Use of a Structured Interview to Evaluate the Validity of the Alcohol and Drug Dependence Scales of the Millon Clinical Multi Axial Inventory II

Bert van Hoek

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Counseling Commons, and the Substance Abuse and Addiction Commons

Recommended Citation

Hoek, Bert van, "Use of a Structured Interview to Evaluate the Validity of the Alcohol and Drug Dependence Scales of the Millon Clinical Multi Axial Inventory II" (1995). Dissertations. 1755.
https://scholarworks.wmich.edu/dissertations/1755

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
USE OF A STRUCTURED INTERVIEW TO EVALUATE THE VALIDITY OF THE ALCOHOL AND DRUG DEPENDENCE SCALES OF THE MILLON CLINICAL MULTIAXIAL INVENTORY II

by

Bert van Hoek

A Dissertation Submitted to the Faculty of The Graduate College in partial fulfillment of the requirements for the Degree of Doctor of Education
Department of Counselor Education and Counseling Psychology

Western Michigan University Kalamazoo, Michigan April 1995
USE OF A STRUCTURED INTERVIEW TO EVALUATE THE VALIDITY OF THE ALCOHOL AND DRUG DEPENDENCE SCALES OF THE MILLON CLINICAL MULTIAXIAL INVENTORY II

Bert van Hoek, Ed.D.
Western Michigan University, 1995

This is an external validation study of the Alcohol and Drug dependence scales of the Millon Clinical Multiaxial Inventory II ([MCMI-II], Millon, 1987). The Structured Clinical Interview for the DSM-III-R ([SCID], Spitzer, Williams, Gibbon, & First, 1990) was administered to a sample of 73 adults who presented for treatment at a substance abuse facility to establish all possible substance abuse diagnoses based on the nosology of the revised Diagnostic Statistical Manual for Mental Disorders (3rd ed. [DSM-III], American Psychiatric Association, 1980). Subsequently, scores on the Alcohol and Drug Dependence scales of the MCMI-II for the present sample were compared to the SCID generated diagnoses to establish sensitivity and specificity for these scales. Additionally, the correlation between the Alcohol and Drug Dependence scales of the MCMI-II was examined to validate their utility as independent measures of Alcohol Dependence and/or Abuse and Drug Dependence and/or Abuse.

An analysis of the data suggests that it is unwarranted for clinicians to use only the Alcohol and Drug Dependence scales of the MCMI-II as measures of current or past substance abuse. The sensitivity of the Alcohol Dependence Scale in the present sample of known substance abusers was .65 with a specificity of .77. The sensitivity of the Drug Dependence Scale for the same sample was .55 with a specificity of .85. Additionally, it was found that the Alcohol Dependence Scale was a better measure of all substance abuse diagnoses than the Drug Dependence Scale.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ACKNOWLEDGMENTS

It is most difficult to adequately acknowledge those individuals who have contributed to the completion of the process that has culminated in this document. Foremost, I owe a great debt of gratitude to my wife Susan, if it had not been for her sacrifice and willingness to carry the extra load, I would not have been able spend the time or energy required to produce a quality product. Finally, it can be her turn! The others who have indulged me are of course my children Megan, Ted, and Elise. Thank you for giving up some of your time with me. Thanks too to my parents Emmy and Kees van Hoek who were always there to lend a hand. Without the emotional and moral support of each of you I could not have persevered.

I'm also indebted to the faculty and friends who, directed, and facilitated this process. First, I want to acknowledge the chair of my committee, Diane Prosser, Ph.D. Your patience, persistence, insight and support were felt each step of the way. A warm and heartfelt "thank you" for your commitment to my success. Thanks too to the members of my committee Robert Betz, Ph.D. and Malcolm Robertson, Ph.D. for your thoughtful feedback and support which was always provided in a constructive manner. This section would not be complete without recognizing Suzanne Hedstrom, Ed.D. without whose gentle prodding I might have given up. Thank you for the kind words and affirmations. To Harry Piersma, Ph.D., I also owe gratitude for the professional direction and guidance which you provided as well as your support as a friend. Finally, to Anne Smit, Ph.D. a word of thanks for helping with the process of having the statistical analysis come together.
Acknowledgments—continued

As I look at this long list, I feel humbled and blessed, because it represents much energy and trust which each of you invested in me. Thank you.

Bert van Hoek
## TABLE OF CONTENTS

ACKNOWLEDGMENTS ...................................................................................... ii

LIST OF TABLES .......................................................................................... viii

LIST OF FIGURES .......................................................................................... ix

CHAPTER

I. INTRODUCTION ....................................................................................... 1

   Historical Background of Problem ....................................................... 1

   Introduction of the MCMI-I ........................................................... 1

   Overview of Core Matrix of Millon's Biopsychosocial
   Theory of Personality ..................................................................... 2

   Relationship of the MCMI-I and the DSM-III ............................... 5

   Introduction of the MCMI-II .......................................................... 6

   Statement of the Current Situation .............................................. 9

   Purpose ................................................................................................ 11

   Delineation of the Study ....................................................................... 12

   Definitions .......................................................................................... 13

   Questions to be Answered ........................................................... 14

   Hypotheses ......................................................................................... 15

II. REVIEW OF THE LITERATURE ............................................................ 17

   The MCMI-I ......................................................................................... 17

   Theoretical Base............................................................................. 18

   Test Construction ........................................................................ 20

   Response to the MCMI-I.......................................................... 24

   The MCMI-II ..................................................................................... 27
Table of Contents--Continued

CHAPTER
Theoretical Base ................................................................. 28
Test Construction ............................................................... 28
From the MCMI-I to the MCMI-II ............................................. 29
Revision of the MCMI-I .......................................................... 30
Significance of Differences Between the MCMI-I and the MCMI-II .................................................. 31
The MCMI-I and MCMI-II and Substance Abuse .................. 33
Theoretical Base ................................................................. 33
Discriminant Validity of the Alcohol and Drug Dependence Scales ........................................... 35
Base Rate Scores ................................................................. 41
Prototypical Categorization ................................................... 41
The Statistic Kappa ............................................................. 42
The Structured Clinical Interview for the DSM-III-R .............. 43
The Development of Operational Diagnostic Criteria .......... 45
The Development of Structured Clinical Interviews ............... 46
Structured Interviews: General Considerations .................... 47
Validation of the SCID ......................................................... 48
Summary ................................................................................. 52

III. METHODOLOGY ............................................................. 54
Description of Method ......................................................... 54
Data Collection .................................................................. 56
Selection and Description of Sample ....................................... 58
Instrumentation ................................................................ 58
Table of Contents--Continued

CHAPTER

The SCID ........................................................................................ 59
The MCMI-II .................................................................................. 62
Limitations of the Present Study................................................. 72
Statistical Hypotheses............................................................... 74
Analysis of Data ........................................................................... 76
Comparing Differences in Sensitivity and Specificity ............ 76
Comparison of BR Scores of the Alcohol and Drug Dependence Scales ................................................. 79

IV. RESULTS ......................................................................................... 82
Comparing Differences in Sensitivity and Specificity ............ 82
Hypothesis 1 ................................................................................ 82
Hypothesis 2 ................................................................................ 84
Hypothesis 3 ................................................................................ 85
Hypothesis 4 ................................................................................ 87
Comparison of BR Scores of Alcohol and Drug Dependence Scales ....................................................... 88
Hypothesis 5 ................................................................................ 88
Hypothesis 6 ................................................................................ 90
Summary ...................................................................................... 92

V. SUMMARY, CONCLUSIONS AND DISCUSSION ...................... 94
Summary ..................................................................................... 94
Findings in Relation to Research Questions ....................... 96
Conclusions ................................................................................. 100
Discussion .................................................................................. 104
Table of Contents--Continued

CHAPTER

Issues Related to General Test Construction Theory ..................... 104
The Polythetic Nature of Categories .............................................. 105
The MCMI-II and DSM-II-R Diagnosis of Substance Abuse ............. 108
Recommendations ......................................................................... 109

APPENDICES

A. Human Subjects Institutional Review Board Approval ................. 112

BIBLIOGRAPHY ............................................................................. 114
# LIST OF TABLES

1. Framework for Millon's Theory-Based Diagnostic Categories ................................................................. 4

2. Revised Framework for Millon's Theory-Based Diagnostic Categories ............................................................. 7

3. Prototypic Items of the Alcohol and Drug Dependence Scales of the MCMI-II ................................................................. 69

4. BR Scores of Alcohol Dependence Scale > BR Scores of Drug Dependence Scale Wilcoxon Matched-Pairs Signed-Ranks Test................................................................................ 89

5. BR Scores of Drug Dependence Scale > BR Scores of Alcohol Dependence Scale Wilcoxon Matched-Pairs Signed-Ranks Test................................................................................ 91

6. Summary of Diagnostic Efficiency of Alcohol Dependence Scale of the MCMI-I and MCMI-II ................................................................. 101

7. Summary of Diagnostic Efficiency of Drug Dependence Scale of the MCMI-I and MCMI-II ................................................................. 101
LIST OF FIGURES

1. Four Cell - Contingency Table .......................................................... 77
2. The z-Test ......................................................................................... 78
3. Contingency Table: Sensitivity and Specificity of Alcohol Dependence Scale .................................................. 84
4. Contingency Table: Sensitivity and Specificity of Drug Dependence Scale .................................................. 87
CHAPTER I

INTRODUCTION

Historical Background of Problem

Psychological tests and inventories have made significant contributions to the evaluation and treatment process. At the same time, as Gibertini (1993) states, "In the field of clinical assessment in psychology, imprecisions are the rule in both the theories underlying the nosological categories and in the instrumentation designed to measure them" (p. 72). In order to ensure high standards, ongoing evaluation of assessment instruments is therefore essential. The most important consideration in this process is the establishment of an instrument's validity or, "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (American Psychological Association, 1985, p. 9). While it is incumbent on the author of new tests to provide data that can be used to assess validity, new instruments need to be subjected to a process of external validation with samples other than those used to construct the test (Butcher & Owen, 1979).

Introduction of the MCMI-1

The Millon Clinical Multiaxial Inventory ([MCMI-I], Millon, 1983), a self-report personality inventory for the assessment of psychiatric patients, was developed by Theodore Millon in the late 1970s. His stated purpose was, "to draw upon the best features of the MMPI, minimize its limitations, and move forward to develop instruments that reflect advances of the past quarter of a century in psychopathology, diagnostic assessment and test construction" (Millon, 1983, p. 1). In accordance with
these goals, the clinical theory base for the instrument was provided by Millon's
Furthermore, the underlying criteria used to develop scales, the format of the
instrument, and scale labels were coordinated with the criteria and schema being
developed for the revised Diagnostic Statistical Manual for Mental Disorders (3rd ed.
[DSM-III], American Psychiatric Association, 1980). Additionally, test construction
included a rigorous and comprehensive process of sequential validation (Greer, 1984;

Overview of Core Matrix of Millon's Biopsychosocial Theory of Personality

The base for the MCMI-I's diagnostic categories is Millon's formulation of
eight basic personality styles (Millon 1969, 1981) and a related set of more transient
symptoms. In developing the descriptive criteria for the personality styles, Millon
adhered to the concept of categorization by prototypes rather than simple traits (Choca,
Shanley, & Van DenBurg, 1992; Millon 1983, 1987). Whereas traditional categories
are viewed as distinct and with rigid boundaries, categories that are based on
prototypes are intended to be more reflective of natural cognitive processes
categories of human thought ... [are] ... 'fuzzy sets' whose spatial (analog)
representation possesses loosely defined category boundaries and a membership that
is probabilistic rather than discrete" (p. 1335). Prototypical members of a category are
those which most closely exemplify the defining characteristics and criteria for a class.
In addition to these prototypical items, there are other descriptive members of the class
whose goodness of fit ranges from close to minimal approximation of the exemplary
items. In utilizing the concept of prototypes, Millon focused on personality as a
cluster of traits and recognized the weighted contribution of each characteristic (Buss,
The resulting personality styles were conceptualized as discrete entities, "that have no intrinsic logical relation to one another" (Choca et al., 1992, p. 10). However, individual descriptive criteria and defining characteristics will, depending on the goodness of fit with the prototypical items, contribute in varying degrees to a number of personality styles (Broughton, 1984; Millon, 1983, 1987).

The theoretical matrix that forms the basis for Millon's typology of basic personality types incorporates two dimensions: reinforcement and basic interpersonal coping style (Millon, 1969, 1981). Reinforcement has four variables: detached, dependent, independent, and ambivalent. The detached person is one who finds little satisfaction in relationship with others or from within himself. Dependent people look for others to determine their level of satisfaction or dissatisfaction. The independent individual is motivated by rewards from within the realm of her own needs, values and desires. Lastly, the ambivalent person has difficulty deciding whether to look to self or others for gratification. Interpersonal coping style has two variables: active and passive. The person with an active style works to influence and control life situations. On the other hand the passive style is characterized by apathy, restrained, and a perceived lack of power and control. This 4 x 2 matrix results in eight combinations that are the basis for the core personality types (see Table 1). Each of these personality styles involves a set of assumptions about self and others, typical ways of thinking and feeling, and specific behavioral repertoires (Choca et al., 1992; Millon 1983). For example, in the MCMI-I test manual the passive-detached type is described as being, "characterized by social impassivity ... [while] ... . Affectionate needs and feelings are minimal, and the patient functions as a passive observer isolated from the rewards and affections, as well as from the dangers of human relationships" (Millon, 1983, p. 33).
### Table 1
Framework for Millon’s Theory-Based Diagnostic Categories

<table>
<thead>
<tr>
<th>Reinforcement</th>
<th>Interpersonal Behavior and Related Personality Style (PS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active</td>
</tr>
<tr>
<td>Detached</td>
<td>Active Detached Avoidant PS</td>
</tr>
<tr>
<td>Dependent</td>
<td>Active Dependent Histrionic PS</td>
</tr>
<tr>
<td>Independent</td>
<td>Active Independent Antisocial PS</td>
</tr>
<tr>
<td>Ambivalent</td>
<td>Active Ambivalent Passive Aggressive PS</td>
</tr>
</tbody>
</table>

Millon, 1983

The core personality types may be adversely affected by life stressors in two ways. First, exposure to continuously stressful conditions may lead to "more serious patterns of personality pathology ... [which] ... are elaborations of one of the basic eight styles" (Millon 1983, p. 34). For example, the description of the extension of the passive-detached personality style includes a reference to cognitive dysfunction and is considered a pathological personality disorder (Schizotypal Personality Disorder). Secondly, the interaction between stress and any of the personality styles may result in sets of transient symptoms which constitute the clinical syndromes. The clinical syndromes are seen as exaggerations or extensions of one or more specific personality styles and should be viewed in the context of the basic personality patterns (Millon, 1983, 1987; Wetzler, 1990). For example, the behavioral manifestation for the individual who has an active-detached (Avoidant PS) style may include anxiety,
dysthymia, and depression, symptoms which may also be exhibited by individuals who have a passive-dependent (Dependent PS) orientation. Alcohol and drug abuse are considered clinical syndromes and are most likely to be associated with the active dependent (Histrionic PS) and the active-independent (Antisocial PS) styles (Millon, 1969, 1981, 1983).

Ongoing evaluation of theory and research in the area of diagnostic criteria for personality disorders led Millon (1987) to introduce a change in the typology. The dimension of reinforcement which consisted of four variables (detached, dependent, independent, ambivalent) was expanded to include discordant as a variable. The discordant person may seek pain instead of pleasure and find satisfaction in turning potentially positively rewarding situations into negative experiences. This may be done both at the expense of self and others. With the addition of discordant as a variable the earlier 4 x 2 matrix evolved into a 5 x 2 matrix (see Table 2) which resulted in the formulation of two new core personality types (Millon, 1987; Millon & Klerman, 1986). This new typology forms the clinical theory base for the revised instrument, the Millon Multiaxial Inventory II (MCMI-II), which Millon introduced in 1987.

**Relationship of the MCMI-I and the DSM-III**

Millon was an active participant in the development of the DSM-III diagnostic criteria, a process which occurred simultaneously with the construction of the MCMI-I. Indeed, the DSM-III schema of diagnosis along axes is reflective of Millon’s personality theory especially in relation to Axis I and Axis II. Each of the eight basic personality disorders, three pathological personality disorders, and nine clinical syndromes that are part of the MCMI-I profile, correspond with the diagnostic labels
of the DSM-III. Additionally, the descriptive criteria that define the syndromes of the MCMI-I are intended to parallel the diagnostic criteria of the DSM-III.

Introduction of the MCMI-II

Although initial review of the MCMI-I was generally favorable (Dana & Cantrell, 1988; Gibertini, Brandenburg, & Retzlaff, 1986; Greer, 1984; Hess, 1985; Widiger, 1985; Widiger, Williams, Spitzer, & Frances, 1985), ensuing research raised concerns about the reliability of the instrument (Choca, Peterson, & Shanley, 1986; McMahon, Flynn, & Davidson, 1985; Piersma, 1986, 1987; Retzlaff & Gibertini, 1987; Widiger & Sanderson, 1987). Consequently, Millon revised the MCMI-I and introduced a new instrument, the MCMI-II (Millon, 1987).

The revision of the MCMI-I was prompted by the following three major factors: (1) the evolution of Millon's theory of personality, (2) the increased interest in the personality disorders and subsequent revision of the current diagnostic nomenclature, and (3) a growing body of research on the MCMI-I (Millon, 1985, 1987). The development of the underlying theory led to a change in the character of some of the original personality disorders and a conceptualization of ten personality styles rather than eight (see Table 2). Concurrently, the data base related to personality disorders grew, allowing for the refinement of descriptive criteria utilized in the development of the scales and in test reports.

Additionally, the revision of the DSM-III served as a catalyst for changes. According to Millon, one of the distinguishing features of the MCMI-I is the unique correlation between both the diagnostic format (Axis I and Axis II) and the terminology of the inventory and the DSM-III. Millon (1987) was committed to optimizing the clinical utility of the inventory through the incorporation of theoretical constructs underlying the revised Diagnostic and Statistical Manual of Mental Health Disorders.
Disorders (3rd edition-revised [DSM-III-R], American Psychiatric Association, 1987). Furthermore, he sought to utilize terminology consistent with the DSM-III-R

Table 2
Revised Framework for Millon's Theory-Based Diagnostic Categories

<table>
<thead>
<tr>
<th>Reinforcement</th>
<th>Interpersonal Behavior and Related Personality Style (PS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active</td>
</tr>
<tr>
<td>Detached</td>
<td>Active Detached</td>
</tr>
<tr>
<td>Dependend</td>
<td>Active Dependent</td>
</tr>
<tr>
<td>Independent</td>
<td>Active Independent</td>
</tr>
<tr>
<td>Ambivalent</td>
<td>Active Ambivalent</td>
</tr>
<tr>
<td>Discordant</td>
<td>Active Discordant</td>
</tr>
</tbody>
</table>

Millon, 1987

in order to promote clarity in communication. In order to meet these objectives, it was necessary, "to broaden the trait and syndromal domains of the MCMI-I" (Millon, 1987, p. 76).

The growing body of research that focused on the reliability and validity of the MCMI-I prompted further evaluation of all the scales as well as the scoring system. Ongoing cross-validation research, by the test construction team, with large samples of psychiatric patients reportedly confirmed the discriminant validity of the scales. These studies also helped to identify a number of items whose endorsement
frequencies were less than optimal and therefore in need of replacement (Millon, 1985, 1987). Results of external validity studies conducted by other researchers raised questions about the efficacy of several scales (Bryer, Martinez, & Dignan, 1990; Calsyn, Saxon, & Daisy, 1990, 1991; Choca, Bresolin, Okonek, & Ostrow, 1988; Gibertini et. al., 1986; McMahon & Davidson, 1986). Additionally, other studies with non-test construction samples produced data that led to criticism of the excessive overlapping of scales (Choca et al., 1986; Lorr, Retzlaff, & Tarr, 1989; Retzlaff & Gibertini, 1987; Widiger & Sanderson, 1987).

Millon (1983, 1987) had anticipated that the MCMI-I would need revision. The new instrument (MCMI-II) shares the guiding principles of the MCMI-I such as the underlying personality theory, the extensive process of theory based and psychometrically sophisticated test construction, and the operating characteristics (Gibertini, 1993). However, the body of knowledge, generated by studies of and clinical experience with the MCMI-I, was indicative of the need for significant changes in the instrument. In order to enhance the diagnostic utility of the new test, Millon (1987) incorporated both the results of research and the clinical data that had been generated into the test construction process of the MCMI-II. The critical review and the revision of the original instrument are consistent with Millon's stated ideal that,

The MCMI has not been cast in stone. It is and will remain an evolving assessment instrument, significantly upgraded and refined perhaps every decade or so, to reflect substantive advances in knowledge, be it from theory, research or clinical experience. We intend to follow this model of regular upgrading, a policy that has been notably lacking among our more established assessment instruments ... many of our one [sic] vaunted techniques would have been better served had their adherents encouraged their periodic refinement and modification. (p. 389)
Statement of the Current Situation

As was the case with its predecessor, the MCMI-I, the MCMI-II has become a widely used instrument in clinical practice. This stature within the community of practitioners has been attributed in part to the thorough, theory based process used in the construction of the instrument (Dana & Cantrell, 1988; Wetzler, 1990). Furthermore, the apparent correlation between the MCMI-II and the DSM-III-R make the instrument attractive to clinicians. Wide acceptance of the MCMI-II may, however, be premature as a review of the literature shows that few external validation studies of the MCMI-II have been conducted. Moreover, the reported results of this type of research for the MCMI-I have been mixed, with a significant number of researchers questioning the effectiveness and discriminant validity of the latter instrument (Bryer et al., 1990; Calsyn et al., 1990, 1991; Choca et al., 1988; Gibertini et al., 1986; McMahon & Davidson, 1986).

Test construction for both the MCMI-I and MCMI-II involved a three step sequential validation process. The studies involved in this process generated extensive reliability and validity data which Millon (1983, 1987) reported in the test manuals. Efficacy of the instruments as well as the individual scales was supported by positive results from studies that focused on: (a) reliability involving test-retest measures of individual scales and profile stability, (b) factor analysis, (c) convergent validity, and (d) external discriminant validity. The validity data was reported in terms of conditional probabilities, sensitivity and specificity. For the MCMI-II, the sensitivity figures for the Alcohol and Drug Dependence scales were reported as 87% and 72% respectively, with specificity of 99% and 98% for the same scales (Millon, 1987). On the basis of these figures, Millon (1987) stated with some confidence that the substance dependence scales of the MCMI-II corresponded extremely well with
diagnoses that were established using the DSM-III-R criteria. However, few external validation studies, with samples other than those used in test construction, have been conducted to corroborate these positive findings. This is especially disconcerting as the effectiveness and discriminant validity of the MCMI-I in general, and the Alcohol and Drug Dependence scales in particular, had been called into question (Bryer et al., 1990; Calsyn et al., 1990, 1991; Choca et al., 1988; Gibertini et al., 1986; McMahon & Davidson, 1986).

External validation research has been plagued by the lack of availability of a reliable assessment standard for comparison (Bryer et al., 1990; Piersma, 1993, Spitzer, 1983). Clinical assessment and diagnosis have been expected to supply this "gold standard" (Spitzer & Williams, 1980). However, the processes of psychiatric diagnosis and assessment, utilizing a non-structured clinical interview, are influenced by numerous confounding variables. These may include the clinician's theoretical orientation, training, biases and skills. Furthermore, according to McReynolds (1989), "such judgments are made difficult by the fact that unlike most diseases, behavioral problems have-for the most part-no identifiable pathogenic agents, and that normal problems in living merge imperceptibly into clearly disturbed behaviors" (p. 85). The result is a generally low reliability for psychiatric diagnosis across clinicians (Spitzer & Fleis, 1974).

The development of the DSM-III based on atheoretical operational criteria was intended to enhance standardized diagnostic practices (McReynolds, 1989; Robins & Helzer, 1986; Spitzer & Williams, 1980; Zimmerman & Coryell, 1989). In their report on the DSM-III field trials, Spitzer, Forman, & Nee (1979) indicated that the inter-rater diagnostic reliability was higher for the DSM-III than the DSM-II. Widiger, Hurt, Frances, Clarkin, and Gilmore (1984) affirmed these findings, though their study revealed "shortcomings of the criteria and fixed rules of DSM-III" (p.
In his presidential address to the American Psychological Association, in which he reviewed the status of assessment, Matarazzo (1990) defended the progress in differential diagnosis represented by the development of descriptive criteria. Similarly, McReynolds (1989) in his review of both the DSM-III and DSM-III-R stated that, "even with all their limitations these systems represent a significant advance in psychopathological taxonomy" (p. 96).

In order to operationalize the new criteria based nosological system, the use of structured interviews in the diagnostic process gained renewed support (McReynolds, 1989; Robins & Helzer, 1986; Spiker & Ehler, 1984; Spitzer, 1983). It was expected that the use of these structured and/or standardized instruments, in addition to the new criteria, would result in more consistent, reliable, and therefore valid assessment practices. The Structured Clinical Interview for the DSM-III-R (SCID), Spitzer, Williams, Gibbon, & First, 1990) was specifically developed to mirror the clinical assessment process suggested in the DSM-III-R. The diagnoses generated by this and other interview schedules have been identified as potential standards to which to compare the performance of personality inventories for the purpose of external validation. (Piersma, 1993).

Purpose

The MCMI-II holds much promise as a diagnostic instrument because it is both a theoretically grounded and comprehensively standardized personality inventory. Its relatively close correlation to the diagnostic system of the DSM-III-R makes this a potentially attractive assessment tool for general clinical practice. Though conceptually similar to the MCMI-I, due to the substantive revisions scientific and clinical principles demand that the new instrument be validated apart from the original (Piersma, 1989). A review of the literature shows that to date very little external
validation data has been reported. In the present study, the effectiveness of the Alcohol and Drug Dependence scales of the MCMI-II will be examined in order to provide external validation of the positive findings reported by Millon (1987). Additionally, the correlation between the Alcohol and Drug Dependence scales will be examined to validate their utility as independent measures of the clinical syndromes of Alcohol Dependence and/or Abuse and Drug Dependence and/or Abuse as reported by Millon (1987).

Delineation of the Study

This is a study which is based on data collected as part of an ongoing research program at a substance abuse treatment center in Lubbock, Texas. The sample consisted of adults (18 and over) who entered treatment for substance abuse in the inpatient or outpatient units of the center during a six month time period. This resulted in the identification of 73 participants. As part of the regular intake procedures at the facility, the participants were administered a battery of eight assessment instruments. This process began immediately after admission unless, for clinical reasons, a person was judged to be incapable of completing the process. No individuals were administered the instruments while involved in detoxification.

The data base for the present study came from two self report instruments which were part of the assessment process at this agency. The first of these instruments, the SCID, was administered to establish a psychiatric diagnosis based on the nosology of the DSM-III-R. The second instrument was the MCMI-II which was administered in accordance with accepted standards of practice. It is the base rate (BR) scores of the Alcohol and Drug Dependence scales which are the focus of the present study. The substance abuse diagnoses established by the SCID serve as the
standard to which the data of the MCMI-II Alcohol and Drug Dependence scales are compared.

Definitions

The following are key terms in the present study which warrant definition.

1. Discriminant validity - the agreement between the scale score and an independent measure of the same construct (Loevinger, 1957; Skinner, 1981). "Poor discriminant validity would be evidenced by the finding that the differentiation among types based on one assessment mode (e.g., self-report data) was largely lost when the classification is attempted with a parallel set of measures (e.g., clinical ratings)" (Skinner, 1981, p. 77).

2. Sensitivity - the proportion of subjects, usually expressed in percentage form, that are identified as true positives in relation to an established criterion (American Psychological Association, 1985; Gibertini et al., 1986).

3. Specificity - the proportion of subjects, usually expressed in percentage form, that are identified as true negatives in relation to an established criterion (American Psychological Association, 1985; Gibertini et al., 1986).

4. Base rate (BR) score - an expression of test data based on the clinically judged prevalence of the disorder in a given population.

5. Polythetic - a term used to describe a set of flexible rules, as opposed to monothetic or fixed and invariant rules, associated with constructing diagnostic systems (Millon, 1988). This set of rules provides for the inclusion of criteria which contribute to and fit in diagnostic categories to differential degrees. When used to describe diagnostic categories this term implies that there are natural relationships between categories and that one may be an extension of another (Millon, 1987).
6. Prototypic - Within the context of a polythetic approach to categories, a term used to describe a criteria of a category which fits most closely with the construct being defined.

Questions to be Answered

The following questions will serve as the guiding principles throughout the course of the present study.

1. Is the discriminant validity of the MCMI-II Alcohol Dependence scale for the present sample, expressed in terms of sensitivity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

2. Is the discriminant validity of the MCMI-II Alcohol Dependence scale for the present sample, expressed in terms of specificity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

3. Is the discriminant validity of the MCMI-II Drug Dependence scale for the present sample, expressed in terms of sensitivity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

4. Is the discriminant validity of the MCMI-II Drug Dependence scale for the present sample, expressed in terms of specificity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

5. When Alcohol Dependence and/or Abuse is the only substance abuse diagnosis on Axis I, are the BR scores of the MCMI-II Alcohol Dependence scale greater than the BR scores of the MCMI-II Drug Dependence scale?
6. When Drug Dependence and/or Abuse is the only substance abuse diagnosis on Axis I, are the BR scores of the Drug Dependence scale greater than the BR scores of the Alcohol Dependence scale?

Hypotheses

The following are the hypotheses for the present study. The direction of the prediction in the first four hypothesis is supported by external validation studies of the MCMI-I Alcohol and Drug Dependence scales which suggest that the discriminant ability of these scales is less than reported by Millon (Bryer, et al., 1990; Calsyn et al., 1990, 1991; Marsh, Stile, Stoughton, & Trout-Landen, 1988, Millon, 1983). The last two hypotheses are stated in different terms because there has been some indication in the literature (Jaffe & Archer, 1987) that the MCMI-I Alcohol Dependence scale may be a better determinant of drug abuse than the MCMI-I Drug Dependence scale.

1. The true positive proportion (sensitivity) on the MCMI-II Alcohol Dependence scale in the present sample will be less than 87%, the sensitivity of this scale reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II.

2. The true negative proportion (specificity) on the MCMI-II Alcohol Dependence scale in the present sample will be less than 99%, the specificity reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

3. The true positive proportion (sensitivity) on the MCMI-II Drug Dependence scale in the present sample will be less than 72%, the sensitivity reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
4. The true negative proportion (specificity) on the MCMI-II Drug Dependence scale in the present sample will be less than 98%, the specificity reported by Millon for the cross validation sample used in the construction of the MCMI-II?.

5. When Alcohol Dependence and/or Alcohol Abuse is the only substance abuse diagnosis, the BR scores of the MCMI-II Alcohol Dependence scale will be greater than the BR scores of the MCMI-II Drug Dependence scale.

6. When Drug Dependence and/or Drug Abuse is the only substance abuse diagnosis, the BR scores of the MCMI-II Alcohol Dependence scale will be greater than the BR scores of the MCMI-II Drug Dependence scale.
CHAPTER II

REVIEW OF THE LITERATURE

The focus of the current study is on the diagnostic accuracy of two scales, the Alcohol and Drug Dependence scales, of a personality inventory, the Millon Clinical Multiaxial Inventory II (MCMI-II). Diagnoses generated by a structured interview, the Structured Diagnostic Interview for the DSM-III-R (SCID), are utilized as the standard for comparison. The MCMI-II was based on an elaborate theory of personality and classification, and the development of the instrument involved an innovative approach to test construction. As Gibertini (1993) states, "Because of theory-and empirically-driven construction procedures of the MCMIs, we are finally in a position to seriously examine the operating characteristics of a major psychodiagnostic inventory in clinical psychology" (p. 72). The SCID is a structured clinical interview which has been developed to standardize the process of making a differential diagnosis. The classification schema on which the SCID is based is similar to the system of classification utilized in the development of the MCMI-II. This review will examine theoretical constructs which guided the development of these instruments. The current research on the two instruments, and their relevant predecessors, as assessment tools of alcohol and drug abuse are also examined.

The MCMI-I

The MCMI-I was developed by Theodore Millon to incorporate developments in the assessment and theory of personality (1987). The instrument was expected to
provide an alternative to the MMPI. Additionally, it was designed to incorporate Millon's emerging theory of personality styles and disorders.

Theoretical Base

The development of the MCMI-I was based on Millon's biopsychosocial theory of personality (Millon, 1969, 1973, 1981). Within this context, personality is viewed as consisting of a complex set of interacting elements based on a person's past and present experiences. In order to adequately account for all contributing factors and the complexity of the relation between the variables of a specific personality style, Millon utilized a model based on clinical prototypes (Millon, 1983, 1987). The prototypical conceptualization of personality is in sharp contrast to theories which view personality characteristics as independent or pure traits. However, as Skinner and Blashfield (1982) in their discussion of cluster analysis conclude, "empirical research has tended to support more complex dimensional or hierarchical models, versus the discrete categories of diagnostic schema" (p. 730). Unlike trait theorists, Millon does not accept the concept of independent personality traits. However, developing a personality inventory, which would address and delineate each of the interactive intricacies of personality in an individualized manner, would be a nearly impossible task (Millon, 1983, 1987). As Millon (1987) states,

> the instrument must limit its focus to certain salient aspects of behavior ... a nucleus of factors must be selected which will capture the most relevant and essential characteristics of patients. This nucleus of attributes that focuses on the core distinguishing features of a syndrome or disorder may be spoken of as a clinical prototype. (p. 16)

Therefore, the MCMI-I was constructed as a tool to identify patients who share the core traits and behavioral manifestations which characterize clinical syndromes and personality disorders.
According to Millon (1983, 1987), the most important purpose of an inventory is to provide clinically useful information. Thus, Millon's primary focus is on the discovery of the nuances of clinical syndromes and the identification of enduring personality patterns. This focus on capturing the complexity of the individual's experience, requires that an inventory utilize a wide range of clinically significant traits and correlated behaviors in order to establish the presence or absence of a disorder. Though the use of prototypes does not account for all differences, it allows for a more accurate description of the individual's experience. Instruments based on these groupings, are therefore expected to provide the clinician with valuable information, perhaps previously unknown, about the patient (Millon, 1983, 1987). When these subtleties are presented in a consistent manner, it will facilitate the processes of differential diagnosis and treatment planning for the patient, and ongoing communication among professional peers.

In order to be consistent with the concept of clinical prototypes, a congruent system of classification would include diagnostic categories which were inclusive, "of any behavior or phenomena that appears promising in terms of its significant correlates" (Zigler and Phillips, 1961, p. 616, cited by Millon, 1987). To assure adherence to this ideal, Millon (1983) utilized three guiding principles in the developing the MCMI-I. First, categories were expected to accommodate the complexity of clinical presentations and to reflect the interactive nature of traits and symptoms rather than focusing on one outstanding behavioral sign. Thus, in constructing the scales of the MCMI-I a wide range of clinical elements were included to, "tap the intricacy and diversity of both personality styles and symptom syndromes" (Millon, 1987, p. 17). Secondly, the categories were conceptualized as being polythetic (see pg. 12) in nature. As Millon (1987) states, "Each diagnostic category should be shown, where appropriate to be a precursor, an extension or a modification
of other clinical categories" (p. 17). Thus, clinical symptoms, as measured by the clinical syndrome scales of the MCMI-I, are conceptualized as being correlated with the basic personality styles. The clinical syndromes may emerge under stress and as transient manifestations, while the basic personality styles, as measured by the personality scales of the MCMI-I, consist of more enduring traits. Lastly, categories were constructed in a manner which allowed for the distinction of levels of severity (Millon, 1987). Accordingly, the clinical syndrome and personality scales were divided into two sets (see pg. 56, Instrumentation) with one group intended to measure moderate severity (Clinical Personality Pattern and Clinical Syndrome scales) and the second marked severity (Severe Personality Pathology and Severe Syndrome scales).

Test Construction

The conceptual model used to construct and standardize the MCMI-I was a three step sequential validation process which was both innovative and extensive (Millon 1983, 1987). Millon based the rationale for this strategy on Loevinger's (1957) and Jackson's (1970) assertion that validation should be an integral aspect of test construction. Loevinger (1957) identified the phases of sequential validation as substantive, structural, and external. Accordingly, in constructing the MCMI-I, the test items were first examined to determine the extent to which they were representative of the organized theoretical framework (substantive validation) (Millon, 1983). Those which were not representative were eliminated. Secondly, the remaining items were judged in terms of fit with the conceptual model of the test (structural validation), particularly as this related to the concepts of prototypical categorization. It was expected that overlap of scales would provide evidence for adherence to this concept. Thirdly, those items which represented both substantive
aspects of underlying theory and were congruent with test construction principles, were checked for correlation with other independent measures of the trait or syndrome under question (external validation).

Theoretical-substantive validation was an integral part of the selection of scale items that would be reflective of the proposed clinical prototypes (Millon 1983, 1987). This process involved the creation of an initial pool of items which was based on the published literature describing each of the syndromes. Sources such as the clinical literature, rating scales, other assessment instruments, as well as the text Modern Psychopathology (Millon, 1969), were reviewed, and this search provided more than 3,500 clinically descriptive statements (Millon, 1987). This initial set of items was used to construct 20 scales. It was reduced to a total of 1,100-plus items through separate screening processes that involved patients and clinicians, in addition to the test construction team. As this initial pool was reduced, there was emphasis on maintaining a range of items which was inclusive of all the traits and behaviors deemed descriptive of the syndromes. The items which remained were used to construct two 566 item provisional forms.

Internal-structural validation ensures that an instrument is consistent with the underlying theory or model (Loevinger, 1957; Millon, 1983, 1987). This component of the validation process for the MCMI-I was therefore guided by the concepts of prototypical categorization as well as the polythetic nature of categories. In accordance with the concept of prototypes, each diagnostic category had core (prototypical) items in addition to a set of items which were not central to the category but which had descriptive value. Items are shared with other scales. Consequently, items for one scale may be keyed for another scale as less perfectly correlated (descriptive) criteria. Millon (1983, 1987) argues that, as a result of this purposeful sharing of items, the scales of the MCMI-I should be expected to overlap and be
highly correlated with other theoretically related scales. Furthermore, Millon (1987) states, that the prototypical approach to categorization suggests a model which, "stresses internal scale consistency but does not require the scale independence that characterizes factorial approaches" (p. 35). For example, the Avoidant and Schizoid Scales are measures of two distinct clinical syndromes, and, therefore, one would expect each to show internal consistency. However, there are a number of items which are prototypical for one scale and descriptive for the other scale. Due to these inherent commonalties many items should overlap both scales. Similarly, the Alcohol and Drug Dependence scale are intended to identify two separate clinical syndromes, but, consistent with the concept of prototypes, there are a significant number of items (16 out of a sum of 77 total items) that are shared. 

The second guiding principle in the process of internal-structural validation was the concept of the polythetic nature of categories. According to the principle of polythetic categories, the clinical syndrome scales are conceptualized as extensions of the personality scales. The clinical syndromes emerge as a person is faced with the vicissitudes and stress of daily life and are seen as disruptions in basic personality styles. Thus, there are also natural commonalties between the clinical syndrome and basic personality scales and many items should be keyed for both sets of scales. An example of this type of relationship is found between the Dependent, basic personality, and Dysthymic, basic clinical syndrome, scales. Each of these scales address issues related to self perceptions and self confidence and passivity or lack of inertia (Millon, 1987). It is more difficult and controversial to view the substance abuse disorders as extensions of personality style, however. A number of studies substantiate consistent clusters of traits associated with both alcoholics (Craig, Verinis, & Wexler, 1985; McMahon & Davidson, 1985) and drug abusers (Calsyn et al., 1990; Craig et al., 1985). However, there is considerable literature which
suggests that the relationship between personality type and substance abuse is complex and therefore difficult to establish (Choca et al., 1992; Nerviano & Gross, 1983). None the less, Millon (1983) hypothesizes that alcohol abuse is primarily associated with histrionic, antisocial, and compulsive personality patterns. The polythetic nature of categories is more readily apparent in the relationship between the basic personality disorder scales and the pathological personality disorder scales. For example, the Avoidant Personality is seen as a precursor to the Schizotypal Personality and accordingly one would expect that the corresponding scales would, "display selective overlap and a high degree of correlation" (Millon, 1983, p. 36).

The process of structural-validation involved three steps aimed at reducing the two 566 item provisional forms, developed in the theoretical-substantive validation phase, to a single instrument (Millon, 1983). The provisional forms were first administered to a diverse clinical sample drawn from a variety of mental health agencies. The selection of the sample included consideration of an appropriate representation of patients based on both demographic and clinical variables. Subsequently, a process of analysis identified those items which were highly correlated with the scale to which they were originally assigned (prototypic items). The remaining items were examined for their level of correlation to other than the primary scale. Only those items which achieved a correlation greater than .30 or -.30 were retained as descriptive items. The remaining item pool was also evaluated to ensure, "adequate representation of each scale's trait diversity or syndrome complexity" (Millon, 1987, p. 41). These procedures reduced the 1,132 provisional items to 289 structurally valid items.

The 289 item Research Form of the MCMI-I was utilized in the final phase of external-criterion validation (Millon, 1983). Two studies were conducted, each of which used experienced clinicians' assessments of psychiatric patients as the criterion
to which the results of the inventory were compared. Consistency of assessment was assured through the use of booklets describing the syndromes, personality patterns, and guidelines for the criteria which needed to be met to make diagnoses. These studies were conducted in a variety of settings throughout the United States and Europe. After the first study (in this third step of the validation process), three of the 20 original scales were identified as lacking clinical utility. The clinicians who participated in the first external validity study had been asked to identify additional scales which would make the instrument a more useful assessment tool. Hypomania, Alcohol Abuse, and Drug Abuse were listed most frequently by all groups of clinicians and were subsequently added to replace three scales which did not appear to add significantly to the process of differential diagnosis. The procedure to develop these three new scales was in essence similar to that employed in constructing the original scales (Millon 1983). The new scales were incorporated into the instrument used for the second external validation study.

Response to the MCMI-I

The potential of the MCMI-I as a diagnostic and research tool was met with considerable enthusiasm and resulted in a great deal of review of the new instrument (Dana & Cattrell, 1988). The positive comments about the inventory as a whole, as opposed to specific scales, fell into five categories. First, many of the reviewers acknowledged and applauded the grounding of the MCMI-I in a coherent theory of psychopathology with particular emphasis on the personality disorders (Butcher & Owen, 1979; Hess, 1985; Wetzler, 1990; Widiger, 1985). Secondly, though the potential difficulties in item sharing among scales was recognized, the resulting relative brevity (175 items total) was seen as a positive aspect for an instrument which was intended for use in general clinical practice (Hess, 1985; Widiger, 1985).
Thirdly, the comprehensive and sophisticated test construction process was identified as a psychometric advance (Butcher & Owen, 1979; Dana & Cantrell, 1988; Greer, 1984; Hess, 1985; Wetzler, 1990; Widiger, 1985; Widiger et al., 1985). Fourth, the association of the MCMI-I with the current diagnostic nomenclature, as found in the DSM-III, was also recognized as a strength (Dana & Cantrell, 1988; Hess, 1985; Wetzler, 1990; Widiger, 1985; Widiger et al., 1985). Lastly, the use of base rate (BR) scores was heralded as an improvement over the use of standard scores (Gibertini et al., 1986; Hess, 1985; Wetzler, 1990; Widiger, 1985).

Though the initial reviews were generally focused on the positive aspects of the MCMI-I, ensuing research and review has been more critical of the instrument. The areas of concern that were identified fall into the following six categories:

1. Item overlap and the consequent correlation of scales led to questions about the distinctiveness of the personality and syndromes identified on the basis of the instrument (Choca et al., 1986; Lorr et al., 1989; Retzlaff & Gibertini, 1987; Widiger & Sanderson, 1987). Although Millon (1983) postulates that there is a relationship between personality styles and clinical syndromes, Choca and colleagues (1986) showed that 65% of the intercorrelation between scales was due to item overlap. Thus, as Wetzler, (1990) states, "interrelationships among scales appear to be artifactual rather than a function of intrinsic conceptual relationships among dimensions" (p. 447).

2. Based on data generated with a test-retest strategy, several authors raised questions about the reliability of the MCMI-I (McMahon et al., 1985; Murphy, Greenblatt, Mozdzierz, & Trimakas, 1990; Piersma, 1986, 1987; Wetzler, 1990). McMahon and colleagues (1985) determined that the time at which the instrument was administered during the course of treatment influenced test-retest reliability. Piersma (1986, 1987), in two separate studies with hospitalized psychiatric patients, reported
that test-retest reliability diminished. He attributed these results to the effect of active treatment. Although one might expect improvement rather than stability over time as a result of treatment, the test-retest reliability reported by Millon (1987) for the composite scales-Kuder-Richardson coefficient value above .80-was also with a sample whose members were actively involved in treatment. In contrast, Piersma (1986) reported the test-retest reliability in his sample as averaging .52.

3. Millon's (1983) distinction between clinical syndromes and personality disorders was intended to convey the difference between transient symptoms and long-standing personality patterns. However, there is some indication that the personality scales do not measure pure trait (Wetzler, 1990) as Millon (1983) suggests. For example, Piersma (1986) reports that one of the personality scales (Borderline) did not remain stable over time though one would not expect that short term treatment would impact personality significantly.

4. Utilizing a factor analytic strategy, Retzlaff and Gibertini (1987) determined that there was a lack of consistent confirmation of Millon's underlying constructs. However, there are a number of other factor analytic studies of the MCMI-I in which consistent factors are identified (Choca et al., 1986; Lewis & Harder, 1990; Lorr et al., 1989; Piersma, 1986; Widiger, 1985). Though such reports could be encouraging, the extent of scale overlap suggests that the identification of consistent factors, "may have been an artifact of item overlap rather than a feature of the factor structure per se " (Wetzler, 1990, p. 450).

5. A number of reviewers questioned the lack of convergent and discriminant validation of the MCMI-I and the DSM-III. Critics questioned Millon's (1983, 1985, 1986) claims that the MCMI-I was correlated with the DSM-III as there were no studies to substantiate this claim (Bonato, Cyr, Kalpin, Pendergast, & Sanhueza,
6. External validation studies have only partially confirmed the diagnostic efficiency of the scales of the MCMI-I. The results of a number of studies suggest that the MCMI-I may over-diagnose personality disorders, and the high sensitivity percentages of the clinical syndrome scales as reported by Millon (1983) have not always been replicated (Bryer et al., 1990; Calsyn et al., 1990, 1991; Choca et al., 1988; Gibertini et al., 1986; McMahon & Davidson, 1986).

Millon (1983) had anticipated that the MCMI-I would need to be updated. The serious concerns about the efficacy of the MCMI I enumerated above, served in part as a catalyst for the evaluation and revision of the instrument (Millon, 1987). The new inventory, the MCMI-II, was introduced in 1987.

The MCMI-II

At the time that Millon (1983) introduced the MCMI-I, he committed himself to revise the inventory on a regular basis in order to accommodate developing theory, research findings, and clinical experience. Therefore, as a result of the growing and substantial critiques of the efficacy (validity and reliability) of the MCMI-I, the evolution of the core matrix of personality styles, and a desire for the instrument to be congruent with the DSM-III-R, the MCMI-II was developed (Choca et al., 1992; Craig & Weinberg, 1993; Millon, 1987). The resulting revisions to the MCMI-I were substantive, but did not include changes in the guiding theory utilized in the construction of the original instrument. Underlying the MCMI-II is therefore a similar emphasis on the concepts of clinical prototypes, the inclusiveness of categories, the polythetic nature of categories, and differentiation according to severity (Millon,
Additionally, a similar process of test construction was used in the development of the new instrument.

**Theoretical Base**

There were two changes in the clinical theory base which resulted in the need to revise the MCMI-I (Millon, 1987). The first of these was the evolution of Millon's conceptualization of the basic personality styles. Based on clinical experience and research, the dimension of reinforcement which consisted of four variables (detached, dependent, independent, ambivalent) was expanded to include discordant as a variable. With the addition of this variable the earlier 4 x 2 matrix evolved into a 5 x 2 matrix to include ten as opposed to eight basic personality patterns. The two new personality styles were characterized by active discordant and passive discordant patterns (Millon, 1987). This led to the introduction of the aggressive/sadistic (active-discordant) and self-defeating/masochistic (passive-discordant) personality disorders. The second development in the clinical theory base, was the introduction of changes in the descriptive criteria of the borderline and antisocial personality disorders (Millon, 1987).

**Test Construction**

As with the MCMI-I, the development of the MCMI-II included an emphasis on sequential validation (Choca et al., 1992; Craig & Weinberg, 1993; Millon, 1987). The two major goals were: (1) to develop items for the two new personality scales, and (2) to develop items for the original scales which would ensure that they corresponded more clearly with Millon's underlying theory of personality and the criteria of the DSM-III-R. The process of substantive validation involved the development of a pool of items for the new scales and a pool of potential replacement...
items. Sources for these pools included DSM-III and DSM-III-R criteria along with traits and attributes which were identified in current theories of personality and psychopathology. A total of 364 new items were developed. Almost all of the proposed items for the clinical syndrome scales were discarded, and the other items were further evaluated for clarity and fit by a group of mental health professionals. The remaining 193 new items were added to the 175 original items of the MCMI-I. This resulted in the provisional form which was used in the structural validation of the new instrument. The form was administered to a group of patients, and once again only those items were retained which showed their highest correlation for the scale for which they were developed. These prototypic items were assessed for correlation with other scales and for congruence with Millon's modified theory of personality. The resulting research form consisted of 289 items of which 111 were new items. This provisional form was used in the external validation process. The procedures were similar to the external validation of the MCMI-I, however, it involved only one study, and the clinicians utilized DSM-III-R criteria to reach diagnoses which were used for comparison. The present study reduced the new item pool from 111 to 45 potential items. An added factor in the construction of the MCMI-II was an increased emphasis on developing norms for minority populations (Choca et al., 1992; Craig & Weinberg, 1993; Millon, 1987).

From the MCMI-I to the MCMI-II

Millon (1981, 1987) had committed himself to revising and updating the inventory as warranted by external validation and clinical experience. After the introduction of the MCMI-II it became necessary to evaluate the relationship and correlation between the MCMI-I and the new instrument.
Revision of the MCMI-I

The MCMI-II represents a comprehensive revision of the MCMI-I which includes the scales and their items, the scoring system, the approach to interpretation and an increased focus on minority issues. Two new personality disorder scales, Aggressive/Sadistic (active-discordant) and Self-Defeating (passive-discordant), were added, and three scales were included to detect potentially distorted response patterns, and subsequent invalid profiles. Furthermore, in an effort to enhance the correlation between the MCMI-II and the DSM-III-R, new descriptors for the Borderline and Antisocial Personality Scales were incorporated. In order to address the concerns about the validity of some of the individual scales, each of the items on the MCMI-I was assessed, utilizing the DSM-III-R criteria. This reevaluation was intended to establish the robustness and consistency with which each item contributed to measuring the intended construct. The process resulted in the identification of 45 redundant items. Therefore, to accommodate the new scales, and to replace the items which were found to be lacking, 45 new items were added (Millon, 1987).

Criticisms generated by external-validation studies of the MCMI-I were directed at unnecessary scale overlap as well as the lack of norms for minority groups (Choca et al., 1986; Lorr et al., 1989; Retzlaff & Gibertini, 1987; Widiger & Sanderson, 1987). To address these as well as concerns related to interpretation, Millon (1987) introduced a new system and a modification in approach to scoring. Weighted scoring as opposed to equal scoring was introduced to reduce the negative effects of scale overlapping. Incorporation of this system was intended to account for the fact that though an item may be a highly effective discriminator for one scale it may only make a limited contribution to the robustness of several other scales (Millon, 1987; Wetzler, 1990). Changes in scoring also included standards for an adjustment
when high depression and anxiety are found, because patients who exhibited high levels of either tended to receive too many personality disorder diagnoses (Wetzler, 1990). Additionally, a distinct alteration in the approach to scoring and interpretation of the basic personality patterns was a shift in emphasis to high point scale profiles as opposed to diagnoses based on individual scale BR cut off points. Finally, Millon (1987) included normative data for a number of minority groups.

Significance of Differences Between the MCMI-I and the MCMI-II

There is some evidence to indicate that the two MCMIs remain conceptually similar. In their extensive review of the MCMI literature, Craig and Weinberg (1993), enumerate three reasons which suggest that knowledge gained from research of the MCMI-I is applicable to the MCMI-II. First, Millon "intended for the MCMI-II to be isomorphic with the MCMI" (Craig & Weinberg, 1993, p. 59). Second, the studies that have addressed the correspondence between the two instruments have identified only minor differences. Third, there is evidence in the research that due to the similarity in some of the scales, the two instruments could be considered parallel forms of the same test (Craig, & Weinberg, 1993). Lorr (1993) confirms this analysis in his review of the factor analytic studies of the MCMIs. Both instruments have been shown to consistently identify four to six similar and consistent personality factors and five symptom scales (Lorr, 1993). Thus the instruments appear to measure similar dimensions, but there continue to be questions about the ability of either instrument to identify all the personality disorders and clinical symptoms suggested by Millon (1983, 1987).

While further research may support the assertion that the MCMI-I and II are essentially similar instruments, there is significant evidence to the contrary (Gibertini, 1993). The development of the MCMI-II involved substantive changes which has led
Gibertini (1993) to conclude that it is a very different instrument from the MCMI-I. The inclusion of two new categories of basic personality structure (active and passive discordant types) has necessitated the introduction of a new set of prototypical items. However, the length of the instrument was not altered, and therefore to accommodate these prototypical items, a significant change resulted in the pool of descriptive items. Furthermore, the introduction of the two new personality disorders actually resulted in six new categories, the two new ones as well as the revision of the four original ones. As Gibertini (1993) states,

In the new instrument there are now three types of aggression to be distinguished: passive (passive-aggressive), active (antisocial), and coercive (aggressive-sadistic). Dependency has a similar configuration: passive (dependent), active (histrionic), and collapsed (self-defeating). (p. 76)

Thus, in the process of interpreting the results and formulating a differential diagnosis, the clinician is faced with having to incorporate significant theoretical changes (Gibertini, 1993).

A number of other factors are suggestive that the MCMI-II is a substantially different instrument. The first of these is a significant change in the external validity studies conducted in the process of the development of the MCMI-II. In the third step of the validation process, clinicians were asked to record up to three diagnoses on Axis I and Axis II for a sample of patients to whom the research forms of the instrument had been administered. The clinicians, who were involved in this part of the sequential validation process for the MCMI-I, were provided with descriptive criteria based on Millon's formulations. However, the clinicians who participated in this step for the MCMI-II, were given descriptive criteria which were essentially those utilized in the DSM-III-R, except for the Sadistic personality disorder which was not included in the official nosological system (Gibertini, 1993). Thus, the inclusion or rejection of prototypical and descriptive items for the MCMI-II was based on different
criteria than used in the validation process of the MCMI-I. Secondly, the raters had a different number of diagnostic categories from which to choose (8 for MCMI-I raters and 13 for the MCMI-II raters) and the additional categories were conceptually new. It is therefore likely that inter-rater reliability suffered (Gibertini, 1993). The potentially significant differences between the two instruments suggests that the MCMI-II will need to be subject to research and validation as an instrument distinct from the MCMI-I (Millon, 1987; Piersma, 1989, 1993).

The MCMI-I and MCMI-II and Substance Abuse

Both the MCMI-I and MCMI-II include scales which are intended to identify individuals who have a history of alcohol and/or drug dependence. The theoretical base as well as the performance of the scales have been questioned as evidenced in the following review of pertinent literature.

Theoretical Base

The MCMI-I and the MCMI-II both have scales which are intended to identify individuals who have a history (current and past) of alcohol and/or drug dependence. Whereas Millon has been articulate and explicit in his theory of personality and related psychopathology, a review of his writings reveals only minimal reference to substance abuse (Millon, 1969, 1973, 1981, 1990). Furthermore, neither the manual for the MCMI-I nor the MCMI-II speaks to the theoretical underpinnings for these diagnostic categories (Millon 1983, 1987). Indeed, the first two draft versions of the MCMI-I, which emerged as part of the sequential validation process, did not include substance abuse scales. These scales were added in the final process of scale validation when clinicians were asked to identify additional or alternate scales which would add to the clinical utility of the MCMI-I. It is therefore difficult to ascertain whether or not the
"substantive-theoretical" step in the validation process was adhered to or violated in the construction of these scales.

Historically, the identification of a substance abuse personality has been the subject of significant debate (Nerviano & Gross, 1983). For example, no research has led to substantiation that all alcoholics fit Freud's "oral" personality (Choca et al., 1992). Several factor analytic studies aimed at clarifying and categorizing the underlying constructs of both instruments, have shown consistency in identifying an acting-out substance abuse dimension (Choca et al., 1986; Lewis & Harder, 1990; Lorr et al., 1989; Lorr, Strack, Campbell, & Lamnin, 1990; Strack, Lorr, Campbell, & Lamnin, 1992). This suggests that substance abuse is one of the basic dimensions assessable by the instruments (Lewis & Harder, 1990). However, as Choca and colleagues (1992) state,

> From a characterological viewpoint alcoholics and substance abusers are more different than they are alike. As sophistication increases in measuring characterological tendencies, the old debate seems obsolete because it becomes clear that, even within one of the addictive groups, the patients are not alike. (p. 123)

This perspective is supported by a number of studies with the MCMI-I which identify substance abuser subtypes (Bartsch & Hoffman, 1985; Craig et al., 1985; Gibertini & Retzlaff, 1988; Mayer & Scott, 1988). First, all of these studies (Bartsch & Hoffman, 1985; Craig et al., 1985; Lorr, 1993; Mayer & Scott, 1988) identified a group of patients who scored high on the Negativistic Personality scale with a tendency to score high on the Borderline and Paranoid Personality scales. Prevailing behavioral characteristics included mood swings, irritability, social aggressiveness, acting out, and suspiciousness. For example, in Mayer and Scott's (1988) sample of alcohol and/or drug abusers, 38% fell in this group-high scores on Negativistic, Borderline and Paranoid Personality scales. A second group of alcohol and/or drug abusers identified in each of these studies (Bartsch & Hoffman, 1985; Craig et al.,
1985; Lorr, 1993; Mayer & Scott, 1988) had elevated scores on the Narcissistic and Histrionic scales and drinking was, "a manifestation of a lifestyle of self-indulgence and thrill seeking" (Bartsch & Hoffman, 1985, p. 711). A third group identified by these studies consisted of alcoholic patients who scored high on the Compulsive Scale and who were likely to use alcohol as permission to express feelings or act out (Bartsch & Hoffman, 1985; Craig et al., 1985). Lastly, all of the studies identified a cluster of alcohol and/or drug abusers who shared schizoid, avoidant, and dependent traits (Bartsch & Hoffman, 1985; Craig et al., 1985; Gibertini & Retzlaff, 1988; Mayer & Scott, 1988). This wide variety in the types of alcohol and/or drug abusers identified by these cluster studies supports Choca's and his colleagues' (1992) view that linking diagnosis of substance abuse with personality may be difficult and inefficient. Of course, this is precisely the intent of an instrument like the MCMI-II which emphasizes the polythetic nature of categories and the connection between personality disorders and symptom disorders such as Alcohol Dependence and/or Abuse and Drug Dependence and/or Abuse. Thus the clinical theory base which Millon utilized to construct the MCMIs, has not been validated by these studies and, may therefore, be inadequate as it pertains to the identification of alcohol and/or drug abusers.

**Discriminant Validity of the Alcohol and Drug Dependence Scales**

The skeptical view of the connection between substance abuse and personality may well be supported by the continued debate of the efficacy of the Alcohol and Drug Dependence scales of the MCMI-I (Craig, 1993; Craig & Weinberg, 1992). Several studies (Flynn & McMahon, 1983, 1984; McMahon et al., 1985), as well as the test construction data reported by Millon (1983), present supportive evidence for the accuracy of these scales. However, more recent studies have not corroborated these
positive findings. Indeed, the data generated by these studies has raised questions about the validity and discriminant capacity of these scales and therefore about the use of the MCMI-I as an assessment tool for substance abuse (Bryer et al., 1990; Calsyn et al., 1990, 1991; Gibertini & Retzlaff, 1988; Jaffe & Archer, 1987; Marsh et al., 1988; McCann, 1990; Miller & Streiner, 1990). Even though the MCMI-I was the focus of most of these studies, a review of this literature is relevant to this current study of the Alcohol and Drug Dependence scales of the MCMI-II. As Bryer and colleagues (1990) state,

> Although the MCMI-II may represent an improvement over the MCMI-I regarding the overall effectiveness of the substance abuse scales, many of the ... issues regarding item content, scale composition and intercorrelation, and the assessment of substance abuse through personality correlation apply to the MCMI-II as well. (p. 440)

Furthermore, Millon (1987), reports that the prototypical items for the substance abuse scales are the same for both instruments.

The strongest evidence for the effectiveness of the Alcohol and Drug Dependence scales was reported by Millon (1983, 1987) in the test manuals. Based on a sample of 978 psychiatric patients, the author found that 88% of the subjects were classified correctly on the Alcohol Dependence scale, and the Drug Dependence scale was reported to correctly classify 94% of the subjects. In a cross validation study consisting of 256 psychiatric patients, Millon (1983) reports that the MCMI-I correctly classified subjects who had a history of alcohol abuse 89% of the time. The Drug Dependence scale was also reported to perform well identifying 94% of the subjects correctly. Though he uses different terminology, Millon (1987) reports similar results, for these scales of the MCMI-II, in a cross validation study which involved a sample of 703 psychiatric patients. He reports that the overall diagnostic power, proportion correctly classified, for the Alcohol Dependence scale is 97% with
a sensitivity, true positive rate, of 87%. The results for the Drug Dependence scale are equally positive with Millon reporting an overall diagnostic power of 94% with a sensitivity of 72%.

Three studies, supportive of the MCMI-I as a measure to assess substance abuse, focus on the stability of scores during the course of treatment and on the score characteristics. Flynn and McMahon (1983) administered the inventory at three intervals (intake, 1 month, 3 months) to two groups of known drug abusers. They reported an acceptable stability coefficient (.74) between the administrations at 1 and 3 months. The correlation of scores between intake and 1 month (group A, .45, group B, .55) and intake and 3 months (both groups, .56) into treatment were lower than expected. However, the authors attribute these latter findings to the effect of the withdrawal from drugs and the adjustment to treatment on the intake scores. They further note that the .74 level of stability between the 1 and 3 months scores replicates the 5 week test-retest coefficient (.74) reported by Millon. In an expanded study to determine scale stability, McMahon and colleagues (1985) found that scores of alcoholics and drug abusers on the Alcohol and Drug Dependence scales of the MCMI-I remained more consistently elevated over time than scores on other scales. Additionally, Flynn and McMahon (1984), using the same sample as in their previous study, report that the characteristic scores of the present sample of known substance abusers corresponds to those found in the standardization sample Millon (1987) used. However, as Calsyn and colleagues (1991) point out, Flynn and McMahon (1984) neglected to report the percent of correct classifications (sensitivity) and, "An examination of the standard deviation (15.88) and the minimum value of 35 they did report suggests that a significant number of cases would obtain scores less than 75 [BR]." (Calsyn et al., 1991). These reliability studies suggest that the Alcohol and Drug Dependence scales measure enduring traits, results that contradict the studies.
cited earlier which indicated that there was no consistent constellation of traits that
defined the alcohol and/or substance abusing individuals (Bartsch & Hoffman, 1985;
Craig et al., 1985; Gibertini & Retzlaff, 1988; Lorr, 1993; Mayer & Scott, 1988).
However, though reliability is necessary before validity can be shown, reliability does

In contrast, a number of other external validity studies offer evidence of
limitations in the efficacy of these scales for identifying subjects with alcohol and
substance abuse histories (Bryer et al., 1990; Gibertini & Retzlaff, 1988; Jaffe &
Archer, 1987; McCann, 1990; Miller & Streiner, 1990). In a study of 163
rehabilitated opiate addicts, Marsh and colleagues (1988) found that both the Drug and
Alcohol Dependence scales were among the elevated scales on the MCMI-I, with the
Drug Dependence scale having the highest mean score. However, on examining
individual data, the authors found that only 49% of the subjects had significant
elevations (BR \geq 75). Thus, the Drug Dependence scale failed to identify 51% of the
present sample of known opiate addicts as having or having had a history of substance
abuse. Bryer and colleagues (1990) report similarly ambivalent data. In a sample of
561 adult psychiatric inpatients, the Alcohol and Drug Dependence scales of the
MCMI-I identified 43% of the alcoholics and 49% of the drug abusers (BR \geq 75) with
false positives for both scales exceeding 50%. Diagnoses generated by psychiatrists
in clinical practice were used as the standard for comparison. The authors
acknowledge the lack of reliability of clinical diagnoses, but conclude that their
findings raise serious questions about the validity of the scales. At the same time,
Miller and Streiner (1990), in a concurrent validity study between the MCMI-I
Alcohol Dependence scale and the MacAndrew Scale of the MMPI, found that the
MCMI-I Alcohol Dependence scale identified only 43% of patients diagnosed with
Alcohol Abuse and/or Dependence based on DSM-III-R criteria. Moreover, Gibertini
and Retzlaff (1988) found that of their sample, drawn from an inpatient alcohol rehabilitation setting, only 17% had elevated scores (BR ≥ 75) on the Alcohol Dependence scale. Using an known sample of drug abusers as well as a sample of psychiatric patients who did not have a diagnosis of substance abuse, Calsyn and colleagues (1990) found that 39.4% (BR ≥ 75) of the drug abusers were identified correctly. Thus 60.6% of the sample of drug dependent individuals did not obtain clinically significant scores (BR ≥ 75). However, false positive diagnoses for the psychiatric sample was only 12%. Although the Drug Dependence scale is therefore unlikely to misclassify psychiatric patients, the authors suggest that the use of the scale as an assessment tool for substance abuse is limited. In a subsequent study, the authors (Calsyn et al., 1991) confirm the questions about the effectiveness of the Drug Dependence scale. Using a sample of 110 veterans who presented for drug treatment, they found that only 49% received scores greater than or equal to 75 (BR) on the Drug Dependence scale.

Further questions about validity have been raised as a result of critical analyses of the scales. The critique in this regard has particularly focused on the content of scale items and the correlation between the scales (Jaffe & Archer, 1987; Marsh et al., 1988; McCann, 1990). Jaffe and Archer (1987) examined the Alcohol and Drug Dependence scales as part of their study of five self-report assessment tools. Even though they report that both scales were, "significantly involved in predicting a variety of individual drug classes and in predicting poly drug use patterns" (Jaffe & Archer, 1987, p. 250), they express concern about the specificity of the performance as well as the independence of the scales. In this study the Alcohol Dependence scale was more likely to identify a variety of drug use classes than the Drug Dependence scale, and the correlation between the two scales was found to be .65 as opposed to the -.08 reported by Millon (1983). It should be noted, however, that this study was
conducted with a sample of "normals" which violates Millon's (1983, 1987) assertion that the MCMIs are only intended to be used with a clinical population. Marsh and colleagues (1988), in the study cited above, also attribute the poor performance of the Drug Dependence scale, in part, to the construction and overlap of the scales. McCann's (1990) study utilized a complex method of factor analyses to examine convergent and discriminant validity for the clinical syndrome scales of the MCMI-II. His findings support the poor discriminant validity of the Alcohol and Drug Dependence scales, but he does not attribute this to scale overlap. "Limited content sampling of constructs, common item artifact, ... [and to a lesser degree] ... item overlap" (McCann, 1990, pp. 474-5), are cited as contributing to these limitations.

The prototypic items for the Alcohol Dependence scale were not changed from the MCMI-I to the MCMI-II. Using the data provided by Millon (1983), Gibertini (1993) calculated the sensitivity for the Alcohol Dependence scale for the MCMI-I (.74). Subsequently, he reported that the Alcohol Dependence scale had increased in sensitivity form 74% to 87%. Furthermore, results from factor analytic studies suggest that the MCMI-II is consistent in identifying personality patterns associated with substance abusing patients (Lorr, 1993). However, the discriminant validity of the Alcohol and Drug Dependence scales lacks consistent support. One concern that has been expressed is the lack of studies which use a reliable standard for comparison (Bryer et al., 1990, Piersma, 1993). Craig and Weinberg (1992) state, "There are contradictory reports concerning the test's predictive validity with drug abusers, and research on this matter is in its infancy" (p. 254). External validation studies which use reliable diagnostic procedures to establish the criterion for comparison are needed to bring clarity to the utility of the MCMI-II.
Base Rate Scores

Prediction and classification of behavior may be approached through either clinical or statistical (actuarial) methods (Meehl, 1954). Traditionally the actuarial approach has involved the procedure of transforming raw scores into standard scores (Millon, 1987). Statistics such as the standard, or T score, assume a normal distribution of the items being measured. As Millon (1987) states, "This assumption is not met when a set of scales is designed to represent either personality 'types' or clinical 'syndromes,' since neither is normally distributed nor of equal prevalence among patient populations" (p. 93). Instead of utilizing standard scores, and the traditional two standard deviations to make a diagnosis, the concept of prevalence base rates was introduced (Hsu, 1985; Meehl, 1954; Meehl & Rosen, 1955). Meehl and Rosen (1955) referred to base rates as "actuarial or experience tables ... [on the basis of which] ... diagnostic and prognostic statements can often be made with a high degree of accuracy" (p. 194). These authors (Meehl & Rosen, 1955) warned that if cut scores for instruments did not take into account the population base rates, the test results may be more inaccurate than classifications which can be made statistically with a 66% to 75% hit rate based on actuarial tables alone. For the MCMI-I and MCMI-II base rate cut off points were calculated from distributions of raw scores in a manner which assured that the proportion of patients who score above this point will match the actual prevalence in a normative, national population (Millon, 1987).

Prototypical Categorization

Prototype theory is a core construct on which the MCMI-I and MCMI-II as well as the DSM-III and DSM-III-R are based. Traditionally categories have been thought of as distinct groupings with clear and rigid boundaries (Broughton, 1984).
Prototypical categories, however, allow for the inclusion of items which do not necessarily adhere to all the descriptive criteria associated with the exemplar items. Though initially developed in the literature of cognitive psychology (Broughton, 1984; Rosch, 1973), Cantor and Mischel (1979) conducted studies which suggested that this type of categorization applies to types of people as well as natural objects.

In subsequent work Cantor, Smith, French, and Mezzich (1980) extended this view to the process of psychiatric diagnosis. The authors reported that a study of the hierarchy of psychiatric categories is consistent with the hierarchy of natural-object categories. As Cantor and colleagues (1980) stated, "while clinical diagnosis may appear 'messy' and disordered from the classical perspective, it seems orderly and principled from the perspective of the prototype view" (p.181). In a study which focused on the different strategies for developing personality scales, Broughton (1984), introduced a prototype strategy as one of the alternatives. Results indicated that this strategy was statistically superior to the other strategies.

The Statistic Kappa

The validity of the diagnosis is inextricably related to reliability and to determine the latter a measure had to be developed to quantify diagnostic agreement. Lord and Novick ([1968] cited in Fleis, Spitzer, Endicott, & Cohen 1972) summarize the dilemma as follows:

In the absence of ultimate criteria for validating psychiatric diagnosis, such as are usually provided by various laboratory tests in other branches of medicine, we are thrown back on determining its reliability, since the degree of agreement between diagnosticians necessarily represents the upper limit of validity. (p. 186)

To address this situation these authors identified a statistical measure to quantify the agreement between psychiatrists who assign diagnoses across a sample of patients. Chi square was rejected as an inadequate measure of reliability because it identifies
association of any kind and doesn't specifically address agreement (Fleis, 1973). The statistic weighted \textit{kappa} was developed as an alternative index of agreement that corrects for chance agreement (Fleis et al., 1972; Helzer et al., 1977), and is described as having the following properties:

(1) it gives partial credit for moderate disagreements; (2) it measures agreement corrected for, i.e., over and above what is expected purely by chance; (3) it is scaled from -1 to +1 so that negative values indicate worse than chance agreement, 0 indicates exactly chance agreement, and positive values indicate better than chance agreement; and (4) it has a well-defined standard error which permits a statistical assessment of the significance of the observed degree of agreement. (p. 187)

Although it has been the statistic that has been used to report levels of inter-rater reliability throughout the literature under review, it has one limitation that diminishes its usefulness. Grove, Andreasen, McDonald-Scott, Keller, and Shapiro (1981) have shown that \textit{kappa} should not be used for a diagnosis which has a frequency of less than 5% in a given sample.

The Structured Clinical Interview for the DSM-III-R

The classification and diagnosis of mental disorders has received attention throughout history (Spiker & Ehler, 1984). This process has taken many different forms and at various times was based on biological, cultural, and spiritual criteria. Modern descriptive psychiatry can be traced to Emil Kraepelin (1855-1926) who introduced a nosological system which divided the major psychoses into two groups, dementia praecox and manic depressive psychosis (Spiker & Ehler, 1984). However, the emphasis on psychiatric diagnosis gained most momentum with the discovery and introduction of psychoactive drugs in the 1950s and 1960s. This was due to the fact that the various psychotropic medications were associated with treatment of specific
conditions making an accurate process of differential diagnosis essential (Spiker & Ehler, 1984).

Studies which evaluated the status of the reliability of psychiatric diagnosis in the middle of the century resulted in a considerable amount of negative critique (Kreitman, 1961; Spitzer & Fleis, 1974; Ward, Beck, Mendelson, Mock, & Erbaugh, 1962; Zubin, 1967). It was found that the process of differential diagnosis was most negatively affected by the fact that psychiatrists utilized different, "formal inclusion and exclusion criteria to summarize patient data into psychiatric diagnoses" (Spiker & Ehler, 1984, p. 295). Spitzer, Endicott, and Robins (1975) labeled this phenomenon "criterion variance." A second major source for the poor reliability of psychiatric diagnosis related to the fact that each clinician would gather and pay attention to different information on which to base the final assessment (Spiker & Ehler, 1984). Spitzer and his colleagues labeled this phenomenon "information variance." These variances occurred because the diagnostic systems were inadequate, and as a result of a lack of standardization in the process of collecting data on which to base assessments (Endicott & Spitzer, 1978; Sholomskas, 1990; Spiker & Ehler, 1984; Ward et al., 1962).

In order to address the need for accurate differential diagnosis, changes were initiated both in the nosological system and approach to the assessment process (Harkavy-Friedman, 1989; Sholomskas, 1990; Spiker & Ehler, 1984). As early as 1967 Wing, Birley, Cooper, Graham, and Isaacs, in reflecting both on the complexity of psychiatric evaluation and the need to standardize the psychiatric interview, stated,

> The components of the diagnostic process include the skill and attitudes of the psychiatrist, this method of examination, the replies and attitudes of the patient, the amount of material obtained, the method of recording and coding data, the rules of classification and how these are applied. It is reasonable to suppose that if each component could be fully standardized, there would be no disagreement at all as to how a case should be categorized. (p. 499)
Therefore, to address the issue of criterion variance, renewed efforts were made to develop more reliable nosological systems (Harkavy-Friedman, 1989; Sholomskas, 1990; Spiker & Ehler, 1984). Additionally, standardization of the diagnostic process was addressed through the development of structured clinical interviews. The development of these new instruments was intended to assure consistent application of the newly developed diagnostic criteria, and to reduce the impact of information variance on the process of differential diagnosis (Endicott & Spitzer, 1978; Harkavy-Friedman, 1989; Helzer, Clayton, Pambakian, & Woodruff, 1978; Page, 1991; Spiker & Ehler, 1984; Sholomskas, 1990; Wixted, Morrison, & Rinaldi, 1993; Weiss, 1993).

The Development of Operational Diagnostic Criteria

As it became more apparent that psychiatric diagnosis was to a large degree subjective, efforts were made to develop operational and consistent diagnostic criteria. The first of a number of diagnostic schemas which included standardized categories were the Feighner criteria (Feighner et al., 1972). This model consisted of diagnostic criteria for 14 psychiatric disorders including distinctions between primary and secondary affective disorders. Like the systems that followed, Feighner and colleagues (1972) utilized a descriptive and atheoretical approach to diagnosis. The criteria were derived from clinical and research studies and provided a uniform language for describing patients' conditions. Spitzer, Endicott, and Robins (1978) expanded and modified the Feighner criteria and developed the Research Diagnostic Criteria (RDC). This system provided criteria for 24 diagnostic categories and provided a glossary of psychiatric definitions. As the use of objective criteria increased the reliability of diagnosis, these concepts and principles were eventually incorporated into the development of the DSM-III (Rubinson & Asnis, 1989;
The Development of Structured Clinical Interviews

Although the development of these nosological schemas provided more uniform sets of diagnostic criteria, the mere presence of these new categories did not ensure that they were being utilized accurately (Harkavy-Friedman, 1989; Rubinson & Asnis 1989; Sholomskas, 1990). As Endicott and Spitzer (1978) state, "To the extent that a procedure for making diagnostic judgments is unreliable, a limit is placed on the validity for any use" (p. 837). Thus, Wing, Cooper, and Sartorious (1974) introduced one of the first structured interviews, the Present State Exam, which included standardized questions but also emphasized, "a flexible clinical style" (Harkavy-Friedman, 1989, p. 84). The need for an objective interview instrument was also driven by the desire to do research on the new nosological systems, and thus the introduction of each new set of criteria generally resulted in the development of one or more structured interviews (Harkavy-Friedman, 1989; Helzer et al., 1978; Page, 1991; Spiker & Ehler, 1984; Sholomskas, 1990; Wipted et al., 1993; Weiss, 1993). Four of these instruments gained wide acceptance. The Feighner criteria were followed by the development of the Renard Diagnostic Interview (Helzer, Robins, Croughan, & Welner, 1981), the RDC were used as a basis for the Schedule of Affective Disorders and Schizophrenia (Endicott & Spitzer, 1978), and the DSM-III was accompanied by the SCID. Additionally, the Diagnostic Interview Schedule (Robins, Helzer, Croughan, & Ratcliff, 1981), based on the Feighner criteria, the RDC and the DSM-III, was developed as a structured interview to facilitate use in epidemiological studies. This instrument could be administered by a trained, non-clinician (Sholomskas, 1990).
Structured Interviews: General Considerations

While there are specific differences in the various interviews, there are a number of defining characteristics that they have in common. In general, a structured interview consists of a set of formal questions which are organized according to, "internally consistent rules that govern the content of questions asked of an interviewee, the order in which topics are covered, and the specific kind of information sought" (Weiss, 1993, p. 179). In addition to the structure of the interview, rules are provided to arrive at diagnoses. In an effort to maintain clinical flexibility, as suggested by Wing and colleagues (1974), many structured clinical interviews are intended to be administered by professionals with clinical training. This allows for the inclusion of open ended questions which require the use of clinical acumen and judgment (Weiss, 1993).

The goals of structured interviews have been well articulated, but there has been resistance from clinicians to using this assessment technique (Harkavy-Friedman, 1989; Rubinson & Asnis, 1989; Spiker & Ehler, 1984; Weiss, 1993). Harkavy-Friedman (1989) summarized the common goals of structured interviews as follows:

(1) to standardize descriptive information gathered via psychiatric interviews; (2) to standardize the scoring of information gathered by psychiatric interview techniques; and/or (3) to gather information in a manner that allows for the application of specific diagnostic systems. (p. 84)

The objections to the use of structured interviews have not focused on these goals, but rather on the impact of the implementation on clinical practice. The concerns tend to take three forms (Rubinson & Asnis, 1989). First, structured interviews are considered as counter therapeutic because it forces the clinician to deviate from accepted clinical practice. The format is perceived as rigid, too focused on
psychopathology, and as a result an impediment to the formation of a therapeutic alliance. Second, some individuals feel that the use of these assessment techniques needlessly devalue the clinician. Third, the use of structured interviews is seen as unnecessary, time consuming, and too costly (Rubinson & Asnis, 1989).

Statistical analyses, as well as verbal reports by clinicians who have used these instruments, support the efficacy of structured interviews as standardized models of diagnostic practice (Endicott & Spitzer, 1978; Harkavy-Friedman, 1989; Helzer et al., 1978; Page, 1991; Spiker & Ehler, 1984; Sholomskas, 1990; Wixted et al., 1993; Weiss, 1993). It has been consistently shown that the structured interview technique enhances the reliability of psychiatric diagnosis. This suggest that the information gathered during a structured interview is replicable and consistent across different interviewers (Harkavy-Friedman, 1989). While, "reliability is a prerequisite for validity" ([Italics in original], Page, 1991), the fact that an instrument is reliable does not assure its validity. In the development of structured interviews the emphasis has been on developing reliable instruments which produce diagnoses that are consistent with the nosological system on which they are based. This has left the issue of validity to those who develop the diagnostic categories (Page, 1991). In spite of this somewhat narrow focus on reliability it is clear that as Weiss (1993) states:

Undoubtedly, if the clinical interview is being undertaken for purposes that will include systematic empirical research, then the use of a structured clinical interview is highly desirable if not an absolute requirement. Certainly, the state of the art is to use a structured interview for making diagnoses. (p. 180)

Validation of the SCID

One of the more recent structured clinical interviews, the SCID, is based on the current nosological system as found in the DSM-III-R. This system of classification is grounded in well researched and validated operational, diagnostic
criteria (Spitzer et al., 1979). Thus, both the form of this particular instrument-a structured clinical interview-and the nosology on which it is based, represent what is currently the most valid and reliable approach to diagnosis (Sholomskas, 1990; Spitzer et al., 1990; Weiss, 1993).

Though the SCID is a new diagnostic tool, it has received a significant amount of attention in the literature. Williams and colleagues (1992) report that, "over 100 published studies ... have used the instrument to select or describe their study samples" (p. 630). An examination of a representative sample of this research provides support for the methodology and design of this project as well as evidence of the reliability of the SCID in generating substance abuse diagnoses. The review of this first set of studies is intended to focus on methodology, as opposed to results, as each uses a SCID generated Axis I (DSM-III and DSM-III-R) diagnosis as the standard for comparison. Stukenberg, Dura, and Kiecolt-Glaser (1990) utilized the SCID in identifying cases of depression. They used the SCID generated diagnoses as the standard to which results from three depression screening scales were compared. Similarly, in a study with a sample of in-patients, Hamer, Sanjeev, Butterworth, and Barczak (1991) utilized the SCID to generate diagnoses to which the results of the Hospital Anxiety and Depression Scale were compared. In an effort to identify potential subtypes of cocaine abusers, Nunes, Quitkin, and Klein (1989) administered the SCID to generate both the substance abuse as well as secondary diagnoses, and Bryant, Rounsaville, Spitzer, and Williams (1992) used SCID generated diagnoses to examine the impact of the concurrent occurrence of substance abuse and psychiatric disorders on the reliability of the diagnostic process. Similarly, in a study of drug abusers, the purpose of which was to establish the relationship between precipitated opiate withdrawal and the concept of a dependence syndrome, Kosten, Jacobson, and Kosten (1989) used the SCID to establish diagnoses and severity of dependence.
Additionally, Toner, Gillies, Prendergast, Cote, and Browne (1992) examined the pattern of substance abuse among a group of chronically mentally ill patients, and used the SCID to generate the substance abuse diagnoses. None of these studies report limitations as a result of the incorporation of the SCID in the research design. Therefore, the review of this representative set of studies suggests that the use of SCID generated Axis I (DSM-III-R) diagnoses as the standard of comparison, is warranted.

The following is a summary of the literature examining inter-rater reliability for the SCID expressed in terms of kappa. In the SCID manual, Spitzer and colleagues (1990) reported on a study in which 506 pairs of interviews were conducted at six sites using a test-retest design. The subjects were selected randomly and the interviewers did not have access to any data other than that collected during the interview. Even under these stringent conditions, the kappas for the Axis I diagnoses were reported to be generally equivalent to the kappas reported for other structured interview schedules. As revealed in the literature review, the reliability of the structured interview schedules has generally been very positive. However, it would have been more helpful if Spitzer and colleagues (1990) had reported the numerical values of the kappa coefficients on which they based their positive evaluation of the performance of the SCID.

In an extensive follow up to the above study, Williams and colleagues (1992) drew subjects from six sites in the United States and one in Germany. A total of 592 subjects were interviewed using a test-retest design. There were 25 interviewers each of whom, for purposes of comparison, were paired in multiple combinations. A number of different SCID forms were used as well. The authors found that the kappas for current diagnoses for most of the clinical symptom scales were above .60. However, two of the drug categories, cannabis dependence and poly-drug...
dependence, were below .60. The authors express disappointment with these reliability coefficients as they are, in general, not better than the values obtained in the DSM-III reliability studies in which non-structured interviews were used. The authors conjecture that, "Perhaps clinicians with an already-high level of expertise and commitment to using diagnostic criteria do not improve their reliability by using a structured interview that has the flexibility of the SCID" (Williams et al., 1992).

A number of studies have examined the efficacy of the SCID in identifying depressive and anxiety disorders. In a study of DSM-III criteria for melancholia and endogenous depression, Copolov and colleagues (1986), found that ratings of SCID interviews conducted with nine patients resulted in a kappa of .79 among eight interviewers. The authors report that this is an improvement over the inter-rater reliabilities found in the DSM-III field trials (kappa = .67) in which non-structured interviews were utilized. In the Federal Republic of Germany, Maier, Phillip, and Buller (1988) studied the effectiveness of the Upjohn Version of the SCID ([SCID-UP], Spitzer & Williams, 1988) with a sample of 40 psychiatric inpatients with either depressive or anxiety syndromes. Patients were interviewed a total of four times, two times by the treating psychiatrist who used a checklist, and two times by the testing psychiatrist using the SCID-UP. They reported kappa coefficients of .70 for major depressive episodes, .83 for panic disorder, and .53 for panic disorder with avoidance behavior. These reliability coefficients are uniformly higher than the kappas, .59, .66, and .48 respectively, which were the inter-rater reliability figures when diagnoses were made on the basis of a non-standardized interview. Utilizing videotaped interviews and paired raters, Riskind, Beck, Berchick, Brown, and Steer (1987) examined the inter-rater reliability of generalized anxiety and major depressive disorder diagnoses generated by the SCID. They found that the overall kappa was .74 with reliability coefficients of .72 for major depressive disorder and .79 for
generalized anxiety disorder. Due to the design of this study it was necessary to complete the entire SCID for each patient, and therefore the authors further report a kappa of .72 for all other diagnoses.

A number of studies have been conducted utilizing the SCID to diagnose substance abuse disorders. Kosten, Bryant, and Rounsaville (1991) evaluated the reliability of the SCID with a substance abuse population both in an inpatient and a non-clinical setting. A test-retest design was used to compare SCID generated diagnoses by two raters. Using a very large sample, kappas were reported to range from .51 to .95 with a kappa of .95 for the diagnosis of drug dependence. The authors conclude, "future studies using the SCID in substance abusing populations seems well warranted" (Kosten et al., 1991, p. 218). Supporting this latter finding were Skre, Onstad, Torgersen, and Kringlen (1991), who reported kappas of .96 and .93 for alcohol abuse or dependence and psychoactive substance use disorder respectively. Citing this excellent inter-rater agreement these authors state, "Our findings confirm that the use of SCID leads to reliable classification, both for broad, specific and combinations of DSM-III-R Axis I diagnoses" (Skre et al., 1991, p. 17). The research application of the SCID is therefore supported both by studies which have incorporated the SCID generated diagnoses into the research design, as well as, by the generally high levels of inter-rater reliability that have been reported for this instrument.

Summary

The review of this literature provides a clear rationale for the present study. The Alcohol and Drug Dependence scales of the MCMI-I functioned inconsistently in identifying patients with substance abuse histories. Though the MCMI-I was revised, the prototypical items used to construct the new Alcohol and Drug Dependence scales
were not changed. Furthermore, even though the MCMI-II is conceptually similar to the MCMI-I, the changes to the instrument were significant. Thus, the MCMI-II needs to be validated as a diagnostic instrument apart from the MCMI-I. Previously conducted external validation studies have been called into question because they utilized diagnoses generated by psychiatrists in general practice as the standard for comparison. These critiques were confirmed by the studies which consistently showed abysmal reliability coefficients for diagnoses generated in this manner. Structured clinical interviews, such as the SCID, represent the state of the art in establishing reliable psychiatric diagnoses. Of particular significance for the present study, the SCID has been shown to be a reliable instrument in the diagnosis of Alcohol and Drug dependence as well as Alcohol and Drug abuse using DSM-III-R criteria. Together these facts support the need for the present external validation study which utilizes the SCID to generate substance abuse diagnoses to which to compare the results from the MCMI-II Alcohol and Drug Dependence scales.
CHAPTER III

METHODOLOGY

As evidenced by the literature review, there continues to be significant concern about the efficacy of the MCMI-I and MCMI-II for diagnostic purposes. With the introduction of operational criteria as the base of the current nosology (DSM-III-R), reliability in diagnosis has increased. It has therefore become feasible to answer questions about the validity of assessment instruments more definitively. The use of structured clinical interviews has further added to the accuracy of diagnosis, and the SCID has been shown to be a reliable instrument when used to generate DSM-III-R, Axis I diagnoses. To date most studies critical of the MCMI-I have relied on the open-ended, unstructured diagnosis made by psychiatrists as the standard for comparison. As Piersma (1993) states, "Given the much lesser reliability of unstructured interviews compared to structured interviews, it hardly seems fair to criticize the MCMI because it does not coincide with a criterion that is highly unreliable" (p. 13). Therefore, in order to provide more reliable external validation of the MCMI-II, the present study will use SCID generated diagnoses as the standard of comparison.

Description of Method

The purpose of the present study is to assess the discriminant validity of the MCMI-II Alcohol and Drug Dependence scales. A pre-test only design was utilized to generate data for a sample of 73 individuals who presented for treatment at a substance abuse center. In addition to seven other assessment instruments, the participants
completed the MCMI-II. They were also assessed by means of a diagnostic interview utilizing the SCID. The BR scores of the MCMI-II and the SCID generated DSM-III-R substance abuse diagnoses comprised the data which was used to examine the performance of the Alcohol and Drug Dependence scales.

The present study focuses on two aspects of the performance of the Alcohol and Drug Dependence scales. First, the sensitivity and specificity of the two scales were examined, and secondly the BR scores of the scales were compared for each of two homogeneous sub-samples: (1) alcohol related diagnoses only, and (2) drug related diagnoses only. According to Gibertini and colleagues (1986), "it is useful to frame the operating characteristics of a test in terms of conditional probabilities" (p. 556). Millon (1987) acknowledged the utility of this approach and reported the validity data which had been generated as part of test construction in terms of two proportions: (1) sensitivity, the probability that the test is positive given the disorder is present; and (2) specificity, the probability that the test is negative given the disorder is absent. In order to establish the sensitivity of the Alcohol and Drug Dependence scales for the present sample, the standard cut off BR score (BR = 75) was used to identify participants who scored positive (BR > 75) on the Alcohol or Drug Dependence scales. The SCID generated diagnoses provided the indicators of presence of Alcohol and Drug Abuse and/or Dependence. The resulting proportions, the sensitivity of the Alcohol and Drug Dependence scales for the present sample (see pg. 74 for calculation procedures), were compared to the proportions reported by Millon (1987) for the normative, test construction sample. The statistical analysis followed the standard procedures for the Chi-Square Goodness of Fit Test. In order to establish the specificity of the Alcohol and Drug Dependence scales for the present sample, the standard cut off BR score (BR = 75) was used to identify participants who scored negative (BR < 75) on the Alcohol or Drug Dependence scales. The
SCID generated diagnoses provided the indicators of absence of Alcohol and Drug Abuse and/or Dependence. The resulting proportions, the specificity of the Alcohol and Drug Dependence scales for the present sample (see pg. 75 for calculation procedures), were compared to the proportions reported by Millon (1987) for the normative, test construction sample. The statistical analysis followed the standard procedures for the Chi-Square Goodness of Fit Test.

Secondly, in an analysis of two homogeneous sub-samples, the results of the scales were compared to determine: (a) if the Alcohol Dependence scale was more effective than the Drug Dependence scale in identifying individuals with a diagnosis of Alcohol Abuse and/or Dependence, and (b) if the Drug Dependence scale was more effective than the Alcohol Dependence scale in identifying individuals with a diagnosis of Drug Abuse and/or Dependence. The first sub-sample was comprised of those participants who only had SCID generated diagnoses of Alcohol Dependence and/or Abuse. Using all the BR scores of the individuals in this sub-sample, the Wilcoxon Matched-Pairs Signed-Ranks Test (Wilcoxon test) was used to determine if the Alcohol Dependence scale provided a more accurate assessment of Alcohol Dependence and Abuse than the Drug Dependence scale. The second sub-sample consisted of those participants who only had SCID generated diagnoses of Drug Dependence and/or Abuse. A similar procedure using the Wilcoxon test was followed to determine if the Drug Dependence scale provided a more accurate assessment of Drug Dependence and Abuse than the Alcohol Dependence scale.

Data Collection

The collection of data for the present study was part of an ongoing research project conducted by faculty and students from Texas Tech University at a private substance abuse treatment center in Lubbock, Texas. As part of this larger project,
individuals who entered treatment were requested to participate in the ongoing research. This involved the completion of a set of eight assessment instruments within the first ten days of treatment. Additionally, a structured interview, the SCID, was administered upon admission. The study which is the focus of this paper utilizes archival data which was collected between October, 1991 and March, 1992. Data collection followed the established protocol for the ongoing research project. Participants' anonymity as well as confidentiality were assured through the assignment of numerical codes. These codes were also used to determine if any patients had been readmitted during the data collection phase. Data from all those who were readmitted was removed for the analysis.

The MCMI-II was administered using accepted group test taking procedures. As designed by Millon (1983, 1987), answer sheets were sent to National Computer Systems for scoring and a printed Profile Report. The SCID was used to generate all possible DSM-III-R, Axis I diagnoses, while a supplement, the Structured Clinical Interview for the DSM-III-R Personality Disorders, ([SCID II], Spitzer et al., 1990), was used to establish DSM-III-R, Axis II diagnoses. The SCID was conducted by a doctoral student in a clinical psychology program as part of her role as graduate assistant to the director of the research project. The research assistant was given instruction in psychopathology and assessment, particularly as these relate to the diagnostic process and the criteria of the DSM-III-R. Additionally, she received instruction in the theoretical base and administration of the SCID (Spitzer et al., 1990). The research protocol included a provision for a monthly inter-rater reliability check which occurred on two occasions during the collection of this data. Procedurally, this involved having a second individual with training in psychopathology, assessment, and administration of the SCID, join the research assistant as she administered the SCID to a participant. The second individual would
complete the SCID at the same time as the research assistant. Formal kappa coefficients were not calculated, because on each occasion the two individuals arrived at identical Alcohol and Drug Dependence and/or Abuse diagnoses.

Selection and Description of Sample

The participants in the present study were patients admitted to a substance abuse treatment facility in Lubbock, Texas. At the time of admission participants were provided with information regarding treatment as well as the research project. Involvement in the research project was voluntary and patients were assured that lack of involvement would not negatively impact their treatment. Participants signed an informed consent form signifying their understanding of the terms of treatment as well as the voluntary nature of their involvement in the research project. A majority of the participants for the present sample were self selected and were seeking treatment voluntarily. Some participants were remanded to treatment by the judicial system. The sample consists of 73 subjects of whom 64 were Caucasian, 6 were Hispanic, and 2 were African American. One person declined to identify an ethnic background. The sample includes 24 females and 49 males whose ages ranged from 18 to 70.

Instrumentation

The instrumentation for the present study consists of a structured clinical interview, the SCID, and a self report personality inventory, the MCMI-II. Both of the instruments are grounded in atheoretical descriptive criteria, and each includes several unique features. Descriptive characteristics as well as the reported reliability and validity data of each of these instruments will be described in the following sections.
The SCID

The Structured Clinical Interview for the DSM-III-R is an instrument that is designed to obtain the data necessary to make a diagnosis according to the DSM-III-R Axis I and Axis II criteria (Skre et al., 1991; Spitzer et al., 1990). Spitzer and colleagues (1990) recommend the use of this instrument in research, "to characterize a study population in terms of current and past psychiatric diagnoses" (p. 624). This suggestion is congruent with the literature on structured interviews (Bryer et al, 1990; Endicott & Spitzer, 1978; Helzer et al., 1978; Robins et al., 1981; Robins, Helzer, Ratcliff, & Seyfried, 1982; Williams et al., 1992; Wing et al., 1967) and endorsed by Piersma (1993) when he states, "There is no doubt that the preferred methodology for future studies is to evaluate the MCMI against diagnostic standards which are based on structured interviews" (p. 14).

The distinguishing feature of the SCID lies in its intended purpose. Spitzer and his colleagues (1990) developed the SCID to replicate, "the differential diagnostic process of an experienced clinician" (p. 1). It is designed to be conducted by a trained mental health clinician. The SCID is divided into modules which match the diagnostic categories of the DSM-III-R. The interview begins with an overview section which allows the clinician to gather enough information to formulate initial diagnostic impressions. The remaining modules consist of questions which are grouped by diagnosis and which incorporate the DSM-III-R criteria. The interviewer rates each symptom according to a four-point scale, per instructions in the Instruction Manual for the SCID (Spitzer et al., 1990), as follows:

\[
\begin{align*}
? & = \text{Inadequate information to rule in or out a diagnosis of the disorder.} \\
1 & = \text{Absent: there is adequate information to judge that the criteria have not been met.}
\end{align*}
\]
2 = Sub threshold: the full criteria have not quite been met, but clinically the disorder seems likely.
3 = Threshold: the full criteria have been met. (p. 5-6)

When a required criterion for a diagnosis is not met the interviewer is instructed to skip the remaining questions ("skip outs") and to proceed to the next diagnostic module. As Spitzer and colleagues (1990) have stated, "Since the DSM-III-R diagnostic criteria are embedded in the SCID and are assessed as the interview progresses, the interviewer is, in effect, continually testing diagnostic hypotheses" (p. 2). To enhance this ongoing assessment the interviewer is encouraged to generate and ask additional probing and clarifying questions. In fact, if the patient denies the presence of a symptom which appears to be present, the interviewer is expected to challenge the denial. Thus, if a patient appears depressed but denies depressed mood, a comment from the interviewer regarding the patient's appearance would be indicated. According to Spitzer, Williams, Gibbon, and First (1992), "The rationale for allowing for clinical judgment in modifying and supplementing the SCID interview questions is that the validity of the interviewer's rating will thereby be increased, although perhaps at the expense of some degree of inter-rater reliability" (p. 625).

As indicated by the literature review, there are a number of researchers who affirm the use of the SCID as the means to establish psychiatric diagnoses as a standard for comparison (Copolov et al., 1986; Bryant, Rounsaville, & Babor, 1991; Kosten et al., 1991; Maier et al., 1988; Riskind et al., 1987; Skre et al., 1991; Spitzer et al., 1990; Williams et al., 1992). Additionally, both the SCID and the MCMI-II are intended to correlate with the DSM-III-R which enhances the attractiveness of the SCID for the present study (Millon, 1983, 1985, 1987; Millon & Klerman, 1986; Spitzer et al., 1990; Spitzer et al., 1992). The fact that the development of the instruments is based on common constructs and clinical criteria, may minimize spurious findings.
However, the flexibility which is a part of the design of the SCID does introduce the possibility of error (Spitzer, et al., 1992). Many structured interviews do not allow for probing questions and require the interviewer to ask each of the questions in order without omitting any. In contrast, the SCID allows for "skip outs" and follow up probes. The interview is therefore partly directed by the type of follow up questions the interviewer asks, whereas in a totally structured interview the written standardized questions direct the process. "Skip outs" introduce the variable of judgment, and with it the potential for error as the clinician may decide incorrectly that a module, containing questions related to a specific area of pathology, need not be examined (Spitzer et al., 1990; Spitzer et al., 1992; Williams et al., 1992). Different styles of interviewing as well as varying skill levels may also effect the reliability of the diagnosis. As mentioned before, a clinician who is more aggressive in probing areas of apparent denial may get confirmation of a criteria needed to make a diagnosis (Spitzer et al., 1992).

Both of these potential sources of error are related to the variable of the interviewer. According to Williams and colleagues (1992), "Maximizing reliability with the SCID clearly requires extensive training in the intent of the of the various diagnostic criteria and insistence that interviewers elicit descriptions of behavior to justify each criterion coded as 'present'" (p. 636). It is therefore essential that the interviewer receives training in the administration of the SCID, and that this training include the directive to ask follow up questions to minimize the effect of information variance. Additionally, researchers are encouraged to conduct studies to determine the reliability of interviewers for a particular clinical population (Williams et al., 1992).
The MCMI-II

The Millon Clinical Multiaxial Inventory II is a self report inventory which is clinical in nature and intended solely for use with a psychiatric population (Millon, 1987). The normative sample for the inventory was drawn from a variety of inpatient and outpatient settings. The purpose for the construction of the test was to provide the practitioner with a practical assessment tool which would help with the process of differential diagnosis. In order to enhance the utility of the instrument for clinical practice, every attempt was made to keep the instrument short while maintaining enough items to be useful in the assessment of a wide range of behaviors. Thus the inventory consists of 175 true-false items and can generally be completed in 20 to 30 minutes (Millon, 1987).

The MCMI-II consists of 22 clinical scales. The scales were constructed to identify acute clinical disorders and more enduring personality characteristics much like Axis I and Axis II which were introduced in the DSM-III. With the addition of severity as a distinguishing factor, the scales can be divided into four groups namely: Clinical Syndrome, Severe Syndrome, Clinical Personality Pattern, and Severe Personality Pathology (Millon, 1987).

Clinical Symptom Scales

These scales are designed to measure symptoms which manifest themselves when an individual is moderately stressed. A brief description (Choca et al., 1992; Millon, 1987) of some of the traits measured by each scale follows:

Scale A-Anxiety: Elevation on this scale is indicative of apprehension, phobic reactions, tension, a specific set of stress related physical complaints, and hypersensitivity to one's physical and social environment.
Scale H-Somatoform: Elevation on this scale reflects a preoccupation with one's health and associated physical complaints which are intended to gain attention.

Scale N- Bipolar: Manic: Elevation on this scale is indicative of hyperactivity, impulsivity, ungrounded elation and self aggrandizement, pressured speech, and irritability.

Scale D-Dysthymia: Elevation on this scale is associated with long term (several years) feelings of discouragement, apathy and guilt. Suicidal ideation may be present as a result of a pessimistic view of the future.

Scale B-Alcohol Dependence: Elevations on this scale indicate a likely history of alcoholism that has negatively effected relationships and employment. This patient tends to be impulsive and have low self esteem.

Scale T-Drug Dependence: Elevations on this scale suggest that the patient has a history of drug use. The individual is likely to be exploitive of others, impulsive and resentful of authority.

Severe Syndrome Scales

These three scales identify individuals who are markedly dysfunctional. A brief description (Choca et al., 1992; Millon, 1987) of some of the traits measured by each scale follows:

Scale SS-Psychotic Thinking: Elevations on this scale suggest that the patient may have confused thought processes and experience delusions and hallucinations. Inappropriate affect, suspicion, isolating and experiences of mental imbalance also characterize this individual.

Scale CC-Psychotic Depression: Elevations on this scale are indicative of a severe affective disorder which prevents the individual from functioning. Suicidal
ideation, hopelessness, agitation, psychomotor retardation and other somatic manifestations are likely to be present.

Scale PP-Psychotic Delusional: Elevations on this scale are associated with the presence of confused, irrational, and delusional thought processes. This patient is likely to be paranoid or grandiose and may be hostile and periodically belligerent.

Clinical Personality Pattern Scales

There are 10 Clinical Personality Pattern Scales which measure the enduring patterns of functioning associated with a personality style. As Choca, and colleagues (1992) state, "a personality style is the set of life-long assumptions that the person holds about the self and the world, together with the typical ways of thinking and feeling, and the behavioral patterns associated with those assumptions" (p. 5). The following are brief descriptions (Choca et al., 1992; Millon, 1987) of the significance of each of these scales when elevated:

Scale 1-Schizoid: Reflective of the passive-detached orientation, these patients are likely to lack desire and an ability to experience depth of affect. They express minimal needs and remain distant and aloof.

Scale 2-Avoidant: Reflective of the active-detached orientation, these patients mistrust others, have low self esteem, and find social situations anxiety producing. Though desirous of relationships they deal with the associated fear by withdrawal and isolating themselves.

Scale 3-Dependent: Reflective of the passive-dependent orientation, these patients lack initiative and self confidence. They lean on others for affection and support and willingly submit to others authority in order to gain affection.

Scale 4-Histrionic: Reflective of the active-dependent orientation, these patients eagerly seek attention, are often impulsive, and are easily bored. Though
outwardly self confident in social situations, this appearance of self assurance masks a need for repeated signs of approval and a fear of autonomy.

Scale 5-Narcissistic: Reflective of the passive-independent orientation, these patients tend to present themselves as arrogant or superior and dislike being controlled. In relationships they are likely to lack empathy and be defensive as well as exploitive.

Scale 6A-Antisocial: Reflective of the active-independent orientation, these patients are often resentful of authority, impulsive, and insensitive. Using people for their own advantage and an inability to cathect with others on an affective level are hallmarks of these individuals.

Scale 6B-Aggressive/Sadistic: Reflective of the active-discordant orientation, these patients take pleasure in the humiliation and pain of others. In relationships they are frequently dominant, hostile, and competitive and may impulsively strike out.

Scale 7-Compulsive: Reflective of the passive-ambivalent orientation, these patients are generally controlled, perfectionistic and compliant. Rigidity, social caution and emotional distance, serve as covers for intense oppositional and angry feelings.

Scale 8A-Passive-Aggressive: Reflective of the active ambivalent orientation, these patients vent their feelings in a covert manner. They tend to vacillate between over compliance and obstinacy as well as explosiveness and feelings of guilt and shame.

Scale 8B-Self-Defeating: Reflective of the passive-discordant orientation, these patients are often obsequious and focus on their worst traits. Interpersonally they are self sacrificing and may seek out situations in which they will be used or rejected.
Severe Personality Pathology Scales

These scales are clearly associated with pathological ways of functioning due to severe defects in personality integration and development. The following are brief descriptions (Choca et al., 1992; Millon, 1987) of the significance of each of these scales when elevated:

Scale S-Schizotypal: These patients are fearful of human contact and form very few real relationships. They may be cognitively confused, ruminative, and others often perceive them as peculiar and strange.

Scale C-Borderline: These patients have developmental deficits which effect their ability to form stable relationships. They experience all affect intensely, may be self-abusive, experience periodic psychotic episodes, and are emotionally labile.

Scale P-Paranoid: These patients are extremely mistrustful and hypersensitive to criticism. They may be irritable, abrasive, judgmental, and fiercely independent.

Each of the clinical scales corresponds with a DSM-III-R diagnostic label. Though this scheme is helpful in the organization of data, these scales are not intended to reflect discrete categories. The reader is reminded that Millon conceptualizes clinical categories as polythetic in nature and clinical syndromes as "disruptions in a patient's basic personality pattern" (Millon 1987, p. 18). The interactions of disorders and syndromes are reflected in the shared items of the scales.

Interpreting the scales of the MCMI-II involves the BR score and established cut off points. Based on the prevalence of a particular characteristic in the normative population, as opposed to the normative sample, Millon identified four points which "would place an individual in the same relative position in the standardizing population as one moves from one scale to the next" (Choca et al., 1992). The median BR score for the non psychiatric population was 35, and the median BR score for psychiatric
patients was 60. A BR score of 75 identified the definite presence of the characteristic being measured, and 85 was the point where the characteristic was the most prominent feature for the individual. These cut off points were established on the basis of prevalence rates established by Millon utilizing a nation wide survey of a wide group of populations. However, if the local prevalence rates are different from those calculated by Millon, these cut off points may result in an inaccurate diagnosis (Choca et al., 1992; Millon, 1987).

Modifier Scales

In addition to the 22 clinical scales, Millon (1987) introduced three scales which he labeled Modifier Indices. These are scales which identify response sets that may negatively effect the accuracy of the results of the test. The Disclosure Level Index (Scale X) was introduced to identify those patients who were either highly defensive, or who were unusually self revealing in regards to personally sensitive material. The Desirability Scale (Scale Y) measures tendencies to put oneself in a positive light, while the Debasement Scale (Scale Z) identifies those individuals who attempt to look bad on the inventory. These three new scales were incorporated into the test to complement the Validity Index which was maintained from the MCMI-I. This last scale consists of four absurd items which are intended to identify those individuals who cannot read or understand items appropriately, or who may be answering in a random (true only) manner.

Interpretation of the Validity Index has been problematic. This is the first of the modifier scales to be addressed in interpretation of the protocol, and, if 2 or more items are endorsed, the test is considered invalid (Millon, 1987). However, if only 1 item is endorsed Millon (1987) accepts the profile and suggests that it is interpretable. Choca and colleagues (1992) question this practice since a person has to endorse an
item which is blatantly absurd. However, Millon (1987) found that many subjects in the normative sample endorsed one of these items and therefore a score of 1 was statistically within an acceptable range. For purposes of the present study, Millon's (1987) standard will be followed and scores of participants who have endorsed one item, or no items on the Validity Index will be included in the analysis of the data.

The following are some of the important psychometric characteristics of the MCMI-II as reported by Millon (1987). Standardization was done with a very large sample of patients which varied widely across the demographic categories of age, gender, race, religion and marital status. The identified patients were assessed in the following settings: outpatient (81.9%), inpatient (9.9%), correctional (2.4%), college center (2.1%), and other (3.8%) (Choca et al., 1992; Millon, 1987). Three external validation studies conducted as part of the test construction process indicate that the sensitivity of the scales ranged from 50% to 79% and the specificity ranged from 93% to 98%. Reliability, reported in terms of stability coefficients, ranged from .80 to .89 for the clinical personality pattern scales, from .79 to .89 for the severe personality pathology scales and from .79 to .91 for all of the clinical syndrome scales (Choca et al., 1992; Millon, 1987).

Millon (1987) provides detailed information about the Alcohol Dependence scale in terms of composition, correlation, reliability, and validity. The scale consists of 46 items all but two of which are significant when answered in the affirmative (true). To increase the utility of the scale, many subtle and indirect items were included. The scale is composed of 6 prototypic items (see Table 3), and 40 descriptive items. For the purposes of weighting and scoring, these descriptive criteria are divided into 9 secondary and 31 tertiary items. When endorsed, the prototypic items receive 3 points, except for the two items which are significant when
answered in the negative (false) direction which receive 2 points. The secondary and tertiary items receive 2 and 1 points respectively. The 6 prototypical items for the

Table 3
Prototypic Items of the Alcohol and Drug Dependence Scales of the MCMI-II

<table>
<thead>
<tr>
<th>Prototypic Item and Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alcohol Dependence Scale</strong></td>
</tr>
<tr>
<td>17. I have a drinking problem that I’ve tried unsuccessfully to end.</td>
</tr>
<tr>
<td>52. Drinking alcohol has never caused me any real problems in my work.</td>
</tr>
<tr>
<td>87. I have an alcohol problem that has made difficulties for me and my family.</td>
</tr>
<tr>
<td>119. I have a great deal of trouble trying to control an impulse to drink to excess.</td>
</tr>
<tr>
<td>122. I have succeeded over the years in keeping my drinking of alcohol to a minimum</td>
</tr>
<tr>
<td>157. On occasion I have had as many as ten or more drinks without becoming drunk.</td>
</tr>
<tr>
<td><strong>Drug Dependence Scale</strong></td>
</tr>
<tr>
<td>35. My drug habits have often gotten me in a good deal of trouble in the past.</td>
</tr>
<tr>
<td>70. Taking so-called illegal drugs may be unwise, but in the past I found I needed them.</td>
</tr>
<tr>
<td>105. My habit of abusing drugs has caused me to miss work in the past.</td>
</tr>
<tr>
<td>140. My use of so-called illegal drugs has led to family arguments.</td>
</tr>
<tr>
<td>144. Frankly, I lie quite often to get out of trouble.</td>
</tr>
<tr>
<td>175. I have had difficulties in the past stopping myself from over-using drugs or alcohol.</td>
</tr>
</tbody>
</table>

Millon, (1987)
Alcohol Dependence scale of the MCMI-II are the same ones utilized for the construction of the Alcohol Dependence scale of the MCMI-I. As with all other scales, the selection of each primary item was based on its, "high point biserial correlation with its substantive scale" (Millon, 1987, p. 128). Therefore, it is expected that the scales will be internally consistent. Millon (1987) reports that, using the Kuder-Richardson Formula 20 (KR), the KR coefficient for the Alcohol Dependence scale is .84. This internal consistency is attained inspite of the fact that the items of the Alcohol Dependence scale overlap 39% with the Drug Dependence scale. Millon (1987) further reports that the scales have a level of intercorrelation of .76. Moreover, when intercorrelations were calculated on the basis of those items which are strictly theory grounded (prototypical items), Millon (1987) reports that there were, "close parallels among scale-to-scale coefficients that conform well to the basic theoretical model. Hence, the removal of overlapping items did not alter fundamental scale correlations, although the magnitudes have been reduced" (p. 132).

In order to establish the reliability of each of the scales, Millon (1987) conducted studies with a non-clinical, a psychiatric outpatient, and a psychiatric inpatient population. A test-retest method was used to establish scale stability over time. In each group, participants were tested and re-tested at various intervals. For the outpatient and inpatient samples, the clinical syndrome scales would be expected to change as the result of treatment or other intervening variables (Millon, 1987; Piersma, 1986). It is therefore most reasonable to compare scores when participants are not in a state of initial crisis and have engaged in treatment for some time and before the process of termination begins. Using these criteria to establish test-retest intervals, Millon (1987) reports reliability coefficients of .88 for the non-clinical sample, .83 for the psychiatric outpatient sample, and .66 for the psychiatric inpatient sample.
The validity of the Alcohol Dependence scale was addressed using two methods. First, the sensitivity and specificity of the scale were calculated as measures of accuracy of the scale when compared to a diagnostic criteria (Millon, 1987). These proportions are reported as .87 and .98 respectively. Secondly, the median BR scores of each of the 22 scales of the MCMI-II for every diagnostic category were established using two samples (Millon, 1987). For those individuals who received a diagnosis of Alcohol Dependence and/or Alcohol Abuse using DSM-III-R criteria, the median BR score on the Alcohol Dependence scale for the first sample was reported as 86 and for the second sample as 85. The median BR score on the Drug Dependence scale for these same individuals was reported as 78 and 73.

Millon (1987) also provides detailed information about the Drug Dependence scale in terms of composition, correlation, reliability, and validity. The scale consists of 58 items all but one of which are significant when answered in the affirmative (true). To increase the utility of the scale, many subtle and indirect items were included. The scale is composed of 6 prototypic items (see Table 3), and 52 descriptive items which, for the purposes of weighting and scoring, are divided into 20 secondary and 32 tertiary items. When endorsed, the prototypic items receive 3 points, and the secondary and tertiary items 2 and 1 point respectively. The 6 prototypical items for the Drug Dependence scale of the MCMI-II were the same ones utilized in the construction of the Drug Dependence scale for the MCMI-I. The KR coefficient, as a measure of internal scale consistency, is reported as .87 even though there is significant scale overlapping with the Alcohol Dependence scale, correlation .76 (Millon, 1987).

The rationale for establishing scale stability figures described earlier was used to establish reliability figures for the Drug Dependence scale. It is most reasonable to compare scores when participants are not in a state of initial crisis and have engaged in
treatment for some time, but before the process of termination from treatment begins. Using these criteria to establish test-retest intervals, Millon (1987) reports reliability coefficients of .85 for the non-clinical sample, .80 for the psychiatric outpatient sample, and .71 for the psychiatric inpatient sample.

The validity of the Drug Dependence scale was addressed using two methods. First, the sensitivity and specificity of the scales were calculated as a measure of accuracy of the scale when compared to a diagnostic criteria (Millon, 1987). These proportions are reported as .72 and .99 respectively. Secondly, the median BR scores of each of the 22 scales of the MCMI-II for every diagnostic category were established using two samples (Millon, 1987). For those individuals who received a diagnosis of Drug Dependence and/or Drug Abuse using DSM-III-R criteria, the median BR score on the Drug Dependence scale for the first sample was reported as 77 and for the second sample as 78. The median BR score on the Alcohol Dependence scale for these same individuals was reported as 72 and 71.

Limitations of the Present Study

As in all scientific investigations, the present study is not without confounding variables that influence the generalizability of the findings. First, the reliability of the diagnosis established with a structured interview is dependent on the reliability of the interviewer. Confidence in the reliability of the interviewer is established through periodic checks of inter-rater reliability, a procedure which was part of the design of the data collection protocol in the present study. However, these checks were not completed on a monthly basis as had been intended. Only two of the five scheduled checks actually occurred. Based on face validity, it appears that there was diagnostic agreement between the two raters in the assessment of substance abuse, however, a \textit{kappa} correlation statistic was not calculated to establish inter-rater reliability based on
all the modules. Secondly, the diagnosis established with a structured interview has been shown to be most accurate when all corroborating data has been considered. However, for the present study the diagnosis is based strictly on the data gathered during the completion of the structured interview. Thirdly, it should be noted that this is a preferential sample due to the fact that this is a privately, as opposed to a publicly, supported treatment center. The availability of financial resources to pay for treatment is therefore likely to distinguish the present sample on the demographic factors of social and economic status as well as race. Fourth, both the MCMI-II and the SCID utilize self-report as the method of data gathering. Self-reports may be negatively affected by factors that create limitations in the patient's ability to be a reliable informant (Millon, 1987; Wetzler, 1990). These factors may include impaired reality testing, agitation, disorganization, language barriers, and reading level. Additionally, the context in which the data is being collected may effect the patient's motivation and result in a distorted response set. In describing these limitations of self-report measures Millon (1987) states that there are,

inherent psychometric limits ... [due to] ... the tendency of similar patients to interpret questions differently, the effect of current affective states on trait measures, the effort of patients to effect certain false appearances and impressions, ... [which] ... all narrow the upper boundaries of this method's potential accuracy. (p. 10)

Lastly, it should be noted that the normative sample which Millon (1983, 1987) used to establish sensitivity, specificity (the comparative proportions on which the statistical hypothesis of the present study are based), and scale intercorrelation statistics, consisted of general psychiatric patients. However, the sample for the present study is drawn from a population of known substance abusers.
Statistical Hypotheses

As evidenced by an analysis of the literature review, the sensitivity and specificity for the Alcohol and Drug Dependence scales for the present sample are expected to be less than the proportions reported by Millon (1987) for these same scales (Bryer et al., 1990; Calsyn et al., 1990, 1991; Marsh et al., 1988). These proportions were calculated for the participants in the present study and were compared to the proportions reported by Millon (1987) in the test manual for the MCMI-II. It is these comparisons that are the focus of the first four hypotheses.

To determine if a significant difference exists between the sensitivity calculated for the Alcohol Dependence scale of the MCMI-II in the present sample, and the sensitivity reported by Millon (1987) for the same scale, the following null hypothesis was tested:

1. No significant difference exists between the sensitivity calculated for the Alcohol Dependence scale for the present sample and the expected sensitivity (.87) reported by Millon for the same scale.

To determine if a significant difference exists between the specificity calculated for the Alcohol Dependence scale of the MCMI-II in the present sample, and the specificity reported by Millon (1987) for the same scale, the following null hypothesis was tested:

2. No significant difference exists between the specificity calculated for the Alcohol Dependence scale for the present sample and the expected specificity (.99) reported by Millon for the same scale.

To determine if a significant difference exists between the sensitivity calculated for the Drug Dependence scale of the MCMI-II in the present sample, and the
sensitivity reported by Millon (1987) for the same scale, the following null hypothesis was tested:

3. No significant difference exists between the sensitivity calculated for the Drug Dependence scale for the present sample and the expected sensitivity (.72) reported by Millon for the same scale.

To determine if a significant difference exists between the specificity calculated for the Drug Dependence scale of the MCMI-II in the present sample, and the specificity reported by Millon (1987) for the same scale, the following null hypothesis was tested:

4. No significant difference exists between the specificity calculated for the Drug Dependence scale for the present sample and the expected specificity (.98) reported by Millon for the same scale.

The overlap between the Alcohol and Drug Dependence scales has led to continued questions about the independence of these two scales. Further review of the literature pointed to the possibility that the Alcohol Dependence scale may be the most adequate measure for all forms of substance abuse (Jaffe & Archer, 1987).

To determine if the BR score for the Alcohol Dependence scale was significantly greater than the BR score for the Drug Dependence scale when Alcohol Dependence and/or Alcohol Abuse were the only substance abuse diagnoses, the following null hypothesis was tested:

5. No significant difference exists between the BR scores of the Alcohol Dependence scale and the BR scores for the Drug Dependence scale when Alcohol Dependence and/or Alcohol Abuse are the only substance abuse diagnoses.

To determine if the BR scores of the Drug Dependence scale were significantly greater than the BR score for the Alcohol Dependence scale when Drug Dependence
and/or Drug Abuse were the only substance abuse diagnoses, the following null hypothesis was tested:

6. No significant difference exists between the BR scores of the Alcohol Dependence scale and the BR scores of the Drug Dependence scale when Drug Dependence and/or Drug Abuse are the only substance abuse diagnoses.

Analysis of Data

Data analysis involved two sets of statistical analyses. The first set involved the calculation and comparison of the true positive and true negative proportions of the Alcohol and Drug Dependence scales in this sample with the figures reported by Millon (1987). The second set involved the comparison of BR scores of two homogenous sub-samples.

Comparing Differences in Sensitivity and Specificity

Testing hypotheses 1, 2, 3, and 4, first requires the determination of the sensitivity and specificity of the MCMI-II Alcohol and Drug Dependence scales for the present sample. Sensitivity is defined as the probability that the test is positive given the disorder is present. This proportion can be calculated utilizing two cells of a four cell contingency table which accounts for positive and negative test results as well as presence and absence of the disorder (see Figure 1). Mathematically this involves calculating a proportion of those participants for whom the disorder is present, but for whom the test results were either positive or negative. Using the symbols of Figure 1, the formula to calculate sensitivity reads \( \frac{a}{a + c} \). For purposes of the present study, \( a \) is the number of participants who fit the following two criteria: (1) had a BR score > 74 on the MCMI-II scale (test result positive) for which sensitivity was being calculated, and (2) who received a SCID generated diagnosis (disorder present) for the
disorder under consideration. On the other hand, \( c \) is the number of participants who meet the following two criteria: (1) had a BR score \( \leq 74 \) on the MCMI-II scale (test result negative) for which sensitivity was being calculated, and (2) who received a SCID generated diagnosis (disorder present) for the disorder under consideration. For example, to calculate the sensitivity for the Alcohol Dependence scale for the present sample, \( a \) represents those participants who had a BR \( \geq 75 \) on the Alcohol Dependence scale of the MCMI-II and who received a SCID generated diagnosis of either Alcohol Dependence or Alcohol Abuse. Similarly, \( e \) represents those participants who had a BR \( < 74 \) on the Alcohol Dependence scale of the MCMI-II but who received a SCID generated diagnosis of either Alcohol Dependence or Alcohol Abuse.

<table>
<thead>
<tr>
<th></th>
<th>Disorder Present SCID</th>
<th>Disorder Absent SCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive MCMI-II</td>
<td>Cell - a, True case</td>
<td>Cell - b, False positive</td>
</tr>
<tr>
<td>Negative MCMI-II</td>
<td>Cell - c, True case</td>
<td>Cell - d, False negative</td>
</tr>
</tbody>
</table>

Figure 1. Four Cell Contingency Table.

Specificity is defined as the probability that the test is negative given the disorder is absent. Like sensitivity, this proportion can be calculated utilizing two cells of a four cell contingency table which accounts for positive and negative test results as well as presence and absence of the disorder (see Figure 1). Mathematically this involves calculating a proportion of those participants for whom the disorder is
absent, but for whom the test results were either positive or negative. Using the symbols of Figure 1, the formula to calculate sensitivity reads \( d/(b + d) \). For purposes of the present study, \( d \) is the number of participants who fit the following two criteria: (1) had a BR score \( \leq 74 \) on the MCMI-II scale (test result negative) for which specificity was being calculated, but (2) who did not receive a SCID generated diagnosis (disorder absent) for the disorder under consideration. On the other hand, \( h \) is the number of participants who meet the following two criteria: (1) had a BR score \( > 74 \) on the MCMI-II scale (test result positive) for which specificity was being calculated, but (2) who did not receive a SCID generated diagnosis (disorder absent) for the disorder under consideration. For example, to calculate the specificity for the Alcohol Dependence scale for the present sample, \( d \) represents those participants who had a BR \( < 74 \) on the Alcohol Dependence scale of the MCMI-II and who did not receive a SCID generated diagnosis of either Alcohol Dependence or Alcohol Abuse. Similarly, \( h \) represents those participants who had a BR \( \geq 75 \) on the Alcohol Dependence scale of the MCMI-II but who did not receive a SCID generated diagnosis of either Alcohol Dependence or Alcohol Abuse. Having established the sensitivity and specificity of the MCMI-II Alcohol and Drug Dependence scales for the present sample, these proportions were compared to the sensitivity and specificity for the same two scales as reported by Millon (1987). The Chi Square Goodness-of-Fit-Test, "can be used to determine whether the observed proportions differ significantly from a priori or theoretically expected proportions" (Hopkins & Glass, 1978, p. 309).

While the Chi Square statistic (\( \chi^2 \)) can accommodate more than two categories simultaneously, when there are only two categories the Chi Square Goodness-of-Fit-Test is equivalent to the \( z \)-test (Hopkins & Glass, 1978). Thus, to establish if there is a significant difference between the sensitivity and specificity of the Alcohol and Drug
Dependence scales for the present sample and the a priori proportions established by Millon, the z test, was utilized (see Figure 2). The level of significance was $p < .05$.

$$z = \frac{p - \pi}{\sigma_p} \quad \text{where} \quad \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- $p$ - observed proportion
- $\pi$ - theoretically established/expected proportion
- $\sigma_p$ - standard error of a proportion

Figure 2. The z-test.

Comparison of BR Scores of the Alcohol and Drug Dependence Scales

In order to test hypotheses 5 and 6, a decision was made to use nonparametric rather than parametric procedures. Parametric statistical techniques make assumptions about the population from which the numerical scores are drawn (Siegel, 1956). The first of these assumptions is that the population from which the scores are drawn is normally distributed. This is not a valid assumption for the population (sample of 73 participants) from which the (sub-) samples for this analysis were drawn. Plotting a frequency distribution for the scores within this population indicated that the distribution was negatively skewed with a majority of the scores falling above the median score (BR 60) reported by Millon for the normative sample of psychiatric patients. In comparison to parametric tests, nonparametric procedures make less stringent assumptions about the nature of the population. As Siegel (1956) states, "These newer 'distribution-free' techniques result in conclusions which require fewer qualifications" (p. 3). Thus, using nonparametric statistics allows for inferences to be made about differences among groups regardless of the shape of the distribution of population scores.
An additional consideration relates to the type of scales of measurement which fit the variables. Siegel (1956) makes the case that the mathematical manipulations which are allowable for a given set of scores is dependent on the level of measurement utilized. This author cogently states:

For example, if a researcher collects data made up of numerical scores and then manipulates these scores by, say, adding and dividing (which are necessary operations in finding means and standard deviations), he is assuming that the structure of his measurement is isomorphic to that numerical structure known as arithmetic. That is, he is assuming that he has attained a high level of measurement. (Siegel, 1956, p. 22)

Within the social sciences it is often difficult to determine if the high level of measurement, variables which can be categorized as interval or ratio data, has been reached (Pfeiffer & Olson, 1981). When faced with this decision many researchers choose to treat the data as interval rather than ordinal measurement, due to statistical advantages which this provides. However, Siegel (1956) argues that when parametric statistics are applied to data which is ordinal in nature, any decisions about hypotheses are questionable. This conclusion is consistent with Millon's (1987) observation that the application of statistical analyses which require the use of interval level to BR scores may be inappropriate. Thus, due to the fact that the assumption of normal distribution of the population had not been met, and that the data for the present study might be ordinal as opposed to interval data, the more conservative approach to these research questions was to use nonparametric statistics.

The Wilcoxon Matched-Pairs Signed-Ranks Test (Wilcoxon test) is a nonparametric procedure which is applicable in situations where two related samples are examined, but one or more of the assumptions of the t test are violated. The Wilcoxon test may be employed in situations where there are two samples of paired scores if the following conditions hold (Pfeiffer & Olson, 1981; Runyon & Haber, 1968; Siegel, 1956).
1. The difference between the two members of each pair can be calculated and assigned direction (negative or positive) thus establishing which member is the greater.

2. The difference between the two members of each pair can be rank ordered. It has been shown, that when these assumptions are met, the efficiency of the Wilcoxon test compared with the t test is 95.5 per cent (Siegel, 1956).

The assumptions of the Wilcoxon test are met by the two sub-samples in this component of the current study. The first sub-sample (hypothesis 5) was defined as those 30 participants who had SCID generated diagnoses of Alcohol Dependence or Alcohol Abuse only. To test the hypothesis that the Alcohol Dependence scale of the MCMI-II provides a more accurate assessment of one of the alcohol related diagnoses in the present sample than the Drug Dependence scale, the Wilcoxon test was used to compare the difference between the BR scores of the Alcohol and Drug Dependence scales. The second sub-sample (hypothesis 6) was defined as those 11 participants who had SCID generated diagnoses of Drug Dependence and/or Drug Abuse only. To test the hypothesis that the Drug Dependence scale provides a more accurate assessment of one of the drug related diagnoses in this sub-sample than the Alcohol Dependence scale, the Wilcoxon test was used to compare the difference between the BR scores of the Drug and Alcohol Dependence scales. The level of significance was p. < .05 for both comparisons.
CHAPTER IV

RESULTS

This is an external validation study which focuses on the Alcohol and Drug Dependence scales of the MCMI-II. Using the SCID as the standard for comparison, data was generated which allowed for the calculation of the sensitivity and specificity of the Alcohol and Drug Dependence scales of the MCMI-II for the present sample. Using standard procedures for the Chi-Square Goodness-of-Fit-Test, these proportions were compared to the sensitivity and specificity reported by Millon (1987) for the normative, test construction sample. Furthermore, the performance of the Alcohol and Drug Dependence scales was examined under conditions when either Alcohol Dependence and/or Alcohol Abuse, or Drug Dependence and/or Abuse was the only substance abuse diagnosis. Both the hypothesis that the Alcohol Dependence scale provided a more accurate assessment of Alcohol Dependence and Abuse than the Drug Dependence scale, and the hypothesis that the Drug Dependence scale provided a more accurate assessment of Drug Dependence and Abuse than the Alcohol Dependence scale, were tested using the Wilcoxon test. The level of significance was p. < .05 for all of the tests.

Comparing Differences in Sensitivity and Specificity

Hypothesis 1

No significant difference exists between the sensitivity calculated for the Alcohol Dependence scale of the MCMI-II for the present sample and the expected sensitivity (.87) reported by Millon for the same scale.
The sensitivity of the Alcohol Dependence scale for the present sample was calculated utilizing the data from cells a and c of a four cell contingency table (see Figure 3). Participants who had a SCID generated diagnosis of Alcohol Dependence or Abuse and who had obtained a BR score greater than 74 on the Alcohol Dependence scale of the MCMI-II, were assigned to cell a. These participants thus filled the condition of true positives where the disorder was present and the test was positive. Participants who had a SCID generated diagnosis of Alcohol Dependence or Abuse but who had obtained a BR score less than or equal to 74 on the Alcohol Dependence scale of the MCMI-II, were assigned to cell c. These participants thus filled the condition of false negatives where the disorder was present but the test was negative. Using the formula suggested by Gibertini and colleagues (1986) to calculate sensitivity, the total number of participants in cell a were divided by the sum of the number of participants in cell a and c. This resulted in a sensitivity of .65 for the Alcohol Dependence scale of the MCMI-II in the present sample.

To determine if there was a significant difference, level of significance .05, between the sensitivity of the Alcohol Dependence scale for the present sample (.65), and the sensitivity reported by Millon (1987) for the same scale (.87), the standard procedures for the Chi Square Goodness-of-Fit-Test were followed. Because there were only two categories, the $\chi^2$ test was equivalent to the z-test (Hopkins & Glass, 1978). Thus, the z-test was used to test hypothesis 1 of the present study. The z value was calculated according to the formula (see Figure 2) suggested by Hopkins and Glass (1978) and found to be -5.069. At a significance level of .05, the null hypothesis is rejected if the absolute value of z, $|z|$, is greater than or equal to 1.96. Therefore, a statistically significant difference exists between the sensitivity of the Alcohol Dependence scale of the MCMI-II for the present sample and the sensitivity
reported by Millon (1987) for the same scale. Thus, the null hypothesis of no significant difference must be rejected.

<table>
<thead>
<tr>
<th>Positive MCMI-II</th>
<th>Present - SCID</th>
<th>Absent - SCID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cell - a 39</td>
<td>Cell - b 3</td>
</tr>
<tr>
<td>Negative MCMI-II</td>
<td>Cell - c 21</td>
<td>Cell - d 10</td>
</tr>
</tbody>
</table>

Sensitivity = 65 Specificity = .77

Figure 3. Contingency Table: Sensitivity and Specificity of Alcohol Dependence Scale.

**Hypothesis 2**

No significant difference exists between the specificity calculated for the Alcohol Dependence scale of the MCMI-II for the present sample and the expected specificity (.99) reported by Millon for the same scale.

The specificity of the Alcohol Dependence scale for the present sample was calculated utilizing the data from cells b and d of a four cell contingency table (see Figure 3). Participants who did not receive a SCID generated diagnosis of Alcohol Dependence or Abuse but who had obtained a BR score greater than 74 on the Alcohol Dependence scale of the MCMI-II, were assigned to cell b. These participants thus filled the condition of false positives where the disorder was absent but the test was positive. Participants who were not assigned a SCID generated diagnosis of Alcohol Dependence or Abuse and who had obtained a BR score less than or equal to 74 on
the Alcohol Dependence scale of the MCMI-II, were assigned to cell d. These participants thus filled the condition of true negatives where the disorder was absent and the test was negative. Using the formula suggested by Gibertini and colleagues (1986) to calculate specificity, the total number of participants in cell d were divided by the sum of the number of participants in cell d and b. This resulted in a specificity of .77 for the Alcohol Dependence scale of the MCMI-II for the present sample.

To determine if there was a significant difference, level of significance .05, between the specificity of the Alcohol Dependence scale for the present sample (.77), and the specificity reported by Millon (1987) for the same scale (.99), the standard procedures for the Chi Square Goodness-of-Fit-Test were followed. Because there are only two categories, the $\chi^2$ test is equivalent to z-test (Hopkins & Glass, 1978). Thus, the z-test was used to test hypothesis 2 of the present study. The z value was calculated according to the formula (see Figure 2) suggested by Hopkins and Glass (1978) and found to be -7.971. At a significance level of .05, the null hypothesis is rejected if the $|z|$ is greater than or equal to 1.96. Therefore, there is a significant difference between the specificity of the Alcohol Dependence scale of the MCMI-II and the specificity reported by Millon (1987) for the same scale. Thus, the null hypothesis of no significant difference must be rejected.

**Hypothesis 3**

No significant difference exists between the sensitivity calculated for the Drug Dependence scale of the MCMI-II for the present sample and the expected sensitivity (.72) reported by Millon for the same scale.

The sensitivity of the Drug Dependence scale for the present sample was calculated utilizing the data from cells a and c of a four cell contingency table (see Figure 4). Participants who received a SCID generated diagnosis of Drug
Dependence or Abuse and who had obtained a BR score greater than 74 on the Drug Dependence scale of the MCMI-II, were assigned to cell a. These participants thus filled the condition of true positives where the disorder was present and the test was positive. Participants who received a SCID generated diagnosis of Drug Dependence or Abuse but who had obtained a BR score less than or equal to 74 on the Drug Dependence scale of the MCMI-II, were assigned to cell c. These participants thus filled the condition of false negatives where the disorder was present but the test was negative. Using the formula suggested by Gibertini and colleagues (1986) to calculate sensitivity, the total number of participants in cell a were divided by the sum of the number of participants in cell a and c. This resulted in a sensitivity of .55 for the Drug Dependence scale of the MCMI-II for the present sample.

To determine if there was a significant difference, level of significance .05, between the sensitivity of the Drug Dependence scale for the present sample (.55), and the sensitivity reported by Millon (1987) for the same scale (.72), the standard procedures for the Chi Square Goodness-of-Fit-Test were followed. Because there were only two categories, the $\chi^2$ test was equivalent to the z-test (Hopkins & Glass, 1978). Thus, the z-test was used to test hypothesis 3 of the present study. The z value was calculated according to the formula (see Figure 2) suggested by Hopkins and Glass (1978) and was found to be -2.456. At a significance level of .05, the null hypothesis is rejected if the $|z|$ is greater than or equal to 1.96. Therefore, there is a significant difference between the sensitivity of the Drug Dependence scale of the MCMI-II for the present sample and the sensitivity reported by Millon (1987) for the same scale. Based on these results, the null hypothesis of no significant difference must be rejected.
Hypothesis 4

No significant difference exists between the specificity calculated for the Drug Dependence scale of the MCMI-II for the present sample and the expected specificity (.98) reported by Millon for the same scale.

The specificity of the Drug Dependence scale for the present sample was calculated utilizing the data from cells b and d of a four cell contingency table (see Figure 4). Participants who did not receive a SCID generated diagnosis of Drug Dependence or Abuse but who had obtained a BR score greater than 74 on the Drug Dependence scale of the MCMI-II, were assigned to cell b. These participants thus filled the condition of false positives where the disorder was absent but the test was positive. Participants who were not assigned a SCID generated diagnosis of Drug Dependence or Abuse and who had obtained a BR score less than or equal to 74 on the Drug Dependence scale of the MCMI-II, were assigned to cell d. These participants thus filled the condition of true negatives where the disorder was absent and the test was negative. Using the formula suggested by Gibertini and colleagues (1986) to calculate specificity, the total number of participants in cell d were divided...
by the sum of the number of participants in cell d and h. This resulted in a specificity of .84 for the Drug Dependence scale of the MCMI-II for the present sample.

To determine if there was a significant difference, level of significance .05, between the specificity of the Drug Dependence scale for the present sample (.84), and the specificity reported by Millon (1987) for the same scale (.98), the standard procedures for the Chi Square Goodness-of-Fit-Test were followed. Because there were only two categories, the \( \chi^2 \) test was equivalent to \( z \)-test (Hopkins & Glass, 1978). Thus, the \( z \)-test was used to test hypothesis 4 of the present study. The \( z \) value was calculated according to the formula (see Figure 2), suggested by Hopkins and Glass (1978), and found to be -5.557. At a significance level of .05, the null hypothesis is rejected if the \( |z| \) is greater than or equal to 1.96. Therefore, there is a significant difference between the specificity of the Drug Dependence scale of the MCMI-II for the present sample and the specificity reported by Millon (1987) for the same scale. Based on these results, the null hypothesis of no significant difference must be rejected.

Comparison of BR Scores of Alcohol and Drug Dependence Scales

**Hypothesis 5**

No significant difference exists between the BR scores of the Alcohol Dependence scale and the BR scores of the Drug Dependence scale when Alcohol Dependence and/or Alcohol Abuse is the only substance abuse diagnosis. The Wilcoxon test was used to determine if there was a significant difference between the BR scores on the Alcohol and Drug Dependence scales when Alcohol Dependence and/or Abuse was the only SCID generated substance abuse diagnosis (see Table 4). There were 30 individuals who fit these criteria and who subsequently comprised the sub-sample
Table 4

BR Scores of Alcohol Dependence Scale > BR Scores of Drug Dependence Scale
Wilcoxon Matched-Pairs Signed-Ranks Test

| Participant code no. | BR Score Alcohol Dependence Scale (B) | BR Score Drug Dependence Scale (T) | Difference (D) | | | Rank of | Rank with Less Frequent Sign |
|----------------------|----------------------------------------|------------------------------------|----------------|---------|-----------------|-----------------|
| 01                   | 60                                     | 41                                 | 19             | 19      | 20.5            |                 |
| 02                   | 96                                     | 81                                 | 15             | 15      | 15.5            |                 |
| 03                   | 75                                     | 60                                 | 15             | 15      | 15.5            |                 |
| 08                   | 79                                     | 66                                 | 13             | 13      | 11.5            |                 |
| 12                   | 65                                     | 61                                 | 4              | 4       | 2.5             |                 |
| 13                   | 92                                     | 87                                 | 5              | 5       | 5               |                 |
| 14                   | 88                                     | 65                                 | 23             | 23      | 25              |                 |
| 15                   | 67                                     | 71                                 | -4             | 4       | 2.5             | (-) 2.5         |
| 16                   | 65                                     | 60                                 | 5              | 5       | 5               |                 |
| 18                   | 94                                     | 71                                 | 23             | 23      | 25              |                 |
| 19                   | 81                                     | 53                                 | 28             | 28      | 27              |                 |
| 21                   | 79                                     | 60                                 | 19             | 19      | 20.5            |                 |
| 23                   | 81                                     | 64                                 | 17             | 17      | 18              |                 |
| 33                   | 78                                     | 63                                 | 15             | 15      | 15.5            |                 |
| 38                   | 61                                     | 55                                 | 6              | 6       | 7.5             |                 |
| 43                   | 60                                     | 37                                 | 23             | 23      | 25              |                 |
| 45                   | 75                                     | 61                                 | 14             | 14      | 13              |                 |
| 47                   | 81                                     | 61                                 | 20             | 20      | 22              |                 |
| 48                   | 61                                     | 55                                 | 6              | 6       | 7.5             |                 |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 4-Continued

<table>
<thead>
<tr>
<th>Participant code no.</th>
<th>BR Score Alcohol Dependence Scale (B)</th>
<th>BR Score Drug Dependence Scale (T)</th>
<th>Difference (D) BR Scores (B - T)</th>
<th></th>
<th>1D</th>
<th>Rank of</th>
<th>Rank with</th>
<th>Less Frequent Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>81</td>
<td>48</td>
<td>33</td>
<td>33</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>93</td>
<td>96</td>
<td>-3</td>
<td>3</td>
<td>1</td>
<td>(-) 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>83</td>
<td>92</td>
<td>-9</td>
<td>9</td>
<td>9</td>
<td>(-) 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>75</td>
<td>60</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>52</td>
<td>39</td>
<td>13</td>
<td>13</td>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>101</td>
<td>89</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>73</td>
<td>44</td>
<td>29</td>
<td>29</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>51</td>
<td>30</td>
<td>21</td>
<td>21</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>71</td>
<td>56</td>
<td>15</td>
<td>15</td>
<td>15.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>69</td>
<td>21</td>
<td>48</td>
<td>48</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>80</td>
<td>62</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N=30  \[ T = \text{Sum} = 12.5 \]

utilized in this aspect of the current study. The BR scores obtained by the 30 participants on the Drug Dependence scale were subtracted from the BR scores obtained by the same 30 participants on the Alcohol Dependence scale. The obtained differences were ranked without respect to algebraic sign. When there were equal differences, an average of the tied ranks was assigned to each of the equal differences.
After the scores had been ranked, the algebraic sign consistent with the original
difference was assigned to the ranks. The ranks with the less frequent sign were
summed. The absolute value of this sum was 12.5 which is the value of the statistic
$T$, the figure used to make a decision about the null hypothesis. At a .05 level of
significance, the null hypothesis is rejected if the $T$ statistic is less than or equal to 151.
Therefore, there was a significant difference between the BR scores on the Alcohol
and Drug Dependence scales of the MCMI-II for this sub-sample when Alcohol
Dependence and/or Abuse were the only SCID generated substance abuse diagnosis.
Based on these results, the null hypothesis of no significant difference must be
rejected.

Hypothesis 6

No significant difference exists between the BR scores of the Drug
Dependence scale and the BR scores of the Alcohol Dependence scale of the MCMI-II
when Drug Dependence and/or Drug Abuse is the only substance abuse diagnosis.

The Wilcoxon test was used to determine if there was a significant difference
between BR scores on the Drug and Alcohol Dependence scales when Drug
Dependence and/or Abuse was the only SCID generated substance abuse diagnosis
(see Table 5). There were 11 individuals who met these criteria and who
subsequently comprised the sub-sample utilized in this aspect of the present study.
The BR scores obtained by the 11 participants on the Alcohol Dependence scale
were subtracted from the BR scores obtained by the same 11 participants on the
Drug Dependence scale. The obtained differences were ranked without respect to
algebraic sign. When there were equal differences, an average of the tied ranks was
assigned to each of the equal differences. After the scores had been ranked, the
algebraic sign consistent with the original difference was assigned to the ranks. The
Table 5

BR Scores of Drug Dependence Scale > BR Scores of Alcohol Dependence Scale
Wilcoxon Matched-Pairs Signed-Ranks Test

| Participant code no. | BR Score Drug Dependence Scale (T) | BR Score Alcohol Dependence Scale (B) | Difference (D) | | D | Rank of | Rank with Less Frequent Sign |
|----------------------|------------------------------------|---------------------------------------|----------------|---|-----|-----------------------------|
| 06                   | 67                                 | 59                                    | 8              | 8 | 3.5 |                               |
| 22                   | 86                                 | 64                                    | 22             | 22| 9   |                               |
| 27                   | 90                                 | 107                                   | -17            | 17| 7   | (-) 7                        |
| 28                   | 64                                 | 73                                    | -9             | 9 | 5   | (-) 5                        |
| 30                   | 77                                 | 62                                    | 15             | 15| 6   |                               |
| 32                   | 60                                 | 42                                    | 18             | 18| 8   |                               |
| 35                   | 44                                 | 41                                    | 3              | 3 | 2   |                               |
| 36                   | 61                                 | 60                                    | 1              | 1 | 1   |                               |
| 44                   | 57                                 | 65                                    | -8             | 8 | 3.5 | (-) 3.5                      |
| 55                   | 106                                | 56                                    | 50             | 50| 11  |                               |
| 57                   | 82                                 | 57                                    | 25             | 25| 10  |                               |

N = 11

T = Sum 15.5

ranks with the less frequent sign were summed. The absolute value of this sum was
15.5 which is the value of the statistic T, the figure used to make a decision about
the null hypothesis. At a $p < .05$ level of significance, the null hypothesis is rejected
if the T statistic is less than or equal to 10. Therefore, there was no significant
difference between the BR scores on the Drug and Alcohol Dependence scales of the
MCMI-II for this sub-sample when Drug Abuse was the only SCID generated
substance abuse diagnosis. Therefore, the null hypothesis of no significant difference must be retained.

Summary

The statistical hypotheses concerning sensitivity and specificity were tested using a z-test. This choice of analysis was indicated as the standard procedure for the use of the Chi Square Goodness-of-Fit-Test states that when there are only two categories the \( \chi^2 \) test is equivalent to the z-test (Hopkins & Glass, 1978). Significant differences were found between the sensitivity of the Alcohol and Drug Dependence scales for the present sample and the sensitivity of the Alcohol and Drug Dependence scales for a normative, test construction sample as reported by Millon (1987). Significant differences were also found between the specificity of the Alcohol and Drug Dependence scales for the present sample and the specificity of the Alcohol and Drug Dependence scales for a normative, test construction sample as reported by Millon (1987).

The hypotheses which addressed the differences in BR scores on the Alcohol and Drug Dependence scales for two sub-samples were tested using the Wilcoxon test. For the first sub-sample, for whom Alcohol Dependence and/or Abuse was the only diagnosis, the difference between the BR scores of the Alcohol and Drug Dependence scales was significant. However, for the second sub-sample, for whom Drug Dependence and/or Abuse was the only diagnosis, the difference between the BR scores of the Drug and Alcohol Dependence scales was not found to be significant.
CHAPTER V

SUMMARY, CONCLUSIONS, AND DISCUSSION

Summary

Psychological tests and inventories continue to be used as part of the process of assessment and treatment planning in clinical practice. When the results of these measures are used to assign people to diagnostic categories or to prescribe treatment, it is essential that the validity of the instrument has been established (Gibertini, 1993). It has become a standard of ethical practice for the author of the test to provide data that can be used to assess the validity of the instrument (American Psychological Association, 1985). Additionally, new instruments need to be externally validated with samples other than the normative sample (Millon, 1987).

The MCMI-II is a self report personality inventory for the assessment of psychiatric patients which was introduced as a comprehensive revision of the original instrument, the MCMI-I. Changes were made in the scales and their items, the scoring system, the approach to interpretation of the personality disorder scales, and there was an increased focus on standardization for minority groups. Though the MCMI-II is a different instrument from the MCMI-I, the emphasis on such concepts as clinical prototypes, the polythetic nature of categories, and the differentiation of disorders according to severity, remains the same.

One of the issues that led to the development of the MCMI-II was concern about the diagnostic efficiency, both in terms of reliability and validity, of the MCMI-I (Choca et al., 1986; McMahon et al., 1985; Piersma, 1986, 1987; Retzlaff & Gibertini, 1987; Widiger & Sanderson, 1987). The results of a number of studies
(Bryer et al., 1990; Calsyn et al., 1990, 1991; Choca et al., 1988; Gibertini et al.,
1986; McMahon & Davidson, 1986) indicated that the MCMI-I over diagnosed
personality disorders, and did not replicate the high sensitivity figures reported by
Millon (1983). This mixed review on the efficacy of the MCMI-I scales also included
the Alcohol and Drug Dependence scales. Several studies (Flynn & McMahon, 1983,
1984; McMahon et al., 1985) along with the test construction data, presented
supportive conclusions about the discriminant capacity of these scales. However, the
results of a substantial number of studies called the use of the MCMI-I as an
assessment tool for substance abuse into question (Bryer et al., 1990; Calsyn et al.,
1990, 1991; Gibertini & Retzlaff, 1988; Jaffe & Archer, 1987; Marsh et al., 1988;
McCann, 1990; Miller & Streiner, 1990). Furthermore, there was some evidence that
the Alcohol Dependence scale was the most effective measure for all forms of
substance abuse (Jaffe & Archer, 1987).

External validation studies of the MCMI-I have been criticized for using non-
standardized psychiatric diagnoses as the standard for comparison (Bryer et al., 1990;
Piersma, 1993; Spitzer, 1983). The generally low reliability for psychiatric diagnosis
across clinicians led to the development of structured interviews for research and
clinical purposes (Endicott & Spitzer, 1978; Harkavy-Friedman, 1989; Helzer et al.,
1978; Page, 1991; Spiker & Ehler, 1984; Sholomskas, 1990; Wixted et al., 1993;
Weiss, 1993). One of the recently developed structured interviews is the SCID, an
interview that is based on the current nosological system as found in the DSM-III-R.

The research application of the SCID is supported by the high levels of inter-
rater reliability that have been reported for the instrument (Copolov et al., 1986;
Bryant et al., 1991; Kosten et al., 1991; Maier et al., 1988; Riskind et al., 1987; Skre
et al., 1991; Spitzer et. al, 1990; Williams et al., 1992). Two of these studies
included the use of the SCID to diagnose substance abuse disorders (Kosten et al.,
Stated in terms of the statistic kappa, the inter-rater reliability for the diagnosis of Alcohol Dependence and/or Abuse and Drug Dependence and/or Abuse was reported as ranging from .93 to .96.

The findings concerning the diagnostic efficiency of the Alcohol and Drug Dependence scales of the MCMI-I have been ambivalent. The SCID on the other hand has been identified as a reliable instrument for the differential diagnosis of DSM-III-R psychiatric disorders. Within this context the current study used the SCID to generate substance dependence and abuse diagnoses as the standard to which to compare the results of the Alcohol and Drug Dependence scales. Furthermore, the BR scores on the Alcohol and Drug Dependence scales were evaluated to determine if each was the most efficient scale to diagnose the disorders that they were designed to identify.

Findings in Relation to Research Questions

Is the discriminant validity of the MCMI-II Alcohol Dependence scale for the present sample, expressed in terms of sensitivity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

The sensitivity of the Alcohol Dependence Scale (.65), for the present sample of individuals who presented themselves for substance abuse treatment, was significantly less than the sensitivity reported by Millon (1987) for the cross validation sample (.87). These results are consistent with similar studies of the MCMI-I substance dependence scales, an analysis of which called into question the ability of the Alcohol Dependence scale to identify individuals who received a diagnosis of Alcohol Dependence or Abuse using DSM-III-R criteria (Bryer et al., 1990; Gibertini & Retzlaff, 1988; Jaffe & Archer, 1987; McCann, 1990; Miller & Streiner, 1990). No external validation studies of the diagnostic efficacy of the MCMI-II Alcohol Dependence...
Dependence scale are reported in the literature to date. The results of this aspect of the current study indicate that the probability that the Alcohol Dependence scale is positive (BR ≥ 75) given the disorder is present is .62. Thus, the scale neglected to identify 38% of the individuals in the present sample who received a SCID generated diagnosis of either Alcohol Dependence and/or Alcohol Abuse. In terms of sensitivity with the present sample, the Alcohol Dependence scale therefore appears to function only marginally better than chance.

Is the discriminant validity of the MCMI-II Alcohol Dependence scale for the present sample, expressed in terms of specificity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

The specificity of the Alcohol Dependence Scale (.77), for the present sample of individuals who presented themselves for substance abuse treatment, was significantly less than the sensitivity reported by Millon (1987) for the cross validation sample (.98). These results are not consistent with studies of the MCMI-I substance dependence scales, an analysis of which generally affirmed that the Alcohol Dependence scale did not identify individuals as having substance abuse difficulties when they did not receive a diagnosis of Alcohol Dependence or Abuse using DSM-III-R criteria (Bryer et al., 1990; Gibertini & Retzlaff, 1988; Jaffe & Archer, 1987; McCann, 1990; Miller & Streiner, 1990). The results of this aspect of the current study indicate that the probability that the Alcohol Dependence scale is negative (BR < 75) given the disorder is absent is .77. Thus, there is a 23% chance that the Alcohol Dependence Scale will identify an individual as having a history of, or current dependence on, alcohol, even though the individual does not warrant a SCID generated diagnosis of Alcohol Dependence and/or Alcohol Abuse. In terms of specificity with the present sample, this indicates the clinician may have only a
marginal level of confidence about the results obtained using the Alcohol Dependence scale.

Is the discriminant validity of the MCMI-II Drug Dependence scale for the present sample, expressed in terms of sensitivity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?

The sensitivity of the Drug Dependence Scale (.55), for the present sample of individuals who presented themselves for substance abuse treatment, was significantly less than the sensitivity reported by Millon (1987) for the cross validation sample (.72). These results are consistent with a substantial number of studies of the MCMI-I substance dependence scales. The accumulation of this data has called into question the ability of the Drug Dependence scale to identify individuals who receive a diagnosis of Drug Dependence or Abuse using DSM-III-R criteria (Bryer et al., 1990; Calsyn et al., 1990, 1991; Marsh et al., 1988; McCann, 1990). No external validation studies of the diagnostic efficacy of the MCMI-II Drug Dependence scale are reported in the literature to date. An analysis of the results of this aspect of the current study indicates that the probability that the Drug Dependence scale is positive (BR ≥ 75) given the disorder is present is .55. Thus, the scale neglected to identify 45% of the individuals in the present sample who received a SCID generated diagnosis of either Drug Dependence and/or Drug Abuse. In terms of sensitivity with the present sample, the Drug Dependence scale therefore appears to function only slightly better than chance.

Is the discriminant validity of the MCMI-II Drug Dependence Scale for the present sample, expressed in terms of specificity, comparable to the values reported by Millon (1987) for the cross validation sample used in the construction of the MCMI-II?
The specificity of the Drug Dependence Scale (.85), for the present sample of individuals who presented themselves for substance abuse treatment, was significantly less than the sensitivity reported by Millon (1987) for the cross validation sample (.99). In general these results are consistent with studies which addressed the specificity of the MCMI-I Drug Dependence scale (Bryer et al., 1990; Calsyn et al., 1990, 1991; Marsh et al., 1988; McCann, 1990). The results of this aspect of the current study indicate that the probability that the Drug Dependence scale is negative (BR < 75) given the disorder is absent is .85. Thus, there is a 15% chance that the Drug Dependence Scale will be positive (BR ≥ 75) when the individual does not receive a SCID generated diagnosis of Drug Dependence and/or Abuse. In terms of specificity, this leaves the clinician with a marginal level of confidence in the Drug Dependence scale as it may misclassify individuals 15% of the time.

When Alcohol Dependence and/or Alcohol Abuse is the only diagnosis on Axis I, are the BR scores of the MCMI-II Alcohol Dependence scale greater than the BR scores of the MCMI-II Drug Dependence scale?

According to the data, the BR scores of the Alcohol Dependence scale, for the sub-sample of participants for whom the diagnosis of Drug Dependence and/or Drug Abuse was the only substance abuse diagnosis, were significantly greater than the BR scores of the same individuals on the Drug Dependence scale. These findings are similar to the study by Jaffe and Archer (1987), the results of which indicated that the Alcohol Dependence scale was the most elevated of the two substance abuse scales, when Alcohol Dependence and/or Abuse was the only substance abuse diagnosis. Thus, the Alcohol Dependence scale appears to function as it was designed, relative to the Drug Dependence scale, when the only substance abuse diagnosis an individual has received is either Alcohol Dependence and/or Abuse.
When Drug Dependence and/or Drug Abuse is the only substance abuse diagnosis on Axis I, are the BR scores of the Drug Dependence scale greater than the BR scores of the MCMI-II Alcohol Dependence scale?

According to the data, the BR scores of the Drug Dependence scale, for the sub-sample of participants for whom the diagnosis of Drug Dependence and/or Drug Abuse was the only substance abuse diagnosis, were not significantly greater than the BR scores of the same individuals on the Alcohol Dependence scale. These findings are similar to the data reported by Jaffe and Archer (1987) who indicated that the Alcohol Dependence scale was the most elevated of the two substance abuse scales when Drug Dependence and/or Drug Abuse was the only substance abuse diagnosis. Thus, data continues to emerge which suggests that the Drug Dependence scale does not function as it was designed, relative to the Alcohol Dependence scale.

Conclusions

An analysis of the data generated by the present study indicates that there is a significant discrepancy between the diagnostic efficacy of the Alcohol and Drug Dependence scales reported by Millon for the normative sample (1987) and a sample of individuals presenting for substance abuse treatment. These finding are consistent with data reported in the literature (See Tables 6 and 7) which had been critical of the manner in which the Alcohol and Drug Dependence scales of the MCMI-I functioned across a wide spectrum of populations (Bryer et al., 1990; Calsyn et al., 1990, 1991; Gibertini & Retzlaff, 1988; Jaffe & Archer, 1987; Marsh et al., 1988; McCann, 1990; Miller & Streiner, 1990). The MCMIs were developed to aid the clinician in the process of differential diagnosis. As Burisch (1984) states, in his astute discussion of validity and the purposes of personality inventories, "it should be clear enough that
### Table 6
Summary of Diagnostic Efficiency of Alcohol Dependence Scale of the MCMI-I and MCMI-II

<table>
<thead>
<tr>
<th>Authors</th>
<th>MCMI Scale</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCann (1990)</td>
<td>Alcohol Dependence</td>
<td>poor discriminant validity</td>
</tr>
<tr>
<td>Gibertini &amp; Retzlaff</td>
<td>Alcohol Dependence</td>
<td>17% of known alcoholics had BR scores &gt; 74</td>
</tr>
<tr>
<td>(1988)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bryer et al. (1990)</td>
<td>Alcohol Dependence</td>
<td>Sensitivity .43</td>
</tr>
<tr>
<td>Miller &amp; Streiner (1990)</td>
<td>Alcohol Dependence</td>
<td>Sensitivity .49</td>
</tr>
<tr>
<td>van Hoek (current)</td>
<td>Alcohol Dependence</td>
<td>Sensitivity .62</td>
</tr>
<tr>
<td>Millon (1987)</td>
<td>Alcohol Dependence</td>
<td>Sensitivity .87</td>
</tr>
</tbody>
</table>

### Table 7
Summary of Diagnostic Efficiency of Drug Dependence Scale of the MCMI-I and MCMI-II

<table>
<thead>
<tr>
<th>Authors</th>
<th>MCMI Scale</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bryer et al. (1990)</td>
<td>Drug Dependence</td>
<td>Sensitivity .49</td>
</tr>
<tr>
<td>Calsyn et al. (1991)</td>
<td>Drug Dependence</td>
<td>Sensitivity .49</td>
</tr>
<tr>
<td>Marsh et al. (1988)</td>
<td>Drug Dependence</td>
<td>Sensitivity .49</td>
</tr>
<tr>
<td>van Hoek (current)</td>
<td>Drug Dependence</td>
<td>Sensitivity .55</td>
</tr>
</tbody>
</table>
criterion and discriminant validity are of critical importance in the assessment and research contexts, but that in a prognosis situation effectiveness is all that is necessary" (p. 218). On the basis of on the results of the current study, the efficacy of the Alcohol and Drug Dependence scales of the MCMI-II is questionable. These findings have considerable implication for the clinician who is faced with a diagnostic decision. On the basis of the data related to sensitivity for the present sample, a clinician who receives a profile, on which there is no clinically significant elevation (BR ≥ 75) on the Alcohol Dependence scale, must recognize that there may be a 35% chance that the disorder is actually present even though the test is negative. Likewise, for the Drug Dependence scale, if there is no clinically significant elevation (BR ≥ 75) the clinician must consider that there may be a 45% chance that the disorder is actually present even though the test is negative. In terms of specificity, the implications for clinical practice are equally troubling. Thus, a clinician who receives a profile that is clinically significant (BR ≥ 75) on the Alcohol Dependence scale must consider that there could be a 23% chance that the disorder is in actuality not present. Similarly, for the Drug Dependence scale, when there is a clinically significant elevation (BR ≥ 75) there could be a 15% chance that the disorder is in actuality not present. Thus, these scales may be of limited value to the clinician whose diagnostic decisions are likely to have a significant impact in such areas as treatment planning or an individual's employment status.

An analysis of the results of the present study indicates that the Drug Dependence scale may be no more effective in identifying individuals who abuse drugs than the Alcohol Dependence scale. Corroborating evidence for the results of the analysis of the BR scores with the Wilcoxon test for both these scales, can be found in a calculation of the median scores for the sub-samples. For the sub-sample of 11 participants whose only SCID generated diagnosis was Drug Dependence and/or...
Abuse, the median BR score for the Drug Dependence scale was 60 while the median BR score for the Alcohol Dependence scale was 67. This is not consistent with the data reported by Millon (1987) for two samples of individuals with a diagnosis of Drug Dependence and/or Drug Abuse. For these samples the reported median BR scores for the Drug Dependence scale were 77 and 78 and the median BR scores for the Alcohol Dependence scales were reported as 72 and 71. A further examination of the individual BR scores of this sub-sample showed that the BR scores of the Drug Dependence scale ranged from a low of 41 to a high of 107. However, of the 11 scores only one, the BR score of 107, was clinically significant (BR ≥ 75). The BR scores of the Alcohol Dependence scale for the same individuals ranged from 44 to 106, and of the 11 scores 5 were clinically significant (BR ≥ 75).

For the sub-sample of 30 participants who received a SCID generated diagnosis of only Alcohol Dependence and/or Abuse, the Alcohol Dependence scale appeared to function as expected in relation to the Drug Dependence scale. A calculation of the median BR scores provides corroborating evidence for this conclusion. The median BR score of the Alcohol Dependence scale was 78 while the median BR score of the Drug Dependence scale was 61. Similarly, the median BR scores of the Alcohol and Drug Dependence scales for two samples of individuals who were diagnosed as having substance abuse difficulties, reported by Millon (1987) were 86 and 85 and 78 and 73 respectively. While the median BR scores for the sub-sample in the current study were less than the scores reported by Millon (1987), the elevations relative to each other are consistent with the data reported in the test manual. Thus, it appears that the results of the present study are consistent with the data reported by Millon (1987) regarding the performance of the Alcohol Dependence scale relative to the Drug Dependence scale. It appears that the Alcohol Dependence scale
identifies individuals with alcohol related diagnoses more accurately than the Drug Dependence scale.

Discussion

It is difficult to reconcile the differences in the reported validity figures (Millon, 1987) and the data generated for the sample in the present study. Further analysis indicates that an examination of the following three areas may provide a context from which to understand these discrepancies. The remainder of this section will therefore focus on the implications of: (a) test construction theory, (b) the incorporation of the polythetic approach to classification and the resulting item selection and scale overlap, (c) the diagnosis of substance abuse using DSM-III-R criteria.

Issues Related to General Test Construction Theory

The MCMIs were developed using a comprehensive test construction theory. Millon (1983, 1987) makes a case that the reliability and the validity of the MCMIs should be expected to be high as a result of the care which was taken to incorporate these as part of the test construction process. However, in extensive reviews of studies that compared the effectiveness of approaches to test construction, Burisch (1984) and Hase and Goldberg (1989) have shown that there is no difference in the effectiveness of the resulting instruments in identifying the constructs which they were intended to measure. Burisch (1984) further argues that the more economical and direct approach of simple, trait, self-rating scales is therefore the preferable choice of design. The analyses provided by these authors diminish Millon’s (1983, 1987) convincing claims of superiority of the method of test construction utilized in the
development of the MCMIs, and by implication the reliability and validity, of the instruments.

An additional consideration involves item selection. Millon (1983, 1987) reported that the item pool for the Alcohol and Drug Dependence scales included a considerable number of subtle items which were expected to increase the efficacy of the scales. However, a number of studies have shown that the utilization of subtle items does not increase the validity of scales and that subtle items do not contribute unique information to the evaluation process (Burisch, 1984; Lanyon, 1984; Paunonen & Jackson, 1985; Worthington & Schlottmann, 1986). Moreover, in the assessment of substance dependence a more direct approach has been advocated (Skinner, 1989).

The results of the present study appear to support the arguments of those who assert that the extensive test validation process does not improve the discriminant validity of theory based instruments (Burisch, 1984; Hase & Goldberg, 1989). While there is some question whether or not "substantive-theoretical" validation occurred for the Alcohol and Drug Dependence scales, the inconsistent performance of these scales suggests that the careful test construction strategy employed by Millon (1987) has not translated into increased discriminant validity of these scales. Furthermore, based on the data generated by the present study, it may be stated that the inclusion of subtle items in the evaluation of substance abusers is not indicated. In this regard the current study provides validation for Skinner's (1989) point of view that a direct approach to the assessment of substance abuse is more effective.

**The Polythetic Nature of Categories**

One of the guiding principles in the construction of the MCMI-I and MCMI-II, was a polythetic approach to the classification of psychiatric disorders (Millon, 1983,
1987). This decision had significant impact on the process of item selection as well as the degree of overlap and correlation between scales. Polythetic rules of classification are operationalized through the selection of prototypical items which in conjunction with more descriptive items comprise the scales. The inclusion of descriptive items is based on the assumption that there are consistent connections between personality and clinical syndrome scales. It is not at all clear that this is a valid assumption for substance dependence and or abuse disorders. Factor analytic studies have consistently identified what appears to be a substance abuse dimension (Lorr, 1993) that is measured by the MCMIs, and Millon (1987) has suggested that there is a correlation between narcissistic, antisocial, and histrionic personality disorders and substance abuse. The descriptive items of the Alcohol and Drug Dependence scales are therefore designed to assess personality characteristics such as impulsivity, extroversion, mania, and paranoia. However, as Strang, Bradley, and Stockwell (1989) state, "Studies of drug use and drug problems are perhaps more dogged by assumptions of causal relationships than any other field. Despite warnings of dangers of confusing causes, correlates and consequences, the warnings are insufficiently heeded" (p. 211). The poor performance of the Alcohol and Drug Dependence scales may therefore partly be attributed to an erroneous assumption about the correlation of personality and substance abuse. Moreover, the descriptive items may also be viewed as subtle items and, as was noted earlier, the contribution of this type of item is questionable.

The limited utility of the descriptive items is exacerbated by the fact that there are only a limited number of prototypic items. As Paunonen and Jackson (1985) state, "the most internally consistent item clusters (a) are those most saturated with trait relevant content, (b) may represent highly salient and concrete behavioral exemplars of the trait, and therefore (c) may be the most valid items with respect to a
criterion" (p. 336). Both the Alcohol and Drug Dependence scales have only 6 prototypic items that directly address aspects of substance abuse. The performance of the scales may therefore be adversely affected by limited content sampling (McCann, 1990). Furthermore, McCann, Flynn, and Gersh, (1992) report that it is possible for an individual to endorse all of the prototypic items for a particular scale and yet not receive a score that is clinically significant (BR ≥ 75).

Another reason for the poor discriminant validity of the Alcohol and Drug Dependence scales appears to be related to the excessive overlap of the scales. The fact that there are a limited number of items from which to draw all the scales of the MCMIs, had raised concerns about the influence of common item artifact (McCann, 1990; Wetzler 1990). In his analysis of the poor discriminant validity of the MCMI-II Clinical Syndrome scales, McCann (1990) reports that regardless of the presence of common item artifact, however that, "the clinical scales appear to be permeated with a general maladjustment factor" (p. 474). Though Millon (1987) responded to the concerns about overlapping of scales, and the resulting common item artifact, by introducing a weighted scoring system, a number of authors (Retzlaff, Sheehan, & Lorr, 1990; Streiner & Miller, 1989; Streiner, Goldberg, & Miller, 1993) have demonstrated that there was no significant difference between the manner in which the scales functioned with or without weighted scoring. Additionally, McCann (1990) reports that some of the MCMI-II scales perform better in overlapping form, but that the Alcohol and Drug Dependence scales were most affected by common item artifact. This suggests that these scales would be most effective as a clinical tool if the diagnosis was based on an expanded pool of prototypic items only (McCann et al., 1992).
The SCID generates substance abuse diagnoses according to the criteria of the DSM-III-R. A major contributing factor to the development of these criteria was the introduction of the concept of a dependence syndrome that has as a focus behavioral and physiological changes separate from the social impact of substance abuse (Edwards, Arif, & Hodgson, 1982). The DSM-III-R criteria therefore focus on such behaviors as: (a) inability to stop drug use, (b) problems in getting through a set period of time without using, and (c) withdrawal signs and symptoms. An examination of the items of the MCMI-II Alcohol and Drug Dependence scales indicates that these are not areas of focus. As Bryer and colleagues (1990) state:

As the dearth of many items with direct substance-abuse content attests, these scales are designed to assess substance abuse through evaluation of personality dynamics and related symptoms. This approach raises questions because the literature on personality style and alcoholism is complex and somewhat inconclusive. (p. 438)

Thus, the poor discriminant validity of the Alcohol and Drug Dependence scales, when compared to SCID generated diagnoses of Alcohol and Drug Dependence and/or Abuse, may be the result of the fact that the two instruments do not measure the same constructs. For example, the SCID may be measuring the physiological and overt signs and symptoms of substance abuse, while the Alcohol and Drug Dependence scale may be identifying personality features and other dynamic issues. However, the prevalence rate of substance abuse in the sample for the present study was expected to be 100%, and the SCID identified all but one of the participants as having a substance abuse diagnosis. It is therefore more likely that the concerns expressed by Bryer and colleagues (1990) are valid and that the criteria on which the Alcohol and Drug Dependence scales are based are inadequate as measures of substance abuse. The
sensitivity and specificity of the Alcohol and Drug Dependence scales in the present study would support this perspective.

Recommendations

The MCMI I was introduced as an instrument that would aid in the process of differential diagnosis. One of Millon's (1983, 1987) intentions was to construct an instrument that would reflect the natural relationships between personality styles and clinical symptoms. In order to accomplish this goal, the concepts of the polythetic nature of categories and prototypes, was used as a guiding principle in the construction of the tests. Furthermore, Millon (1983, 1987) made a commitment to follow an extensive process of ongoing validation as part of test construction and to update the instrument periodically to incorporate new developments in the theories of personality and assessment as well as changes in the nosological system. The development of the MCMI-II also occurred within this context and in an effort to address concerns raised about the MCMI-I.

To date there have been few studies that focus on the validation of the individual scales of the MCMI-II (Craig, 1993). The current study of the Alcohol and Drug Dependence scales was intended to be a part of the ongoing process of scale validation. The findings from this investigation do not affirm the efficacy of either of these two scales. Though the Alcohol Dependence scale generally outperformed the Drug Dependence scale, neither consistently classified individuals accurately when SCID generated diagnoses were used as the standard for comparison. Additionally, the data from the present study was consistent with an earlier study (Jaffe & Archer, 1987) which indicated that the Alcohol Dependence scale was more effective in identifying all forms of substance dependence and/or abuse. Recognizing the limits of this and other personality assessment instruments, Millon (1987) states:
No diagnostic procedure, be it self-report, structured interview, or projective test, should be gauged against a criterion of absolute accuracy, since all procedures incur some measure of error. Rather, the magnitude of error produced with each instrument must be quantified and compared with alternative procedures. (p. 201)

The current investigation supports Millon's (1987) perspective on the lack of efficacy of diagnostic instruments. Furthermore, the magnitude of the inaccuracies were such that neither the Alcohol or the Substance Dependence scales performed much better than chance. Having quantified these differences it appears indicated to suggest that the Alcohol and Drug Dependence scales are likely to be of limited usefulness to the clinician engaged in the process of differential diagnosis as it pertains to substance abuse.

There are a number of areas that warrant exploration and investigation to maximize the utility of the MCMI-II in the area of substance abuse assessment. First, there appear to be a number of factors which point to the lack of effectiveness in using the concept of polythetic categories in the assessment of substance abuse. The use of scale overlapping appears to have a detrimental effect on the performance of both the Alcohol and Drug Dependence scales. In fact, when all but the prototypic items are eliminated, the scales are more effective (McCann et al., 1992). Serious consideration needs to therefore be given to the development of a substance abuse scale(s) which consists of prototypic items only, an approach which is consistent with the concept of substance dependence underlying the DSM-III-R diagnosis. Secondly, limited content sampling along with the excessive overlapping of the scales, contribute to the fact that the Alcohol Dependence scale is the most effective measure of all substance abuse disorders. It therefore appears warranted to eliminate the Drug Dependence scale and to develop one scale that would provide a general screen for substance abuse. Third, the design of a general substance abuse screening scale needs to include
items that reflect the concept of dependence, a central component in establishing a substance abuse diagnosis using DSM-III-R criteria. Lastly, there is a need for studies that will evaluate the convergent validity of the Alcohol and Drug Dependence scale and measures of substance dependence that focus on a direct approach to identifying individuals with substance abuse issues.

Substance abuse assessment is particularly difficult. Due to the many facets of substance abuse and the variety of personality factors that are involved when assessing substance abuse, it may be unrealistic to expect one instrument to account for all of these variables. Sound clinical practice may require the use of a combination of several instruments in contrast to a single test or scale to establish substance abuse diagnoses. An analysis of the data from the current study, indicates that it is unwarranted for clinicians to use only the Alcohol and Drug Dependence scales of the MCMI-II as measures of current or past substance abuse. Additionally, an elevation on the Alcohol Dependence scale requires the investigation of possible poly-substance abuse. In this regard the current study adds to the literature that questions the discriminant validity of the clinical syndrome scales of the MCMIs and specifically the performance of the Alcohol and Substance Dependence scales.
Appendix A

Human Subject Institutional Review Board Approval
Date: February 6, 1995
To: van Hoek, Bert
From: Richard Wright, Interim Chair
Re: HSIRB Project Number 95-01-10

This letter will serve as confirmation that your research project entitled “Use of a structured interview to evaluate the validity of the alcohol and drug dependency scales of the Millon Clinical Multiaxial Inventory II” has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you must seek specific approval for any changes in this design. You must also seek reapproval if the project extends beyond the termination date. In addition if there are any unanticipated adverse or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: February 6, 1996

xc: Prosser, CECP
BIBLIOGRAPHY


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Lanyon (Eds.), *Readings in Personality Assessment* (pp. 281-304). New York: Wiley.


T. E. Moore (Ed.). Cognitive Development and the Acquisition of Language. 

assessment for clinicians (pp. 45-66). Washington, D.C.: American 
Psychiatric Press.

Reading, M.A.: Addison-Wesley.

(Eds.), Depressive disorders facts theories and treatment methods (pp. 231-

McGraw-Hill.

Journal of Abnormal Psychology, 90, 68-87.

Wetzler, (Ed.), Measuring Mental Illness: Psychometric Assessment for 

research: The case of psychiatric classification. Journal of Consulting and 
Clinical Psychology, 50,727-735.

for the Structured Clinical Interview for DSM-III-R Axis I. Acta Psychiatrica 
Scandinavica, 84, 167-173.

G. Goldstein & M. Hersen (Eds.), Handbook of Psychological Assessment 

Comprehensive Psychiatry, 24, 399-411.


