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.

U·M·I

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

Order Number 9238149

**The consistency of scores and configural patterns between the
MMPI and MMPI-2**

Chojnacki, Joseph T., Ph.D.

The Ohio State University, 1992

U·M·I

**300 N. Zeeb Rd.
Ann Arbor, MI 48106**

THE CONSISTENCY OF SCORES AND CONFIGURAL PATTERNS
BETWEEN THE MMPI AND MMPI-2

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in the
Graduate School of the Ohio State University

by

Joseph T. Chojnacki, B.A., M.A.

* * * * *

The Ohio State University

1992

Dissertation Examination Committee:

Approved by

W. B. Walsh

S. H. Osipow

R. K. Russell

Adviser

Department of Psychology

ACKNOWLEDGEMENTS

A project such as this unfairly bears the name of only one person, yet without the help and support of other persons, such a project would never have come into existence. I would like to express my sincere appreciation to Dr. Samuel H. Osipow, Dr. Richard K. Russell, and Dr. James V. Wigtil for their suggestions and guidance while serving as members of my dissertation committee. To my advisor, Dr. W. Bruce Walsh, I offer deep gratitude for his never tiring support and encouragement during this project. I extend a special thanks to Deb Serling, Todd Manolis, Dr. Terry A. Carlson, Zeina Daoud, Jean Chagnon, Lynnea Erikson, Eileen McMahon, Anita Davena, Kathy Ingram, John Beery, Eric George, and Cindy Roush whose timely advice, comforting support, and insightful perspective were essential to the successful completion of such a project. Foremost, I would like to thank my family -- whose endless, lifelong love and support makes possible endeavors such as this.

VITA

August 6, 1965 Born - Elmhurst, Illinois

1987 B.A., Cornell College,
Mount Vernon, Iowa

1987-1988 University Fellow,
The Ohio State University

1988-1991 Graduate Teaching Associate,
Department of Psychology,
The Ohio State University

1991 Adjunct Faculty
Otterbein College
Westerville, Ohio

1991-1992 Psychology Intern
Student Counseling Center
Illinois State University
Normal, Illinois

PUBLICATIONS

Chojnacki, J. T., & Walsh, W. B. (1990). Reliability and concurrent validity of the Sternberg Triangular Love Scale. Psychological Reports, 76, 219-224.

FIELDS OF STUDY

Major Field: Psychology

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
INTRODUCTION	1
REVIEW OF LITERATURE	9
MMPI: Developments and Issues	9
MMPI-2: The Changes	25
Empirical Investigation of Consistency	32
The Present Study	43
METHOD	46
Subjects	46
Instruments	47
Procedure	51
Scoring	52
Data Analyses	53
RESULTS	62
Demographic Information	62
Correlational Analyses	64
Configural Analyses	75

DISCUSSION	90
Correlational Analyses	90
Configural Analyses	96
Implications of Analyses	101
Limitations and Future Research	105
Summary	107
REFERENCES	108
APPENDICES	115
A. Demographic Questionnaire	115
B. Scales of the MMPI and MMPI-2	117
C. Written Instructions - First Administration	119
D. Written Instructions - Second Administration	121

LIST OF TABLES

TABLE	PAGE
1. Demographic variables by gender and condition.	63
2. Correlations for the basic scales of MMPI test-retest (O-O condition) and between MMPI and MMPI-2 (O-R condition) by gender	65
3. Correlations for the supplemental scales of MMPI test-retest (O-O condition) and between MMPI and MMPI-2 (O-R condition) by gender	67
4. Correlations for the Harris-Lingoes subscales of MMPI test-retest (O-O condition) and between MMPI and MMPI-2 (O-R condition) by gender	68
5. Interscale correlations for scales L-4 with L-0 of the MMPI and MMPI-2 in the O-R condition for males (<u>n</u> =92)	71
6. Interscale correlations for scales 5-0 with 5-0 of the MMPI and MMPI-2 in the O-R condition for males (<u>n</u> =92)	72
7. Interscale correlations for scales L-4 with L-0 of the MMPI and MMPI-2 in the O-R condition for females (<u>n</u> =96)	73
8. Interscale correlations for scales 5-0 with 5-0 of the MMPI and MMPI-2 in the O-R conditions for females (<u>n</u> =96)	74
9. Configural analyses using 8 clinical scales (scales 0 and 5 deleted)	76
10. Configural analyses using all 10 clinical scales	79

11.	Basic statistics and mean differences of the basic scales for each test administration by condition for males	83
12.	Basic statistics and mean differences of the basic scales for each test administration by condition for females	86

LIST OF FIGURES

FIGURE	PAGE
1. Graphic representation of the means of the basic scales of MMPI and MMPI-2 for males in the O-R condition (<u>n</u> =92)	85
2. Graphic representation of the means of the basic scales of MMPI and MMPI-2 for females in the O-R condition (<u>n</u> =96)	88

Introduction

The Minnesota Multiphasic Personality Inventory (MMPI) is a standardized, objective, personality inventory containing 556 self reference items which can be responded to as either true or false. These items are scored, compared to a normative group, and result in a quantitative assessment of an individual's level of emotional adjustment and test taking attitude. The MMPI contains 4 validity scales, 10 clinical scales, and a number of supplemental and research scales.

The MMPI is the most widely used and researched personality inventory in the United States (Lubin, Larsen, & Matarazzo, 1984). Over 8,000 research studies have been published using the MMPI, and some 500 research scales have been developed for use with it (Groth-Marnat, 1990; Mitchell, 1985). As Walsh and Betz (1990) wrote, "there is little question that the Minnesota Multiphasic Personality Inventory ... is the most useful psychological test available in clinical and counseling settings for assessing the degree and nature of emotional upset" (p.117).

Not only has the MMPI been preeminent in the assessment of personality in applied and research psychology in general, but it has also played an important role in the special-

ty area of counseling psychology. A national survey of 1,000 members of APA Division 17, the Division of Counseling Psychology, by Watkins, Campbell, and McGregor (1988) found that the MMPI was the most frequently used personality inventory across various work settings for counseling psychologists and ranked highest on a list of personality inventories in which members felt that counseling psychologists should be competent. Over 90% of the instructors primarily responsible for the teaching of psychological assessment in 40 of the 47 APA approved counseling psychology programs in 1989 rated the MMPI as moderately or very useful for training and clinical practice, and over 70% rated it similarly for research uses (Watkins, Campbell, McGregor, & Godin, 1989). The MMPI has been an integral component in personality assessment, both in the specialty of counseling psychology and in the field of psychology as a whole.

Much of the contemporary value and utility of the MMPI is derived from the myriad of validity studies that have been conducted with it (Anastasi, 1988; Graham, 1987, 1990; Groth-Marnat, 1984; Kaplan & Saccuzzo, 1982). No other structured inventory has had the amount of research devoted to its validation as the MMPI (Groth-Marnat, 1990; Kaplan & Saccuzzo, 1982). Contemporary interpretive manuals (Duckworth & Anderson, 1986; Graham, 1987, 1990; Greene, 1980, 1990; Lachar, 1974; Marks, Seeman, & Haller, 1974) are based

upon extensively researched empirical correlates of various configural patterns of MMPI scores. Configural patterns refer to how a score on an individual profile stands in relation to other scores on the same profile. Investigations of configural patterns have traditionally included analyses of single scale elevations, two and three point elevations, low scores, and overall elevation, shape, and scatter of a profile (Greene, 1980, 1990).

Since its inception, the MMPI has remained popular despite a number of problems. Problems have included its failure to fulfill its original purpose, severe psychometric limitations, and controversies over the appropriateness of its normative group. For each of these problems, it has been the existence of a vast number of validity studies of various configural patterns which has kept the MMPI useful. For example, Groth-Marnat (1984), writing about problems with the normative group, stated that "the subsequent research performed on empirical code profiles is far more useful than the original normative scale comparisons, it is the extensiveness of research in this area which represents what is probably the strongest asset of the MMPI" (p.258).

In 1989 a revised version of the MMPI, the MMPI-2, was released (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Changes made in the revised version included: 1) the deletion, revision, and addition of items, 2) the use of a contemporary normative group, and 3) the use of a new method

of distributing T scores called the "uniform T" distribution. Developers of the MMPI-2 asserted that "the results of years of research on the MMPI component scales and their pattern of interrelationship are available to guide interpretation of MMPI-2 test scores" (Butcher et al., 1989, p. 1). Can the validity studies, which kept the MMPI useful for so many years, be transferred to MMPI-2? The validity studies are applicable to MMPI-2 only if scores and their resulting configural patterns remain consistent between MMPI and MMPI-2. The purpose of the present study is to empirically investigate the consistency of scores and configural patterns between MMPI and MMPI-2.

Ben-Porath and Butcher (1989a) have attempted to investigate the comparability of MMPI and MMPI-2. These investigators administered the MMPI twice during a time span of 1-2 weeks to one half of a group of 377 undergraduate college students. The other half were administered the MMPI followed 1-2 weeks later by the administration of the MMPI-2. Ben-Porath and Butcher analyzed scores from the validity scales, clinical scales, and 8 of the frequently used supplemental scales which the MMPI and MMPI-2 have in common. Test-retest correlations for the individual scales and analyses of configural patterns indicated, with the exception of two scales, that score changes between the MMPI and MMPI-2 were no different than the score changes that would normally occur when the MMPI is administered twice. Ben-

Porath and Butcher cite an investigation by Butcher, Graham, Dahlstrom, and Bowman (1990) showing similarity in MMPI-2 results between a national sample of college students and the MMPI-2 normative sample in order to assert that their study, showing the comparability of MMPI and MMPI-2, likely generalizes to other normal populations.

The investigation by Ben-Porath and Butcher (1989a) was limited. It did not use MMPI-2 in its investigation, but used MMPI-AX, an experimental form of the MMPI-2 used in the MMPI restandardization project (Butcher et al., 1989). Further, in using test-retest correlations to compare MMPI results with MMPI-2 results, Ben-Porath and Butcher did analyze potential changes in scores between MMPI and MMPI-2 due to item changes, but this type of analysis did not account for changes between MMPI and MMPI-2 due to MMPI-2 utilization of a contemporary normative sample and a uniform T distribution. Finally, configural analyses, conducted by Ben-Porath and Butcher, which do assess changes between the MMPI and MMPI-2 that result from all three of the major changes (item changes, use of contemporary normative group, and the uniform T distribution) were limited in scope and used only 8 of the 10 clinical scales in the analyses.

A study reported by Butcher et al. (1989) indicated that the contemporary normative sample and the use of a uniform T distribution may well affect scores and configural patterns. Butcher et al. (1989) administered the MMPI-2 to

424 psychiatric patients and normed their raw scores on the linear T distribution based on the original Minnesota normative group, on the linear T distribution based on the contemporary normative group, and on the uniform T distribution based on the contemporary group.

Butcher et al. (1989) found changes in the scores and configural patterns of MMPI-2 raw scores depending on the distribution on which they were normed. The study indicated the possible effect of the contemporary normative group and the use of a uniform T distribution on scores and configural patterns. The study by Butcher et al. (1989) was limited, as it used MMPI-AX, did not assess changes in scores and configural pattern resulting from item changes, and had other methodological confounds.

The present study corrects for shortcomings of Ben-Porath and Butcher (1989a) and Butcher et al. (1989). It does not use MMPI-AX, but the actual MMPI-2. This study, like Ben-Porath and Butcher, assesses changes between the MMPI and MMPI-2 due to item changes by the use of test-retest correlations of the validity scales, clinical scales, and 8 of the frequently used supplemental scales which the MMPI and MMPI-2 have in common. The present study expands such an analyses to include the Harris-Lingoes subscales. No empirical data yet exist indicating that Harris-Lingoes subscale scores from MMPI are comparable to scores for MMPI-2, yet Butcher et al. (1989) and Graham (1990) use the

validity information collected for the Harris-Lingoes subscales for the MMPI in their descriptions of the meanings of the Harris-Lingoes subscales for MMPI-2. Further, the present study also expands the correlational analysis by not only assessing the effect of item changes on individual scales, but also the effect of these changes upon how the scales interrelate with each other by making a comparison between the interscale correlations of the MMPI and MMPI-2.

Beyond the correlational analyses, the present study also assesses changes in configural patterns between MMPI and MMPI-2. It uses all configural pattern analyses performed by Ben-Porath and Butcher (1989), which included analyses of consistency in high point elevation, two point elevation, within-normal-limits profiles, number of elevated scales, and the calculation of D^2 , used as a global measure of overall changes in configuration. In the present study analyses are conducted not only on 8 of the clinical scales, but on all 10 of them. Further, the configural analyses are expanded to analyze validity scale configurations, three point elevations, low point consistency, consistency of well defined code types, the D statistic, and overall scale elevation.

Much controversy and debate about the merits of MMPI-2 has surfaced in the literature (Adler, 1990; Edwards & Weissman, 1990; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1990). Such attention and scrutiny is certainly

prudent given the widespread use and popularity of MMPI. MMPI utility rests upon the numerous validity studies conducted with it. Members of the committee that revised the MMPI claim that this research is transferable to MMPI-2, a claim which has yet to be empirically confirmed. This study has been designed as an empirical assessment of that claim.

Selected Review of the Literature

The review of literature will be divided into four sections. The first section will review the MMPI, its development and issues surrounding its use. The second section will focus on the MMPI-2 and how it differs with the MMPI. The third section will review existing data on the consistency of scores and configural patterns between MMPI and MMPI-2. The final section will outline the present study.

MMPI: Development and Issues

This section will begin by briefly reviewing the development of the MMPI. This will be followed by a review of three general problems and issues surrounding MMPI use: 1) the failure of the MMPI to fulfill its original purpose, 2) severe psychometric limitations of the MMPI, and 3) controversies over the appropriateness of the MMPI normative group. These three problems are not mutually exclusive, nor are they a complete list of issues surrounding MMPI use. They are chosen to be representative of MMPI controversies and also for their use in providing background information relevant to the discussion of MMPI-2. Each overview of an MMPI problem will include a discussion of how the problem has been overcome or circumvented by the vast number of

validity studies of MMPI scale and configural correlates.

MMPI Development. Starke R. Hathaway and J. Charnley McKinley of the University of Minnesota first published the MMPI in 1943. The MMPI was originally developed with the purpose of aiding in the assignment of psychodiagnostic labels to psychiatric patients. Hathaway and McKinley used empirical criterion keying in order to construct the MMPI scales. Using this method, items were included on a scale not based on face validity, but based on whether they were able to differentiate between two groups of subjects (Hathaway & McKinley, 1940).

One group of subjects, the "Minnesota normals", consisted of 724 friends and relatives of patients at the University of Minnesota hospitals. The second group of subjects, the "clinical subjects", were psychiatric patients at the University of Minnesota Hospitals. The clinical subjects were divided into subgroups based on their diagnosis of either hypochondriasis, depression, hysteria, psychopathic deviate, paranoia, psychasthenia, schizophrenia, or hypomania. Test items (developed from previous instruments, clinical reports, case histories, psychiatric interviewing manuals, and personal clinical experience) were generated and given to the normal and clinical subjects. An item which the normal group and a particular clinical subgroup responded to differently was included on a scale named after that particular clinical subgroup. The resulting scales

were then cross-validated by selecting a new group of normal subjects, consisting of various groups of recent high school graduates, Work Progress Administration workers, and Medical patients at the University of Minnesota Hospitals, and comparing their responses with new subgroups of clinical subjects. Items that continued to differentiate between the two groups were used to form eight clinical scales: hypochondriasis (Hs), depression (D), hysteria (Hy), psychopathic deviate (Pd), paranoia (Pa), psychasthenia (Pt), schizophrenia (Sc), and hypomania (Ma) (Hathaway 1956; Hathaway & McKinley, 1940; Hathaway & McKinley, 1942; McKinley & Hathaway, 1940; McKinley & Hathaway, 1942; McKinley & Hathaway, 1944).

Two additional clinical scales were also developed and included in the MMPI. The masculinity-femininity (Mf) scale was originally developed to differentiate male homosexuals from male heterosexuals. Not enough items differentiated between these two groups, and so the scale was expanded to include items that were differentially responded to by males and females and also items from the Terman and Miles (1936) I scale. The second additional clinical scale added to the MMPI, the social introversion (Si) scale, was developed by Drake (1946) using empirical criterion keying to differentiate between female college students who participated extensively in extracurricular activities with those who did not.

In addition to these 10 clinical scales, 4 validity

scales were also developed (Hathaway & McKinley, 1942; McKinley, Hathaway, & Meehl, 1948; Meehl & Hathaway, 1946). These scales were developed to assess different test taking attitudes which could potentially invalidate the results of the clinical scales. The cannot say (?) scale is simply the total number of unanswered questions. The lie (L) scale attempts to detect naive attempts by subjects to present themselves favorably. Items for the L scale were rationally developed and designed to indicate a reluctance to admit to even minor weaknesses in character. The infrequency (F) scale is composed of items that less than 10% of the Minnesota normal group endorsed in a particular direction. The scale is designed to detect subjects who are faking pathology or not following instructions. The correction (K) scale was designed to detect clinical defensiveness. Items were selected that discriminated a group of known psychiatric patients who produced normal MMPI profiles (clinically defensive) from a group of normal subjects who produced normal MMPI profiles. Later the K scale was also used as a correction factor to adjust the scores of some of the clinical scales which may have been artificially lowered as a result of clinical defensiveness.

Together these scales form the four validity scales and ten clinical scales of the MMPI, often referred to as the "basic scales". Scores on these scales, with the exception of the ? scale, are converted to T scores by norming raw

scores with the distribution of scores obtained by the Minnesota normal group on these scales. These scores are often plotted together on a graph and represent a subject's MMPI profile.

A number of supplemental and research scales have also been developed for the MMPI (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975). Some of the most popular and commonly scored of these scales have been the Supplemental scales, the Wiggins (1969) Content scales, the Harris-Lingoes (1955, 1968) subscales, and the Wiener-Harmon (1946) subtle-obvious scales. In addition to these scales, lists of "critical items" are also often scored (Koss & Butcher 1973; Lachar & Wrobel, 1979).

The MMPI has a number of formats. The format most commonly used and recommended for research is known as the "group form" (Megargee, 1979). A number of short forms of the MMPI have been developed, but many researchers have concluded that the limitations of MMPI short forms are too great to use them for clinical prediction (Butcher & Hostetler, 1990).

Failure in Original Purpose. The original purpose of the MMPI was to develop an instrument that could effectively be used for the psychodiagnosis of patients, but as Graham (1990) wrote, "after a decade of clinical use and additional validity studies it became apparent that the MMPI was not adequately successful for its original purpose" (p.6). The

clinical scales were found not to be pure measures of the diagnostic categories they had been named after, as patients with particular diagnoses were scoring high on more than one scale, and many normal subjects were obtaining high scores on one or more scales. Much of the failure of MMPI to differentially diagnose patients has been attributed to psychometric shortcomings (eg. high interrelatedness of the subscales) of the instrument.

Researchers such as Meehl (1951) and Welsh (1948) began advocating that scale names, since they were not measuring pure types, were misleading and that scales should be labeled using numbers instead of psychodiagnosis labels (Hs as scale 1, D as scale 2, Hy as scale 3, Pd as scale 4, Mf as scale 5, Pa as scale 6, Pt as scale 7, Sc as scale 8, Ma as scale 9, and Si as scale 0). Further, Meehl suggested that empirical correlates of each scale needed to be examined, and that these correlates need not be generated solely for individual scales, but also for combinations of scales. Meehl suggested using high two point elevations of scales, and other configural patterns, to empirically determine MMPI correlates. Meehl, in effect, suggested a new purpose for the MMPI, a purpose based on the empirical study of individual MMPI scale and configural correlates.

Meehl's (1951) suggestion inspired a great number of studies investigating characteristics of persons with certain profile configurations. Actuarial investigation of the

MMPI flourished during the next three decades. Elaborate "cookbook" systems were developed using various populations, such as college students (Drake & Oetting, 1959), male veterans (Gilberstadt & Duker, 1965), adolescents (Marks & Seeman, 1963), and psychiatric patients (Gynther, Altman, & Sletten, 1973; Marks & Seeman, 1963). These systems would define typical profile configurations and then describe the empirical correlates of such configurations.

Greene (1980) and Mauger (1980) described some of the problems of these early actuarial systems. These systems often contained extremely specific and "cumbersome configurational rules" (Greene, 1980, p.938). In some settings only one fourth to one fifth of the profiles were able to be categorized into strictly defined profile configuration types. There was difficulty generalizing the results of some systems to populations other than the one the system had been based.

Contemporary interpretive systems for the MMPI have "improved and broadened" (Greene, 1980, p.937), both in terms of classification rules and generalizability across populations. Such current systems are typified by books by Duckworth and Anderson (1986), Graham (1987, 1990), Greene (1980, 1990), Lachar (1974), and Marks, Seeman and Haller (1974). These books report empirical correlates of single scale elevations, two and three point elevations, low scores, and validity scale configuration. An elevated score

is typically defined as a scale with a T score above 70. Each of these books emphasizes the importance of not considering scale scores in isolation, but in relationship to each other on any one profile.

The interpretation of any one scale will be modified by the other scales that are elevated along with it ... the second and third highest scales may modify the highest scale in ways that change rather markedly what is said about an individual (Duckworth & Anderson, 1986, p.9).

For example, an elevation of scale 4 and scale 9 is interpreted in these systems differently from an elevation of scale 4 and scale 3. Greene (1980) discusses not only the importance of high points, but low points as well. He writes:

Frequently, clinicians who are learning to interpret the MMPI focus solely on the analysis of the high point pair or some other feature of the profile and miss other significant data, usually the low points in the profile. This is unfortunate since there are several scales for which low points are particularly significant (Greene, 1980, p. 145).

For example, a high scale 2 and a low scale 9 can be interpreted as indicating greater depression than just a high scale 2.

Within the past 25 years the role of the MMPI has been further expanded by applying it not only to pathological populations, but researching its use with normal populations (Duckworth & Anderson, 1986; Duckworth & Barley, 1988; Graham, 1990; Graham & McCord, 1985; Kuncie & Anderson, 1976, 1984). Interpretive systems have been developed for MMPI

use in personnel selection (Graham, 1990; Graham & McCord, 1985). Duckworth and Anderson's (1986) interpretation manual is devoted to presenting MMPI information for persons who "are basically normal" (p.iii), defined as persons who are not hospitalized and do not require extreme alteration of personality. One of the main goals of the book is to identify positive aspects of moderate elevations on MMPI scales, defined as T scores in the range of 60 to 70, thereby making the MMPI a broader personality assessment instrument and utilizing "its richness and complexity to the fullest" (Duckworth & Anderson, 1986, p.iii). Profiles in which all of the basic scales have a T score below 70 are termed "within normal limits" (WNL) profiles (Duckworth & Barley, 1988). A number of guidelines have been developed for indicating if score elevations may merit more positive interpretations (Duckworth & Anderson, 1986).

The MMPI failed in its original purpose as a psychodiagnostic measure. It retained its usefulness through a wealth of validity studies looking at the empirical correlates of configural patterns of MMPI profiles. Contemporary approaches to interpretation emphasize analysis of profile configuration, both within severely disturbed and normal populations. The validity studies and the interpretation systems resulting from them form the basis of contemporary MMPI utility. As Groth-Marnat (1984) writes, "it is primarily because of these extensive validity studies that the

MMPI continues to be used" (p.256).

Psychometric Limitations. The MMPI has been described as a "psychometric nightmare" (Rodgers, 1972). The basic scales of the MMPI appear to lack adequate reliability. The varied content of MMPI items has led to low split half reliability, .05 to .96, and low internal consistency, -.05 to .96 (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975). Test-retest reliability for normals ranged from .29 to .92 for 2 week intervals and from .13 to .73 for 1 year intervals. Test-retest reliability for psychiatric cases ranged from .61 to .94 for 2 week intervals and from .22 to .72 for 1 year intervals (Graham 1990). Anastasi (1988) noted that certain MMPI scales, such as scale 2, assess behavior that is likely to vary over time, leading to low test-retest reliability.

A number of the scales of the MMPI have an imbalance in the ratio of items that are scored if answered true versus items that are scored if answered false. For example, 29 of the 30 items on the K scale are scored if answered false, and scales 7, 8, and 9 have a 3:1 ratio of true to false items. This led researchers such as Edwards (1957, 1964) and Messick and Jackson (1961) to argue that MMPI results do not reflect psychological traits as much as generalized test taking attitudes such as acquiescence. Koss (1979) in a review of the literature on this issue concluded that response sets can and do exist, but that a subject's tendency

to respond to content is stronger.

Many of the subscales are highly intercorrelated (eg. correlation of scale 7 and scale 8 has been reported as high as .87) (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975). The interrelatedness of the scales has also been demonstrated by factor analysis studies indicating that only two main factors account for most of the variance (Dahlstrom, Welsh, Dahlstrom, 1972, 1975). Scale interrelatedness is most likely a product of MMPI scale construction. Items were chosen for a particular scale because they discriminated between normal and clinical subjects, not because they discriminated between various classifications of clinical subjects. Information on the clinical groups themselves is scarce (Colligan, Osborne, Swenson, & Offord, 1983). Typically, no demographic information was reported for the clinical groups (Dahlstrom & Dahlstrom, 1980). The numbers of some of these groups were quite small. For example, a clinical sample of 20 was used to develop scale 7 (McKinley & Hathaway, 1942). No information on the sample size of the clinical sample used to develop scale 4 is available (Colligan, et al., 1983; McKinley & Hathaway, 1944). Hathaway and McKinley claimed that clinical subjects were representative of pure diagnostic categories, but no interrater reliability has been provided concerning the clinical subjects diagnosis (Dahlstrom & Dahlstrom, 1980). There is also a large amount of item overlap between scales. For example, of the 78

items that make up scale 8, only 16 items are unique to that scale only. As Groth-Marnat (1984) wrote, "all of this strongly suggests that there is a high degree of redundancy regarding what the scales measure"(p. 255). The high scale interrelatedness implies that the MMPI scales do not measure unidimensional traits.

Problems with reliability, scale development, the imbalance between true and false items, and item overlap have resulted in the MMPI being termed a "psychometric nightmare." In spite of its psychometric shortcomings, the MMPI has remained popular due to the extensive validity studies of it. It has been these extensive validity studies which, as Kaplan and Saccuzzo (1982) wrote, "have allowed the MMPI to survive despite its poor showing when subjected to rigorous psychometric analysis" (p.328).

Appropriateness of Normative Group. Not unrelated to psychometric considerations is concern about the use of the Minnesota normal group as the basis for the MMPI norms, both in terms of the representativeness of the normative group and in terms of how the normative distributions were calculated. The normative group used was not based on all 724 subjects of the original Minnesota normal group. Subjects who had incomplete records or faulty background information were eventually dropped from the normative group in a "purifying" of the group. (Hathaway & Briggs, 1957). It is this purified group that the T score distributions are

based.

The Minnesota normative group was all white, typically married, residing in a rural area, about 35 years of age, and had about 8 years of education (Dahlstrom, Welsh, & Dahlstrom, 1972). The demographic representativeness of the Minnesota normal group has long been an issue (Greene, 1980; Gynther & Greene, 1989; Pritchard & Rosenblatt, 1980). "It has been demonstrated that age, sex, race, place of residence, intelligence, education, and socioeconomic status are all related to MMPI scales" (Groth-Marnat, 1984, p.259). Graham (1979, 1987) advocated using adjustments to the scores of certain scales depending on the age, race, and educational level of the subject. For example, Graham indicated that college students score, on the average, 5 to 10 T score points on scales 8 and 9 as compared to the general population. Gynther (1989) in the latest of a number of reviews on the impact of demographic variables on MMPI scores reviewed the literature on the scores of Blacks on the MMPI.

Demographic representativeness is just part of the concern. Others have written that the society has changed as a whole since the early 1940s, resulting in normal subjects today responding in different patterns than the original Minnesota normative group to items on the MMPI. Colli-gan et al. (1983) wrote that "obviously there have been dramatic changes in our society in the past 45 years, but

there has not been a systematic reexamination of the original normative data" (p.xiv). They pointed to changes in the standard of living, moral and ethical views, the nuclear family, and urbanization as just a few of the many ways in which the society has changed.

Another concern about the appropriateness of the Minnesota normative group is the difference in the instructional set of the MMPI used when the normative group was tested versus the current instructional set used with the MMPI. Hathaway and McKinley (1940) allowed subjects to not answer items, which a number of persons in the normative group did. Current instructions (Hathaway & McKinley, 1983) encourage subjects to answer all items. Subjects encouraged to answer all items are likely to endorse more items scored on the basic scales, leading to higher scores. Parenthetically, some have objected to the content of a number of the items themselves, seeing them as unnecessarily intrusive or offensive (Butcher & Tellegen, 1966). Concerns about changes in society, the demographic representativeness of the Minnesota group, and the difference in instructional sets have led to concern about the appropriateness of the Minnesota normative group.

Not only have the characteristics of the Minnesota normative group been scrutinized, but the way in which their scores were transformed into T scores has also been questioned. Colligan, Osborne, and Offord (1980) called atten-

tion to how the T score distributions for the normative sample were developed. When referring to T score distributions researchers and clinicians traditionally assume that the scores are normally distributed, with $M=50$ and $SD=10$. One would assume that the traditional point of significance on the MMPI, a T score of 70, would fall at the 97.7 percentile of the distribution. The scales of the MMPI are not normally distributed, however. The raw score distributions for these scales are skewed, usually positively. The T score distributions of the MMPI are based on a simple linear transformation of raw scores to T scores using the following equation: $T = 50 + 10 (\underline{x}_i - \underline{x}) / \underline{s}$ ($\underline{x}_i$ = raw score to be transformed, $\underline{x}$ = mean of the distribution, $\underline{s}$ = standard deviation of the distribution). The positively skewed distributions of raw scores are maintained by using a linear transformation of raw scores to T scores. Colligan et al. (1980) analyzed the T score distributions and found statistically significant degrees of skewness and kurtosis (degree of flatness or spike) for most of the distributions of scores for the validity and clinical scales. An example of the impact of skewness can be seen in the distribution of scale 2 for males. With a normal distribution 2.3% of the scores would be expected to fall above a T score of 70. However, with the skewed, male score distributions on scale 2, 8.0% of the males fall above a T score of 70. Positively skewed distributions were noted by Colligan et al. (1980) on

scales 1, 2, 7, 8, and 9 for both males and females. With various scale score distributions "scale elevations of similar degree in terms of T scores ... do not have equivalent meaning in terms of deviations from the mean (that is percentile rank)" (Colligan et al., 1983, p. xiv). Colligan et al. (1980) argued that these discrepancies could lead to the erroneous assignment of code types with high scale scores not necessarily indicating the greatest deviation from the mean. Colligan et al. (1983) developed a new set of normalized T scores based on a sample of 1,711 subjects in 1983. The potential impact of changing the distributions of T scores is evident when Colligan et al. (1983) wrote that "the change to normalized T scores implies the destruction or invalidation of years of research" (p.285).

Problems with the characteristics of the Minnesota normative group and how T scores were derived has lead some to argue that the original standardization sample should be used as a nonnormative, fixed reference group, thereby circumventing the problems it has as a normative sample (Anastasi, 1988; Groth-Marnat, 1984). Again, it is the validity studies of the MMPI that are invoked as the reason for its continued use. As Groth-Marnat (1984) wrote, "the subsequent research performed on empirical code profiles is far more useful than the original normative scale comparisons" (p. 258).

Whether the MMPI has failed its original purpose, has

poor psychometric properties, or has used an inappropriate normative sample, it has been the power of the numerous validity studies which has allowed the MMPI to continue to be the "sin qua non in the psychologists armamentarium of psychometric aids" (King, 1978, p.938).

THE MMPI-2: The Changes

In response to some of the problems of the MMPI cited above, particularly those dealing with the normative group, the MMPI was revised, and a new version, the MMPI-2, was released in 1989 by Butcher, Dahlstrom, Graham, Tellegen and Kaemmer. Many changes were made in the revision of the MMPI. Three major changes that could influence the scores and configural patterns of MMPI-2 are: 1) the deletion, revision, and addition of items, 2) the development of a contemporary normative sample, and 3) the use of a new method of distributing T scores. These major changes subsume a number of secondary changes such as the revision, deletion, and addition of supplemental scales, the use of a T score of 65 as a cutoff for defining an elevation, the reordering of items, and the requirement of an eighth grade reading level for subjects. Butcher et al. (1989) asserted that despite these changes the vast amount of validity studies done on the MMPI are applicable to MMPI-2. "The results of years of research on the MMPI component scales and their patterns of interrelationship are available to guide interpretation of MMPI-2 test scores" (Butcher et al.,

1989, p.1)). The following sections will explore, in detail, the three major changes in the MMPI-2 that could affect score and configural patterns. Accompanying secondary changes will also be discussed in each section.

It is important to note that all empirical information to date about the MMPI-2 was based upon data gathered using the MMPI-AX. The MMPI-AX was an experimental form used for restandardizing the MMPI. The MMPI-AX contained the original 550 items of the MMPI, less the 16 duplicate items on the MMPI used only to help with machine scoring. These 550 items contained the new wording revisions planned for MMPI-2. The MMPI-AX also contained an additional 154 provisional items, mainly designed to assess new content areas, bringing the total number of items to 704. All 567 items of the MMPI-2 were on MMPI-AX. All analyses of the "MMPI-2" have been based on an extraction of only those items on the MMPI-AX that are on the MMPI-2.

Changes in Items. Because of problems such as sexist wording, unfamiliar idiomatic expressions, and objectionable items, 71 of the 386 items of the MMPI basic scales were revised. Of the changes made, 13 items were deleted and 68 items were reworded. Wording changes were classified into four types: 1) elimination of sexist wording, 2) modernization of idioms and usage, 3) grammatical clarification, and 4) simplifications (Ben-Porath & Butcher, 1989; Butcher et al., 1989).

Butcher et al. (1989) reported that the psychometric effect of the wording changes have been tested by Ben-Porath and Butcher (1989b). Ben-Porath and Butcher (1989b) administered the MMPI twice to half of a group of 377 college undergraduates. They administered the MMPI and MMPI-2 to the other half of this group. Results showed significant differences by gender in the responses to 9 of the rewritten items. None of these significant differences were consistent across gender. No item-scale correlations changed significantly as a result of the rewritten items.

Specific information about items that were not a part of the basic scales was not given by Butcher et al. (1989), however, a number of completely new items did replace a number of these items. These new items were mainly designed to assess new content areas for supplemental and content scales.

As a result of item changes, commonly scored supplemental and content scales were revised, deleted, and added. Eight of the commonly used supplemental scales (A, R, MAC-R, Es, Do, Re, O-H, Mt) were revised for the MMPI-2. These scales had a number of items deleted (0, 3, 16, 4, 3, 2, 3, and 2, respectfully). The 4 items deleted on the MAC-R scale were replaced by 4 new items. No information was given about wording changes of these scales. Changes in items of the Harris-Lingoes subscales were similar to the changes of the items of basic scales, as the Harris-Lingoes

subscales are content subscales of clinical scales 2, 3, 4, 6, 8, and 9. In the Harris-Lingoes subscales of the MMPI some items in a particular subscale were not items found on its corresponding clinical scale but were from other scales. These items were deleted on the Harris-Lingoes subscales of the MMPI-2. In spite of changes in the items of the supplemental scales and Harris-Lingoes subscales Butcher et al. (1989) and Graham (1990) implied, by using MMPI research on these scales as the basis of interpretive statements for MMPI-2, that the scales are comparable between MMPI and MMPI-2. Graham (1990) wrote that "research data from the original MMPI were utilized to generate interpretive statements. Because these scales are essentially the same in the two versions of the test, this approach seems appropriate" (p.142).

The Wiener-Harmon subtle obvious scales were carried over with alterations to the MMPI-2, but Butcher et al. (1989) did not endorse their use, as they asserted that there is no consensus on the use and interpretation of these scales, either on MMPI or MMPI-2. Critical items lists developed by Koss and Butcher (1973) and Lachar and Wrobel (1979) were modified and are available to use on MMPI-2. Because of item deletions, the Wiggins content scales were dropped from the MMPI-2, as were some of the commonly scored supplemental scales. Using items specifically developed to assess new content areas, new content scales and some new

supplemental scales were developed for MMPI-2.

An analysis of the reading difficulty of the MMPI-2 items was performed. Based on this analysis of items changes, Butcher et al. (1989) recommended that subjects be able to read at an eighth grade level. A sixth grade reading level had been recommended for MMPI (Hathaway & McKinley, 1983). Also, items on the MMPI-2 were ordered such that all items of the basic scales were located within the first 370 items of the inventory. In the group form of the MMPI items of the basic scale are located throughout the inventory.

Contemporary Normative Sample. The MMPI-2 normative sample consisted of 1138 males and 1462 females ranging in age from 18-90 years. Subjects were recruited from seven states across the country using a variety of recruitment methods. Butcher et al. (1989) wrote that a forthcoming report will show no effect of recruitment method or location on MMPI-2 scores. Subjects from a federal Indian reservation and the military were also proportionally added to the normative sample.

Butcher et al. (1989) attempted to create a normative sample which was demographically similar to the 1980 census on such variables as age, ethnicity, marital status, educational level, and socioeconomic status. Butcher et al. reported achieving this similarity with a few exceptions. Hispanics and Asians were underrepresented in the MMPI-2

normative sample, as were persons of lower educational attainment and lower socioeconomic status. Butcher et al. responded to these shortcomings by asserting that educational level and socioeconomic status do not greatly affect MMPI-2 scores, and that the normative sample is actually closer on these demographic variables to the typical subject taking the MMPI-2. No empirical evidence was given to substantiate either of these claims.

Originally about 2900 subjects were recruited for the normative sample. This group was cut to 2600 subjects because of incomplete records or invalid profiles. Invalid profiles were defined as ones which had a score of 40 or more on the ? scale or a raw score of 20 or more on scale F or scale Fb on the MMPI-AX. Scale Fb, or the back F scale, was created for use with the MMPI-2. It was developed in a manner similar to the F scale, containing items that are responded to by less than 10% of normal subjects. These items are located at the back of the inventory, as F scale items are located only in the first 370 items.

As expected, the contemporary normative group responded to items differently than did the Minnesota normative group (Colligan et al., 1983). In general, the contemporary normative group responded to more items in the scored direction. In general, Butcher et al. (1989) assert that subjects will score 5 T score points lower on each of the MMPI-2 scales than on the MMPI scales due to the pattern of

responding of the normative group. The contemporary normative group tended to respond to more items in the scored direction than the original MMPI normative sample. Because of the different patterns of responding by the contemporary normative group, Butcher et al. recommended that a T score of 65, as opposed to the traditional T score of 70, be used as the cutoff for defining an elevated scale.

Uniform T Distribution. Like the distribution of raw scores for the MMPI, the normative distributions of raw scores for the MMPI-2 vary from scale to scale in skewness and kurtosis. In order to make the distribution of T scores consistent across scales, Butcher et al. (1989) developed the uniform T distribution. The uniform T distribution was designed for clinical scales 1-4 and 6-9. Conceptually, the uniform T distribution was developed by converting the variously skewed raw score distributions of the 8 clinical scales into linear T distributions and then "averaging" these linear T distributions together. This resulted in an average T score distribution of the 8 clinical scales called the "uniform T" distribution. This distribution, although developed on the 8 clinical scales, is used as the distribution for the newly developed content scales as well. Scales L, F, K, 5, and 0 were not included because the distributions of these scales were so different from the other scales. Graphs of the various linear distribution of all of the basic scales are presented in Butcher, Graham, Dahl-

strom, and Bowman (1990).

Statistically, Butcher et al. (1989) did not include the specific calculations used to average the distributions. Butcher et al. also did not provide a specific description of the characteristics of the uniform T distribution. Graham (1990) did provide some information about the distribution. The uniform T distribution is not a normal distribution with $M=50$ and $SD=10$. The uniform T distribution is positively skewed. Graham (1990) reported that a T score of 70 is in the 96th percentile, a T score of 50 is in the 55th percentile, and a T score of 30 is in the 1st percentile of the distribution.

Changes in items, the use of contemporary normative group, and the use of a uniform T distribution are three of the significant changes made in MMPI-2. It has been changes such as these which have led investigators to question the consistency between MMPI and MMPI-2.

Empirical Investigation of Consistency

Many concerns about the use of MMPI-2 have begun to surface in the literature (Adler, 1990; Butcher, Dahlstrom, Graham, Tellegen, Kaemmer, 1990; Edwards & Weissman, 1990). Among these concerns is whether there is consistency between MMPI and MMPI-2 such that, as Graham (1990) claimed, "the research base that supported the validity of the original MMPI also supports the validity of the MMPI-2" (p. 180). There has been anecdotal evidence concerning consistency.

Duckworth (personal communication, June 28, 1990) indicated that for the approximately 80 subjects to whom she had given the MMPI and MMPI-2 there appeared to be a number of disparities between scores. Graham (1990) on the other hand, reported in his book that he has found MMPI and MMPI-2 scales to all correlate at .98 or better, but gives no details on how he arrived at these numbers.

There are two systematic, published studies concerning the consistency of scores and configural patterns between MMPI and MMPI-2, one by Ben-Porath and Butcher (1989a) and one by Butcher et al. (1989). Each of these studies will be described separately.

Ben-Porath and Butcher (1989a). Ben-Porath and Butcher (1989a) recruited 403 undergraduate college students to participate in their study. Of these 403 students, 6 failed to complete the entire study and 20 students were excluded because of scoring greater than 15 on scale ? or having a raw score greater than 20 on scale F or Fb. The responses of 377 students, 199 women and 178 men, were used.

About half of the subjects were administered the MMPI twice, the original to original condition (O-O). The group form of the MMPI was used. The other half received the MMPI and MMPI-2, the original to revised condition (O-R); the order of administering these two instruments was counterbalanced. The average interval between administration of the two inventories was 9 days, with a range of 1 to 2 weeks.

Test-retest correlations of the basic scales and 8 supplemental scales shared by MMPI and MMPI-2 indicated, for the most part, that there was no significant difference in scale scores between the O-O and O-R conditions at the .01 level of significance. In other words, the correlations of the scales of the MMPI and MMPI-2 were not significantly different from the test-retest correlations which normally occur when MMPI is administered twice. There were scores on two scales, F and Es, in which women had significantly lower correlation for the O-R condition as compared to the O-O condition.

Ben-Porath and Butcher (1989a) also conducted a set of analyses to look at configural consistency between MMPI and MMPI-2. They looked at changes in five profile characteristics between the O-O and O-R condition. They found that the percentage of profiles with the same high point, same two highest points, same number of elevated scales, and the same classification as a within normal limits profiles did not differ significantly between the O-O condition and the O-R condition. Elevated scales were defined as scales with a T score greater than 70 for the MMPI and greater than 65 for the MMPI-2. An analysis of the D^2 statistic (Wiggins, 1973) also showed no difference between changes in overall profile configuration of the O-O condition and the O-R condition. The D^2 statistic is computed by summing the squares of the differences between the two administrations of the scores of

scales 1-4 and 6-9 for each subject. Five points were added to the T score of scales of the MMPI-2 "to reflect the change in the criterion for elevation from 70 in the MMPI to 65 in the MMPI-2" (Ben-Porath & Butcher, 1989a, p. 347).

Ben-Porath and Butcher (1989a) concluded from these results that there is substantial congruence in scores and profiles between MMPI and MMPI-2. Further, they asserted that these findings may generalize beyond the college student population, by citing a study by Butcher, Graham, Dahlstrom, and Bowman (1990) indicating that college student perform similarly to the normative group on the MMPI-2. Butcher, Graham, Dahlstrom, and Bowman (1990) surveyed 1312 college students from four universities and found that for most scales, college student norms were within 1-3 T score points of the contemporary normative sample. On scales 7, 8, and 9 T scores ranged from 4-7 points higher for college students.

The study by Ben-Porath and Butcher (1989a) is limited in a number of ways. First, it made use of MMPI-2 results extracted from MMPI-AX. The order of MMPI-2 items is different from the order of these items on MMPI-AX. The first 550 items of the MMPI-AX are the 550 original MMPI items. These 550 items on the MMPI-AX contain the wording changes for the MMPI-2, however being exposed to these items in a similar order across the MMPI and MMPI-AX may have enhanced consistency in responding. The MMPI-AX is also a longer

instrument. Butcher et al. (1990) reported that subjects, given periodic breaks, took approximately 3 hours to complete the MMPI-AX. These differences could have an impact on scores (Megargee, 1979).

Second, Ben-Porath and Butcher (1989a) made use of test-retest correlations to conclude that scores between the MMPI and MMPI-2 are consistent. Although these correlations are helpful in analyzing consistency, it must be acknowledged, and was not by Ben-Porath and Butcher, that this analysis only accounts for potential differences due to item changes. Ben-Porath and Butcher made use of raw scores in their test-retest correlations. These scores were not standardized with the new normative group using the uniform T distribution. Even if standardized scores had been used by Ben-Porath and Butcher, a correlational analysis does not account for score changes due to a new normative group and the uniform T scores. When raw scores are converted to standardized scores the transformation is linear, and as such, no score changes in its relative rank to other scores within its group. Pearson product moment correlations are sensitive to relative ranking only, and because relative rank does not change when raw scores are standardized, test-retest correlations using raw scores are no different than ones using T scores (Cattell, 1949, Cohen & Cohen, 1983). Correlational analyses are not sensitive to changes caused by the new normative group or the uniform T distribu-

tion. Evidence, reviewed below, by Butcher et al. (1989) indicated that the new normative group and the use of the uniform T distribution could have an effect on scores. The use of correlational analyses by Ben-Porath and Butcher did assess the variance in scores between MMPI and MMPI-2 due to item changes (ie. the deletion of items, wording changes, the presence of new items); it did not, however, assess variance due to the contemporary normative group or uniform T distribution.

Third, the configural analyses that Ben-Porath and Butcher (1989a) used were sensitive to changes due to a new normative sample and the use of the uniform T distribution as well as items changes. These configural analyses were based on only 8 of the 10 clinical scales of the MMPI, however, as scales 5 and scale 0 were omitted from the analysis. These scales might possibly have been omitted because they were developed separately from the other 8 scales when the MMPI was originally developed, although no formal rationale for the omission was reported in the study. The omission of these scales in the analysis is important, however, as all major interpretive manuals include sections on scale 5 and scale 0 interpretation, both as individual scales and in relations to other scales (Duckworth & Anderson, 1986; Graham, 1987, 1990; Greene, 1980, 1990; Marks, Seeman, & Haller, 1874, Lachar, 1974).

Fourth, the types of configural analyses that Ben-

Porath and Butcher (1989a) used seem appropriate, with the exception of D^2 which may have been a weak indicator of configural stability. Within group variance is so great with this indicator that between group variance is difficult to detect. For example, in the O-O condition for males $M=501.62$ and in the O-R condition $M=619.54$, suggestive of possible differences between conditions. However $SD=421.75$ and $SD=650.93$ in each of these conditions respectively, resulting in a t test showing nonsignificant differences. A further problem with the D^2 statistic, as indicated by the reported means and standard deviations reported, is that its distribution is not normal; it is highly skewed. Problems of exaggerated variance and a skewed distribution with the D^2 statistic were described by Cronbach and Gleser (1953) who recommend the use of the D statistic.

Fifth, Ben-Porath and Butcher's (1989a) configural analyses were limited in scope. Their analyses failed to assess configural characteristics described in major interpretive manuals such as three point elevations, low scores, and validity scale configurations (Duckworth & Anderson, 1986; Graham, 1987, 1990; Greene, 1980, 1990; Marks, Seeman, & Haller, 1874, Lachar, 1974). Ben-Porath and Butcher's analysis also failed to examine potential changes in relative elevations of T scores on the MMPI as compared to the MMPI-2. Butcher et al. (1989) asserted that the new cutoff on the MMPI-2 be reduced from 70 to 65, as T scores on each

of the basic scales of the MMPI-2 should be about 5 points lower than on the MMPI. Ben-Porath and Butcher do not report mean scores for the basic scales of the MMPI and MMPI-2 and so the question of whether the overall configuration of MMPI-2 scores are, on the average, 5 points lower than on the MMPI remains unanswered.

Butcher et al. (1989). Empirical data cited by Butcher et al. (1989) in the MMPI-2 manual indicated that scores on the MMPI-2 may be affected by the new normative sample and the use of the uniform T distribution. Butcher et al. reported data collected on 425 psychiatric inpatients in hospitals in Ohio and Minnesota using the MMPI-2. Raw scores on the clinical scales of the MMPI-2 were plotted on the new uniform T distribution (scales L, F, K ,0 and 5 were excluded) based on the contemporary group, the linear T distributions based on the contemporary normative group, and the linear T distributions based on the Minnesota normative group. Adjustments were made in the raw scores when plotted with the Minnesota normative group because of the 16 items that were deleted from the basic scales when the MMPI-2 was developed.

A number of differences in scores and scale configuration occurred when the raw scores were normed on the various distributions. For example, when MMPI-2 raw scores for males were normed using the Minnesota normative group, 44% had scale 8 as their highest elevation. Only 29% of the

males had scale 8 as their highest scale when their raw scores were normed using the linear or uniform T distributions of the contemporary normative group. This difference, from 44% to 29% demonstrates the potential impact of the contemporary normative group on relative T score elevations. For males 3.9% and 4.7% had scale 7 as their highest elevation when normed on the Minnesota normative group and the linear T distribution based on the contemporary normative group respectively, but when these raw scores were plotted on the uniform T distribution, 8.6% of the males had scale 7 as their highest scale. This difference, from 3.9% and 4.7% to 8.6% demonstrates the potential impact of using the uniform T distribution as opposed to a linear T distribution. No test was conducted to see if these percentages were significantly different, but they are suggestive of the impact of the new normative group and the use of the uniform T distribution on scores.

Butcher et al. (1989) also analyzed two point code agreement on this sample based on whether the raw scores were normed using the linear T distribution of the Minnesota normative group or the uniform T distribution of the contemporary normative group. Irrespective of the cutoff used to define a two point elevation (ie. both T scores over 70, 80, 90) about one-third of the two point codes did not agree when normed on the two different distributions. Common two point code types such as 6/2, 6/4, 9/6 and 6/8 ranged from

45-56% agreement between the linear T distribution of the Minnesota normative group and the uniform T distribution of the contemporary normative group.

This study showed, by holding item content constant as subjects took only the MMPI-2, that the normative group and the method of T score distribution, linear versus uniform, can influence scale scores and profile configurations. These results were confounded by the fact that the Minnesota normative group distributions were not based upon responses to the rewritten items of the MMPI-2. Results were further limited by the use of MMPI-AX in the study. The results were suggestive, however, of substantial changes in scores and configurations from MMPI to MMPI-2 because of the contemporary normative group and the use of a uniform T distribution. Butcher et al. (1989) admitted that there is a difference in profile configuration depending upon the method of T score distribution employed as they wrote, "comparisons of the traditional linear and the uniform scaling methods show that the prevalence of high point codes (and high point pairs) invoking scales 1, 3, 6, and 7 is significantly increased by the norm of the uniform T values" (p.36).

Summary of Empirical Findings. When both the MMPI and MMPI-2 are given there are four potential sources of variation in scores between the two inventories. These sources of variation are: A) normal test-retest variance, B) vari-

ance due to item changes, C) variance due to differential responding patterns of the new normative group, and D) variance due to the use of uniform T distributions. In the Ben-Porath and Butcher (1989a) study, test-retest variance and variance due to item changes were considered by the analyses of raw scores. Ben-Porath and Butcher found some effect due to item changes in two scales, but no effect on overall configuration. Their analyses of individual scale scores did not consider variance due to the new normative group or the use of a uniform T distribution, but their analysis of configural patterns did. Their analysis of configural pattern was limited, however. The study described by Butcher et al. (1989) did consider the variance due to the new normative group and the use of the uniform T distribution. It did not consider test-retest variance or variance due to item changes, since only the MMPI-2 was given. Results from this study gave indications that the new normative group and the use of the uniform T distribution may substantially impact scores and configurations of the MMPI-2 as compared to the MMPI. Taken together these studies suggest that scores and configural pattern may not be consistent between MMPI and MMPI-2. Further, they suggest that differences could be due to: 1) item changes, 2) the use of a contemporary normative group, and/or 3) the use of a uniform T distribution.

The Present Study

The present study was designed to assess differences between the MMPI and MMPI-2 with respect to scores of the individual scales, and with respect to configural patterns of the basic scales.

The present study seeks to determine if scores on the basic scales of the MMPI differ significantly from the scores of the basic scales of the MMPI-2. This was investigated, similar to the Ben-Porath and Butcher (1989a) design, by dividing a sample of college undergraduates into two groups, one receiving the MMPI twice (O-O condition) and the other receiving the MMPI and MMPI-2 (O-R condition). Test-retest correlations of MMPI scores for the basic scales were calculated (O-O condition). These were compared to the correlations of MMPI basic scales with MMPI-2 basic scales (O-R condition). Significantly lower correlations in the O-R condition, are used to indicate that scores between MMPI and MMPI-2 are different.

The present study assessed changes in scores of the frequently used supplemental scales common to both MMPI and MMPI-2. It also assessed changes in scores of the Harris-Lingoes subscales. No data have yet been reported on the consistency of Harris-Lingoes subscale scores between MMPI and MMPI-2. Because of the number of significance tests conducted, resulting in greater chances of investigation-wise, Type I error, analyses of the supplementary scales and

Harris-Lingoes subscales are considered exploratory.

In addition to these analyses of changes in individual scales, changes in how scales interrelate with each other were also assessed by comparing the interscale correlations of the MMPI with the interscale correlations of the MMPI-2 in the O-R condition. Significance tests were used to determine if MMPI interscale correlations significantly differ from MMPI-2 interscale correlations. Because of the number of significance tests conducted, resulting in greater chances of investigationwise, Type I error, analyses of interscale correlations are considered exploratory. All analyses described above, because they are correlational, only assess differences between the MMPI and MMPI-2 due to item changes.

Configural analyses, which account for all major differences between the MMPI and MMPI-2, were also utilized. Like Ben-Porath and Butcher (1989a) percentages of profiles with the same high point, same two highest points, same number of elevated scales, and the same classification as a within normal limits profile were calculated within the O-O₂ and O-R condition. D₂ was also be calculated. Significance tests are used to determine if the two conditions differ on these characteristics.

Configural analyses in the present study were conducted on all 10 of the clinical scales as well as on just the 8 scales used by Ben-Porath and Butcher (1989a). Configural

analyses were also expanded to include analysis of three point elevations, low scores, validity scale configuration, and "well defined" code types and the D statistic. Well defined code type is included because of Graham's (1987, 1990) assertion that well defined codes are the only type of code for which code type interpretation is relevant.

In addition to these configural analyses, T score means on the basic scales of the MMPI were compared to T score means of the MMPI-2 to assess the Butcher et al. (1989) recommendation that the T score cutoff be lowered to 65. In this analysis, 5 T score points were added to all the basic scales of the MMPI-2. Mean T scores between the basic scales of the MMPI and MMPI-2 should be similar after the 5 point correction. If differences exist, these differences suggest scales causing potential configural shifts between the MMPI and MMPI-2.

Separate analyses for males and females were conducted on all analyses of scores and configural patterns. All analyses in this study tested the null hypothesis that there are no significant differences between the scores and configurations of MMPI and MMPI-2.

Method

Subjects

Four hundred students (200 males and 200 females) from introductory psychology courses at The Ohio State University received class credit for participation during the fall quarter of the 1990-1991 academic year. Initially, 444 students volunteered to participate, however 44 failed to return for the second of two required testing sessions. A low attrition rate of 10% is attributed largely to the requirement that subjects attend both testing sessions to receive credit. Half of the students were assigned to the O-O condition and half to the O-R condition. Each condition consisted of 100 males and 100 females. Beyond assuring a gender balance between conditions, assignment to conditions was random.

Decision rules for classifying profiles as invalid were the same as those used by Ben-Porath and Butcher (1989a). Profiles containing raw scores of greater than 15 on scale ? or greater than 20 on scales F or Fb were considered invalid and were not used in the study. These exclusionary criteria applied to 32 profiles leaving a total of 86 males and 94 females in the O-O condition and 92 males and 96 females in the O-R condition. Subjects' ages ranged from 16 to 45 years, with a mean of 19.5 years.

Instruments

Demographic Questionnaire. Demographic information was collected on subjects' gender, age, ethnicity, class standing, academic major, marital status, highest level of education attained by each parent, and occupation of each parent (Appendix A).

MMPI. The group form of the MMPI was used. The following scales of the MMPI were scored: the basic scales, the 8 frequently used supplemental scales shared by MMPI and MMPI-2, and the Harris-Lingoes subscales. A listing of all scale names and abbreviations for the MMPI and MMPI-2 can be found in Appendix B. Psychometric information about these scales of the MMPI is outlined below.

MMPI Basic Scales. Of the 4 validity and 10 clinical scales, the ? scale is seldom included in the statistical analyses or analyses of configural patterns. The ? scale is most often used simply, as a binary variable, to determine profile validity. Aside from the ? scale, the number of items of the basic scales range from 15 to 78. Internal consistency correlations of the basic scales ranged from -.05 to .96. Factor analyses studies of the MMPI have indicated that two factors account for most of the variance. Many of the scales are highly intercorrelated (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975). Test-retest reliability for normals ranged from .29 to .92 for 1-2 week intervals (Graham, 1987). The validity of the basic scales of the

MMPI has been well established with numerous studies examining empirical correlates of the scales (Duckworth & Anderson, 1986; Graham, 1990, 1987; Greene, 1990, 1980; Lachar, 1974; Marks, Seeman, & Haller, 1974).

MMPI Supplemental Scales. The 8 frequently used supplemental scales shared by the MMPI and MMPI-2 contain from 28 to 68 items on each scale. Scales A and R were developed using factor analytic methods. All other scales were developed using empirical criterion keying. Internal consistency correlations ranged from .56 to .91 but are not available for the MAC-R and Mt scales (Graham, 1987). Six week test-retest correlations ranged from .56 to .90 for college students (Moreland, 1985). In Graham's (1987) review of the validity of the supplemental scales he stated that the "validity of many of these scales has not been adequately demonstrated" (p. 194). Graham indicated that scales A, R, Es, and MAC have more validity data than scales O-H, Do, Re, and Mt.

MMPI Harris-Lingoes Subscales. The Harris-Lingoes subscales contain 28 rationally constructed subscales based on an analysis of the content of clinical scales 2, 3, 4, 6, 8, and 9. The number of items on the subscales range from 6 to 32. Kuder Richardson 21 internal consistency coefficients were above .45 for all but 6 of the subscales (Gocka, 1965). There is item overlap between subscale items resulting in some high interscale correlations (Harris & Lingoes,

1968). Six week test-retest correlations ranged from .46 to .86 for college students (Moreland, 1985). Validity research of the Harris-Lingoes subscales of the MMPI was summarized by Graham (1987) and, although limited, is generally supportive of the dimensions developed by Harris and Lingoes (1955, 1968).

MMPI-2. The following scales of the MMPI-2 were scored: the basic scales, the 8 frequently used supplemental scales shared by MMPI and MMPI-2, and the Harris-Lingoes subscales. Psychometric information about these scales of the MMPI-2 is outlined below.

MMPI-2 Basic Scales. The number of items on the basic scales range from 15 to 78. Internal consistency correlations based on the contemporary normative sample ranged from .34-.87. Factor analysis resulted in factors similar to those typical of the MMPI, although the factors did vary a bit between genders. Also, like MMPI, many of the scales of MMPI-2 were highly intercorrelated. Test-retest correlations based on a mean interval of 8.5 days on a subgroup of 193 subjects from the normative sample ranged from .58-.92 (Butcher, et al., 1989).

Butcher et al. (1989) assessed validity of the MMPI-2 basic scales by giving all subjects in the normative sample a modified Recent Life Events Survey (Holmes & Rahe, 1967). Couples living together or married for at least one year who were a part of the normative sample were give a modified

Katz Adjustment Scale (Katz, 1968) to rate their partner on various dimensions. For the most part, the basic scales of the MMPI-2 correlated with appropriate correlates, although most correlations were fairly low, with no correlations over .35. High interscale relatedness often resulted in more than one scale correlating highly with a predicted correlate. Graham (1990) also provided validity data on a clinical sample of 425 patients who were given behavioral ratings by their attending psychiatrist or psychologist. Graham concluded that the behavioral correlates for each scale were similar to those expected if MMPI had been given.

MMPI-2 Supplemental Scales. The 8 scales contain from 25 to 52 items. Cronbach alpha internal consistency coefficients from the contemporary normative group ranged from .24 to .90. Test-retest correlations based on a mean interval of 8.5 days of a subgroup of 193 subjects from the contemporary normative group ranged from .62 to .91 (Butcher et al., 1989). Butcher et al. (1989) reported correlations between the modified Recent Life Events Survey and Katz Adjustment Scale with the MMPI-2 supplemental scales. Most correlations appeared appropriate, but many were low with no correlation over .39.

MMPI-2 Harris-Lingoes Subscales. The number of items on the subscales range from 6 to 32. Cronbach alpha internal consistency coefficients based on the contemporary normative group was above .45 for all but 9 of the sub-

scales. High interscale correlations were found between many subscales. Test-retest correlations based on a mean interval of 8.5 days of a subgroup of 193 subjects from the contemporary normative group ranged from .58 to .88 (Graham, 1990). No validity information is available on the Harris-Lingoes subscales of the MMPI-2.

Procedure

In the O-O condition subjects were administered the MMPI. Subjects returned 1 week later to take the MMPI again. In the O-R condition half of the subjects were administered the MMPI, followed 1 week later by the MMPI-2, and the other half were administered the MMPI-2 followed 1 week later by the MMPI. All subjects received a demographic questionnaire which was completed before their initial testing session.

Subjects were tested in groups ranging in size from 15-50 persons. Upon arrival to the first testing session all subjects received a sheet of written instructions (Appendix C). In addition to these written instructions, subjects were orally reminded of the requirement to return in 1 week, to read all instructions before beginning, and to put all materials in an envelope that was provided to them when finished with the testing. Further, subjects were reminded not to use their name anywhere on the test materials, but to make sure they included the last four digits of their social security number which was use to match sub-

jects' responses across the two testing sessions.

Upon arrival to the second testing session, all subjects received another sheet of written instructions (Appendix D). Subjects were given oral instruction similar to those given during the first testing session. In addition, subjects were told that the test they received during this session may be similar to the test they had received during the first testing session. This instruction was given to normalize any concern subjects may have had about taking the same test, or a similar test, twice. Subjects were also reminded that a full debriefing was available after this testing session. The experimenter had a list of professional referral sources during both testing session should any subject have felt a degree of emotional upset after taking the MMPI or MMPI-2. No subject indicated emotional upset after taking either instrument.

Scoring

Most scales of the MMPI and all scales of the MMPI-2 were computer scored by National Computer Systems (NCS) under the Research Assistance Program. In exchange for results of the research, NCS provided a discount for materials and computer scoring.

The NCS Research Assistance Program applied to all scales of the MMPI-2. It did not apply to all scales of the MMPI, however. The Harris-Lingoes subscale and the supplemental scales of Mt and O-H were not scored for the MMPI

under the Research Assistance Program. Because these scales were not scored under the program, these scales were scored by hand.

The answer sheets provided by NCS for the MMPI were a special form of answer sheets used for computer scoring only. No commercial templates existed for these answer sheets. Using Graham's (1987) interpretation manual, templates were constructed by the primary investigator. The primary investigator hired two research assistants to hand score the Harris-Lingoes subscales and the Mt and O-H scales of the 600 MMPIs used in the research.

The primary investigator randomly inspected 10% of the scales scored by the research assistants to assess scoring accuracy. The two research assistants had error rates of 1.0% and 1.8% respectively, with no error ever greater than 1 raw score point.

Data Analyses

All data analyses were conducted separately by gender. All significance tests assess the null hypothesis that no differences exist between the O-O condition and O-R condition.

Demographic Information. Means and frequencies for demographic variables were calculated to insure that no major demographic differences exist between subjects in the two conditions. Academic majors were classified into one of six Holland categories using a classification system of

academic majors at The Ohio State University developed by Gordon (1986). Parental occupation was classified into one of nine occupational categories as classified by the Dictionary of Occupational Titles (U.S. Department of Labor, 1977). For age, t tests were used to determine significant differences between conditions within gender. Chi square analyses of proportions were used for the other demographic variables.

Correlational Analyses. Pearson product moment correlations were calculated between the raw scores of the basic scales of the MMPI taken twice (O-O condition) and between the raw scores of the basic scales of the MMPI and MMPI-2 (O-R condition). Two tailed Fisher's r to Z transformations were calculated at the .05 and .01 level to determine if these correlations differ significantly from each other. The .05 level of significance is included to provide a less conservative test of the null hypothesis.

The same analysis using T scores instead of raw scores would not result in different correlations, as when raw scores are converted to T scores they undergo a linear transformation to which correlational data is not sensitive (Cattell, 1949, Cohen & Cohen, 1983). Because correlational data is not sensitive to how raw scores are normed, the effect of the new normative group and uniform T scores was not assessed by this analysis. Significantly lower correlations in the O-R condition are hypothesized to show differ-

ences between the MMPI and MMPI-2, differences due to item changes.

Correlational analyses were conducted on the supplemental scales and were similar to those conducted on the basic scales, with the exception that these analyses are considered exploratory because of the increasing probability of familywise, Type I error resulting from the number of significance tests conducted.

Correlational analyses were conducted on the Harris-Lingoes subscales and were similar to those used for the basic scales and supplemental scales. These analyses are also considered exploratory because of the increasing probability of familywise, Type I error.

For the O-R condition, interscale correlations of the MMPI were calculated and compared, using the Fisher's r to Z transformation, to the interscale correlations of the MMPI-2. This analysis is used to assess whether the scales of the MMPI-2 relate with each other in a way that is similar to the MMPI. Significant differences between interscale correlations are hypothesized to suggest that the scales of the MMPI related to each other differently than the scales of the MMPI-2. This correlational analysis, like the others, is sensitive only to item changes between the MMPI and MMPI-2. Because of the number of significance tests, 78 for each gender, such an analysis must be considered exploratory. Familywise, Type 1 error rates are approximately .50

for each analysis by gender when a .01 level of significance is used (Cohen & Cohen, 1983).

Configural Analyses. The consistency of a number of configural characteristics in the O-O condition and O-R condition were analyzed. An analysis of consistency in profile configuration is sensitive to all 3 major sources of variation between the MMPI and MMPI-2 (item changes, new normative sample, and uniform T distribution). Percentages of the number of profiles in each condition in which there was consistency in the configural characteristic were calculated. A two tailed Z test of the significance of the difference between two proportions at the .05 level was used to determine if the proportions in the two conditions were significantly different from each other. Should the O-R condition show a significantly lower percentage of configural consistency on a given characteristic than the O-O condition, it is hypothesized that this difference is due to differences between the MMPI and MMPI-2. The only exception to this type of analysis was the calculation of the D^2 and D^2 statistics, for which means were calculated and two tailed t -tests were conducted at the .05 level. Should the O-R condition show a significantly higher D^2 or D^2 statistic than the O-O condition it is hypothesized that this difference is due to differences between the MMPI and MMPI-2.

For all configural characteristics analyzed, which include the five characteristics analyzed by Ben-Porath and

Butcher (1989a), analyses were conducted using 8 of the 10 clinical scales so as to facilitate comparison with the Ben-Porath and Butcher study. In addition, these analyses were also conducted using all 10 of the clinical scales. The only exception to this being the analysis of validity configuration which utilizes the 3 validity scales. The various configural characteristics are defined below.

High point consistency occurred when a subject's most elevated scale was the same scale on both administrations. In the event two scales were elevated equally (ie. there was a tie between highest T scores) both scales were considered high points. For example, if a subjects's most elevated scale on one instrument was scale 4 and on the second instrument scales 2 and 4 were tied, with both scales having the highest T scores on the profile, both scales 2 and 4 were considered high points and the subject would have been classified as having high point consistency. In this characteristic, and in the five following analyses, ties on relevant scales were classified as consistent. A proportion was calculated by totaling the number of subjects within a condition having consistency in high point code in relation to the total number of subjects within the condition.

Two point consistency occurred when a subject's two most elevated scales were the same over both administrations. Ben-Porath and Butcher (1989a) did not specifically state whether the order of these two scales need to be the

same across administrations in order to be defined as consistent. According to Graham (1987), order should not be considered for two point consistency. It is assumed that this was the definition adopted by Ben-Porath and Butcher and was the definition used in the present study.

Three point consistency was similar to two point consistency but utilized the highest three scales. Low point consistency was similar to high point consistency but utilized the lowest scale.

Validity configural consistency occurred when a the order of a subject's three validity scales, from highest T score to lowest T score, was consistent across both test administrations.

Well defined code types (ie., at least 5 T score points between the scales in the code type and the next highest scale in the profile) are recommended by Graham (1987, 1990) as the only type of codes for which code type interpretation is relevant. Given this recommendation an analysis of consistency of well defined code types was conducted. A subject was classified as having a well defined code type on the MMPI if he/she had a one point, two point, or three point code that was at least 5 T score points above the next highest scale. If a subject had more than one well defined code type, the simplest code type was used (eg. if a subject had a one point and a three point well defined code, his/her simplest code, in this case the one point code, was used).

If a subject was classified as having a well defined code, his/her profile on the other instrument (the MMPI in the O-O condition or the MMPI-2 in the O-R condition) was analyzed. If the top scales on the second instrument matched the well defined code type on the original MMPI analyzed, the subject was classified as having well defined code consistency. Like for two and three point consistency, the order of the top scales were not considered for consistency. The proportion was calculated by totaling the number of subjects within a condition having well defined code consistency in relation to the total number of subjects within that condition who had a well defined code on the MMPI.

A within normal limits (WNL) profile is defined as all clinical scales below a T score of 70 for the MMPI and a T score of 65 for the MMPI-2. WNL profile consistency occurred when a subject had a WNL profile on both test administrations. Ben-Porath and Butcher (1989a) did not indicate how they calculated WNL profile consistency proportions. The proportion was calculated in the present study by totaling the number of subjects within a condition having matching WNL profiles in relation to the total number of subjects within that condition who had at least one WNL profile.

An elevated scale is any clinical scale with a T score above 70 on the MMPI and a T score above 65 on the MMPI-2. A subject having the same number of elevated scales on both administrations of the test was classified as having con-

sistency in the number of elevated scales.

The D^2 statistic was calculated by summing the squares of the differences between scores on all of the clinical scales across the two administrations for each subject (Wiggins, 1973). Similar to the Ben-Porath and Butcher (1989a) study, 5 T score point were added to each of the clinical scale scores of the MMPI-2 given the Butcher et al. (1989) assertion that MMPI-2 scores will be about 5 points lower than MMPI scores.

The D statistic was calculated by taking the square root of the sum of the squared differences between scores on all of the clinical scales across the two administrations for each subject (Cronbach & Gleser, 1953; Osgood & Suci, 1958). Again, 5 T score points were added to the clinical scales scores of the MMPI-2.

As a final configural analysis, means and standard deviations were calculated for each of the basic scales on each of the instruments given within a condition. For the O-O condition, means on the basic scales for the first administration of the MMPI were compared to means on the second administration. Differences between means on the first administration and the second administration were calculated. Given MMPI basic scale reliability data and standard error information, it is predicted that mean scores should differ in the O-O condition anywhere from 0-3 T score points (Dahlstrom, Welsh, & Dahlstrom, 1972, 1975). Mean

differences in the O-O condition were calculated for the purpose of providing a baseline comparison. Means, standard deviations, and mean differences were also calculated in the O-R condition. For the MMPI-2, 5 T score points were added to the scores of each of the basic scales. Mean differences greater than 3 T score points are hypothesized to be a result of differences between the MMPI and MMPI-2. Means of scores on the MMPI and MMPI-2 from the O-R condition were plotted on a graph to illustrate any difference between mean scale scores. Large differences between means on given scales are suggestive of scales which may cause configural shifts in the pattern of results between MMPI and MMPI-2.

Results

Demographic Information

Demographic information by gender and condition is presented in Table 1. For purposes of economy of presentation, the six categories for father's education and mother's education were collapsed into three categories: "high school," "college," and "advanced degree." Of the nine primary occupational categories of the Dictionary of Occupational Titles (U.S. Department of Labor, 1977), eight were collapsed into the category "non-professional." The other remaining category titled "Professional, Technical, and Managerial Occupations" by the Dictionary of Occupational Titles is labeled as "professional" in Table 1. Subjects were classified as having professional parents if one or both parents had an occupation categorized as "professional." The ethnic categories of "Asian," "Native American," "Black," and "Hispanic" were collapsed to form the category of "non-White."

Table 1

Demographic Variables by Gender and Condition

Variable	Males		Females	
	O-O	O-R	O-O	O-R
N	86	92	94	96
Age				
M	19.9	19.2	18.9	19.7
range	18-31	16-31	17-38	17-45
College Major				
Undecided	11	14	17	18
Realistic	3	2	2	0
Investigative	31	33	25	16
Artistic	7	5	9	7
Social	17	13	31	37
Enterprising	14	18	8	14
Conventional	2	2	1	3
Marital Status				
Single	83	89	91	92
Married	3	3	2	4
Class Standing				
Freshperson	50	61	71	69
Sophomore	26	18	19	23
Junior	4	7	5	3
Senior	6	4	0	1
Ethnicity				
White	75	86	84	87
Non-White	11	6	10	9
Parental Occupation				
Professional	50	52	58	59
Non-Professional	29	36	29	22

Table 1 (cont.)

Variable	Males		Females	
	O-O	O-R	O-O	O-R
Father's Education				
High School	27	37	40	33
College	43	41	34	42
Advanced Degree	16	13	19	20
Mother's Education				
High School	45	42	43	46
College	37	43	40	42
Advanced Degree	4	7	2	1

T tests on age and chi square analyses of proportions on the other demographic variables indicate no significant difference between conditions within gender. For chi square analyses the categories of "sophomore," "junior," and "senior" for class standing were collapsed into one category, as were "college" and "advanced degree" for father and mother's educational level. College major was not included in chi square analyses.

Correlational Analyses

Basic Scales. Pearson product moment correlations calculated between the raw scores of the basic scales of the MMPI taken twice (O-O condition) and between the raw scores of the basic scales of the MMPI and MMPI-2 (O-R) condition are presented by gender in table 2. Two tailed Fisher's r to Z transformations at the .05 and .01 level of significance are used to test differences between correlations in the O-O condition and O-R condition.

Table 2

Correlations for the Basic Scales of MMPI Test-Retest (O-O Condition) and Between MMPI and MMPI-2 (O-R Condition) by Gender

Scale	Males		Females	
	O-O (<u>n</u> =86)	O-R (<u>n</u> =92)	O-O (<u>n</u> =94)	O-R (<u>n</u> =96)
L	.75	.55*	.82	.66*
F	.79	.60*	.72	.73
K	.83	.68*	.72	.81
1	.69	.70	.73	.59
2	.74	.76	.80	.76
3	.72	.68	.74	.66
4	.76	.59*	.69	.73
5	.84	.70*	.76	.70
6	.77	.58*	.61	.70
7	.72	.65	.70	.75
8	.85	.64**	.78	.77
9	.78	.77	.72	.75
0	.86	.88	.89	.88

* $p < .05$.

** $p < .01$.

For males, correlations for scales L, F, K, 4, 5, and 6 are significantly lower in the O-R condition ($Z=2.33$, $Z=2.47$, $Z=2.35$, $Z=2.08$, $Z=2.32$, $Z=2.35$, respectively, $p<.05$). Scale 8 is also significantly lower in the O-R condition ($Z=3.26$, $p<.002$). For females, the correlation for scale L is significantly lower in the O-R condition ($Z=2.47$, $p<.05$).

Supplemental Scales. Pearson product moment correlations calculated between the raw scores of the supplemental scales of the MMPI taken twice (O-O condition) and between the raw scores of the supplemental scales of the MMPI and MMPI-2 (O-R condition) are presented by gender in table 3. Two tailed Fisher's r to Z transformations at the .05 and .01 level of significance are used to test differences between correlations in the O-O condition and O-R condition.

Table 3

Correlations for the Supplemental Scales of MMPI Test-Retest (O-O Condition) and Between MMPI and MMPI-2 (O-R Condition) by Gender

Scale	Males		Females	
	O-O (<u>n</u> =86)	O-R (<u>n</u> =92)	O-O (<u>n</u> =94)	O-R (<u>n</u> =96)
A	.86	.76	.89	.89
R	.81	.57**	.80	.65*
ES	.77	.69	.84	.84
MAC	.67	.72	.69	.73
Do	.77	.65	.64	.72
Re	.85	.79	.70	.78
O-H	.70	.48*	.58	.60
Mt	.88	.76*	.87	.83

* $p < .05$.

** $p < .01$.

For males, correlations for scales O-H, and Mt are significantly lower in the O-R condition ($Z=2.25$, $Z=2.49$, respectively, $p < .05$). Scale R is also significantly lower in the O-R condition ($Z=3.19$, $p < .002$). For females, the correlation for scale R is significantly lower in the O-R condition ($Z=2.20$, $p < .05$).

Harris-Lingoes Subscales. Pearson product moment correlations calculated between the raw scores of the Harris-Lingoes subscales of the MMPI taken twice (O-O condition) and between the raw scores of the Harris-Lingoes subscales of the MMPI and MMPI-2 (O-R condition) are presented by gender in table 4. Two tailed Fisher's r to Z transformations at the .05 and .01 level of significance are used to test differences between correlations in the O-O condition and O-R condition.

Table 4

Correlations for the Harris-Lingoes Subscales of MMPI Test-Retest (O-O Condition) and Between MMPI and MMPI-2 (O-R Condition) by Gender

Scale	Males		Females	
	O-O ($n=86$)	O-R ($n=92$)	O-O ($n=94$)	O-R ($n=96$)
D1	.74	.76	.84	.87
D2	.69	.58	.73	.49**
D3	.56	.61	.68	.52
D4	.77	.68	.78	.84
D5	.72	.69	.81	.79
Hy1	.77	.84	.87	.80
Hy2	.76	.66	.75	.77
Hy3	.81	.74	.80	.74
Hy4	.81	.75	.81	.74
Hy5	.76	.52**	.66	.55

Table 4 (cont.)

Scale	Males		Females	
	O-O ($n=86$)	O-R ($n=92$)	O-O ($n=94$)	O-R ($n=96$)
Pd1	.74	.66	.84	.87
Pd2	.48	.68*	.61	.63
Pd3	.73	.72	.76	.82
Pd4	.79	.61*	.78	.69
Pd5	.79	.67	.76	.77
Pa1	.82	.62*	.70	.67
Pa2	.72	.53*	.70	.72
Pa3	.66	.60	.67	.68
Sc1	.86	.51**	.81	.83
Sc2	.69	.55	.65	.59
Sc3	.77	.68	.77	.82
Sc4	.85	.74*	.81	.81
Sc5	.71	.65	.78	.76
Sc6	.71	.69	.70	.73
Ma1	.69	.52	.55	.64
Ma2	.67	.62	.72	.67
Ma3	.65	.66	.70	.64
Ma4	.69	.57	.54	.68

* $p < .05$.** $p < .01$.

For males, correlations for scales Pd4, Pa1, Pa2, and Sc4 are significantly lower in the O-R condition ($\underline{Z}=2.37$, $Z=2.83$, $Z=2.08$, $Z=2.01$, respectively, $p<.05$). Scales Hy5 and Sc1 are also significantly lower in the O-R condition ($\underline{Z}=2.75$, $p<.01$; $\underline{Z}=4.78$, $p<.0001$; respectively). The correlation of scale Pd2 is significantly higher in the O-R condition ($\underline{Z}=2.01$, $p<.05$). For females, the correlation for scale D2 is significantly lower in the O-R condition ($\underline{Z}=2.66$, $p<.01$).

Interscale Correlations. Interscale correlations of the MMPI and MMPI-2 calculated using the Pearson product moment correlation are presented in Tables 5 and 6 for males in the O-R condition. The presentation of the interscale correlation matrix is divided into two tables because of space constraints. Table 5 shows the interscale correlations of scales L-4 with scales L-0. Table 6 shows the interscale correlations of scales 5-0 with scales 5-0. Between these two tables all possible paired combinations between scales are presented. Interscale correlations for the MMPI are listed above interscale correlations for the MMPI-2. A two tailed Fisher's r to $\underline{Z}$ transformation is used to test whether each of the interscale correlations of the MMPI differ significantly from the interscale correlations of the MMPI-2. Tables 7 and 8 present similar data for females in the O-R condition.

Table 5

Interscale Correlations for Scales L-4 with L-0 of the MMPI and MMPI-2 in the O-R Condition for Males (n=92)

Scale	Scale						
	L	F	K	1	2	3	4
L	-- --						
F	.10 -.08	-- --					
K	.07 .37*	-.41 -.39	-- --				
1	-.01 .16	.42 .33	.03 .22	-- --			
2	.11 .07	.39 .31	-.44 -.33	.44 .49	-- --		
3	.04 .22	.08 .15	.21 .26	.57 .78**	.15 .34	-- --	
4	-.24 -.16	.27 .39	.15 .02	.37 .46	.17 .36	.36 .45	-- --
5	-.15 .01	.08 .10	-.10 -.02	.21 .28	.29 .36	.16 .28	.08 .24
6	-.02 -.07	.57 .45	-.23 -.14	.44 .36	.33 .23	.16 .33	.33 .42
7	-.29 -.13	.41 .44	-.24 -.27	.59 .54	.53 .66	.34 .38	.47 .53
8	-.09 -.14	.66 .67	-.13 -.18	.65 .54	.41 .45	.36 .39	.51 .58
9	-.14 -.23	.33 .31	-.09 -.09	.09 .12	-.27 -.21	.01 .10	.29 .29
0	-.13 -.12	.40 .34	-.56 -.54	.23 .08	.69 .59	-.11 -.10	.06 .11

Note: Interscale correlations for the MMPI are listed first. Interscale correlations for the MMPI-2 are listed below those for the MMPI.

*p<.05 **p<.01

Table 6

Interscale Correlations for Scales 5-0 with 5-0 of the MMPI and MMPI-2 in the O-R Condition for Males (n=92)

Scale	Scale					
	5	6	7	8	9	0
5	-- --					
6	.25 .19	-- --				
7	.37 .43	.58 .48	-- --			
8	.30 .34	.68 .57	.80 .76	-- --		
9	-.08 -.01	.21 .19	.16 .18	.32 .37	-- --	
0	.17 .26	.37 .18	.50 .51	.36 .35	-.23 -.34	-- --

Note: Interscale correlations for the MMPI are listed first. Interscale correlations for the MMPI-2 are listed below those for the MMPI.

* $p < .05$ ** $p < .01$

Table 7

Interscale Correlations for Scales L-4 with L-0 of the MMPI and MMPI-2 in the O-R Condition for Females (n=96)

Scale	Scale						
	L	F	K	1	2	3	4
L	-- --						
F	-.26 -.38	-- --					
K	.28 .46*	-.63 -.62	-- --				
1	.02 .03	.31 .21	.02 .06	-- --			
2	-.30 -.16	.53 .42	-.48 -.42	.28 .40	-- --		
3	-.10 .10	.26 .05*	.12 .17	.62 .68	.34 .35	-- --	
4	-.23 -.19	.37 .26	-.08 .02	.27 .29	.41 .30	.50 .44	-- --
5	.09 -.18**	.29 .29	-.10 -.17	.08 -.06	-.16 -.14	-.06 -.16	-.09 .03
6	-.19 -.29	.43 .39	-.22 -.24	.29 .31	.36 .33	.32 .28	.39 .19*
7	-.32 -.42	.43 .52	-.29 -.34	.33 .39	.69 .67	.37 .39	.45 .33
8	-.33 -.37	.65 .64	-.35 -.33	.45 .35	.49 .40	.39 .28	.40 .32
9	-.06 -.34**	.39 .54	-.27 -.45*	.06 .07	-.01 .10	.08 .10	.20 .14
0	-.21 -.29	.46 .37	-.63 -.63	.15 .01	.69 .58	-.01 -.19	.13 .00

Note: Interscale correlations for the MMPI are listed first. Interscale correlations for the MMPI-2 are listed below those for the MMPI.

* $p < .05$ ** $p < .01$

Table 8

Interscale Correlations for Scales 5-0 with 5-0 of the MMPI and MMPI-2 in the O-R Condition for Females (n=96)

Scale	Scale					
	5	6	7	8	9	0
5	-- --					
6	.01 -.10	-- --				
7	-.18 -.11	.48 .48	-- --			
8	.18 .20	.53 .48	.69 .73	-- --		
9	.24 .15	.20 .31	.24 .33	.51 .53	-- --	
0	-.04 -.01	.18 .12	.48 .47	.36 .29	-.07 .04	-- --

Note: Interscale correlations for the MMPI are listed first. Interscale correlations for the MMPI-2 are listed below those for the MMPI.

* $p < .05$ ** $p < .01$

For males, there is a significant difference in interscale correlations between the MMPI and MMPI-2 on the following pairs of scales: L/K ($Z=2.53$, $p < .05$), 1/3 ($Z=3.75$, $p < .001$). For females, there is a significant difference in interscale correlations between the MMPI and MMPI-2 on the following pairs of scales: L/K ($Z=2.02$, $p < .05$), L/5 ($Z=2.62$, $p < .01$), L/9 ($Z=2.85$, $p < .01$), F/3 ($Z=2.08$, $p < .05$), K/9 ($Z=2.01$, $p < .05$), 4/6 ($Z=2.12$, $p < .05$).

Configural Analyses

Analyses Using 8 Clinical Scales. The configural analyses presented in Table 9 are based on the scores of 8 of the 10 clinical scales, with scales 0 and 5 deleted from the analyses. All numbers presented in the chart are percentages, with the exception of D^2 and D statistics for which means and standard deviations are presented. A two-tailed Z test of the significance of the difference between two proportions is used between the two conditions for each configural characteristic to test if the percentages are significantly different from each other within gender. This significance test is used for all configural analyses with the exception of the D^2 and D statistic in which a two tailed t test is used to determine significant differences. Significant differences are reported at the .05 and .01 levels.

The number of subjects classified as having well defined code types on the MMPI were 63, 70, 63, and 65 in the male O-O condition, male O-R condition, female O-O condition, and female O-R condition, respectively. The number of subjects having at least one WNL profile were 41, 53, 59, and 64 in the male O-O condition, male O-R condition, female O-O condition, and female O-R condition, respectively.

Table 9

Configural Analyses Using 8 Clinical Scale (Scales 0 and 5 Deleted)

Characteristic	Males		Females	
	O-O (<u>n</u> =86)	O-R (<u>n</u> =92)	O-O (<u>n</u> =94)	O-R (<u>n</u> =96)
High Point Consistency	64.0	52.2	57.4	58.3
Two Point Consistency	41.9	34.8	44.7	26.0**
Three Point Consistency	43.0	42.4	44.7	25.0**
Low Point Consistency	44.2	50.0	54.3	40.6
Well Defined Code Consistency	66.7	55.7	70.0	61.5
Same Number of Elevations	38.4	40.2	54.3	53.1
Same WNL Profile	46.3	47.2	66.1	59.4
2				
D Statistic				
<u>M</u>	515.62	666.97	456.12	453.44
<u>SD</u>	458.44	621.19	419.18	327.46
D Statistic				
<u>M</u>	20.93	23.92*	19.63	20.19
<u>SD</u>	8.87	9.78	8.46	6.82

Note: All numbers reported in the table are percentages with the exception of the D2 and D statistics as indicated.

* $p < .05$

** $p < .01$

For males, the D statistic is significantly higher in the O-R condition ($df=176$, $t=2.13$, $p<.05$). For females, two point consistency and three point consistency are significantly lower in the O-R condition ($Z=2.75$, $p<.01$; $Z=2.91$, $p<.01$; respectively). Low point consistency approaches significance ($Z=1.91$, $p<.06$).

In light of the D statistic showing significantly different results between the O-O condition and the O-R condition a follow up statistic was calculated. In the O-O condition T scores for males seem to vary less between the two instruments than in the O-R condition, where it appears that, on the average, T scores change to a greater extent between the two instruments. It was thought useful, therefore, to see how much greater, on the average, T scores change in the O-R condition across an entire profile as compared to the O-O condition. With this purpose in mind, a third statistic, referred to as the "modified D statistic" was calculated. The modified D statistic was calculated in a manner similar to the D statistic except that differences between scores were not squared. Instead, absolute difference between scales were calculated and summed across the 8 clinical scales. Osgood and Suci (1958) reported that using absolute differences in this way results in a statistic with the same characteristics as the D statistic. Further, such a statistic will indicate the amount of T score points that a subject's profile varies from one administration to the

other across the 8 clinical scales.

For males $M=49.17$, $SD=22.14$ and $M=56.20$, $SD=24.28$ for the modified D statistic in the O-O and O-R condition, respectively. The mean modified D statistic for the O-R condition is significantly larger than for the O-O condition ($df=176$, $t=2.09$, $p<.05$), and indicates that, on the average, T scores vary about 7 points more across an entire profile between instruments in the O-R condition as compared to the O-O condition. For females $M=45.49$, $SD=21.85$ and $M=46.98$, $SD=16.09$ for the modified D statistic in the O-O and O-R condition, respectively. The difference between these two conditions is not significant.

Analysis Using 10 Clinical Scales. Analyses using all 10 clinical scales, presented in Table 10, were conducted in a manner similar to those using only 8 scales with the exception that scale 5 and scale 0 were included in the analyses. Validity configural consistency, which is not based on clinical scales but on the 3 validity scales, is also presented in Table 10.

The number of subjects classified as having well defined code types on the MMPI using all 10 clinical scales were 59, 68, 52, and 63 in the male O-O condition, male O-R condition, female O-O condition, female O-R condition, respectively. The number of subjects having at least one WNL profile were 38, 52, 56, and 57 in the male O-O condition, male O-R condition, female O-O condition, female O-R condition, respectively.

Table 10

Configural Analyses Using All 10 Clinical Scales

Characteristic	Males		Females	
	O-O (<u>n</u> =86)	O-R (<u>n</u> =92)	O-O (<u>n</u> =94)	O-R (<u>n</u> =96)
High Point Consistency	62.8	43.5**	55.3	44.8
Two Point Consistency	44.2	28.3*	31.9	18.8*
Three Point Consistency	37.2	26.1	31.9	16.6*
Low Point Consistency	53.5	40.2	50.0	41.7
Well Defined Code Consistency	72.9	47.0**	78.8	38.1**
Same Number of Elevations	30.2	37.0	54.3	46.9
Same WNL Profile	36.8	42.3	62.5	47.4
Validity Configuration Consistency	65.1	45.7*	64.9	43.8**
2				
D Statistic				
<u>M</u>	653.74	825.99	558.10	636.82
<u>SD</u>	587.52	702.41	471.05	395.08
D Statistic				
<u>M</u>	23.64	26.89*	21.99	24.20
<u>SD</u>	9.80	10.19	8.68	7.20

Note: All numbers in Table 10 are percentage with the exception of the D2 and D statistics as indicated.

* $p < .05$

** $p < .01$

For males, high point consistency, two point consistency, validity configuration consistency, and well defined code consistency are significantly lower in the O-R condition ($Z=2.63$, $p<.01$; $Z=2.23$, $p<.05$; $Z=2.56$, $p<.05$; $Z=3.09$, $p<.002$; respectively) with low point consistency approaching significance ($Z=1.79$, $p<.08$). The D statistic is significantly higher in the O-R condition ($df=176$, $t=2.16$, $p<.05$) with D^2 approaching significance ($df=176$, $t=1.77$, $p<.08$).

For females, two point consistency, three point consistency, validity configuration consistency, and well defined code consistency are significantly lower in the O-R condition ($Z=2.41$, $p<.05$; $Z=2.50$, $p<.05$; $Z=2.99$, $p<.01$, $Z=4.88$, $p<.00001$; respectively). The higher D^2 statistic in the O-R condition approached significance ($df=188$, $t=1.91$, $p<.06$).

Using the 10 clinical scales, the modified D statistic statistic was also calculated in this analysis. For males, $M=57.78$, $SD=24.13$ and $M=70.25$, $SD=27.93$ for the modified D statistic statistic in the O-O and O-R condition, respectively. The mean modified D statistic for the O-R condition is significantly larger than for the O-O condition ($df=176$, $t=3.17$, $p<.002$), and indicates that, on the average, T scores vary about 12 points more across an entire profile between instruments in the O-R condition as compared to the O-O condition. For females, $M=53.48$, $SD=23.13$ and $M=61.64$, $SD=18.41$ for the modified D statistic statistic in the O-O

and O-R condition, respectively. The mean modified D statistic for the O-R condition is significantly larger than the O-O condition ($df=188$, $t=2.70$, $p<.01$), and indicates that, on the average, T scores vary about 8 points more across an entire profile between instruments in the O-R condition as compared to the O-O condition.

Additionally, the modified D statistic was calculated including the 3 validity scales along with the 10 clinical scales (the 13 basic scales of the MMPI and MMPI-2). For males, $M=70.24$, $SD=27.29$ and $M=88.83$, $SD=33.94$ for the modified D statistic in the O-O and O-R condition, respectively. The mean modified D statistic for the O-R condition is significantly larger than for the O-O condition ($df=176$, $t=4.01$, $p<.0001$), and indicates that, on the average, T scores vary about 12 points more across an entire profile between instruments in the O-R condition as compared to the O-O condition. For females, $M=65.08$, $SD=26.21$ and $M=79.78$, $SD=21.86$ for the modified D statistic in the O-O and O-R condition, respectively. The mean modified D statistic for the O-R condition is significantly larger than the O-O condition ($df=188$, $t=4.20$, $p<.0001$), and indicates that, on the average, T scores vary about 8 points more across an entire profile between instruments in the O-R condition as compared to the O-O condition.

Basic Statistics and Mean Differences. Table 11 presents the means and standard deviations of the basic scales for

each of the instruments administered to males within each condition. Table 11 also presents the differences between the means of each scale within condition. The means for the basic scales of the MMPI and MMPI-2 in the O-R condition are presented in the form of a graph in Figure 1. Table 12 and Figure 2 present similar data for females.

Table 11

Basic Statistics and Mean Differences of the Basic Scales
for Each Test Administered by Condition for Males

Scale	O-O Condition (n=86)			O-R Condition (n=92)		
	MMPI	MMPI	Mean Difference	MMPI	MMPI-2	Mean Difference
L						
<u>M</u>	46.2	46.4	+0.2	45.0	49.0	+4.1
<u>SD</u>	6.9	6.3		5.9	7.0	
F						
<u>M</u>	59.1	60.1	+1.0	58.4	60.1	+1.7
<u>SD</u>	9.7	10.2		11.1	12.2	
K						
<u>M</u>	50.3	51.7	+1.4	51.0	49.3	-1.7
<u>SD</u>	9.7	9.5		7.9	8.6	
1						
<u>M</u>	54.5	53.5	-1.0	52.5	52.3	-0.2
<u>SD</u>	10.2	10.1		8.3	9.5	
2						
<u>M</u>	58.5	58.2	-0.3	55.1	52.6	-2.5
<u>SD</u>	10.8	12.9		12.8	11.4	
3						
<u>M</u>	56.2	54.4	-2.3	53.5	50.6	-2.9
<u>SD</u>	8.7	9.8		7.1	9.6	

Table 11 (cont.)

Scale	O-O Condition (n=86)			O-R Condition (n=92)		
	MMPI	MMPI	Mean Difference	MMPI	MMPI-2	Mean Difference
4						
<u>M</u>	61.0	61.9	+0.9	58.9	56.3	-2.6
<u>SD</u>	8.8	10.9		9.8	9.4	
5						
<u>M</u>	61.2	60.3	-1.0	60.7	52.5	-8.2
<u>SD</u>	10.2	10.4		9.5	8.1	
6						
<u>M</u>	58.1	57.5	-0.6	56.5	54.8	-1.7
<u>SD</u>	11.1	11.4		10.3	11.6	
7						
<u>M</u>	62.9	62.7	-0.2	63.0	59.7	-3.3
<u>SD</u>	11.8	12.8		12.5	12.1	
8						
<u>M</u>	67.0	67.9	+0.9	65.5	60.3	-5.2
<u>SD</u>	15.5	15.6		14.4	11.9	
9						
<u>M</u>	67.9	68.4	+0.4	66.8	63.4	-3.2
<u>SD</u>	11.5	11.4		12.0	12.2	
0						
<u>M</u>	52.6	52.5	-0.1	52.2	55.3	+3.1
<u>SD</u>	9.9	9.9		11.1	11.2

Figure 1. Graphic representation of the means of the basic scales of MMPI and MMPI-2 for males in the O-R condition (n=92).

Table 12

Basic Statistics and Mean Differences of the Basic Scales
for Each Test Administered by Condition for Females

Scale	O-O Condition (n=94)			O-R Condition (n=96)		
	MMPI	MMPI	Mean Difference	MMPI	MMPI-2	Mean Difference
L						
<u>M</u>	45.6	46.5	+0.9	46.3	52.3	+6.0
<u>SD</u>	6.3	7.2		5.3	8.6	
F						
<u>M</u>	57.4	56.6	-0.8	56.6	58.9	+2.3
<u>SD</u>	8.4	8.8		9.1	12.0	
K						
<u>M</u>	50.0	50.8	+0.8	51.2	51.4	+0.2
<u>SD</u>	7.2	7.9		9.1	10.9	
1						
<u>M</u>	53.4	50.8	-2.6	52.4	54.4	+2.1
<u>SD</u>	9.6	9.6		8.2	9.6	
2						
<u>M</u>	55.0	53.4	-1.6	54.1	53.4	-0.6
<u>SD</u>	9.7	11.0		9.8	8.6	
3						
<u>M</u>	54.2	51.6	-2.6	55.7	53.6	-2.1
<u>SD</u>	7.9	8.7		7.8	9.4	

Table 12 (cont.)

		O-O Condition ($n=94$)			O-R Condition ($n=96$)		
Scale		MMPI	MMPI	Mean Difference	MMPI	MMPI-2	Mean Difference
4	<u>M</u>	57.8	56.4	-1.4	58.3	56.1	-2.2
	<u>SD</u>	10.7	9.9		10.9	9.1	
5	<u>M</u>	52.1	53.5	+1.4	51.3	61.5	+10.2
	<u>SD</u>	8.3	8.8		8.7	10.1	
6	<u>M</u>	57.8	55.4	-2.3	57.1	54.6	-2.5
	<u>SD</u>	9.3	9.3		10.3	10.1	
7	<u>M</u>	59.0	58.5	-0.5	57.0	59.2	+2.1
	<u>SD</u>	9.5	9.6		9.4	10.2	
8	<u>M</u>	59.9	59.6	-0.2	59.4	58.8	-0.6
	<u>SD</u>	11.1	12.0		9.8	10.3	
9	<u>M</u>	62.8	64.8	+2.1	62.4	62.2	-0.3
	<u>SD</u>	10.6	11.8		10.6	10.2	
0	<u>M</u>	53.2	52.4	-0.8	52.7	53.3	+0.6
	<u>SD</u>	9.2	9.5		9.6	9.6

Figure 2. Graphic representation of the means of the basic scales of MMPI and MMPI-2 for females in the O-R condition ($n=96$).

For males and females mean differences in the O-O condition are in the predicted range of 1-3 T score points. For males in the O-R condition, scale 5 scores average about 8 T score points lower on the MMPI-2 as compared to the MMPI. Scores on scales 8 average about 5 T score points lower on the MMPI-2 as compared to the MMPI for males. Scores on scale L for males are about 4 T score points higher on the MMPI-2 than on the MMPI. Figure 1 graphically represents these differences.

For females in the O-R condition, scale 5 is about 10 T score points higher on the MMPI-2 as compared to the MMPI. Scores on scale L for females are about 6 T score points higher on the MMPI-2 as compared to MMPI. Figure 2 graphically represents these differences.

Discussion

This section will include a description of: 1) correlational analyses, 2) configural analyses, 3) implications of the analyses, 4) suggestions for future research, and 5) a summary of the findings.

Correlational Analyses

Chi square analyses and t tests of the demographic data suggest no significant demographic differences between the O-O and O-R conditions within gender. This finding decreases the likelihood that differences found between conditions in correlational and configural analyses are a result of demographic differences. Correlational analysis of the basic scales will be described first, followed by the supplemental scales and Harris-Lingoes subscales. Correlational analyses consider only item changes between the MMPI and MMPI-2.

The correlational analysis of the basic scales indicates significantly lower correlations in the O-R condition for males on scales L, F, K, 4, 5, 6, and 8 and for females on scale L. Such a result suggests that item changes may cause differences in scores between the MMPI and MMPI-2 on these scales. Alternate explanations for these significant differences could be the liberal p value used for

the significance tests and/or spuriously high correlations in the O-O condition. Correlations obtained in the O-O condition of the present study appear comparable to those obtained by Ben-Porath and Butcher (1989a), helping to rule out the latter explanation. Also, significantly lower correlations for scale K for men may have resulted in lower correlations on scales 4 and 8, as raw scores on these scales are partially calculated from scale K scores.

Because of the liberal nature of the significance tests on the basic scales, discussion of the findings will focus on convergent evidence from multiple sources. Scales L, F, 5, and 8 will be described in detail.

Correlations for scale L are significantly lower in the O-R condition for both genders. Although the analysis of interscale correlations is exploratory, the trends that emerge, particularly with scale L, are noteworthy. Of the 8 interscale correlations that differed significantly within gender between the MMPI and MMPI-2, 4 of them involve scale L, indicating that scale L may not be interrelating with other scales in the same way on the MMPI as on the MMPI-2. Butcher et al. (1989) reported that only 2 items were changed on scale L. Graham (1987) indicated that college students often obtain a raw score of 0 or 1 on scale L on the MMPI. A review of the distribution of scale L scores on the MMPI and MMPI-2 in the present study indicates that for both genders about 20%-30% of the subjects scored 0 or 1 on

scale L. With such a floor effect in scores, small differences in responses due to the 2 changed items and/or the new ordering of items on the MMPI-2 may have resulted in lower correlations.

The correlation for scale F is significantly lower in the O-R condition for males in the present study. Ben-Porath and Butcher (1989a) found scale F to be significantly lower at the .01 level in the O-R condition for females. Of all of the basic scales, scale F underwent the most item changes. Butcher et al. (1989) reported that 12 scale L items were reworded and 4 were deleted. It seems that these item changes may result in changes in scores between the MMPI and MMPI-2 on scale F.

The correlation for scale 5 is significantly lower in the O-R condition for males in the present study. A reanalysis of Ben-Porath and Butcher's (1989a) data by the present investigator shows that the correlation for males in the O-R condition of that study was significantly lower at the .05 level. Butcher et al. (1989) reported that 6 items were reworded and 4 were deleted from scale 5. These changes may have resulted in changes in scores on scale 5 between the MMPI and MMPI-2 for males.

The correlation for males on scale 8 is significantly lower in the O-R condition at $p < .002$. The correlation dropped from .85 in the O-O condition to .64 in the O-R condition. Butcher et al. (1989) reported that 13 items

from scale 8 were reworded, more wording changes than any other scale. These wording changes may have effected scores on scale 8 between the MMPI and MMPI-2 for males.

Item changes for scales L and F for both genders and scales 5 and 8 for males appear to produce different scores on the MMPI as opposed to the MMPI-2. Significantly lower correlations for males in the O-R condition are also noted for scales K, 4, 7, and 9, although these scales did not undergo any item deletions and few item changes (1, 4, 2, and 7 item changes, respectively) (Butcher et al. 1989). Significantly lower correlations for these scales may be a result of these few wording changes, may be spurious results, and/or may be the result of a more general change in response pattern in reaction to wording changes, deletions, and differences in item order between the MMPI and MMPI-2 as a whole.

Correlational analysis of the supplemental scales, although exploratory, is suggestive of problems with scale R. Correlations in the O-R condition are significantly lower for both genders, suggesting that item changes may cause changes in scores between the MMPI and MMPI-2 on this scale. Reanalysis of the Ben-Porath and Butcher (1989a) data indicates that correlations in the O-R condition of that study were significantly lower for females at the .05 level. Graham (1990) reported that 3 items were deleted from scale R; there is no information on the number of

wording changes for scale R items. Scale R was originally constructed from a factor analysis of the MMPI (Welsh, 1956). Two factors emerged from the analysis. The first factor, accounting for the most variance, became scale A with the second factor becoming scale R. Welsh's description of scale R items showed it to measure diverse areas, more diverse than scale A. Scale R appears to be a less cohesive scale and may be more sensitive to item deletions and wording changes.

Correlational analysis of the Harris-Lingoes subscales is considered exploratory, but is interesting when considered in the context of an article by Levitt (1990), published subsequent to data collection in the present study. Levitt did not describe wording changes in the items of the Harris-Lingoes subscales, but indicated on which subscales item deletions occurred. D2 had 1 item deleted. The correlation for females in the O-R condition on D2 is significantly lower. The only other Harris-Lingoes subscales to have items deleted were subscales Pd1-Pd5, the subscales of scale 4 (Pd). When these subscales were developed by Harris and Lingoes (1955, 1968), they contained items that were not a part of scale 4 (Pd). Butcher et al. (1989) dropped any items in subscales Pd1-Pd5 that were not a part of the parent scale, scale 4. Two of 11, 3 of 11, 6 of 12, 5 of 18, and 3 of 15 items were deleted from subscales Pd1, Pd2, Pd3, Pd4, and Pd5, respectively. Results of the correla-

tional analysis indicates that only Pd4 had a significantly lower correlation in the O-R condition, and this occurred only for males. Pd2 had a significantly higher correlation in the O-R condition, illustrating the possibility of spurious findings when comparing a large number of correlations. These results suggest that item deletions for subscales Pd1-Pd5 have little effect on scores on the subscales between the MMPI and MMPI-2, with the possible exception of subscale Pd4. Correlational analysis shows Pa1, Pa2, Sc1 and Sc4 also to have significantly lower correlations in the O-R condition for males, reflective of the lower correlations in the O-R condition for males on the parent scales of scale 6 (Pa) and scale 8 (Sc).

Correlational analyses of the present study stand in contrast to those of Ben-Porath and Butcher (1989a). Ben-Porath and Butcher found scales F and Es significantly lower for females in the O-R condition at the .01 level. Results of the present study show scales 8 and R significantly lower for males in the O-R condition at the .01 level. The results of the present study show scales correlating significantly lower in the O-R condition for males more often than for females. The cause of this gender difference is open for speculation. Differences between the results of the Ben-Porath and Butcher study and the present study may be attributable to the difference between using the MMPI-AX in place of the MMPI-2. The differences between results might

also suggest the need for a larger sample size when investigating a large number of scales (Megargee, 1979).

Configural Analyses

Configural analyses consider all major changes between the MMPI and MMPI-2: items changes, new normative sample, and the uniform T distribution. Configural analyses replicating those of Ben-Porath and Butcher (1989a) are consistent with their findings. Of the five types of characteristics analyzed by Ben-Porath and Butcher using 8 clinical scales, only two point consistency is significantly lower for females in the O-R condition in the present study. Ben-Porath and Butcher had found no significant differences in their analyses, but found two point consistency for females to be approaching significance with $p < .06$. The similarity in results suggest that definitions of configural consistency were similar across studies.

The results of the present study demonstrate the conservative nature of the Ben-Porath and Butcher analyses, especially in light of the additional analyses conducted here. The D and modified D statistic for males are significantly higher in the O-R condition, suggesting more variability in scores between the MMPI and MMPI-2, and therefore, less configural consistency. Two and three point consistency are lower in the O-R condition for females, also suggesting less configural consistency.

A decrease in configural consistency between the MMPI

and MMPI-2 is more striking when all 10 clinical scales are included in the analyses. For males, high point, two point, well defined code, and validity configuration consistency are all significantly lower in the O-R condition. For females, two point, three point, well defined code, and validity code consistency are all lower in the O-R condition. High point consistency drops for males from about 60% when the MMPI is given twice (O-O condition) to about 40% when the MMPI and MMPI-2 are given (O-R condition). Three point consistency drops for females from about 30% to about 15%. For both genders validity configural consistency drops from about 65% to about 45%. Of particular note is the drop in well defined code consistency. Well defined code consistency for males is approximately 70% when the MMPI is given twice and drops to approximately 50% when the MMPI and MMPI-2 are given. For females, the percentages drop from approximately 80% consistency to 40% consistency. D , D and modified D statistics all approach significance or are significantly higher in the O-R condition for both genders with the exception of the D statistic for females. As a whole, configural analyses suggest a dramatic decrease in configural consistency between the MMPI and MMPI-2 on many characteristics. This decrease in consistency is particularly striking when scales 0 and 5 are included in the analyses.

Analysis of the means of MMPI and MMPI-2 scores in the

O-R condition support the notion that, in general, MMPI-2 scores are about 5 T score points lower than MMPI scores. Means on the various scales of the MMPI and MMPI-2 are, for the most part, more similar when 5 points are added to MMPI-2 scores.

Not all means of scales of the MMPI-2 matched those of the MMPI, however. Males scored about 5 points lower on scale 8 and about 10 points lower on scale 5 of the MMPI-2 as compared to the MMPI. Males and females both scored about 5 points higher on scale L when using the MMPI-2. Females scored about 10 points higher on scale 5 when using the MMPI-2.

Scores on scales L, 5, and 8 for males and scales L and 5 for females are likely to have contributed to the lack of configural consistency between the MMPI and MMPI-2. Scores on scale L, 5 points higher on average on the MMPI-2, likely contributed to the significantly lower percentages of validity configuration consistency for both genders. Scores on scale 8, 5 points lower on average on the MMPI-2 for males, likely contributed to decreased configural consistency for males. Scale 5 certainly appears to have caused a decrease in configural consistency, with males scoring an average 10 points lower on scale 5 and females scoring an average 10 points higher on scale 5 when using the MMPI-2. The set of configural analyses including scale 5 evidence a great decrease in consistency as compared to the set of analyses

excluding scale 5. Of the 20 significance tests conducted on configural characteristics, 5 are significant or approach significance when scales 0 and 5 are excluded from the analyses; 13 are significant or approach significance when scales 0 and 5 are included. Well defined code consistency drops about 10% between the O-O and O-R conditions when scales 0 and 5 are excluded and drops 20% and 40% for males and females respectively when scales 0 and 5 are included. Mean differences between the MMPI and MMPI-2 for scales L, 5, and 8 are suggestive of scales in which scores varied between the two instruments resulting in configural shifts.

Results of configural consistency thus far indicate a decrease in configural consistency between the MMPI and MMPI-2 (O-R condition) as compared to a test-retest of the MMPI (O-O condition). Further, analyses suggest that scales L, 5, and 8 for males and scales L and 5 for females likely play a large role in the configural shifts between the MMPI and MMPI-2.

Given that these configural shifts exist, it becomes important to investigate how great the variability is between the configurations of the MMPI and MMPI-2. The mean of the modified D statistic suggests that in MMPI test-retest T scores vary a total of about 58 T score points spread across all 10 clinical scales for males. The mean of the modified D statistic suggests that for males taking the MMPI and MMPI-2 T scores vary a total of about 70 T score

points across all 10 clinical scales. On the average, about 12 T score points in variance are added to a male's profile when going from the MMPI to the MMPI-2. For females on all 10 clinical scales the average added variance is 8 T score points. The average added variance across the 13 basic scales is 18 points for males and 14 points for females.

The modified D statistic suggests that variance is added in going from the MMPI to the MMPI-2, but that this variance may not be as great as expected given the configural analyses. Across the 10 clinical scales 8-12 T score points in variance are added, on average, when going from the MMPI to MMPI-2. Another 6 points in variance is added when including the validity scales. The variance seems even less striking when considering that it may be concentrated on a few scales; scales L, 5, and 8 for males and scales L and 5 for females.

The sample used in this study may be a reason for the differences in what the configural analyses suggest as compared to the modified D statistic. The sample, volunteer college students, are from a "normal" population. About half of the subjects in each condition produced at least one WNL profile. The profiles produced by these students were often not highly differentiated, meaning that there was often little variability between scale scores on a given profile. Because the profiles were not highly differentiated, profile configuration characteristics could be easily

effected by small changes in scores. Well defined code consistency, which may be argued is less sensitive to small changes in scores, is not significantly lower in the O-R condition unless scale 5 is included in the analysis. Small changes in scores, particularly concentrated on a few scales, seem to have resulted in shifts in configural characteristics on often undifferentiated profiles.

Implications of Analyses

Are there changes between scores and configural patterns between the MMPI and MMPI-2? Configural analyses suggest that changes do occur. The modified D statistic indicates, however, on average, a moderate 15 T score points in variance are added across the basic scales when using the MMPI-2. Further, analyses suggest that this variance seems to be concentrated on a few scales: L, 5, and 8 for males and L and 5 for females. These analyses suggest that, overall, the MMPI-2 seems to produce scores and configural patterns that are similar to the MMPI, with the possible exception of the above mentioned scales.

The source of the score changes for scales L, 5, and 8 between the MMPI and MMPI-2 could be the item changes and/or the normative changes (new normative group and uniform T scores) between the two instruments. Knowing the source of the changes has important implications. If the source of changes in scores is item changes (wording changes, deletions, changes in item presentation order), then the reli-

ability and validity of the scales become suspect. If the source of changes in scores is normative changes, then scoring adjustments may be employed to correct for discrepancies between the MMPI and MMPI-2.

Correlations are used in this study to analyze the effect of item changes. Test-retest reliability of the MMPI (O-O condition) is compared, in essence, to the parallel forms reliability of the MMPI and MMPI-2 (O-R condition). Lower correlations in the O-R condition suggest that the same subjects taking parallel forms of the same scale rank differently in relation to each other depending upon which form of the scale is used. Lower correlations not only indicate low parallel forms reliability of the scale, but lower validity as well, as reliability sets the upper limits on validity (Walsh & Betz, 1990). The application of the vast amount of validity research on the MMPI to the MMPI-2 becomes questionable. Scale L for both genders and scales 5 and 8 for males all had lower correlations in the O-R condition and all resulted in different mean scores between the MMPI and MMPI-2. The parallel forms reliability and the validity of these scales is, therefore, suspect. The only scale to result in different mean scores without having lower correlations in the O-R condition is scale 5 for females. It is likely that the difference in mean scores for scale 5 for females between the MMPI and MMPI-2 is not due to item changes but normative changes.

An article by Munley (1991), released subsequent to the collection of data in the present study, suggested that all of the scales in question (L, 5, and 8) may have been effected by normative changes between the MMPI and MMPI-2. Munley provided a direct comparison of MMPI and MMPI-2 norms, noting the absence of such information in the MMPI-2 manual. Given a raw score on a scale, Munley showed, in chart form, the corresponding T scores on the MMPI and MMPI-2. Munley's analysis did not consider item changes between the instruments. Munley described complex relationships between the scores of the MMPI and MMPI-2, noting that "a few MMPI-2 scales show enhancement, others show attenuation across the full range of the scales, and others show attenuation in the middle and lower ranges of the scales with enhancement in the upper ranges" (Munley, 1991, p. 87). The results of Munley's comparisons between MMPI and MMPI-2 T scores are consistent with the findings of the present study, with T scores on the MMPI-2 comparatively higher on scale L and lower on scales 5 and 8 for males and with T scores of the MMPI-2 comparatively higher for scales L and 5 for females.

It seems then, given the results of the present study and Munley's (1991) article, that the scores for scale 5 for females seem to change because of normative changes between the MMPI and MMPI-2. Scoring changes for scale L for both genders and scales 5 and 8 for males could be the result of

item changes, normative changes, or an interaction between the two.

The question of whether scale scores change because of item changes versus normative changes is an important one. Normative changes may be coped with by employing score adjustments. For example, the results of this study suggest that if MMPI-2 scores for both genders on scale L were reduced by about 5 T score points, they would be comparable to MMPI scores. Further, if scale 5 was increased by 10 points and scale 8 by 5 points for males and if scale 5 was reduced by 10 points for females, profiles, on average, would be comparable. The use of scoring adjustments to compensate for differences in the MMPI and MMPI-2 would be inappropriate for scales with low parallel forms reliability and correspondingly low validity, however.

The issue of score adjustments can become complicated, especially if one considers the demographic characteristics of the sample used in this study. Interpretive manuals for the MMPI (Duckworth & Anderson, 1986; Graham, 1981, 1987; Greene, 1980) routinely suggest scoring adjustments for college students taking the MMPI. Many of the scoring adjustments suggested by these manuals for college students on the MMPI are already made on the MMPI-2 through its score discrepancies with the MMPI. The implication here is that for practitioners who are in the habit of making scoring adjustments for college students on the MMPI, no scoring

adjustments may be needed on scales L, 5, and 8 to "translate" an MMPI-2 profile to an MMPI profile.

Complications with score adjustments aside, it is important to distinguish between score changes due to item changes versus normative changes. Scores for females on scale 5 appear to be due to normative changes, whereas scores on scale L for both genders and scales 5 and 8 for males seem to be at least in part due to item changes, but may also be effected by normative changes.

Limitations and Future Research

The present study is limited in a number of ways, and it is from these limitations that suggestions come for future research on the consistency between the MMPI and MMPI-2. The sample size in this study is small considering the number of scales under analysis. Psychometric standards suggest 10 subjects for each scale analyzed (Megargee, 1979). Correlational analysis of the basic scales ideally require 130 subjects for each condition within gender. This may explain, in part, why results differed between the present study and Ben-Porath and Butcher (1989a). Given the importance of determining if item changes cause score changes on scales, it would seem important that such an analysis utilize a larger sample size. Future research should further increase sample size if other scales (supplemental, Harris-Lingoes, etc.) are included in the analysis. Score adjustments suggested by this study must be considered

tentative, as they are based only on data from the O-R condition (92 males, 96 females). Continued investigation of score changes between the MMPI and MMPI-2 with college students is warranted to firmly establish the pattern of changes.

Another limitation of the study is its generalizability given the population analyzed. The sample in this study is from a "normal" population resulting in many undifferentiated profiles. Replication with samples drawn from more clinical populations would be important. It would be interesting to see if configural consistency increased between the MMPI and MMPI-2 given more highly differentiated profiles. Also, it would be interesting to see if patterns in the correlational analyses are similar for a clinical population.

Caution must be used in generalizing the results of this study to populations outside that of college students. College students are known to differ in their responses to the MMPI from the general population (Duckworth & Anderson, 1986; Graham, 1981, 1987; Greene, 1980). Item changes may affect different populations in different ways. Score adjustments to translate MMPI-2 scores to MMPI scores suggested by this study may be applicable only to college students. Further study would be needed to investigate score adjustments with other adult populations.

Summary

In broad terms, results of the present study are favorable toward MMPI and MMPI-2 consistency, showing that variance in scores between the two inventories exists but is not very large. Results suggest, for the most part, that past validity studies and resulting interpretive information of the MMPI can be applied to the MMPI-2. These results, within the context of MMPI-2 advantages such as less offensive items, contemporary normative data, and new content scales, suggest that the MMPI-2 may be a useful assessment device.

Optimism about the MMPI-2 is tempered, however, by the potential psychometric shortcomings of a few scales and by the limited generalizability of the present study. Results indicate scoring changes on scales L, 5, and 8 between the MMPI and MMPI-2, the nature of which warrants further investigation. Further investigation is also needed to determine whether the results of this study apply to populations other than college students.

References

- Adler, T. (1990, April). Does the 'new' MMPI beat the classic'? APA Monitor, pp. 18-19.
- Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.
- Ben-Porath, Y. S., & Butcher, J. N. (1989a). The comparability of MMPI and MMPI-2 scales and profiles. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1, 345-347.
- Ben-Porath, Y. S., & Butcher, J. N. (1989b). The psychometric stability of rewritten MMPI items. Journal of Personality Assessment, 53, 645-653.
- Butcher, J. M., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). Minnesota Multiphasic Personality Inventory (MMPI-2): Manual for administration and scoring. Minneapolis: University of Minnesota Press.
- Butcher, J. M., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1990, July). Tangible improvements [Letter to the editor]. APA Monitor, p. 2.
- Butcher, J. M., Graham, J. R., Dahlstrom, W. G., & Bowman, E. (1990). The MMPI-2 with college students. Journal of Personality Assessment, 54, 1-15.
- Butcher, J. N., & Hostetler, K. (1990). Abbreviating MMPI item administration: What can be learned from the MMPI for the MMPI-2? Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 12-21.
- Butcher, J. N., & Tellegen, A. (1966). Objections to MMPI items. Journal of Consulting Psychology, 30, 527-534.
- Cattell, R. B. (1949). R_p and other coefficients of pattern similarity. Psychometrika, 14, 279-298.
- Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Colligan, R. C., Osborne, D., & Offord, K. P. (1980). Linear transformation and the interpretation of MMPI T scores. Journal of Clinical Psychology, 36, 162-165.

Colligan, R. C., Osborne, D., Swenson, W. M., & Offord, K. P. (1983). The MMPI: A contemporary normative study. New York: Praeger.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. Psychological Bulletin, 50, 456-473.

Dahlstrom, W. G., & Dahlstrom, L. (Eds.). (1980). Basic readings on the MMPI: A new selection on personality measurement. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., & Welsh, G. S. & Dahlstrom, L. E. (1972). An MMPI handbook: Vol. I. Clinical interpretation. Minneapolis: University of Minnesota Press.

Dahlstrom, W. G., & Welsh, G. S. & Dahlstrom, L. E. (1975). An MMPI handbook: Vol. II. Research applications. Minneapolis: University of Minnesota Press.

Drake, L. E. (1946). A social I.E. scale for the MMPI. Journal of Applied Psychology, 30, 51-54.

Drake, L. E., & Oetting, E. R. (1959). An MMPI codebook for counselors. Minneapolis: University of Minnesota Press.

Duckworth, J., & Anderson, W. (1986). MMPI interpretation manual for counselors and clinicians. Muncie, IN: Accelerated Development.

Duckworth, J., & Barley, W. D. (1989). Within-normal-limits profiles. In R. L. Greene (Ed.), The MMPI: Use with specific populations (pp. 278-315). Philadelphia: Grune & Stratton.

Edwards, A. L. (1957). The social desirability variable in personality assessment and research. New York: Dryden.

Edwards, A. L. (1964). Social desirability and performance on the MMPI. Psychometrika, 29, 295-308.

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

Gilberstadt, H., & Duker, J. (1965). A handbook for clinical and actuarial MMPI interpretation. Philadelphia: Saunders.

Gocka, E. (1965). American Lake norms for 200 MMPI scales. Unpublished materials, Veterans Administration Hospital, American Lake, WA.

Graham, J. R. (1979). Using the MMPI in counseling and psychotherapy. In J. Butcher, G. Dahlstrom, M. Gynther, & W. Schofield (Eds.), Clinical notes of the MMPI (No. 1). Minneapolis: National Computer Systems.

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

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

Graham, J. R., & McCord, G. (1985). Interpretation of moderately elevated MMPI scores for normal subjects. Journal of Personality Assessment, 49, 477-484.

Greene, R. L. (1980). The MMPI: An interpretive manual. New York: Grune & Stratton.

Greene, R. L. (1990). The MMPI: An interpretive manual (2nd ed.). New York: Grune & Stratton.

Groth-Marnat, G. (1984). Handbook of psychological assessment. New York: Van Nostrand Reinhold.

Groth-Marnat, G. (1990). Handbook of psychological assessment. (2nd ed.). New York: Van Nostrand Reinhold.

Gynther, M. D. (1989). MMPI comparisons of blacks and whites: A review and commentary. Journal of Clinical Psychology, 45, 878-883.

Gynther, M. D., Altman, H., & Sletten, I. W. (1973). Replicated correlates of MMPI two-point types: The Missouri Actuarial System. Journal of Clinical Psychology (Suppl. 39).

Gynther, M. D., & Green, S. B. (1980). Accuracy may make a difference, but does a difference make for accuracy? A response to Pritchard and Rosenblatt. Journal of Consulting and Clinical Psychology, 48, 268-272.

Harris, R., & Lingo, J. (1955). Subscales for the Minnesota Multiphasic Personality Inventory. Mimeographed materials, The Langley Porter Clinic.

Harris, R., & Lingo, J. (1968). Subscales for the Minnesota Multiphasic Personality Inventory. Mimeographed materials, The Langley Porter Clinic.

Hathaway, S. R., & Briggs, P. F. (1957). Some normative data on new MMPI scales. Journal of Clinical Psychology, 13, 364-368.

Hathaway, S. R. (1956). Scales 5 (masculinity-femininity, 6 (paranoia), and 8 (schizophrenia). In G. S. Welsh & W. G. Dahlstrom (Eds.), Basic readings of the MMPI in psychology and medicine. Minneapolis: University of Minnesota Press.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. Journal of Psychology, 10, 249-254.

Hathaway, S. R., & McKinley, J. C. (1942). A multiphasic personality schedule (Minnesota): III. The measurement of symptomatic depression. Journal of Psychology, 14, 73-84.

Hathaway, S. R., & McKinley, J. C. (1983). The Minnesota Multiphasic Personality Inventory manual. New York: Psychological Corporation.

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. Journal of Psychosomatic Research, 11, 213-218.

Kaplan, R. M., & Saccuzzo, D. P. (1982). Psychological testing: Principles, applications, and issues. Monterey, CA: Brook/Cole.

Katz, M. M. (1968). A phenomenological typology of schizophrenia. In M. M. Katz, J. O. Cole, & W. E. Barton, (Eds.), The role and methodology of classification in psychiatry and psychopathology (Public Health Service Publication #1584). Washington, D. C.: U. S. Government Printing Office.

King, G. D. (1978). Minnesota Multiphasic Personality Inventory. In O. K. Buros (Ed.), Eighth mental measurements yearbook (pp. 935-938). Highland Park, NJ: Gryphon.

Koss, M. P. (1979). MMPI item content: Recurring issues. In J. N. Butcher (Ed.), New developments in the use of the MMPI (pp. 3-38). Minneapolis: University of Minnesota Press.

Koss, M. P., & Butcher, J. N. (1973). A comparison of psychiatric patients' self-report with other sources of clinical information. Journal of Research in Personality, 7, 225-236.

- Kunce, J., & Anderson, W. (1976). Normalizing the MMPI. Journal of Clinical Psychology, 32, 776-780.
- Kunce, J., & Anderson, W. (1984). Perspectives on uses of the MMPI in nonpsychiatric settings. In P. McReynolds & G. J. Chelune (Eds.), Advances in Psychological Assessment. San Francisco: Jossey-Bass.
- Lachar, D. (1974). The MMPI: Clinical assessment and automated interpretation. Los Angeles: Western Psychological Services.
- Lachar, D., & Wrobel, T. A. (1979). Validating clinicians' hunches: Construction of a new MMPI critical item set. Journal of Consulting and Clinical Psychology, 47, 277-284.
- Levitt, E. E. (1990). A structural analysis of the impact of MMPI-2 on MMPI-1. Journal of Personality Assessment, 55, 562-577.
- Lubin, B., Larsen, R. M., & Matarazzo, J. B. (1984). Patterns of psychological test usage in the United States: 1935-1982. American Psychologist, 39, 451-454.
- Marks, P. A., & Seeman, W. (1963). Actuarial description of abnormal personality. Baltimore: Williams & Wilkins.
- Marks, P. A., Seeman, W., & Haller, D. L. (1974). The actuarial use of the MMPI with adolescents and adults. Baltimore: Williams & Wilkins.
- Mauger, P. A. (1980). Predicting response to treatment using the MMPI. In J. Butcher, G. Dahlstrom, M. Gynther, & W. Schofield (Eds.), Clinical notes of the MMPI (No. 5). Minneapolis: National Computer Systems.
- McKinley, J. C., & Hathaway, S. R. (1940). A multiphasic personality schedule (Minnesota): II. A differential study of hypochondriasis. Journal of Psychology, 10, 255-268.
- McKinley, J. C., & Hathaway, S. R. (1942). A multiphasic personality schedule (Minnesota): IV. Psychasthenia. Journal of Applied Psychology, 26, 614-624.
- McKinley, J. C., & Hathaway, S. R. (1944). The Minnesota Multiphasic Personality Inventory: V. Hysteria, hypomania, and psychopathic deviate. Journal of Applied Psychology, 28, 153-174.

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. (1948). The MMPI: VI. The K scale. Journal of Consulting Psychology, 12, 20-31.

Meehl, P. E. (1951). Research results for counselors. St. Paul, MN: State Department of Education.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the MMPI. Journal of Applied Psychology, 30, 525-564.

Megargee, E. I. (1979). How to do publishable research with the MMPI. In J. Butcher, G. Dahlstrom, M. Gynther, & W. Schofield (Eds.), Clinical notes of the MMPI (No. 2). Minneapolis: National Computer Systems.

Messick, S., & Jackson, D. N. (1961). Acquiescence and the factorial interpretation of the MMPI. Psychological Bulletin, 58, 299-304.

Mitchell, J. V., Jr. (Ed.). (1985). The ninth mental measurements yearbook. Lincoln, NE: The Buros Institute of Mental Measurements.

Moreland, K. L. (1985). Test-retest reliability of 80 MMPI scales. Minneapolis: National Computer Systems.

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

Osgood, C. E., & Suci, G. J. (1958). A measure of relation determined by both mean difference and profile information. Psychological Bulletin, 49, 251-262.

Pritchard, D. A., & Rosenblatt, A. (1980). Racial bias in the MMPI: A methodological review. Journal of Consulting and Clinical Psychology, 48, 263-267.

Rodgers, D. A. (1972). The MMPI: A review. In O. K. Buros (Ed.). Seventh mental measurements yearbook (pp. 243-250). Highland Park, NJ: Gryphon.

Swenson, W. M., Pearson, J. S., & Osborne, D. (1973). An MMPI source book: Basic item, scale, and pattern data for 50,000 medical patients. Minneapolis: University of Minnesota Press.

Terman, L. M., & Miles, C. C. (1936). Sex and personality: Studies in masculinity and femininity. New York: McGraw-Hill.

U.S. Department of Labor. (1977). Dictionary of Occupational Titles (4th ed.). Washington, DC: U.S. Government Printing Office.

Walsh, W. B., & Betz, N. E. (1990). Tests and assessment (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Watkins, C. E., Campbell, V. L., & McGregor, P. (1988). Counseling psychologists' uses of and opinions about psychological tests: A contemporary perspective. The Counseling Psychologist, 16, 476-486.

Watkins, C. E., Campbell, V. L., McGregor, P., & Godin, K. (1989). The MMPI: Does it have a place in counseling psychology training? Journal of Personality Assessment, 53, 413-417.

Welsh, G. S. (1948). An extension of Hathaway's MMPI profile coding system. Journal of Consulting Psychology, 12, 343-344.

Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: University of Minnesota Press.

Wiggins, J. S. (1969). Content dimensions in the MMPI. In J. N. Butcher (Ed.), MMPI: Research developments and clinical applications. New York: McGraw-Hill.

Wiggins, J. S. (1973). Personality and prediction: Principles of personality assessment. Reading, MA: Addison-Wesley.

Wiener, D. N., & Harmon, L. R. (1946). Subtle and obvious keys for the MMPI: Their development (Advisement Bulletin No. 16). Minneapolis: Regional Veterans Administration Office.

Appendix A
Demographic Questionnaire

DEMOGRAPHIC QUESTIONNAIRE

AGE: _____

SEX: _____

COLLEGE MAJOR: _____

LAST FOUR DIGITS OF
SOCIAL SECURITY NUMBER: _____MARITAL STATUS (CIRCLE ONE):
Single MarriedCLASS STANDING (CIRCLE ONE):
Freshman Sophomore Junior SeniorETHNICITY (CIRCLE ONE):
White (non-Hispanic) Black
Asian Hispanic
Native AmericanOCCUPATION(S) OF PARENT(S): _____
_____HIGHEST LEVEL OF
EDUCATION ACHIEVED BY
FATHER (CIRCLE ONE):

did not complete high school	completed high school degree	completed associates degree or trade school	completed bachelors degree	completed masters degree	completed Ph. D. or professional degree
---------------------------------------	------------------------------------	--	----------------------------------	--------------------------------	--

HIGHEST LEVEL OF
EDUCATION ACHIEVED BY
MOTHER (CIRCLE ONE):

did not complete high school	completed high school degree	completed associates degree or trade school	completed bachelors degree	completed masters degree	completed Ph. D. or professional degree
---------------------------------------	------------------------------------	--	----------------------------------	--------------------------------	--

Appendix B
Scale Names and Abbreviations
for the MMPI and MMPI-2

SCALE NAMES AND ABBREVIATIONS OF THE MNPI AND MMPI-2

Basic Scales

- L - L Scale
- F - F Scale
- K - K Scale
- 1 - Scale 1 (Hypochondriasis)(Hs)
- 2 - Scale 2 (Depression)(D)
- 3 - Scale 3 (Hysteria)(Hy)
- 4 - Scale 4 (Psychopathic Deviate)(Pd)
- 5 - Scale 5 (Masculinity-Femininity)(Mf)
- 6 - Scale 6 (Paranoia)(Pa)
- 7 - Scale 7 (Psychasthenia)(Pt)
- 8 - Scale 8 (Schizophrenia)(Sc)
- 9 - Scale 9 (Hypomania)(Ma)
- 0 - Scale 0 (Social Introversion)(Si)

Supplemental Scales

- A - Anxiety
- R - Repression
- Es - Ego Strength
- MAC - MacAndrew Alcoholism
- Do - Dominance
- Re - Social Responsibility
- O-M - Overcontrolled-Hostility
- Nt - College Maladjustment

Harris-Lingoes Subscales

- D1 - Subjective Depression
- D2 - Psychomotor Retardation
- D3 - Physical Malfunctioning
- D4 - Mental Dullness
- D5 - Brooding
- Hy1 - Denial of Social Anxiety
- Hy2 - Need for Affection
- Hy3 - Lassitude-Malaise
- Hy4 - Somatic Complaints
- Hy5 - Inhibition of Aggression
- Pd1 - Familial Discord
- Pd2 - Authority Problems
- Pd3 - Social Imperturbability
- Pd4 - Social Alienation
- Pd5 - Self Alienation
- Pa1 - Persecutory Ideas
- Pa2 - Poignancy
- Pa3 - Naivete
- Sc1 - Social Alienation
- Sc2 - Emotional Alienation
- Sc3 - Lack of Ego Mastery, Cognitive
- Sc4 - Lack of Ego Mastery, Conative
- Sc5 - Lack of Ego Mastery, Defective Inhibition
- Sc6 - Bizarre Sensory Experiences
- Ma1 - Amoralisity
- Ma2 - Psychomotor Acceleration
- Ma3 - Imperturbability
- Ma4 - Ego Inflation

Appendix C
Written Instructions -
First Administration

PLEASE READ ALL INSTRUCTIONS CAREFULLY

THANK YOU FOR YOUR PARTICIPATION IN THIS STUDY.

IN ORDER TO RECEIVE CREDIT IN THIS STUDY YOU MUST BE ABLE TO RETURN FOR TESTING EXACTLY 1 WEEK FROM TODAY.

IN THIS STUDY YOU WILL BE TAKING A PSYCHOLOGICAL INVENTORY WHICH WILL BE DISTRIBUTED IN A FEW MINUTES. THE INVENTORY WILL CONTAIN A LARGE NUMBER OF QUESTIONS. PLEASE ANSWER ALL ITEMS AS OPENLY AND HONESTLY AS POSSIBLE. THE INSTRUMENT IS DESIGNED TO DETECT RANDOM RESPONDING, SO PLEASE ANSWER THE ITEMS CAREFULLY IN ORDER TO AVOID PRODUCING INVALID DATA.

YOU WILL NOT BE ASKED FOR YOUR NAME ON THIS INVENTORY. ALL RESPONSES WILL BE MADE ANONYMOUSLY.

IF ANYONE WOULD LIKE TO TALK ABOUT THE INVENTORY AFTER TAKING IT, SEE THE TEST PROCTOR. THE PROCTOR WILL GIVE A FULL DEBRIEFING OF THE STUDY AFTER THE SECOND TEST SESSION, 1 WEEK FROM TODAY.

PLEASE FILL OUT THE DEMOGRAPHIC QUESTIONNAIRE ATTACHED TO THIS INSTRUCTION SHEET. THE INVENTORY WILL BE PASSED OUT SHORTLY. ONLY #2 PENCILS CAN BE USED ON THE INVENTORY. IF YOU DO NOT HAVE A #2 PENCIL, THE PROCTOR WILL PROVIDE YOU WITH ONE.

IF YOU HAVE ANY QUESTIONS, PLEASE ASK THE PROCTOR

Appendix D
Written Instructions
Second Administration

PLEASE READ ALL INSTRUCTIONS CAREFULLY

IN THIS PHASE OF THE STUDY YOU WILL BE TAKING AN INVENTORY WHICH WILL BE DISTRIBUTED IN A FEW MINUTES. THE INVENTORY WILL CONTAIN A LARGE NUMBER OF QUESTIONS. SOME OF YOU MAY NOTICE THAT THIS IS A SIMILAR INVENTORY TO THE ONE THAT YOU TOOK LAST WEEK. CONTINUE TO RESPOND TO ITEMS OPENLY AND HONESTLY, IN ORDER TO HELP AVOID INVALID RESULTS.

YOU WILL NOT BE ASKED FOR YOUR NAME ON THIS INVENTORY. ALL RESPONSES WILL BE MADE ANONYMOUSLY.

THE INVENTORY WILL BE PASSED OUT SHORTLY. ONLY #2 PENCILS CAN BE USED ON THE INVENTORY. IF YOU DO NOT HAVE A #2 PENCIL, THE PROCTOR WILL PROVIDE YOU WITH ONE.

FOR ANYONE INTERESTED, THERE WILL BE A FULL DEBRIEFING OF THE EXPERIMENT IMMEDIATELY FOLLOWING TODAY'S TESTING SESSION.

IF YOU HAVE ANY QUESTIONS, PLEASE ASK THE PROCTOR