs excessively

Neutral:

acts depressed  
avoids competitive situations  
behaves in a mature manner  
behaves oddly  
believes she/she is successful  
can easily stay focused on a task  
cries frequently  
criticizes other people  
describes self as having few problems  
develops health problems when under stress  
discusses committing suicide  
discusses problems openly  
does not dwell on physical problems  
does things that most people think are immoral  
exaggerates emotional problems  
fakes illness to receive attention or other benefits  
generally tells the truth  
has excessive fears or phobias  
is overconcerned about the reactions of others  
is perfectionistic  
is realistic  
is seeking help  
is self-indulgent  
is shy  
is unlikely to abuse medications  
makes excessive demands of others  
needs to be more assertive  
shows good judgment  
takes responsibility for own problems  
tries to conform

Uncharacteristic:

expresses emotion in healthy ways  
feels capable of overcoming difficult life situations  
feels extremely happy without cause

---

---

feels misunderstood by others  
finds it easy to put aside troubles  
hides angry feelings  
is not overly concerned about minor matters  
is organized  
is responsible  
is submissive  
is unlikely to benefit from counseling  
needs less affection than most people  
needs to achieve  
obeys authority  
problems are likely to change over time  
rarely gets upset  
reports good physical health  
tries to appear capable  
worries about being accepted by others  
would see a counselor without encouragement from others

**Somewhat Uncharacteristic:**

has problems adjusting to life circumstances  
is aggressive  
is easy to impress  
is hopeful  
is liked by most people who know her/him  
is more emotional than most other people  
is optimistic  
is self-centered  
rarely daydreams  
sees self as better than most other people

**Most Uncharacteristic:**

acts relaxed  
has severe family problems  
is angry  
is likely to develop trust in a counselor  
needs attention

---

**Procedures**

The following steps outline procedures for data collection:

- 1) Archival, deidentified MMPI-2 results based on protocols completed by 65 clients at a Midwestern mental health center were obtained for use in this study. The MMPI-2

protocols were administered during 2010 and included 42 male and 23 female clients, with a mean age of 35.65 years. Additionally, descriptive statements ranked (according to how well they characterize the client) by the mental health clinicians who provided treatment to these 65 clients were obtained. These were presented as documents containing statements from most descriptive of the client to least descriptive of the client, based on the therapist rankings.

- 2) Three raters (advanced clinical psychology doctoral students and recent clinical psychology doctoral graduates) were provided with the documents containing ranked descriptive statements and the raters completed Diagnostic Q-sorts (based on the ranked statements) for each of the 65 cases. This resulted in three sets of Diagnostic Q-sort ratings for 65 cases which were considered to be criterion diagnostic ratings for examining diagnostic validity (due to being based on descriptive information obtained from clinicians). The raters emailed their Q-sorts to the investigator.
- 3) Another set of three raters (advanced clinical psychology graduate students) were provided with CBTI reports generated by each of the four commercially available MMPI-2 interpretive programs for the 65 cases, with a total of 260 (4x65) reports being provided. These three raters then used the Diagnostic Q-sort (on their personal computer) to discern the most appropriate diagnostic labels associated with each CBTI report, based on the information provided in the reports. These reports were edited (uniform font, no names of programs, and no MMPI-2 scale scores) and presented to each rater in random order. A sample CBTI report is attached as Appendix A.

## Analysis

This study aimed to examine the diagnostic validity CBTIs for the MMPI-2. Firstly, it questioned whether the diagnostic suggestions offered by the CBTIs are valid with respect to information provided by treating clinicians. The answer to this question was found by correlating mean criterion diagnostic Q-sorts (averaged across three raters) completed by raters based on descriptive information by treating clinicians with the mean sort of the three graduate student raters for the same case, for each of the four CBTI programs. The correlation coefficients provided an index of similarity between the diagnostic suggestions offered by the CBTIs and diagnostic judgments made by treatment providers. So, for a total of 65 MMPI-2 protocols there were 65 Q-correlation coefficients (one for each protocol) obtained. The average of these 65 Q-correlations provided an index of overall agreement and standard error of the mean provided an estimate of variability.

Ancillary to the question on diagnostic validity, the second question raised in this study was whether specific diagnostic categories obtain higher diagnostic validity coefficients, indicating greater CBTI agreement with information based on descriptions provided by treating clinicians. This was examined by comparing mean diagnostic validity coefficients across the 20 categories in the Diagnostic Q-sort.

Thirdly, this study aimed to examine how CBTIs compare with each other with respect to their diagnostic validity. The comparative validity of these CBTI reports was evaluated by examining the mean values of validity coefficients (averaged for all the profiles) obtained for all the programs to identify whether there were any significant differences in their values.



## CHAPTER III

### RESULTS

The results of the study will be presented with respect to the three questions raised regarding overall diagnostic validity of CBTIs, examining validity for specific diagnostic categories, and comparing the validity of CBTIs with respect to each other.

#### Data Screening

In this study, diagnostic validity was measured by examining the correlation between diagnostic Q-sorts completed by a set of raters using information obtained from treating clinicians (designated as criterion Q-sorts) with diagnostic Q-sorts completed by another set of raters using information gleaned from CBTI reports (designated as CBTI-based Q-sorts). As the criterion for examining the diagnostic validity of CBTIs was diagnostic Q-sorts based on information provided by clinicians, it was important to screen profiles based on the reliability of these sorts to gauge the extent to which they provided consensual information. Table 4 provides mean inter-correlations between diagnostic Q-sorts completed by three raters as well as the coefficient alpha values (by treating raters as items) for each of the 65 profiles. Table 4 also provides information on all the clinical scales (for each of the 65 MMPI-2 profiles) which had a T-score greater than 64, presented in descending order of elevation. Based on examining the alpha values in Table 4, profiles 3, 20, and 45 were deleted from the sample as they obtained negative alpha coefficients, indicating that their criterion sorts present minimally reliable information. This brought the total number of profiles to 62. These profiles had a mean alpha value of .62 and a mean inter-rater correlation coefficient (for diagnostic sorts by three raters) of .38. A possible source contributing to unreliability of criterion sorts includes lack of clarity in

ranked descriptive statements provided by clinicians which may have caused raters to rely on their own judgments during the diagnostic sorting tasks, leading to lower consensus.

Table 4. Reliability of Criterion Sorts Based on Clinician Information

Profile No.	Clinical Scales with T>64	Mean Correlation Between Diagnostic Sorts By Three Raters	Coefficient Alpha
1	72048	.55	.79
2	2406781	.57	.80
3	WNL	-.05	-.17
4	432	.27	.52
5	8670	.22	.47
6	2	.47	.72
7	42	.23	.48
8	13	.23	.48
9	WNL	.55	.79
10	WNL	.35	.62
11	WNL	.55	.79
12	9	.59	.81
13	23	.24	.49
14	8	.12	.30
15	WNL	.34	.61
16	27	.37	.64
17	WNL	.08	.21
18	7	.47	.73
19	WNL	.17	.39
20	43	<b>-.12</b>	<b>-.50</b>
21	5	.57	.80
22	4	.60	.82
23	WNL	.43	.70
24*	87463210	.40	.67
25	WNL	.22	.45
26	WNL	.27	.53
27	WNL	.28	.54
28	WNL	.35	.62
29	13	.29	.55
30	27	.36	.63
31	2	.42	.68
32	6	.50	.75
33	WNL	.37	.64
34	WNL	.34	.61
35	WNL	.22	.45
36	31	.32	.58

37	WNL	.26	.51
38	1	.57	.80
39	3	.11	.27
40	34	.24	.49
41	WNL	.33	.60
42	WNL	.41	.67
43	468371	.28	.54
44	WNL	.39	.66
45	WNL	<b>-.09</b>	<b>-.34</b>
46	47638	.42	.69
47	WNL	.60	.82
48	WNL	.43	.70
49	WNL	.34	.61
50	6	.37	.63
51	327148	.32	.59
52	WNL	.57	.80
53	413	.47	.72
54	WNL	.28	.54
55	321674	.32	.59
56	WNL	.47	.73
57	7346281	.33	.60
58	420	.60	.82
59	5	.43	.70
60	4	.61	.82
61	5	.74	.90
62	WNL	.42	.69
63	642	.57	.80
64	376254	.28	.54
65	782630	.09	.23
Mean	-	.36	.58
Mean with deletions	-	.38	.62

*Note.* Values in bold indicate correlations significant at the .01 level. Profiles marked \* were deemed invalid by some, but not all, CBTIs

Additional screenings were performed with respect to diagnostic reliability (inter-CBTI reliability) of CBTIs as well as the inter-interpretor reliability for CBTI reports to examine whether there were any cases which evidenced marked divergence with respect to diagnoses based on CBTI reports, similar to the screening for criterion sorts described above. Firstly, inter-CBTI diagnostic reliability, reported in Table 5, was calculated for the entire sample by

obtaining mean Q-sorts, aggregated across three raters, for each profile-CBTI combination, resulting in a total of 260 mean sorts (65 profiles multiplied by four CBTI programs). Next, the mean sorts obtained were grouped according to the 65 MMPI-2 profiles, with four different mean sorts representing the four CBTIs for each profile. Each of these 65 groupings was used to provide a coefficient alpha value, which would indicate the reliability (similarity) in diagnostic suggestions offered by the four CBTIs for the same MMPI-2 profile.

Table 5. Inter-CBTI Reliability

Profile No.	Coefficient Alpha	AAA	PA	PS	PAR
1	.92	.86	.90	.92	.90
2	.73	.63	.67	.63	.75
3	.94	.91	.91	.92	.95
4	.93	.92	.90	.90	.92
5	.82	.65	.72	.95	.67
6	.94	.90	.90	.96	.91
7	.94	.91	.90	.94	.93
8	.89	.83	.88	.87	.85
9	.93	.93	.88	.90	.93
10	.94	.90	.92	.91	.96
11	.95	.93	.93	.94	.92
12	.90	.83	.85	.87	.94
13	.87	.82	.80	.91	.80
14	.86	.81	.76	.83	.88
15	.94	.90	.89	.91	.96
16	.94	.93	.90	.94	.94
17	.86	.81	.79	.79	.87
18	.95	.93	.92	.94	.94
19	.87	.77	.74	.84	.93
20	.93	.94	.89	.90	.89
21	.96	.93	.95	.96	.95
22	.95	.94	.96	.94	.93
23	.94	.91	.91	.90	.94
24	.91	.91	.85	.87	.92
25	.97	.96	.96	.95	.98
26	.93	.88	.89	.93	.94
27	.90	.87	.83	.81	.95
28	.90	.84	.84	.88	.93
29	.87	.83	.81	.92	.75

30	.95	.95	.92	.97	.92
31	.93	.90	.93	.91	.91
32	.87	.83	.91	.79	.78
33	.93	.87	.90	.87	.96
34	.90	.82	.81	.82	.98
35	.95	.92	.92	.94	.96
36	.80	.67	.86	.74	.65
37	.87	.87	.77	.89	.78
38	.77	.71	.67	.65	.81
39	.95	.92	.92	.96	.93
40	.94	.94	.90	.92	.90
41	.89	.96	.80	.80	.83
42	.91	.90	.87	.85	.89
43	.75	.80	.58	.67	.71
44	.40	.48	.28	.42	.08
45	.54	.46	.36	.58	.49
46	.88	.84	.84	.84	.86
47	.90	.84	.82	.85	.94
48	.94	.89	.91	.89	.96
49	.94	.93	.91	.93	.93
50	.81	.75	.85	.69	.73
51	.88	.85	.84	.87	.79
52	.96	.93	.94	.94	.97
53	.89	.81	.88	.87	.86
54	.92	.90	.86	.87	.94
55	.87	.79	.79	.91	.86
56	.94	.92	.92	.93	.91
57	.90	.88	.83	.93	.85
58	.91	.89	.89	.84	.90
59	.87	.78	.83	.81	.90
60	.92	.88	.92	.86	.93
61	.93	.94	.88	.92	.90
62	.95	.93	.92	.92	.97
63	.84	.77	.78	.71	.90
64	.93	.92	.91	.91	.91
65	.93	.89	.91	.90	.92
Mean	.89	.85	.84	.86	.87

*Note.* Reliability coefficients for CBTI programs indicate alpha-if-CBTI-deleted values.

Results indicated overall high reliability coefficients across the 65 MMPI-2 profiles, with the mean alpha value across 65 profiles being .89. The three profiles with negative criterion sort alpha values were included for the purpose of examining inter-CBTI reliability and obtained

alphas equal to .54, .93, and .94, respectively. These results indicate that, overall, CBTIs provide consensual diagnostic information. Hence, no profiles were screened on the basis of inter-CBTI reliability.

Secondly, inter-interpretor reliability for CBTI reports was calculated to measure the extent of diagnostic agreement among the three raters and is reported in Table 6. Inter-interpretor reliability was calculated by obtaining alpha coefficients based on Q-sort ratings from all three raters for each of the 260 (65 profiles by 4 programs) Q-sorts. The average of 260 coefficient alpha values obtained was calculated as the average inter-interpretor reliability and provided an index of the degree of similarity in the raters' comprehension of diagnoses offered by the CBTI reports.

Table 6. Inter-interpretor Reliability for CBTI-based Diagnostic Q-sorts

Profile No.	PA	AAA	PAR	PS	Mean
1	.83	.81	.78	.76	.79
2	.83	.75	.69	-.14	.53
3	.57	.49	.70	.72	.62
4	.66	.86	.79	.90	.80
5	.95	.87	.89	.57	.82
6	.69	.88	.79	.49	.71
7	.82	.84	.62	.77	.76
8	.74	.73	.72	.75	.73
9	.48	.74	.77	.62	.65
10	.64	.62	.49	.82	.64
11	.70	.82	.73	.62	.72
12	.71	.62	.49	.33	.54
13	.84	.78	.77	.87	.81
14	.74	.32	.49	.68	.56
15	.85	.77	.50	.85	.74
16	.84	.85	.58	.71	.74
17	.80	.69	.73	.64	.71
18	.82	.82	.72	.52	.72
19	.58	.48	-.26	.86	.41
20	.58	.76	.59	.82	.69
21	.62	.54	.65	.76	.64

22	.83	.85	.82	.82	.83
23	.47	.62	.27	.72	.52
24	-.05	.09	.86	-.05	.21
25	.55	.52	.58	.54	.55
26	.55	.56	.21	.74	.51
27	.54	.73	.30	.58	.54
28	.72	.63	-.17	.76	.48
29	.93	.72	.83	.27	.69
30	.75	.71	.73	.87	.76
31	.55	.74	.60	.70	.65
32	.60	.05	.52	.53	.42
33	.52	.39	.21	.63	.44
34	.57	.58	.43	.59	.54
35	.45	.55	.62	.78	.60
36	.14	.77	.63	.56	.52
37	.88	.85	.61	.54	.72
38	.63	.79	.45	.28	.54
39	.86	.90	.76	.83	.84
40	.81	.92	.89	.88	.87
41	.61	.25	.60	.56	.50
42	.64	.76	.67	.51	.64
43	.70	.87	.72	.71	.75
44	.76	.69	.77	.73	.74
45	-.05	.66	.37	.63	.40
46	.78	.51	.85	.53	.67
47	.80	.16	.11	.87	.48
48	.80	.55	.18	.53	.51
49	.73	.40	.27	.77	.54
50	.66	.60	.49	.72	.62
51	.87	-.11	.82	.58	.54
52	.45	.52	.73	.72	.60
53	.90	.88	.72	.79	.82
54	.61	.82	.71	.82	.74
55	.79	.72	.77	.64	.73
56	.39	.61	.66	.72	.59
57	.87	.86	.73	.67	.78
58	.86	.69	.69	.82	.76
59	.81	.57	.77	.42	.64
60	.78	.80	.90	.77	.81
61	.74	.81	.54	.74	.71
62	.60	.56	.74	.53	.61
63	.84	.71	.72	.73	.75
64	.86	.79	.76	.73	.78
65	.74	.84	.70	.81	.77
Mean	.68	.65	.60	.65	.65

As seen in Table 6, the overall mean inter-rater reliability for all four CBTIs was found to be .65. The highest inter-rater reliability was obtained for the Pearson CBTI and the lowest inter-rater reliability value was obtained for the PAR CBTI. The three profiles with negative criterion sort alpha values (numbers 3, 20, and 45) were included for the purpose of examining inter-CBTI reliability and obtained mean alphas equal to .62, .69, and .40, respectively. These results indicate that, overall, diagnostic information provided by CBTIs was understood consensually by the raters. Hence, no profiles were screened on the basis of inter-interpretor reliability.

### Diagnostic Validity

The first question raised in this study was regarding the diagnostic validity of CBTIs for the MMPI-2 which involved an examination of the level of consensus between diagnoses based on information provided by clinicians with diagnoses based on CBTI reports. With this screened data set of 62 profiles, the first step in calculation of diagnostic validity entailed computing mean Q-sorts which were aggregated across three raters to obtain mean Q-sort ratings for each of the 62 criterion sorts, resulting in 62 mean Q-sorts. The second step involved using the same process of obtaining mean CBTI-based sorts, aggregated across three raters, for each program, per profile. For example, obtaining the mean Q-sort (based on sorts by three raters) for the first profile interpreted by the AAA program, followed by the mean Q-sort (based on sorts by three raters) for the second profile interpreted by the AAA program, and so on till mean Q-sorts for each profile-CBTI combination were obtained. This resulted in a total of 248 mean Q-sorts, with 62 Q-sorts per program ( $4 \times 62 = 248$ ). The third step involved correlating the mean Q-sort for each profile-CBTI combination with the respective mean criterion Q-sort obtained in the first step. For example, the mean criterion Q-sort for profile one was correlated with each of the four



mean CBTI Q-sorts (one each for AAA, PA, PS, and PAR CBTIs) to obtain four correlation coefficients. This step was repeated for each of the 62 profiles.

Table 7 provides correlation coefficients between mean criterion diagnostic Q-sorts and mean CBTI-based diagnostic Q-sorts for the four CBTI programs, along with the overall mean value. The overall correlation (across the four programs) between criterion sorts and CBTI-based sorts was found to  $.29 \pm .04$ , with the PS CBTI obtaining the highest value of .28 and the PAR CBTI obtaining the lowest value of .23. Statistically significant positive correlations (at .01 level of significance) were found for profiles numbered 8, 9, 10, 19, 21, 24, 28, 37, 44, 45, 46, 51, and 58. Profiles 5 and 32 were found to have a statistically significant (at .01 level of significance) negative correlation between criteria-based and CBTI-based diagnostic Q-sorts. Profiles obtaining a negative correlation coefficient included numbers 6, 12, 15, 18, 22, 35, 39, 40, 41, 42, 43, 47, 52, 55, 59, and 60.

Table 7. Diagnostic Validity of CBTIs

Profile	Clinical Scales with	Criterion Diagnosis	AAA	PA	PS	PAR	Mean
1	72048	GAD	.47	<b>.67</b>	.51	.27	.53
2	2406781	NPD	.45	.27	.02	-.34	.14
3	432	SA, ND	.36	.54	.25	.45	.43
4	8670	SPD	.27	.38	-.06	.24	.28
5	2	SCAF	<b>-.69</b>	-.50	-.21	<b>-.69</b>	<b>-.58</b>
6	42	FAC	-.01	.04	.24	.02	.06
7	13	GAD, SA	.04	-.12	-.13	-.15	-.10
8	WNL	HPD	<b>.65</b>	<b>.60</b>	<b>.59</b>	.44	<b>.62</b>
9	WNL	GAD	<b>.60</b>	.48	.46	.17	.47
10	WNL	MDD	<b>.66</b>	<b>.58</b>	<b>.59</b>	<b>.61</b>	<b>.66</b>
11	9	GAD	.47	.17	.21	.43	.36
12	23	SCZ,	-.06	-.09	<b>.57</b>	.12	.16
13	8	SCAF, SA ND	.21	.04	.05	-.34	-.01
14	WNL	FAC	.26	.32	.31	.48	<b>.36</b>
15	27	SCZ	-.24	-.44	-.23	-.31	-.33
16	WNL	MDD,	.20	.07	.22	.41	.27

17	7	FAC	.33	.31	.27	.34	.34
18	WNL	FAC	-.23	-.18	-.07	.05	-.13
19	5	HPD, ND	<b>.63</b>	<b>.67</b>	.50	<b>.71</b>	<b>.66</b>
20	4	FAC	.40	.26	.37	.38	.38
21	WNL	HPD	<b>.78</b>	<b>.73</b>	<b>.77</b>	.51	<b>.77</b>
22	87463210	NPD	-.17	-.35	-.28	-.50	-.38
23	WNL	OCD	.36	.47	.41	.40	.43
24	WNL	GAD, OCD	<b>.63</b>	<b>.65</b>	.56	.44	<b>.63</b>
25	WNL	OCD	.46	.36	.39	.02	.36
26	WNL	ASPD, HPD,	.48	.31	.46	.51	.49
27	13	BPD, AN	.10	.33	.17	.37	.29
28	27	SA	<b>.65</b>	<b>.71</b>	<b>.69</b>	<b>.78</b>	<b>.75</b>
29	2	SA	.53	.35	<b>.64</b>	<b>.58</b>	<b>.57</b>
30	6	BPD, SCZ, SCAF	.39	.07	.04	.12	.17
31	WNL	HPD,	.24	.46	.42	.15	.36
32	WNL	PPD	<b>-.68</b>	<b>-.71</b>	<b>-.75</b>	-.10	<b>-.65</b>
33	WNL	GAD	<b>.64</b>	.50	<b>.69</b>	.54	<b>.64</b>
34	31	FAC, SA	.46	.06	.20	<b>.67</b>	.46
35	WNL	BPD, SA	-.02	.05	.31	.21	.15
36	1	HPD	.47	.56	.01	-.37	.26
37	3	FAC	<b>.64</b>	<b>.66</b>	<b>.59</b>	.46	<b>.63</b>
38	34	SA, ASPD	.47	<b>.58</b>	<b>.69</b>	<b>.62</b>	<b>.64</b>
39	WNL	BPD	-.16	.00	.12	.1	.02
40	WNL	HPD	-.08	.39	.27	.23	.22
41	468371	FAC	-.45	-.25	-.09	-.17	-.32
42	WNL	FAC	-.05	-.33	-.05	-.16	-.25
43	47638	FAC	-.03	.23	.29	.16	.19
44	WNL	GAD	<b>.72</b>	<b>.66</b>	<b>.66</b>	.36	<b>.69</b>
45	WNL	ND	<b>.75</b>	<b>.67</b>	<b>.72</b>	<b>.62</b>	<b>.75</b>
46	WNL	PAPD, ND	<b>.69</b>	<b>.73</b>	<b>.72</b>	<b>.72</b>	<b>.77</b>
47	6	BPD	-.19	-.12	.34	-.07	.13
48	327148	SOM	.25	.49	.07	.23	.31
49	WNL	FAC, HPD	.48	.50	.55	.54	.55
50	413	SA	.23	.18	.29	.49	.34
51	WNL	HPD	<b>.67</b>	<b>.73</b>	<b>.73</b>	<b>.60</b>	<b>.76</b>
52	321674	BPD	-.11	-.16	.14	.19	.02
53	WNL	PAPD, ND	.18	.22	.45	.19	.29
54	7346281	PAPD	.30	.16	.23	.34	.29
55	420	FAC	-.21	.32	.05	-.14	.02
56	5	SA	.46	<b>.69</b>	<b>.64</b>	.22	<b>.59</b>
57	4	FAC	.50	.16	.40	.35	.39
58	5	FAC	<b>.75</b>	<b>.58</b>	.47	<b>.56</b>	<b>.65</b>

59	WNL	SOM, BPD, SCZ	-.45	-.45	-.47	-.17	-.41
60	642	FAC	-.33	.08	-.18	-.61	-.31
61	376254	GAD, OCD, SA	.42	.47	.51	<b>.64</b>	.56
62	782630	BPD, FAC	.34	.29	.34	.34	.36
Mean	-	-	.26	.26	.28	.23	.29
<b>Correlation</b>							

*Note.* Values in bold indicate statistically significant validity coefficients at the .01 level of significance. Mean correlations have not been transformed to Z-scores.

A more nuanced picture of the differences in diagnostic ratings obtained by comparing criterion-based and CBTI-based diagnostic Q-sorts may be achieved by examining the most highly rated diagnostic category for both types of Q-sorts, alongside the elevated ( $T > 64$ ) clinical scales from the respective MMPI-2 profile. For this purpose, mean Q-sorts described in steps one to three above were used to discern the diagnostic category (Q-set item number) receiving the highest average rating per sort. For example, examining the mean criteria sort for profile one showed that the category of Generalized Anxiety Disorder obtained the highest mean rating of 2 which further indicated that this category was rated as the most likely diagnosis by the average of three raters, based on their reading of information provided by treating clinicians.

Table 8 provides a comparison of the most highly rated diagnostic category from mean criterion Q-sorts as well as for mean Q-sorts based on each of the four CBTI programs, for a total of 62 MMPI-2 profiles.

**Table 8. Most Highly Rated Diagnostic Category for Each Profile Combination**

Profile	Clinical Scales with $T > 64$	Criterion Sorts	AAA	PAR	PA	PS
1	72048	GAD	OCD, MDD, GAD	OCD, MDD	MDD	GAD
2	2406781	NPD	SA	SPD	MDD, SA, ASPD	MDD, SA, SPD

3	432	SA, ND	SA	SA	SA	MDD, ASPD
4	8670	SPD	SA	SCZ-P, SA	SCZ-P, SCZ	GAD, MDD
5	2	SCAF	MDD, GAD	GAD, MDD	MDD, SA	MDD, GAD
6	42	FAC	MDD, SA, ASPD	MDD, SA	MDD, SA, ASPD	MDD, SA, ASPD
7	13	GAD, SA	SOM, CON	CON	SOM, CON	CON
8	WNL	HPD	PPD, ND	BPD, HPD	ND	SA, ND
9	WNL	GAD	ND	SA, PPD, ND	ND	NPD
10	WNL	MDD	GAD, MDD, ND	GAD, SPD, ND	ND	GAD, MDD
11	9	GAD	SA, HPD	SOM	BPD, SA	SA, HPD, NPD
12	23	SCZ, SCAF, SA	SOM, MDD	SOM, MDD	SOM, MDD	SOM
13	8	ND	SPD	SA	Schizoid PD, SPD	SPD
14	WNL	FAC	GAD, ND	SOM	ND	GAD, ND
15	27	SCZ	GAD, OCD, MDD	OCD, MDD	GAD, OCD, MDD	OCD, MDD
16	WNL	MDD, SCAF	MDD, GAD, PAPD	PPD	GAD, ND	GAD, MDD
17	7	FAC	GAD, SA	GAD, OCD	GAD, OCD, SA	GAD, MDD, SA
18	WNL	FAC	ND	SA	ND	ND
19	5	HPD, ND	ND	HPD	ND	SA, PAPD
20	4	FAC	SA, ASPD, NPD	SA, NPD	SA, ASPD, SCZ	ASPD
21	WNL	HPD	ND	SA	ND	ND
22	87463210	NPD	MDD	SCZ-P, SCZ	MDD	SA
23	WNL	OCD	ND	HPD	ND	ND
24	WNL	GAD, OCD	ND	SA	ND	NPD

25	WNL	OCD	GAD, ND	SA	ND	ND
26	WNL	ASPD, HPD, PAPD	ND	SA, NPD	GAD, ND	MDD
<b>27</b>	13	BPD, AN	SOM, CON	SOM	SOM, CON	MDD
28	27	SA	GAD, OCD, MDD	MDD	GAD, OCD, MDD	GAD, OCD, MDD
29	2	SA	GAD, MDD	GAD, MDD	MDD	GAD, MDD
<b>30</b>	6	BPD, SCZ, SCAF	PPD	PPD, NPD	PPD, Schizoid PD	PPD
31	WNL	HPD, PAPD	ND	SA	GAD, ND	ND
<b>32</b>	WNL	PPD	ND	Paranoid PD	ND	NPD, PPD
33	WNL	GAD	ND	BPD, SA, HPD	ND	GAD
34	31	FAC, SA	SOM	SOM, SA	MDD	SOM, SA
<b>35</b>	WNL	BPD, SA	GAD, ND	SOM	ND	MDD, SPD
<b>36</b>	1	HPD	SOM	SA	SOM	SA
37	3	FAC	CON	HPD	SOM, GAD	SA, HPD
38	34	SA, ASPD	CON, HPD, PAPD	HPD	PAPD	NPD
<b>39</b>	WNL	BPD	ASPD	HPD, NPD, ND	ND	ND
<b>40</b>	WNL	HPD	ND	NPD	ND	NPD
<b>41</b>	468371	FAC	SCZ – P	SA	MDD, SPD	SOM, SA
<b>42</b>	WNL	FAC	GAD, PAPD, ND	SPD	ND	SOM
<b>43</b>	47638	FAC	GAD, MDD	MDD, BPD, SPD	ND	GAD, MDD
44	WNL	GAD	ND	SA	GAD, ND	GAD, ND
45	WNL	ND	ND	PAPD	GAD, ND	ND
46	WNL	PAPD, ND	ND	HPD, NPD	ND	SA

<b>47</b>	6	BPD	SA	PPD, NPD	SA, PPD	SA, PPD
48	327148	SOM	MDD	OCD	GAD, MDD	SPD
49	WNL	FAC, HPD	ND	SA, HPD, NPD	ND	NPD
50	413	SA	SOM, ASPD	SOM, SA	SOM, SA	PAPD
51	WNL	HPD	GAD, ND	BPD, NPD	ND	HPD, ND
<b>52</b>	321674	BPD	MDD	GAD	MDD,GAD	SOM
<b>53</b>	WNL	PAPD, ND	GAD	Schizoid PD	ND	OCD, MDD
<b>54</b>	7346281	PAPD	GAD, OCD, MDD	GAD, OCD, SOM	GAD	MDD, SOM, SPD
<b>55</b>	420	FAC	MDD, SA	MDD, SA	MDD, SA	MDD
56	5	SA	ND	SA	ND	SA
57	4	FAC	MDD	MDD, SA	SA, ASPD	ASPD, NPD
58	5	FAC	ND	ND	ND	GAD, MDD, SPD, HPD, ND
<b>59</b>	WNL	SOM, BPD, SCZ	ND	SA, ASPD, NPD	ND	ND
<b>60</b>	642	FAC	ASPD	PPD	PPD	ASPD, PPD
61	376254	GAD, OCD, SA	SOM, GAD	GAD, MDD	GAD	SOM, GAD, OCD, MDD
62	782630	BPD, FAC	SPD	OCD	GAD, MDD	SCAF, SPD, SA

*Note.* NPD = Narcissistic PD, SPD = Schizotypal PD, SA = Substance Abuse, BD = Bipolar Disorder, FAC = Factitious Disorder, SCAF = Schizoaffective Disorder, PPD = Paranoid PD, SCZ = Schizophrenia, SCZ-P = Schizophrenia – Paranoid Type, ND = No Diagnosis, SOM = Somatization Disorder, CON = Conversion Disorder, PAPD = Passive Aggressive PD, HPD = Histrionic PD, AN = Anorexia Nervosa. Profiles numbered in bold obtained diagnostic validity coefficients lower than .30.

Significantly different diagnostic differences between Q-sorts based on information provided by therapists and those based on CBTI reports (as seen by correlation coefficients lower

than .30) were seen for profiles numbered 2, 4, 5, 6, 7, 12, 15, 16, 18, 22, 27, 30, 32, 35, 36, 39, 40, 41, 42, 43, 47, 52, 53, 54, 55, 59, and 60.

A second question raised in this study was whether the diagnostic validity is higher for specific diagnostic categories. To study this further, profiles having criterion sorts with an alpha value of .69 or higher (designation as high criterion reliability profiles) were examined as a subset of the sample. A total of 22 profiles met this cut-off alpha value for criterion sort reliability. Table 9 provides information on these profiles with respect to elevated MMPI-2 clinical scales ( $T > 64$ ), most highly rater diagnostic category based on mean criterion sorts, inter-CBTI reliability values, and diagnostic validity values (Q-correlations between criterion sorts and CBTI-based sorts completed by raters) for each CBTI-profile combination.

Table 9. Diagnostic Validity for Profiles with High Criterion Sort Reliability Values

Profile	Clinical Scales with $T > 64$	Criterion Diagnosis	Inter-CBTI Reliability	AAA	PA	PS	PAR	Mean Validity
1	72048	GAD	.92	.47	<b>.67</b>	.51	.27	.53
2	2406781	NPD	.73	.45	.27	.02	-.34	.14
5	<b>2</b>	<b>SCAF</b>	.94	<b>-.69</b>	-.50	-.21	<b>-.69</b>	<b>-.58</b>
8	<b>WNL</b>	<b>HPD</b>	.93	<b>.65</b>	<b>.60</b>	<b>.59</b>	.44	<b>.62</b>
10	<b>WNL</b>	<b>MDD</b>	.95	<b>.66</b>	<b>.58</b>	<b>.59</b>	<b>.61</b>	<b>.66</b>
11	9	GAD	.90	.47	.17	.21	.43	.36
17	7	FAC	.95	.33	.31	.27	.34	.34
19	<b>5</b>	<b>HPD,</b>	.96	<b>.63</b>	<b>.67</b>	.50	<b>.71</b>	<b>.66</b>
20	4	FAC	.95	.40	.26	.37	.38	.38
21	<b>WNL</b>	<b>HPD</b>	.94	<b>.78</b>	<b>.73</b>	<b>.77</b>	.51	<b>.77</b>
30	6	BPD, SCAF, SCZ	.87	.39	.07	.04	.12	.17
36	1	HPD	.77	.47	.56	.01	-.37	.26
44	<b>WNL</b>	<b>GAD</b>	.90	<b>.72</b>	<b>.66</b>	<b>.66</b>	.36	<b>.69</b>
45	<b>WNL</b>	<b>ND</b>	.94	<b>.75</b>	<b>.67</b>	<b>.72</b>	<b>.62</b>	<b>.75</b>
49	WNL	FAC, HPD	.96	.48	.50	.55	.54	.55

50	413	SA	.89	.23	.18	.29	.49	.34
53	WNL	PAPD, SA	.94	.18	.22	.45	.19	.29
55	420	FAC	.91	-.21	.32	.05	-.14	.02
56	5	SA	.87	.46	<b>.69</b>	<b>.64</b>	.22	<b>.59</b>
57	4	FAC	.92	.50	.16	.40	.35	.39
58	<b>5</b>	<b>FAC</b>	.93	<b>.75</b>	<b>.58</b>	.47	<b>.56</b>	<b>.65</b>
60	642	FAC	.84	-.33	.08	-.18	-.61	-.31
Mean Correlation	-	-	.90	.39	.38	.35	.23	.38

*Note.* Correlation values in bold indicate statistical significance at the level of .01. Mean correlations have not been transformed to Z-scores.

As seen in Table 9, the overall mean diagnostic validity for profiles with high criterion sort reliability was found to be  $.38 \pm .07$ . In general, the WNL profiles obtained high validity coefficients, except for profile number five. The diagnostic category of Factitious Disorder comprised almost 32% of this subset. Other diagnostic categories with high representation included Generalized Anxiety Disorder and Histrionic Personality Disorder. Profiles 2 and 60 obtained negative mean validity coefficients.

The WNL profiles comprised almost 40% of the total profiles included in this sample and warrant a more detailed examination. Table 10 presents characteristic of the WNL profiles including diagnostic validity values, inter-CBTI reliability values, and the most highly rated diagnostic category from mean criterion Q-sorts as well as for mean Q-sorts based on each of the four CBTI programs. Profiles 18, 42, and 59 obtained negative correlations between mean criterion and mean CBTI diagnostic sorts, but received moderate to high inter-CBTI reliability values. Diagnostic validity coefficients for these profiles ranged from  $-.65$  to  $.77$ , with an average value of  $.35$ .



Table 10. Characteristics of WNL Profiles

Profile	Criterion Sort Reliability	Diagnostic Validity	Inter-CBTI Reliability	Criterion Sorts	AAA	PAR	PA	PS
8	.55	.62	.93	HPD	PPD, ND	BPD, HPD	ND	SA, ND
9	.35	.47	.94	GAD	ND	SA, PPD, ND	ND	NPD
10	.55	.66	.95	MDD	GAD, MDD, ND	GAD, SPD, ND	ND	GAD, MDD
14	.34	.36	.94	FAC	GAD, ND	SOM	ND	GAD, ND
16	.08	.27	.86	MDD, SCAF	MDD, GAD, PAPD	PPD	GAD, ND	GAD, MDD
<b>18</b>	.17	<b>-.13</b>	<b>.87</b>	<b>FAC</b>	<b>ND</b>	<b>SA</b>	<b>ND</b>	<b>ND</b>
<b>21</b>	.43	<b>.77</b>	<b>.94</b>	<b>HPD</b>	<b>ND</b>	<b>SA</b>	<b>ND</b>	<b>ND</b>
23	.22	.43	.97	OCD	ND	HPD	ND	ND
24	.27	.63	.93	GAD, OCD	ND	SA	ND	NPD
25	.28	.36	.90	OCD	GAD, ND	SA	ND	ND
26	.35	.49	.90	ASPD, HPD, PAPD	ND	SA, NPD	GAD, ND	MDD
31	.37	.36	.93	HPD, PAPD	ND	SA	GAD, ND	ND
<b>32</b>	.34	<b>-.65</b>	<b>.90</b>	<b>PPD</b>	<b>ND</b>	<b>Paranoid PD</b>	<b>ND</b>	<b>NPD, PPD</b>
33	.22	.64	.95	GAD	ND	BPD, SA, HPD	ND	GAD
35	.26	.15	.87	BPD, SA	GAD, ND	SOM	ND	MDD, SPD
39	.33	.02	.89	BPD	ASPD	HPD, NPD, ND	ND	ND
40	.41	.22	.91	HPD	ND	NPD	ND	NPD
<b>42</b>	.39	<b>-.25</b>	<b>.40</b>	<b>FAC</b>	<b>GAD, PAPD, ND</b>	<b>SPD</b>	<b>ND</b>	<b>SOM</b>
44	.60	.69	.90	GAD	ND	SA	GAD, ND	GAD, ND

<b>45</b>	.43	<b>.75</b>	<b>.94</b>	<b>ND</b>	<b>ND</b>	<b>PAPD</b>	<b>GAD,</b> <b>ND</b>	<b>ND</b>
<b>46</b>	.34	<b>.77</b>	<b>.94</b>	<b>PAPD,</b> <b>ND</b>	<b>ND</b>	<b>HPD,</b> <b>NPD</b>	<b>ND</b>	<b>SA</b>
49	.57	.55	.96	FAC, HPD	ND	SA, HPD, NPD	ND	NPD
<b>51</b>	.28	<b>.76</b>	<b>.92</b>	<b>HPD</b>	<b>GAD,</b> <b>ND</b>	<b>BPD,</b> <b>NPD</b>	<b>ND</b>	<b>HPD,</b> <b>ND</b>
53	.47	.29	.94	PAPD, ND	GAD	Schizoid PD	ND	OCD, MDD
<b>59</b>	.42	<b>-.41</b>	<b>.95</b>	<b>SOM,</b> <b>BPD,</b> <b>SCZ</b>	<b>ND</b>	<b>SA,</b> <b>ASPD,</b> <b>NPD</b>	<b>ND</b>	<b>ND</b>

---

*Note.* Values in bold indicate either negative correlation coefficients or coefficients having a value greater than .69. Mean correlations have not been transformed to Z-scores.

## CHAPTER IV

### DISCUSSION

The first question raised in the present study pertained to the diagnostic validity of four commercially available CBTIs for the MMPI-2. Specifically, this study examined the extent to which diagnostic suggestions based on descriptive statements on clients provided by treating clinicians are similar to diagnostic suggestions made based on computerized interpretive reports available for the MMPI-2. The second question raised in this study entailed examining whether specific diagnostic categories showed higher consensus with respect to comparing diagnostic information based on client descriptions provided by treating clinicians and diagnostic information from CBTI reports. This study also examined the validity and reliability of WNL profiles, which comprised 40% of the total profiles included in this study. The third and final question pertained to the relative diagnostic validity of CBTIs and examined whether specific programs performed significantly better or worse than others.

#### Reliability of Criterion Q-Sorts

The source used in this study for criterion diagnostic Q-sorts was ranked descriptive statements on clients obtained from Midwestern Q-sorts completed by the clinicians providing treatment to those clients. Before comparing these criterion sorts with CBTI-based sorts, it was necessary to evaluate the extent to which descriptive information provided by clinicians was reliably (or consensually) understood by the three raters who used it to complete the criterion diagnostic sorts. Three profiles were removed from the sample due to having negative reliability values, indicating that raters mostly disagreed on their diagnostic judgments for these three profiles.

As an example, one of the profiles that was deleted from the sample during screening for unreliability had a negative mean inter-rater reliability coefficient (for the criterion sorts) of  $-.09$  and an alpha value of  $-.34$ , indicating significant differences in rankings made by the three raters for this profile. The most highly ranked statements provided by therapists for this profile included ‘feels misunderstood by others,’ ‘is involved in a variety of activities,’ ‘is perfectionistic,’ ‘needs to be more assertive,’ and ‘shows good judgment’ whereas the lowest rated statements (least characteristic of the client) included ‘discusses committing suicide,’ ‘needs attention,’ ‘needs less affection than most people,’ ‘obeys authority,’ and ‘rarely daydreams.’ It is possible that raters were confused by conflicting information provided in similarly ranked statements (such as someone who does not obey authority yet needs to be more assertive). It is interesting to note that inter-CBTI reliability values for two of the three profiles were  $.93$  and  $.94$ , indicating that CBTIs were highly consensual with respect to diagnostic suggestions offered across programs. However, the third deleted profile had a lower inter-CBTI reliability value of  $.54$ . Further examination showed that this profile obtained a diverse range of highest-ranked diagnoses based on mean sorts, including categories such as Schizophrenia – Paranoid Type, OCD, GAD, and Antisocial Personality Disorder.

The remaining 62 profiles showed moderate agreement between the three raters and, in general, had moderate reliability values. Further analyses were completed using this sample of 62 profiles.

### Diagnostic Validity

To address the first question regarding diagnostic validity of CBTIs for the MMPI-2, results indicated that the overall value for correlations between diagnostic criterion sorts and

CBTI-based sorts was significantly low in this sample. There was also some variation in terms of high and low correlations coefficients obtained for specific profiles which warrants further examination.

Profiles 5 and 32 obtained statistically significant negative correlations with their respective criterion sorts. Upon examining the diagnostic suggestions suggested for profile 5 in Table 8, the highest rated category for mean criteria sorts was Schizoaffective Disorder whereas the highest rated categories for CBTIs varied from MDD and GAD for AAA, PAR, and PS, and MDD and Substance Abuse for PA. In this case, the CBTIs provided similar diagnoses with respect to one another. An examination of the most highly rated descriptive statements for this profile (provided to raters completing criterion sorts) shows that they included ‘acts coarse and tough,’ ‘behaves oddly,’ ‘sees or hears things that do not exist,’ ‘is hopeful,’ and ‘avoids responsibility because of health problems.’ A diagnosis of MDD and GAD appears to be less probable based on these statements, as compared to a diagnosis of Schizoaffective Disorder.

For profile 32, the highest rated category for mean criterion sorts was Paranoid PD whereas the highest rated categories for CBTIs varied from No Diagnosis for AAA and PA, Paranoid PD for PAR, and Narcissistic PD and Paranoid PD for PS. An examination of the most highly rated descriptive statements for this profile (provided to raters completing criterion sorts) shows that they included ‘acts on impulse,’ ‘is irritable,’ ‘develops health problems when under stress,’ ‘makes decisions with ease,’ and ‘values intellectual activities.’ It is likely that this profile obtained a low correlation due to there being differences in ranked diagnoses throughout the Q-set categories, as opposed to solely the highest ranked ones.

Previous research on validity of CBTIs for the MMPI-2 has found that the correlation between subjects’ CBTI reports based on MMPI/MMPI-2 scale scores and their therapists’ Q-

sort ratings ranges from .07 (Harrington, 2000) to .36 (Crumpton, 1974; Moreland, 1983). Results from the present study obtained validity coefficients in the same low to moderate range as previous studies. It is probable that the validity values obtained in this study were reduced by the low reliability of criterion sorts due to ambiguities in ranked descriptive statements provided by treating clinicians. Examples of potentially confusing information found in ranked descriptive statements across profiles include ‘has little energy’ and ‘is involved in a variety of activities’; ‘reasons logically’ and ‘believes in things that are obviously not true’; ‘acts relaxed’ and ‘trembles, sweats, and shows other physical signs of anxiety’; ‘acts depressed’ and ‘is hopeful’; ‘is hopeful’ and ‘discusses committing suicide’; and ‘is realistic’ and ‘believes in things that are obviously not true.’ Having such opposing statements ranked similarly is likely to have significantly contributed towards lowering inter-rater reliability for the criterion Q-sorts as raters are likely to have resorted to their own clinical judgment to make diagnostic decisions.

To examine the effect of criterion reliability on diagnostic validity, a subset of 25 profiles with high criterion reliability values (coefficient alpha greater than .69) was selected for comparison with respect to elevated MMPI-2 clinical scales ( $T > 64$ ) and examined on the basis of most highly rated diagnostic category according to mean criterion sorts, inter-CBTI reliability values, and diagnostic validity values (Q-correlations between criterion sorts and CBTI-based sorts completed by raters) for each CBTI-profile combination. As seen in Table 9, this subset had an overall mean diagnostic validity of .38, which is moderately higher than the validity coefficient of .29 for the entire sample, confirming some effect of low reliability of diagnostic criterion sorts on the diagnostic validity of CBTIs in this study.

Examining the mean validity values in the last column of Table 9 helps to answer the second question raised in this study and provides information on specific diagnoses that may

have obtained greater consensus between client descriptions provided by treating clinicians and information provided by CBTI reports on MMPI-2 scores. Specifically, Histrionic Personality Disorder (HPD) and Generalized Anxiety Disorder (GAD) were found to obtain comparatively higher validity coefficients. It is also evident from Table 9 that a few of these profiles had MMPI-2 clinical scales scores that were within normal limits. Among the Diagnostic Q-sort items, GAD and HPD are likely to be rated in the most likely category for WNL profiles as they constitute diagnostic categories that are lower in terms of serious psychopathology, as compared to some of the other items in the Q-set. Additionally, the Midwestern Q-sort has items that are quite clearly indicative of these diagnostic categories, for example, ‘acts anxious,’ ‘trembles, sweats, or shows other physical signs of anxiety,’ ‘worries about being accepted by other,’ and ‘needs attention.’ This may have resulted in higher validity ratings for these two categories.

The most frequent highly rated diagnostic category for criterion sorts with high reliability values was Factitious Disorder, and it comprised almost 32% of that subset. As shown in Table 9, there is no clear pattern with respect to elevations in MMPI-2 clinical scales for these profiles which had elevations on clinical scales including Psychasthenia, Psychopathic Deviate, Masculinity-Femininity, and combinations of scales including 642 and 420. It is important to note that these profiles obtained high inter-CBTI reliability values and Table 8 shows that the CBTIs generally identified them with diagnostic categories which would be more aligned with their elevated MMPI-2 clinical scale(s). For example, profile 17 had a highest-rated criterion diagnosis of Factitious Disorder, had an elevation on clinical scale 7 (Psychasthenia) on the MMPI-2, and obtained highest-rated CBTI diagnoses of Generalized Anxiety Disorder, Obsessive Compulsive Disorder, Major Depressive Disorder, and Substance Abuse across the four CBTI programs. The highest-ranked CBTI diagnoses appear to be more aligned with the

obtained MMPI-2 elevation on a clinical scale assessing symptoms of anxiety. This trend of highest-rated CBTI diagnostic categories being more aligned with MMPI-2 clinical scale elevations was also seen for profiles 20, 55, 57, and 58, which also received a highest-rated diagnosis of Factitious Disorder based on criterion sorts. These findings further indicate that low reliability of criterion sorts is likely to have impacted the validity of CBTIs, resulting in lower values.

Low reliability of criterion sorts could have been influenced by a number of factors. As discussed previously, ambiguous information provided as similarly ranked statements by treating clinicians may have caused the raters to use their own clinical judgment in completing the sorts. Additionally, the format in which information on clients was provided to raters completing criterion sorts may have been somewhat more tenuous, as compared to the CBTI reports which usually have a section on diagnostic suggestions. Many statements in the Midwestern Q-sort could be applied to descriptions of normal personality traits and may not relate to psychodiagnostic categories in a straightforward way. Further, the distribution of the Midwestern Q-sort includes 30 statements in the 'neutral' category and only five statements each in the 'most descriptive' and 'least descriptive' categories, which could have made it difficult to discern most likely diagnostic categories. These factors may have augmented raters' reliance on idiosyncratic patterns in ranked statements to make diagnostic decisions, thereby negatively affecting reliability.

This sample of 62 profiles also contained a significant number of WNL profiles, which comprised almost 40% of the total profiles. Examining characteristics of these profiles in Table 10 indicates that four profiles had negative diagnostic validity coefficients, with profile 32 having the lowest value of  $-.62$ . An examination of the most highly rated descriptive statements



for this profile (provided to raters completing criterion sorts) shows that they included ‘accepts advice and suggestions,’ ‘is involved in a variety of activities,’ ‘is irritable,’ ‘sees or hears things that do not exist,’ and ‘worries about own sexual feelings.’ The mean correlation between the three criterion diagnostic sorts for this profile was .32 and it received a highest-rated mean diagnosis of Paranoid Personality Disorder. An examination of the validity scales of this MMPI-2 profile did not show any clinically significant elevations which may have indicated defensive responding on the MMPI-2. Thus, it is unclear why the clinician-rated statements were indicative of psychopathology for this WNL profile.

Inter-CBTI reliability values for WNL profiles were found to be significantly high in general. Each program identified a WNL profile as being within normal limits in the interpretive report, however, there was some variability among CBTI programs with respect to reports generated for WNL profiles. For example, the Pearson CBTI did not provide any diagnostic information for WNL profiles and included information on content scales that were elevated. On the other hand, the WNL report by PAR provided interpretive information based on a best-fit model with prototypical profiles and included diagnostic suggestions. Given this information, it is impressive that the CBTIs obtained such high reliability coefficients. One possible reason for this may be that the usual diagnostic suggestions by programs for WNL profiles typically included various forms of Adjustment Disorder which is a category that is not included in the Diagnostic Q-sort, leaving the raters to sort based on the narrative which usually highlighted the absence of psychopathology. Additionally, in this study, one of the raters completing diagnostic Q-sorts based on CBTI reports used the same sort for all explicitly identifiable WNL reports (based on information in the report which described it as being within normal limits). This may have increased reliability values to a small extent for some profiles.

Results obtained in this study indicated overall high inter-CBTI reliability values across the four programs, even though the reliability coefficients were, in general, lower than those obtained in a previous study by Pant and Weed (2008). Examining the most highly ranked diagnostic categories across the four CBTIs for these profiles in Table 8 indicates a few cases of differences between diagnostic suggestions offered by CBTI programs for the same profile. For example, profile 2 interpreted by the AAA CBTI has Substance Abuse as the most highly ranked diagnostic category whereas the PAR program for the same profile has Schizotypal PD as the most highly ranked category. The two other CBTIs (PA and PS) include either or both categories of Substance Abuse and Schizotypal PD, which clarifies why the coefficient alpha would still be high (being the average of all possible split half combinations).

The final question regarding relative diagnostic validity of CBTIs was answered by examining the last row in Table 7 which provides mean validity coefficients across the four programs. The PAR CBTI had the lowest overall validity (.23) and the PS CBTI had the highest overall validity (.28) with respect to the criterion information. The range of validity coefficients was from .23 to .29, indicating that CBTIs do not differ significantly with each other. The CBTI by PAR obtained significantly lower correlations with criterion sorts, as compared to other programs, for a few profiles. For example, the PAR CBTI obtained a correlation of .17 with the criterion sort for profile 9 whereas other programs obtained correlations above .45. For profile 25, the PAR CBTI obtained a correlation of .02 whereas the other programs obtained correlations above .35. It is interesting to note that both these profiles are within normal limits. A possible explanation of lower values obtained by the PAR program for these profiles is that it provides an interpretive report based on a 'best-fit prototype' as opposed to a scale-by-scale or MMPI-2 codetype based interpretation which is used by other programs.

## Limitations

A limitation of this study is that the diagnostic information for criteria Q-sorts was obtained from descriptive statements which were then used by raters to generate Q-sorts. A more direct criterion could be diagnostic Q-sorts completed by treating clinicians themselves so as to avoid potentially reducing the reliability of diagnostic judgments which are to be used as criterion measures. Given the low inter-rater reliability values for some of the profiles, criterion sorts completed by treating clinicians may present more reliable diagnostic judgments. However, it is also possible that using criterion sorts based on one person's clinical judgment may increase error, which may have otherwise been minimized by aggregating across multiple raters.

Another limitation is the sole use of external criteria to determine validity. CBTI validity studies also include "customer satisfaction" type research which asks users of CBTIs to rater computerized interpretive reports on multiple dimensions. Including both types of designs in the same study would provide a more comprehensive view of diagnostic validity.

Additionally, generalizability of this study is limited by the use of raters who are clinical psychology graduate students, instead of clinicians who are more likely to commonly use MMPI-2 CBTIs to conceptualize cases.

The Diagnostic Q-set was not modified from its original version for use in this study and lacks data on whether it accurately captures the breadth of diagnostic suggestions offered by the CBTIs. There were some diagnostic categories on some of the CBTI reports which were not included in the Q-set. Some of these include Adjustment Disorder, Delusional Disorder, NOS categories, and Panic Disorder with Agoraphobia. Additionally, the Q-set utilizes a fixed rectangular distribution with equal number of items for each category. This may be difficult to use in tasks involving diagnostic judgments as the number of neutral diagnostic categories for a

particular client is likely to be greater than the number of most likely and least likely diagnostic categories, suggestive of a more normal distribution shape.

#### Future Directions

This study examined diagnostic validity of four commercially available CBTI programs by using information provided by treating clinicians as the criterion. A more comprehensive diagnostic validity study may be designed to include psychologists as participants and asking them to submit deidentified MMPI-2 protocols on recent/current clients, as well as completing the Diagnostic Q-sort for each MMPI-2 protocol submitted. Additionally, they can be asked to complete a demographic questionnaire and a Likert-type rating scale on aspects of the CBTI report(s) they received, such as diagnostic accuracy, presence of irrelevant information, presence of contradictory diagnoses etc. Barnum effects (vague personality descriptions that seem to apply to everyone due to their ambiguity) can be controlled by providing either authentic or dummy reports (based on generic MMPI-2 profiles) to participants and comparing their ratings. This study would include both “external criteria” and “customer satisfaction” designs, thereby countering some of the limitations of the present study.

Additionally, further research on the Q-set regarding effects of type of distribution used, sort-resort reliability, and utility of adding additional diagnostic categories would also be useful. For example, it would be interesting to examine the effect of distribution shape, specifically, using a rectangular versus quasi-normal distribution, on validity coefficients.

To summarize, the answer to the first question regarding diagnostic validity of CBTIs indicates that the four commercially available CBTIs evaluated in this study obtained overall low to moderate validity coefficients across the 62 MMPI-2 profiles. Three profiles were excluded

from the sample due to having negative reliability values for criterion Q-sorts. The overall mean Q-correlation between criterion diagnostic sorts based on therapist ratings and those based on CBTI reports was .29. As an answer to the second question regarding higher validity of specific diagnostic categories, profiles receiving highest-ranked criterion diagnoses of Histrionic Personality Disorder and Generalized Anxiety Disorder had a higher representation among a subset of profiles, for which membership was based on high criterion sort reliability values (greater than .69). Finally, with respect to the third question regarding comparative validity of the four programs used in this study, the PAR CBTI had the lowest overall correlation (.23) with the criterion sorts and the PS CBTI had the highest overall correlation (.28) with the criterion sorts.

## APPENDIX A

### SAMPLE OF A CBTI REPORT PROVIDED TO RATERS COMPLETING CBTI-BASED Q-SORTS

Name: John 56  
Age: 44  
Sex: M

To aid in diagnosis and treatment planning, Mr. 56 was administered the Minnesota Multiphasic Personality Inventory-2.

The following test findings are based on Mr. 56's responses to a widely used standardized psychological test. As with all such tests, the validity of test results is limited by Mr. 56's honesty and self-awareness. The report below should be taken as generalized probability statements that are made without benefit of clinical interview or history. Further clinical verification is needed to assist in the interpretation of test findings in light of Mr. 56's unique history and present circumstances.

Since the MMPI-2 is a complicated test with multiple scales that measure similar constructs, at times inconsistencies in test results may occur due to Mr. 56's different elevations on similar scales. When this occurs, clinical investigation to evaluate his true status is suggested.

As psychological tests were designed primarily for diagnosis and treatment planning purposes, the findings below focus on problems, deficits and pathology and de-emphasize Mr. 56's strengths. Because of this, use without collaboration, other than for the clinical screening purposes for which they are intended, may be misleading.

The following is a CONFIDENTIAL REPORT meant for qualified Mental Health, Correctional and Substance Abuse professionals. While feedback of test findings to clients is highly encouraged and should be an integral part of therapy and treatment planning, clients should not be given copies of this report as they are probable to misunderstand report contents and their tentative nature.

#### VALIDITY OF TEST RESULTS

In testing, Mr. 56 did not appear to defensively deny having common human faults and weaknesses and was willing to indicate his shortcomings.

He showed a lack of defensiveness in testing and made few attempts to portray himself in a positive light. Mr. 56 may not feel that he is as well off as others or may be experiencing situational difficulties that leave him pessimistic.

Mr. 56 did not present himself in an grandiose way or claim an unrealistic degree of positive characteristics.

Testing shows a strong "fake bad" response set in which Mr. 56 exaggerated and distorted his problems. This limits the validity of the test findings, as Mr. 56's true level of problems/symptoms is probable to be less than what the test results indicate. Such a response set could be due to poor reading ability, emotional interference, mental confusion, and/or a cry for help, though conscious malingering also needs to be ruled out. Mr. 56 moderately over emphasized pathology in later portions of the MMPI-2, which would primarily increase Content, and Supplementary scales. He did not endorse unusual items that are not frequently acknowledged, enhancing the probable validity of test findings.

Mr. 56 was consistent in answering test questions similarly throughout the test, which enhances the probable validity of the test findings given below.

## INTELLECTUAL FUNCTIONING

Significant levels of emotional upset are reported which may interfere with memory, concentration, abstraction and judgment.

Mr. 56 is not reflective or thoughtful which can limit insight and judgment. He does not try to understand the world in cognitive, rational ways.

Concentration difficulties are probable with Mr. 56 being distractible, preoccupied, and inattentive. This may cause Mr. 56 to miss important environmental cues leading to decreased judgment and coping.

Mr. 56 is likely to be concrete in his thinking due to personality factors despite his potential level of intellectual functioning since personality factors predispose Mr. 56 to overly focus on detail and miss general trends.

Mr. 56 likes clear-cut situations and has trouble dealing with ambiguity, novelty and change.

He is an extremely cognitively rigid individual who has fixed ideas from which he has trouble deviating. Mr. 56 may fail to take in additional information or alter his opinion once an idea is formed. Poor judgment and situational misperceptions can result from reacting in terms of these fixed beliefs without seeing if they match the current situation. At this level, delusional ideation may occur.

Due to a lack of self-confidence, Mr. 56 may be indecisive and have problems with decision-making.

Severe obsessive ruminations and worries are reported that are very likely to disrupt Mr. 56's cognitive efficiency. Levels of brooding over problems exist to the point where he may lose control of his thought processes.

Mr. 56 reports severe feelings of mental dullness to the point where he now feels incapable of processing information and he does not trust his own judgment.

Significant Schizotypal features were evident in testing.

Mild to moderate levels of mental confusion are reported. Unusual thoughts and sensory experiences are not reported. Significant ego-alien ideas and feelings of unreality are reported that distress Mr. 56 and make him feel as if he is "losing his mind".

## EMOTIONAL FUNCTIONING

In testing, Mr. 56 reports having significant levels of depression that may be of clinical significance. Many subjective feelings of sadness and dejection exist, as Mr. 56 feels hopeless, helpless, and discouraged. This level of depression may be due to situational factors, may indicate Dysthymia or may show an adjustment to a chronic long-term clinical Depression that the person has in part learned to live with. Mr. 56 does not report being dysphoric or anhedonic. Many physical symptoms/signs of depression are endorsed which suggest a Major Depression. A preoccupation with his physical state may also exist as Mr. 56 denies good health and makes a wide variety of somatic complaints.

Mr. 56's level of true depression is in line with what he subjectively experiences without minimizing or focusing on feelings. Significant subjective depression is noted with Mr. 56 being quite despondent.

Such severe levels of anxiety are reported that Generalized Anxiety Disorder, Panic Attacks, Phobias, ADHD, Mania and PTSD need to be clinically ruled out.

Mr. 56 reports being so nervous that he has trouble dealing with everyday stress, pressure, and demands. He easily feels panicky, distraught, and vulnerable due to over evaluating objective danger, as Mr. 56 feels threatened by people or events commonly seen as of little or no concern. Mr. 56 is an over-ruminative worrier. Physical symptoms due to autonomic over-arousal can occur. Levels of anxiety reported are probable to interfere with coping skills, increase impulsivity and lead to aversive consequences that produce more anxiety.

Mr. 56's level of anxiety is significantly more than what he is subjectively experiencing. Strong defenses may shield him from acknowledging underlying anxiety. Mr. 56 reports a moderate subjective experience of anxiety and nervousness.

Testing does not evidence phobias and/or tendencies to develop phobias. An average level of generalized fearfulness is reported. An average number of specific fears are reported.

Significant Post Traumatic Stress Disorder symptoms are endorsed.

Reported levels of current anger are said to be average. Mr. 56 describes his present anger level as being well modulated.



Mr. 56's level of anger is similar to what he reports subjectively experiencing with him neither minimizing nor overly focusing on angry feelings. Mr. 56 now experiences average levels of subjective anger and hostility. He has average difficulty dealing with hostile and aggressive feelings and to an average degree tries to inhibit them.

He reacts with anger to a normal degree.

Mr. 56 has an average ability to control his anger once evoked. Mr. 56 reports acting in an aggressive manner to an average degree.

He consciously mildly represses feelings and does not like to deal with unpleasant topics. Mr. 56 has difficulty showing feeling and due to repression may seem formal, logical, and cautious.

Mr. 56 can accept and deal with his own feelings and does not see them as strange or foreign to himself.

Impulse control is questionable. While under many situations he can control his impulses, under stress he may not be able to do so.

Mr. 56 subjectively feels very unable to control his impulses and is extremely worried about being controlled by and/or acting out on his feelings.

Mr. 56's reported energy level is in the Normal range. Excessive hyperactivity and labile affect are not likely. A significant lack of subjective drive and ambition is reported. Mr. 56 feels a lack of energy and so may not want to engage in activities or complete tasks. Psychomotor hyperactivity was not described. Psychomotor retardation is not said to exist.

#### ALCOHOL AND USE

The MMPI-2 reveals significant addiction proneness and possible substance use. Mr. 56 tends to be extroverted, sensation seeking, and impulsive which may lead to acting out behavior including substance use. Further evaluation for possible alcohol and drug use is indicated. Mr. 56 does not acknowledge having an unusual number of personality characteristics and life situations that are often found among substance dependent individuals. Mr. 56 does not acknowledge that his substance use has had significant detrimental affects on his life.

#### SOMATIC FUNCTIONING

Current somatic concerns on a wide variety of physical problems are reported which may indicate an over concern about his physical condition. Severe GI symptoms are reported by Mr. 56 in testing. Significant neurological problems were not described.

Mild to moderate hypochondriacal complaints, often without a clear organic basis, are probable as Mr. 56 reports numerous physical symptoms. A history of vague physical signs and a preoccupation with bodily functioning is probable. Mr. 56's symptoms may in part be stress-related as he overly focuses on minor illnesses which causes more concern than would be expected.

He now feels moderately below par mentally and physically.

No significant tendencies for psychosomatic problems to develop under stress exist. Mr. 56 does not usually use physical symptoms for secondary gain. Conversion symptoms are unlikely.

## INTERPERSONAL FUNCTIONING

Mr. 56 is a mildly eccentric individual who may be lacking in social skills. His unusual behaviors and thought patterns can make it hard for others to relate to him.

Mr. 56 is mildly introverted and tends to be more comfortable when alone. Mr. 56 described a moderate degree of interpersonal avoidance and withdrawal.

Social withdrawal secondary to Mr. 56's depression may occur as he may have lost interest in daily activities and have low energy.

Mr. 56 reports experiencing significant levels of social discomfort and anxiety. He does not report significant introversion or extroversion due to social discomfort, with him being not being especially shy. Mr. 56 reports below average levels of social confidence. He questions his ability to effectively deal with and relate to others.

Mr. 56 may at times display a limited sense of empathy.

Mr. 56 reports having average needs for the attention of others.

Mr. 56 reports having below average needs for attention and may shy away from social recognition and high profile activity.

Average needs for affection, love and intimacy are reported. Mr. 56's level of suspiciousness and/or social withdrawal may interfere with Mr. 56's ability to meet his needs for love.

Mr. 56 is extremely passive and submissive. He does not want to take responsibility and can become lost without someone telling him what to do.

His dependency needs are in the average range with him having the ability to be independent and dependent as the situation requires.

Mr. 56 does not feel he is different from others. He feels misunderstood by others.

Average levels of interpersonal suspiciousness were described. Mr. 56 can blame others for his problems and generally sees the world as a threatening and unfair place. He does not especially view others as manipulating him to gain their goals. Mr. 56 has difficulty seeing the positives in people to an average degree. His level of cynicism does not usually lead to interpersonal mistrust.

Mr. 56 can be overly judgmental and critical.

He is a very self-righteous, moralistic, individual who feels that he lives up to high ethical standards. Mr. 56 often has hostile feelings when others do not live up to his standards.

He rigidly attempts to follow social rules and conventions to the letter of the law. Mr. 56 can be so rigid and rule oriented that he is judgmental, critical and authoritarian.

Average levels of family problems are reported. Mr. 56 largely feels loved and supported by his family. Conflictual family relationships and anger are not likely. He reports below average amounts of family discord. Mr. 56 is not alienated from his family. He reports below average amounts of family discord. Mr. 56 is not alienated from his family.

Moderate marital problems are reported with his marriage being conflictual and a source of stress.

Mr. 56 reports having unproductive work attitudes and behaviors that may interfere with work performance and leave him feeling unable to perform his job. An average focus on work was described with Mr. 56 not seeing himself as overly preoccupied with work. He is a patient individual who is competitive to an average degree.

## SELF IMAGE

Mr. 56 is lacking in ego strength, as he often feels overwhelmed and unable to cope with his problems. He is rigid and has trouble adapting to situations. Feelings of failure may cause low self-confidence. Feelings of self-efficacy are poor.

Mr. 56 feels unable to cope and easily withdraws into fantasy.

He is exceptionally self-critical and often focuses on the negative without acknowledging the positive in himself.

Mr. 56 has an average sensitivity to what others think of him.

He has an extremely poor self-esteem and feels unattractive and useless. Mr. 56 easily feels rejected as he projects his own feelings of being unattractive and useless.

Mr. 56 has a low opinion of himself and lacks self-esteem. He easily feels rejected as he projects his own sense of feeling unattractive and useless.

Mr. 56 has poor self-esteem and projects his own sense of feeling useless. Lowered self-esteem and feelings of being incompetent, uselessness and inadequate may in part be secondary to depression. Extreme self-doubt was apparent in testing. Mr. 56 level of self-esteem is extremely likely to result in submissiveness.

Mr. 56 feels somewhat uncomfortable and unhappy with himself. Much guilt and regret are probable.

Mr. 56 appraises his own abilities as average.

Mr. 56 appears to be questioning his value system and is exploring different values.

In testing, Mr. 56 subjectively identifies himself as having Average levels of antisocial values and past antisocial acting out. Average levels of antisocial attitudes and average levels of antisocial behavior were reported in the MMPI-2. He accepts authority and social standards. Mr. 56 does not justify that his ends justifies any means. He does not act opportunistically in his dealing with others.

Mr. 56 is questioning his value system.

## DEFENSES

Rationalization  
Intellectualization

## PERSONALITY FUNCTIONING

Mr. 56 attempts to live an orderly, over structured lifestyle and follows rigid patterns of thoughts, feelings, and actions. Because he is obsessed with doing things the right way, he often is rigid, judgmental, and moralistic. His fear of making mistakes can result in indecision and procrastination, which interfere with task completion. Preoccupation with efficiency and organization results in loss of spontaneity, stiffness, and over concern with technical details.

Mr. 56 is afraid of making mistakes and not living up to his standards, which results in feelings of worthlessness. His high standards can lead Mr. 56 to become over burdened and lead to anxiety, depressions, and feelings of helplessness, insecurity and inadequacy. Obsessive ruminations and disruptive feelings hurt memory, concentration and problem solving causing a maladaptive spiral of greater despair and lack of functioning. Stress induced somatic symptoms are probable.

Mr. 56 is over intellectualized; racing thoughts, cognitive rituals, preoccupation, and "black or white" dichotomous thinking are probable. Mr. 56 can be so strongly focused on details

that he "can't see the forest for the trees." His attention is rigid and narrow with him having a strong need for closure.

He may be very pessimistic and have many negative self-statements and obsessional worries. Mr. 56 is likely to be fearful, slow and cautious, and have trouble initiating actions. Lowered self-esteem and feelings of incompetence, uselessness and inadequacy are probable.

He is anxiety prone and often emotionally over reacts to situations. Mr. 56 is typically tense and joyless because he over controls most emotions. Intense righteous indignation occurs when others do not live up to his standards. Mr. 56 can apply similar standards to himself and can suffer from guilt. Anxiety about possible embarrassment is likely as he fears being found inadequate.

Much chronic depression and anxiety occur with Mr. 56 feeling agitated and overwhelmed by his perceived failures. This subjective discomfort is debilitating and further reduces functioning. Suicidal ideation can develop as he is highly self-punishing and easily feels hopeless.

Out of fear of failure, he can be overly rigid and attempt to follow set patterns which results in a series of nonproductive, compulsive behaviors. Passive aggressive behavior frequently can occur as Mr. 56 feels unable to directly influence his environment and seeks to escape from perceived demands without further lowering his self-esteem.

Relationship problems often exist because of Mr. 56's being critical and perfectionistic. Mr. 56 demands that others do things his way and resists control by others.

#### HARKNESS AND McNUTTY PERSONALITY PSYCHOPATHOLOGY 5 SCALES

Aggressiveness: BELOW AVERAGE

Psychoticism: AVERAGE

Disconstraint: BELOW AVERAGE

Negative Emotionality/Neuroticism: MODERATELY ABOVE AVERAGE

Introversion/Low Positive Emotionality: MILDLY ABOVE AVERAGE

#### DIAGNOSTIC CONSIDERATIONS

RULE OUT

AXIS I

Factitious Disorder, Primarily Psychological

Obsessive Compulsive Disorder

Attention Deficit Hyperactivity Disorder

Panic Attacks without Agoraphobia

Dissociative Disorder

Generalized Anxiety Disorder  
Adjustment Disorder with Mixed Emotional Features  
Major Depression, Moderate  
Phobia  
Post Traumatic Stress Disorder  
AXIS II

POSSIBLE PERSONALITY FEATURES SUGGESTED BY THE MMPI-2  
Borderline Personality features  
Obsessive features

POSSIBLE SECONDARY PERSONALITY PATTERNS INDICATED BY THE MMPI-2

The MMPI-2 does not indicate secondary personality patterns.

#### TREATMENT MOTIVATION

Mr. 56 reports having an extremely poor motivation for psychotherapy. He has marked difficulty with self-disclosure and is very reluctant to share his thoughts, feelings, and history.

#### TREATMENT CONSIDERATIONS

Based on Mr. 56's self-report, the following corrective treatment approaches are recommended. Care should be taken to ensure that these suggestions match Mr. 56's clinical presentation and history. If test invalidity indicators have been raised (see validity section), these recommendations may not reflect Mr. 56's true clinical needs.

DUE TO SIGNIFICANT LEVELS OF DEPRESSION REPORTED, CLINICAL INVESTIGATION OF POSSIBLE SUICIDAL IDEATION SHOULD OCCUR WITH NECESSARY INTERVENTIONS TAKEN.

PSYCHIATRIC REFERRAL FOR EVALUATION FOR PSYCHOTROPIC MEDICATIONS IS WARRANTED INCLUDING MEDICATION FOR:

DEPRESSION  
ANXIETY  
ADHD

As it is likely that Mr. 56's emotions are interfering with his cognitive processing, immediate interventions to alleviate emotional distress are suggested.

Mr. 56 must become more flexible and learn how to develop and choose among alternative ways of thinking, feeling, and behaving. Stress reduction techniques may help Mr. 56 deal with underlying anxiety and free him to experiment with new behaviors.

Mr. 56's attempts to dominate and control are prime therapeutic issues. Mr. 56 must develop faith in his ability to cope with situations over which he has little control and gain insight into the historical causes of his power and control issues.

Rationalizing and intellectualizing must be challenged as Mr. 56 needs to learn that what he does is much more important than his reasons and intentions.

As Mr. 56 blames others for his problems, therapists should encourage him to be responsible and accountable for his actions and not allow him to fall into Victimstance.

Mr. 56 needs to learn to regulate his moods through use of Cognitive Behavioral techniques and/or medication. AODA use may be a cause of Mr. 56's moodiness, though conversely, AODA use may be an attempt to self-medicate his emotional lability.

A Cognitive Behavioral approach to teach Mr. 56 how to acknowledge and then detach from his feelings is necessary since he is prone to acting directly on emotions without thinking. Mr. 56 needs education about the nature of emotions and must learn ways of not immediately responding once feelings arise.

Mr. 56 needs to increase impulse control and learn to see his feelings as "red flags" that call for problem solving rather than as imperatives upon which he must act. Use of Cognitive Behavioral techniques to increase cognitive mediation, to teaching problem-solving skills, increase frustration/stress tolerance through stress inoculation training, and discover impulse triggers is suggested.

Significant Depression is reported which may require Behavioral and Cognitive Behavioral treatment as well as Antidepressant medications. The role of Depression and/or Dysthymic Victimstance in Mr. 56's maladaptive behavior should be established.

High levels of anxiety are reported that may require Mental Health evaluation/treatment if they are clinically seen. Stress management techniques and alternate ways of coping with anxiety and anxiety producing situations should be taught.

Specific treatment for Post Traumatic Stress Disorder is needed if clinically seen.

A Social Learning component is suggested since Mr. 56 must learn positive, prosocial skills to replace current maladaptive patterns. Mr. 56 now relies on maladaptive tactics to meet his needs with him having few alternative prosocial coping skills. Referral for education, volunteer work, job training, etc. should occur once Mr. 56 develops the social and cognitive skills necessary to be successful.

Signs of substance abuse or proneness to abuse exist which may require AA or educational programming.

## SUMMARY

The following summarizes Mr. 56's responses to psychological tests given to him. Results are without benefit of clinical interview or history and should be taken accordingly. Please refer to the validity section to help interpret the accuracy of these test results. The Recommendation section above suggests ways of dealing with Mr. 56's problems.

## INTELLECTUAL FUNCTIONING

Significant levels of emotional upset are reported which may interfere with memory, concentration, abstraction and judgment.

Mr. 56 is not reflective or thoughtful which can limit insight and judgment. He does not try to understand the world in cognitive, rational ways.

Concentration difficulties are probable.

Mr. 56 is likely to be concrete in his thinking due to personality factors.

Mr. 56 likes clear-cut situations and has trouble dealing with ambiguity, novelty and change.

He is extremely cognitively rigid.

Due to a lack of self-confidence, Mr. 56 may be indecisive and have problems with decision-making.

Severe obsessive ruminations and worries are reported.

Mr. 56 reports severe feelings of mentally dullness.

Significant Schizotypal features were evident in testing.

Mild to moderate levels of mental confusion are reported. Unusual thoughts and sensory experiences are not reported. Significant ego alien ideas and feelings make him feel as if he is "losing his mind".

## EMOTIONAL FUNCTIONING

In testing, Mr. 56 reports significant levels of Depression. Mr. 56 does not report being dysphoric or anhedonic. Many physical symptoms/signs of depression are endorsed which suggest a Major Depression.



Mr. 56's level of true depression is in line with what he subjectively experiences with him not minimizing nor focusing on feelings. Significant subjective depression is noted with Mr. 56 being quite despondent.

Such severe levels of anxiety are reported that Generalized Anxiety Disorder, Panic Attacks, Phobias, ADHD, Mania and PTSD need to be clinically ruled out.

Mr. 56's level of anxiety is significantly more than what he subjectively experiences. Strong defenses may shield him from acknowledging underlying anxiety. Mr. 56 reports a moderate subjective experience of anxiety and nervousness.

Testing does not evidence phobias and/or tendencies to develop phobias. An average level of generalized fearfulness is reported. An average number of specific fears are reported.

Significant Post Traumatic Stress Disorder symptoms are endorsed.

Reported levels of current anger are said to be average.

Mr. 56 level of anger is in line with what he reports subjectively experiencing. Mr. 56 now experiences average levels of subjective anger and hostility. He has average difficulty dealing with hostile and aggressive feelings.

He reacts with anger to a normal degree.

Once evoked, Mr. 56 has an average ability to control his anger. Mr. 56 reports acting in an aggressive fashion to an average degree.

He mildly consciously represses feelings.

Mr. 56 can accept and deal with his own feelings and does not see them as strange or foreign to himself.

Impulse control is questionable. While under many situations he can control his impulses, under stress he may not be able to do so.

Mr. 56 subjectively feels very out of control of his impulses.

Mr. 56's reported energy level is in the Normal range. A significant lack of subjective drive and ambition is reported. Psychomotor hyperactivity was not described. Psychomotor retardation is not said to exist.

## ALCOHOL AND USE

MMPI testing reveals significant addiction proneness and possible substance use.

## SOMATIC FUNCTIONING

Current Somatic concerns about a wide variety of physical problems are reported. Severe GI symptoms are reported by Mr. 56 in testing. Significant neurological problems were not described.

Mild to moderate hypochondriacal complaints, often without a clear organic basis, are probable.

He now feels moderately below par mentally and physically.

No significant tendencies for psychosomatic problems to develop under stress exist. Conversion symptoms are unlikely.

Current severe levels of stresses and/or stress proneness can cause real stress-related physical problems to develop.

## INTERPERSONAL FUNCTIONING

Mr. 56 is a mildly eccentric individual who may be lacking in social skills.

Mr. 56 is mildly introverted.

Social withdrawal secondary to Mr. 56's depression may occur.

Mr. 56 reports experiencing significant levels of social discomfort and anxiety. Mr. 56 reports below average levels of social confidence. He questions his ability to effectively deal with others.

Mr. 56 may at times display a limited sense of empathy.

Mr. 56 reports having average needs for the attention of others.

Average needs for affection, love and intimacy are reported. Mr. 56's level of suspiciousness and/or social withdraw may interfere with Mr. 56 meeting his needs for love.

Mr. 56 is extremely passive and submissive.

His dependency needs are in the average range.

Mr. 56 does not feel he is different from others. He feels misunderstood by others.

Average interpersonal suspicion was described. Mr. 56 can blame others for his problems and generally sees the world as threatening. He does not especially view others as manipulating him.

Mr. 56 can be overly judgmental and critical.

He is a very self-righteous, moralistic, individual.

He rigidly attempts to follow social rules and conventions.

Average levels of family problems are reported.

Moderate marital problems are described.

Mr. 56 reports having unproductive work attitudes and behaviors. An average focus on work was described.

#### SELF IMAGE

Mr. 56 is lacking in ego strength, as he often feels overwhelmed and unable to cope.

Mr. 56 withdraws into fantasy a good deal.

He is exceptionally self critical.

Mr. 56 has an average sensitivity to what others think of him.

He has extremely poor self-esteem. Lowered self-esteem may in part be secondary to depression.

Mr. 56 feels somewhat uncomfortable and unhappy with himself.

Mr. 56 appraises his own abilities as average.

Mr. 56 appears to be questioning his value system and is exploring different values.

Mr. 56 subjectively identifies himself as having as having average levels of antisocial values and past antisocial acting out. Average levels of antisocial attitudes and average levels of antisocial behavior were reported in the MMPI-2. He accepts authority and social standards. Mr. 56 does not justify that his ends justifies any means.

Mr. 56 is questioning his value system.

#### DEFENSES

Rationalization  
Intellectualization

## PERSONALITY FUNCTIONING

### POSSIBLE PRIMARY FEATURES

Borderline Personality features

Obsessive features

### POSSIBLE SECONDARY PERSONALITY PATTERNS SUGGESTED BY THE MMPI

The MMPI-2 does not indicate secondary personality features.

## REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- Bagby, R. M., Marshall, M. B., Basso, M. R., Nicholson, R. A., Bacchioni, J., & Miller, L. S. (2005). Distinguishing bipolar depression, major depression and schizophrenia with the MMPI-2 clinical and content scales. *Journal of Personality Assessment, 84*, 89-95.
- Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of MMPI-2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment: A Journal of Clinical and Consulting Psychology, 3*, 634-640.
- Block, J. (1956). A comparison of forced and unforced Q-sorting procedures. *Educational and Psychological Measurement, 16*, 481-493.
- Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research*. Springfield, IL: Charles Thomas.
- Butcher, J.N., Graham, J.R., Ben-Porath, Y.S., Tellegen, A., Dahlstrom, W.G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration, scoring and interpretation (Revised Edition)*. Minneapolis, MN: University of Minnesota Press.
- Butcher, J. N. (2005). User's Guide for The Minnesota Report: Adult Clinical System - Revised, Fourth Edition [Computer Software Manual]. Minneapolis, MN: The Regent of the University of Minnesota.
- Butcher, J. N., Perry, J. N., & Atlis, M. M. (2000). Validity and utility of computer-based test interpretation. *Psychological Assessment, 12*, 6-18.

- Butcher, J. N. (1995). How to use computer-based reports. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches*. New York: Oxford University Press.
- Chase, L. S. (1974). An evaluation of MMPI interpretive systems (Doctoral dissertation, University of Minnesota, 1974). *Dissertation Abstracts International*, 35, 3009.
- Crumpton, C. A. (1974). An evaluation and comparison of three automated MMPI interpretive reports (Doctoral dissertation, University of Texas, 1974). *Dissertation Abstracts International*, 35, 6090.
- Deskovitz, M. A. (2005). *Interpretive reliability of two common MMPI-2 codetypes*. Unpublished master's thesis, Central Michigan University, Mt. Pleasant.
- Deskovitz, M. A., Weed, N. C., & Williams J. E. (2005) *Interpretive reliability of six MMPI-2 computer-based test interpretation program*. Manuscript in preparation, Central Michigan University.
- Eyde, L. D., Kowal, D. M., & Fishburne, F. J. (1991). The validity of computer based test interpretations of the MMPI-2. In T. B. Gutkin, S. L. Wise, & J.C. Conoley (Eds.), *The Computer and the Decision-Making Process* (pp. 75-123). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fowler, R. D., & Miller, M. L. (1969). Computer interpretation of the MMPI: Its use in clinical practice. *Archives of General Psychiatry*, 21, 502-508.
- Fowler, R. D. & Butcher, J. N. (1986). Critique of Matarazzo's views on computerized testing: All sigma and no meaning. *American Psychologist*, 41, 94-96.
- Graham, J. R. (2006). *MMPI-2: Assessing Personality and Psychopathology* (4th ed.). New York: Oxford.

- Green, C. J. (1982). The diagnostic accuracy and utility of MMPI and MCMI computer interpretive reports. *Journal of Personality Assessment*, *46*, 359-365.
- Greene, R. L. (2000). *The MMPI-2: An Interpretive Manual* (2nd ed.). Needham Heights, MA: Allyn & Bacon.
- Greene, R. L., Brown, R. C., & Kovan, R. E. (2006). MMPI-2 Adult Interpretive System for Windows, Professional Manual (Version 3.0) [Computer Software Manual]. Odessa, FL: Psychological Assessment Resources, Inc.
- Harrington, A. R. (2000). Relative validity of computer based interpretations of the MMPI-2 (Doctoral dissertation, University of Mississippi, 2000). *Dissertation Abstracts International*, *60*, 6364.
- Hedlund, J. L., Morgan, D. W., & Master, F. D. (1972). The Mayo Clinic automated MMPI program: Crossvalidation with psychiatric patients in an army hospital. *Journal of Clinical Psychology*, *28*, 505-510.
- Jackson, D. N. (1985). Computer-based personality testing. *Computers in Human Behavior*, *1*, 255-264.
- Jemelka, R. P., Wiegand, E. A., Walker, E. A., Trupin, E. W. (1992). Computerized Offender Assessment: Validation study. *Psychological Assessment*, *4*, 138-144.
- Lachar, D., Dahlstrom, W. G., & Moreland, K. L. (1986). Relationship of ethnic background and other MMPI 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.
- Marks, P. A. & Seeman, W. (1963). *The actuarial description of personality: An atlas for use with the MMPI*. Baltimore, MD: Williams & Wilkins.

- Matarazzo, J. D. (1986). Computerized clinical psychological test interpretations: Unvalidated plus all mean and no sigma (and replies - same issue). *American Psychologist, 41*, 14-24.
- McMinn, M. R., Ellens, B. M., & Soref, E. (1999). Ethical perspectives and practice behaviors involving computer-based test interpretation. *Assessment, 9*, 71-77.
- McNeal, T. P., & Weed, N. C. (1999). *The relative validity of blind clinical and computer based interpretation of the MMPI-2*. Unpublished master's thesis, University of Mississippi, Oxford.
- Meehl, P. E. (1946). Profile analysis of the Minnesota Multiphasic Personality Inventory in differential diagnosis. *Journal of Applied Psychology, 30*, 517-524.
- Meehl, P. E. (1956). Wanted – A good cookbook. *American Psychologist, 11*, 263-272.
- Moreland, K. L. (1983). Diagnostic validity of the MMPI and two short forms. *Journal of Personality Assessment, 47*, 492-493.
- Moreland, K. L. (1985). Validation of computer-based test interpretations: Problems and prospects. *Journal of Consulting and Clinical Psychology, 53*, 816-825.
- Moreland, K. L. (1987). Computer-based test interpretations: Advice to the consumer. *Applied Psychology: An International Review, 36*, 385-399.
- Morey, L. C., Waugh, M. H., & Blashfield, R. K. (1985). MMPI scales for DSM-III personality disorders: Their derivation and correlates. *Journal of Personality Assessment, 49*, 245-251.
- 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.



- Noland, K. A., & Weed, N. C. (1995). The Mississippi Q-sort. [Computer Software]. Oxford, MS: Authors.
- Ozer, D. J. (1993). The Q-sort method and the study of personality development. In D.C. Funder & R. D. Parke (Eds.), *Studying lives through time: Personality and development* (pp. 147-168). Washington, DC: American Psychological Association.
- Pancoast, D. L., Archer, R. P., & Gordon, R. A. (1988). The MMPI and clinical diagnosis: A comparison of classification system outcomes with discharge diagnoses. *Journal of Personality Assessment*, 52, 81-90.
- Pant, H. & Weed, N. C. (2008). *Diagnostic reliability of computer based test interpretations for the MMPI-2*. Manuscript in preparation, Central Michigan University.
- Rainwater, G. D. (1996). MMPI-2 Report Manual (Version 4.0) [Computer Software Manual]. Melbourne, FL: Psychometric Software, Inc.
- Roid, G. H. (1985). Computer-based test interpretation: The potential of quantitative methods of test interpretation. *Computers in Human Behavior*, 1, 207-209.
- Snyder, D. K. (2000). Computer-assisted judgment: Defining strengths and liabilities. *Psychological Assessment*, 12, 52-60.
- Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: The University of Chicago.
- Strassberg, D. S., & Cooper, L. M. (2005). MMPI/MMPI-2 Adult Interpretive Report, Software Manual (Version 10.0P) [Computer Software Manual]. Salt Lake City, UT: Automated Assessment Associates.
- Walters, G. D. (1985). Scale 4 (Pd) of the MMPI and the diagnosis antisocial personality. *Journal of Personality Assessment*, 49, 474-476.

- Webb, J. T., Miller, M. L., & Fowler, R. D., Jr. (1970). Extending professional time: A computerized MMPI interpretation service. *Journal of Clinical Psychology, 26*, 210-214.
- Williams, J. E. (2002). *The relative validity of eight computer-based test interpretations for the MMPI-2*. Unpublished doctoral dissertation, University of Mississippi, Oxford.
- Williams, J. E. & Weed, N. C. (2004). Review of computer-based test interpretation software for the MMPI-2. *Journal of Personality Assessment, 83*, 78-83.
- Williams, J. E. & Weed, N. C. (2004). Relative user ratings of MMPI-2 computer-based test interpretations. *Assessment, 11*, 316-329.