Comparative Validity of the MMPI-2 Restructured Clinical Scale Scores of Native Hawaiian and Other Pacific Islander and Caucasian Mental Health Center Clients

Brian Scott Kehoe
COMPARATIVE VALIDITY OF THE MMPI-2 RESTRUCTURED CLINICAL
SCALE SCORES OF NATIVE HAWAIIAN AND OTHER PACIFIC ISLANDER
AND CAUCASIAN MENTAL HEALTH CENTER CLIENTS

By

BRIAN SCOTT KEHOE

A Dissertation submitted to the
Department of Educational Psychology and Learning Systems
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2010
The members of the Committee approve the Dissertation of Brian Scott Kehoe
defended on April 28, 2009.

__________________________________
Frances Prevatt
Professor Directing Dissertation

__________________________________
Lee Stepina
Outside Committee Member

__________________________________
Briley Proctor
Committee Member

__________________________________
Gary Peterson
Committee Member

Approved:

________________________________________________________________________
Akihito Kamata, Chair, Department of Educational Psychology and Learning
Systems

________________________________________________________________________
Marcy P. Driscoll, Dean, College of Education

The Graduate School has verified and approved the above named committee
members.
I would like to express my gratitude for the support that my family, friends and faculty provided for me over the past several years. My dissertation could not have been written without their guidance and professional expertise. I would like to thank Drs. Frances Prevatt, Gary Peterson, Briley Proctor and Lee Stepina who served as members of my doctoral committee.

Special thanks to Dr. Prevatt, for the time and effort she put into my research. Her supervision, knowledge, encouragement and professionalism were invaluable. I will always be grateful for her support.

I would like to thank Dr. Kubiak who helped me to appreciate and value the practical utility of the MMPI-2. His guidance and professional supervision motivated me to strive for excellence.

I am thankful for Dr. John Graham, who first helped me to visualize the potential need for this research. His knowledge and expertise in the field of personality assessment was much appreciated.

I would like to thank the Hawaii Department of Health and the Hawaii Adult Mental Health Division for valuing the importance of my research.

Thank you to all of my friends who helped me through the dissertation process. Special thanks to Tam and Annie, Lilian and Joy, BK, Marvin, Dan, Turi, Matt, Sid, Kolo, Eddie, the Chase family, and the Han family for their moral support. I also want to thank all of the guys that I played soccer with these past few years as well as Florida State University’s athletic teams for the entertainment that I enjoyed from their consistent success.

I would especially like to thank my parents, Dan and Carol Kehoe for being such great examples for me throughout my life. Their dedication to God, family, and friends has helped me to value each of these in my own life. Their unwavering support and love have always helped me to reach my dreams. I also want to thank my sisters, Mandy, Michelle and Darcie for their encouragement and love. Their support helped me to remain focused throughout the dissertation process.
Finally, I thank my lovely wife, Jennifer for always supporting my interests, including Florida State football. I will always enjoy the football that she got for me, personalized and autographed by Coach Bobby Bowden, as a celebration of the end of my dissertation. Her kindness towards others will always be appreciated and valued. Her sacrifice and encouraging approach often helped me to persevere through the more difficult times of my dissertation. I also want to thank my dog Uncle Sam, for his energy and companionship. He always knows how to make me smile.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>viii</td>
</tr>
</tbody>
</table>

1. Introduction .................................. 1
   - Development of MMPI .......................... 1
   - MMPI Item Selection ......................... 4
   - MMPI Clinical Scale Development .......... 5
   - MMPI Response Sets ........................... 7
   - Widespread Use of the MMPI ................. 10

2. Literature Review ............................ 12
   - Concerns about the MMPI ..................... 12
   - Restandardization ............................ 14
   - Preparation of the National Norms .......... 15
   - Finalizing MMPI-2 Item Selection .......... 19
   - Derivation of Standardized Scores ......... 21
   - Establishing Pathology ...................... 23
   - New Validity Measures ....................... 24
   - Cross Cultural Use of the MMPI-2 .......... 24
   - Current Overview of the MMPI-2 ............. 25
   - Problems Associated with the MMPI-2 ....... 28
   - Interscale Correlation ....................... 29
   - Heterogeneity ................................. 32
   - A Solution ..................................... 33
   - RC Scales Development ....................... 33
   - Improved Psychometric Properties .......... 38
   - Need for Validity Studies across Racial Group 41
   - Population Trends ............................. 44
   - Native Hawaiian and Other Pacific Islanders 45
   - NHOPIs as a Unique Racial Group ........... 47
   - Mental Health and NHOPIs .................... 48
   - Racial Group Comparisons Using the RC Scales 50
   - Statement of Problems and Research Question 51

Research Questions ............................ 53

Research Question 1a: Do NHOPI people tend to score significantly different than Caucasian people on the following nine MMPI-2 RC scales: Demoralization (Rcd), Somatic Complaints (RC1), Low Positive Emotions (RC2), Cynicism (RC3), Antisocial Behavior (RC4), Ideas of Persecution (RC6), Dysfunctional
Negative Emotions (RC7), Aberrant Experiences (RC8), and Hypomanic Activation (RC9)? ......................... 53
Research Question 1b: How are the demographic variables gender, education and age related to mean RC scale differences?................................................ 53
Research Question 2: Is there evidence of test bias using the MMPI-2 RC scale scores............................................. 53

3. Methodology ............................................................................. 54

Statement of the Hypotheses ............................................ 54
Participants................................................................................ 55
Instruments........................................................................ 60
Procedure.......................................................................... 62
Data Analyses ........................................................................ 62
  Research Question 1a ......................................................... 62
  Research Question 1b ......................................................... 63
  Research Question 2 ......................................................... 63
Selection of Predictor and Criterion Variables ................... 63
Slope and Intercept Bias ................................................... 66

4. Results ..................................................................................... 69

Research Question 1a................................................................. 69
Research Question 1b................................................................. 71
Research Question 2................................................................. 74
  Restructured Clinical Scale 4 ....................................... 78
  Restructured Clinical Scale 1 ....................................... 78
  Restructured Clinical Scale 2 ....................................... 78
  Restructured Clinical Scale 3 ....................................... 78
  Restructured Clinical Scale 7 ....................................... 78
  Restructured Clinical Scale 8 ....................................... 78
  Restructured Clinical Scale 9 ....................................... 79

5. Discussion ................................................................................. 80

APPENDIX .................................................................................. 87

IRB/IACUC Approval Letter and Informed Consent Form . 87
REFERENCES .............................................................................. 91

BIOGRAPHICAL SKETCH ............................................................ 101
## LIST OF TABLES

1. Demographic Characteristics of the Sample ............................................... 59
2. Predicted Associations between Restructured Clinical Scales and Patient Description Form Scales .......................................................... 65
3. Individual Means for Racial Group, Gender and Age on the MMPI-2 RC Scales ............................................................................... 70
4. Three-way Means for Racial Group, Gender, and Age on the MMPI-2 RC Scales ............................................................................... 73
5. Correlations between MMPI-2 RC Scale Scores and Therapist Ratings for Caucasian and Native Hawaiian and Other Pacific Islanders ............................................................................... 75
6. Regression Analyses of Patient Description Form Variables on Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Restructured Clinical Scales, Racial Background, and Scale x Racial Background Interaction Term (n = 232) ............................................................................... 77
ABSTRACT

This study evaluated ethnic differences on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) in 126 Caucasian and 106 Native Hawaiian and Other Pacific Islander (NHOPI) mental health center outpatients. MMPI-2 Restructured Clinical (RC) scale scores were first evaluated to determine the extent or influence that the demographic variables of race, gender, age, and education have on identified group MMPI-2 RC scale differences. When assessing for racial group mean differences, five scales studied were identified to be of significance. When assessing for age mean differences, two scales were observed to have been of significance, although there was no significant interaction between race and age for either of these scales. No differences in gender were observed and there were no gender by race interactions. The findings suggest that gender and age had less effect on how a participant would respond to MMPI-2 questions than race. Because preliminary analyses identified a racial group difference by education, this variable was controlled for through a MANCOVA. With the exception of RC scale RC8, racial group differences were not confounded by education. These results were generally consistent with the observed racial group differences on five of the MMPI-2 RC scales and the available extratest criterion data. To assess for test bias, a step-down hierarchical multiple regression procedure was used in which relevant clinical criteria were predicted from the RC scales with race as a potential moderator. Of the 19 scale criterion comparisons, nine demonstrated statistically significant intercept bias of which five achieved a small effect size. Two scales from the regression analyses revealed statistically significant slope bias, both of which demonstrated a small effect size. Although several of the MMPI-2 RC scales suggested some overprediction and underprediction of psychopathology in Native Hawaiian and Other Pacific Islanders, it was determined that racial group membership was not a substantive moderator between MMPI-2 scales scores and clinically relevant phenomena because the effect size of race on prediction was uniformly small. It was recommended that the MMPI-2 RC scales not be
interpreted differently for Caucasian and Native Hawaiian and Other Pacific Islanders, as the magnitude of the predictor error did not warrant separate interpretive guidelines.
CHAPTER 1

INTRODUCTION

Almost 65 years ago, the Minnesota Multiphasic Personality Inventory (MMPI) was introduced by Hathaway and McKinley (1940) as a means of understanding the psychological problems of patients in psychiatric and medical settings (Butcher, 2006). The MMPI test was originated by Starke Hathaway, a medical psychologist, and Jovian McKinley, a neuropsychiatrist, who at the time were working at the University of Minnesota Hospitals (Meyer & Weaver, 2007). According to Starke Hathaway (1965), sheer frustration led him and J.C. McKinley to begin research in 1939 that eventually resulted in the publication of the MMPI in 1943. (Butcher, 2006; Butcher & Williams, 2000). They developed the MMPI to assist themselves and others at the University of Minnesota Hospitals in the routine tasks of assessing and diagnosing patients with mental disorders (Butcher & Williams, 2000). During the 1930s and 1940s a primary function of psychologists and psychiatrists was to assign appropriate psychodiagnostic labels to individual cases (Graham, 2006). Although an individual interview or mental status exam were ordinarily used with each patient, Hathaway and McKinley believed that a group-administered paper-and-pencil personality inventory would provide a more efficient and reliable way of arriving at appropriate psychodiagnostic labels (Graham, 2006).

Development of the MMPI

Hathaway and McKinley believed that the best way to learn what was troubling an individual was to ask him or her (Butcher & Williams, 2000). Consequently, they chose for their inventory statements with which the client could agree or disagree, using a “True” or “False” response (Butcher & Williams, 2000). This belief led Hathaway and McKinley to begin working on the test in the late 1930s and in 1940 they published their first article on the inventory (Hathaway & McKinley, 1940), initially called the Minnesota Personality Schedule (Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom & Kaemmer, 2001). In this
article they summarized the steps they had followed in writing and editing the
items, having gained ideas for potential items from many sources, including a
number of texts on psychiatric interviewing and differential diagnosis, a large
collection of self reference statements from textbooks, psychological reports,
social and emotional attitudes, and personality processes (Meyer & Weaver,
2007; Butcher et al., 2001; Greene, 2000). To this end, Hathaway and McKinley
(1940) assembled more than 1,000 items and after deleting duplicate items and
items that they considered relatively insignificant for their purposes they arrived
at a pool of 504 (Greene, 2000). The content of these original items reflected the
range of psychiatric, medical and neurological disorders in which Hathaway and
McKinley were interested (Butcher et al., 2001). Although Hathaway and
McKinley did not provide a rationale for deleting potentially useful items, this
procedure was acceptable at the time because they had used an empirical
method of item selection (Greene, 2000). These 504 items were written in the
form of a statement of some personal experience, belief or attitude which they
originally assumed to be relatively independent of one another (Meyer & Weaver,
2007; Butcher et al., 2001). The content of these original items reflected the
range of psychiatric, medical and neurological disorders in which the
investigators were interested (Butcher et al., 2001). At that point, they had no
preconceived notion of whether a particular item was related to the constructs of
interest (Butcher & Williams, 2000).

During this process Hathaway and McKinley (1940) arbitrarily classified
the 504 items under 25 headings as a convenience in handling and in an effort to
avoid duplication. However, they did not attempt to obtain any particular number
of items for a category or to ensure that an item was actually properly classified
in a category (Greene, 2000). Hathaway and McKinley assumed that patients
who endorsed similar symptoms or items in their MMPI pool were diagnostically
more alike than they were different (Butcher & Williams, 2000). For example, an
individual endorsing many symptoms related to having a depressed mood was
likely to be more similar to other depressed patients than to other clinical groups
(Butcher & Williams, 2000). Using this logic, Hathaway and McKinley (1940) next
constructed a series of quantitative scales that could be used to assess various categories of psychopathology. To quantify this relationship between number of psychological symptoms and diagnostic similarity, they developed scales by which individuals could be compared on particular variables (Butcher & Williams, 2000).

This being the case, the next step was to select appropriate criterion groups (Graham, 2006). The inventory was developed on groups of patients and nonpatients (visitors to the wards and clinics who volunteered to take the test during the time spent waiting for friends or relatives receiving medical treatment) (Butcher et al., 2001). The nonpatients criterion group, referred to as the Minnesota normals, consisted of 724 relatives and visitor patients in the University of Minnesota Hospitals (Graham, 2006). This group also included 265 recent high school graduates who were attending pre-college conferences at the University of Minnesota, 265 Work Progress Administration workers, and 254 medical patients at the University of Minnesota Hospitals who were hospitalized for some form of physical disease (Graham, 2006). These hospitalized individuals were determined to have no obvious psychiatric symptomatology (Greene, 2000). This nonpatients criterion group was considered representative of the adult population of the state of Minnesota during the 1930’s: mostly married, ranging in age from 16 to 65 and averaging in their mid thirties, living in small towns or rural areas, with an eighth grade education (Butcher et al., 2001). In addition, all of the persons in this primary normative group were of Caucasian descent because very few members of any ethnic minority other than American Indian resided in Minnesota at that time (Dalhstrom, Welsch & Dalhstrom, 1975). The second major group, referred to as the clinical participants, was made up of psychiatric patients at the University of Minnesota Hospitals (Graham, 2006). This second group included 221 patients representing all of the major psychiatric categories being used clinically at the time of the construction of the test (Graham, 2006).

In constructing the test, Hathaway and McKinley considered the selection of scale items based on face validity, the general practice of test developers at
the time, to be too subjective (Butcher & Williams, 2000). Instead, they developed the MMPI on the basis of a more empirical approach of item and scale validity, that is, they required that any item on a scale be assigned to a scale only if it objectively discriminated a given criterion group (e.g., individuals with depression) from their normative sample (Butcher & Williams, 2000). This meant that items had to be answered differentially by the nonpatient criterion group compared with the clinical inpatient group (Greene, 2000). In distinguishing groups based on presenting symptomatology, subgroups were first created by dividing participants into categorical groupings of discrete diagnostic samples according to their clinically determined diagnostic labels (Graham, 2006; Meyer & Weaver, 2007). Whenever there was any doubt about a patient’s clinical diagnosis or when more than one diagnosis was given, the patient was not included in this clinical reference group (Graham, 2006). The various subgroups of clinical participants formed were, Hypochondriasis, Depression, Hysteria, Psychopathic Deviate, Paranoia, Psychasthenia, Schizophrenia, and Hypomania (Graham, 2006; Meyer & Weaver, 2007; Butcher et al., 2001).

**MMPI Item Selection**

Once the criterion group and other normative groups were established, the process of item selection began (Greene, 2000). This next step in scale construction involved administering the original 504 test items to the Minnesota normals and to the patients in each of the clinical groups (Graham, 2006). For the criterion group and each of the normal groups, the frequency of “true” and “false” responses was calculated for each item (Greene, 2000). Since Hathaway and McKinley’s approach was strictly empirical and no theoretical rationale was posited as the basis for accepting or rejecting items on a specific scale, it is not always possible to discern why a particular item distinguishes the criterion group from normal groups (Greene, 2000). In spite of this reality, an item analysis was conducted separately for each of the clinical groups in order to identify the items in the pool of 504 that differentiated groups significantly (Graham, 2006). An item was considered significant and was tentatively selected for a scale if the difference in frequency of response between the criterion group and the
normative groups was at least twice the standard error of the proportion of “true”/“false” responses of the two groups being compared (Greene, 2000). Hathaway and McKinley considered significant any percentage difference of at least twice the standard error of the independent proportions, or any Z value equal to or greater than +2 (Greene, 2000). Since a Z of +2 has a probability slightly less than .05 using a two-tailed test, they essentially selected only items that were significant beyond the .05 level (Greene, 2000). Although this process proved to be successful in initially identifying group differences, Hathaway and McKinley established additional criteria that had to be met before any item would become part of a particular scale. They first established that the frequency of the criterion group’s response had to be greater than the normative group’s response by 10 percent for nearly all items (Greene, 2000). This standard meant that some of the item response rates from the criterion group were excluded even if they were significant statistically because they represented so few criterion cases (Greene, 2000). Items whose responses appeared to reflect biases on variables such as marital status or socioeconomic status were additionally excluded from test construction (Greene, 2000). Any remaining items that after a rationale inspection were concluded to not be germane to the construct of the particular scale in question were also excluded from subsequent construction (Greene, 2000). Although this process was first used in the construction of Scale 1 (Hypochondriasis), the procedure for developing this scale typifies the process by which most of the clinical scales were created. As a result, most of the individual MMPI items that were identified by this procedure were included in the resulting MMPI scale for that clinical group (Graham, 2006). For several scales (Hypochondriasis and Depression) additional procedures were utilized to try to ensure that the scales would differentiate between patients with the diagnosed clinical condition and other persons who had some of the symptoms of the condition but were not assigned to the corresponding diagnosis (Graham, 2006).

**MMPI Clinical Scale Development**

Soon after the development of Scale 1 (Hathaway & McKinley, 1940), five other clinical scales were developed: 2 (Depression) (Hathaway & McKinley,
1942); 7 (Psychasthenia) (McKinley & Hathaway, 1942); and 3 (Hysteria), 4
(Psychopathic Deviate) and 9 (Hypomania) (McKinley & Hathaway, 1944). It was
not until several years later and after considerable preliminary work on the
schedule that Hathaway and McKinley added new items to cover the gender role
characteristics and a defensive style of self-presentation (Butcher et al., 2001).
These additions included scales 5 (Masculinity-Femininity), 6 (Paranoia), and 8
(Schizophrenia) which were not published until 1956 (Hathaway, 1956), yet had
been used routinely for more than a decade (Greene, 2000).

Scale 5 (Masculinity-Femininity) was developed somewhat differently than
the other clinical scales (Graham, 2006; Greene, 2000). After the initial MMPI
item pool had been collected from the original normative sample, 55 items some
of which were from the Terman and Miles Attitude Interest Analysis Test (1936)
which were mostly related to sexual orientation were added to the item pool
(Greene, 2000). Being that the criterion group of male homosexuals who were
used in developing scale 5 could not be contrasted with the normative group on
these items, 54 male soldiers were used as the normative group for this scale,
and items that distinguished them from the male homosexuals were included on
Scale 5 (Greene, 2000). Because of difficulties in identifying adequate numbers
of items that subsequently differentiated between these two groups, Hathaway
and McKinley broadened their approach in the construction of Scale 5 (Graham,
2006). These variations subsequently included incorporating items that were
differentially endorsed by men and women in the normal samples (Graham,
2006; Greene, 2000).

Later a tenth scale, 0 (Social Introversion), developed at the University of
Wisconsin by L.E. Drake (Drake, 1946) was added to the MMPI, completing the
standard MMPI clinical profile (Greene, 2000). Scale 0 was also constructed
differently from other clinical scales (Greene, 2000). Drake selected items for
Scale 0 by contrasting item response frequencies that differentiated 50 female
college students who scored above the 65th percentile on the Minnesota T-S-E
Inventory (Benton, 1949; Evans & McConnell, 1941) from 50 female students
who scored below the 35th percentile (Greene, 2000; Graham, 2006). This cross
validation included comparing scores of females who had participated in many extracurricular activities with those who participated in few or no extracurricular activities (Graham, 2006). The Minnesota T-S-E Inventory used in creating this scale assesses introversion-extroversion in three areas: thinking (T), social (S), and emotional (E) (Greene, 2000). Although interested in all three areas of the inventory Drake limited most of his investigation of introversion-extroversion only in the social area as assessed by the Minnesota T-S-E Inventory. Although Drake conducted much of his analysis with women, the subsequent utilization of this scale into the standard MMPI profile was later extended to men as well as women (Graham, 2006).

These scale additions brought the original number of items to 550 although some versions of the MMPI at that time included the duplication of 16 items to facilitate the machine scoring available at the time, which brought the total number of items to 566 (Butcher et al., 2001).

**MMPI Response Sets**

In answering these items Hathaway and McKinley were aware that individuals sometimes fail to provide a veridical self-report in responding to personality inventory items. Greene (2000) has identified several reasons for an individual’s inaccurate self description. First, although persons constructing the test items generally assume that each item has essentially the same meaning to all persons taking the test, this assumption is not always appropriate (Greene, 2000; Lewak, Nichols, & Webb, 2001). Second, although self-ratings provided through item responses can be useful because direct observations of behavior are often impractical, impossible, or inefficient, individuals vary in their self-awareness and in their ability or willingness to report the appropriate behaviors. Third, the rationale method of test construction also requires that the test developer be knowledgeable about the relationship between person’s responses to individual items and the construct being assessed. Although these issues unquestionably exist in the interpretation of the content of individuals items on the MMPI, they do not invalidate it (Greene, 2000). The empirical approach to item selection used by Hathaway and McKinley, in fact, freed them of these
problems because it assumes that the client’s self-report is just that and makes no a priori assumptions about the relationships between the clients self-report and the clients behavior. As previously stated, items were selected for inclusion in a specific scale only because the criterion group answered the items differently than the normative group irrespective of whether the item content was actually an accurate description of the criterion group (Butcher et al., 2001; Butcher, 2000; Greene, 2000).

Because Hathaway and McKinley (1940) developed the MMPI under the banner of empiricism, they recognized that the honesty or frankness with which the client responds to the items needs to be assessed empirically each time the MMPI is administered rather than blithely assume that the client has answered the items appropriately (Greene, 2000). In response to this awareness that test takers could falsify or distort their responses to the items in self-report inventories, McKinley, Hathaway and Meehl (1948) additionally developed four scales, hereafter referred to as the validity scales, to detect deviant test taking attitudes (Graham, 2006). In assessing the test-taking attitudes of a client it was understood that a client might adopt a test taking attitude other than that desired by the test developer, thus creating a need to measure a clients perceived tendency to do this (Greene, 2000). Three of the scales designed to detect this potential were included in the initial publication of the MMPI in 1942, which provided the user with indicators of the validity of the answers of any given test taker (Butcher et al., 2001). These initial validity scales included a measure designed to record the number of items in the inventory that were left unanswered (Cannot Say score), a measure of defensive role playing (L scale), and a measure of extremely deviant or random responding to the test (F scale) (Graham, 2006; Butcher et al., 2001). More specifically, the Cannot Say scale recognized that the omission of large number of items, would likely lower the scores on the clinical scales, thus calling into question the interpretability of the whole resulting profile of scores (Graham, 2006). The L scale, originally called the Lie scale of the MMPI, was designed to detect rather unsophisticated and naïve attempts on the part of test takers to present themselves in an overly
favorable light (Graham, 2006). The Infrequency (F) scale of the MMPI was
designed to detect individuals whose approach to the test-taking task is different
from that intended by the test authors (Graham, 2006). These indicators together
helped the test user evaluate the possibility that the test record was spoiled or
marred by the respondent’s failure to comply in one fashion or another with the
test instructions (Butcher et al., 2001).

The need for these measurements was first acknowledged by Meehl and
Hathaway (1946) as they became convinced of the necessity of assessing two
dichotomous categories of test taking attitudes: defensiveness (“faking good”)
and plus-getting, or (“faking bad”) (Greene, 2000). A faking good response style
has the effect of lowering clinical scales; in many assessment contexts,
respondents may be consciously or unconsciously motivated by personal goals
to underreport symptoms of psychopathology (Bagby, Marshall, Bury, Bachiochi
& Miller, 2006). In contrast, a faking bad response style has the effect of
elevating clinical scales which are typically intentional and engaged with the full
awareness of the respondent (Bagby et al., 2006). Given that underreporting and
overreporting response styles have the potential to invalidate a protocol and that
such responding likely occurs in a variety of important assessment contexts, the
assessment and accurate detection of response biases are critical aspects of
MMPI interpretation (Bagby et al., 2006).

Although Meehl and Hathaway (1946) considered a number of
approaches to assess these two categories of test-taking attitudes which all
helped in the development of the aforementioned validity scales, later, a fourth
validity indicator, the K (correction) scale (McKinley, Hathaway & Meehl, 1948),
was added. This scale was designed to appraise the possibility that the test-taker
had answered the MMPI with a more subtle but pervasive tendency either to
cover up (high scores) or to exaggerate (low scores) his or her problems and
difficulties (Butcher et al., 2001). The K (correction) scale of the MMPI was thus
created to identify clinical defensiveness in test takers (Graham, 2006). In
developing the K scale, Meehl and Hathaway (1946) attempted to reduce the
number of false negatives produced by the MMPI scoring (i.e., indicating no
evidence of psychopathology when it exists) (Bagby et al., 2006). It was noted that some clearly abnormal persons who took the MMPI obtained scores on the clinical scales that were not as elevated as would be expected given their clinical status (Graham, 2006). As a result, the K scale was constructed selecting only those items that distinguished persons exhibiting significant pathology and whose overall profiles were within the normal range of scores from individuals not exhibiting psychopathology who produced elevated profiles (Nichols, 2001). As such the K scale is now recognized as a more subtle measure of impression management then the L scale (Butcher et al., 2001).

As an extension to its role in the set of validity indicators, the K scale score was also used later in adding corrective weights to five of the scales in the clinical set (Butcher et al., 2001). Meehl and Hathaway reasoned that if the effect of a defensive test-taking attitude, as reflected by a high K score, is to lower the scores on the clinical scales, it might be possible to determine the extent to which the scores on the clinical scales should be raised in order to reflect more accurately a person’s behavior (Graham, 2006). By comparing the diagnostic efficiency of each clinical scale with various portions of K-scale scores added as a correction factor, Meehl and Hathaway empirically determined the proportions of K, that when added to a clinical scale, would maximize the discrimination between the criterion group and the normative group (Greene, 2000). In doing this certain clinical scales were not K corrected because the simple raw score on those clinical scales seemed to result in the most accurate prediction about a person’s clinical condition while other scales have proportions of K, ranging from .2 to 1.0, which are added to adjust the clinical scales appropriately (Graham, 2006).

**Widespread Use of the MMPI**

By 1950 the basic format for the MMPI appeared to be set (Butcher, 2001). Acceptance of the test grew steadily in the United States and in translation throughout the world (Butcher et al., 2001). Despite the MMPI’s origins in a single psychiatric service in Minnesota, it became the most widely used and researched objective personality inventory in the world (Meyer &
Weaver, 2007; Lees-Haley, Smith, Williams & Dunn, 1996). In fact, 50 years after its first publication Butcher and Rouse (1996) found over 4,300 references to the MMPI. Shortly after its initial development the MMPI was commonly employed with patients in general medical settings, adolescents in schools, inmates in correctional facilities, individuals in drug and alcohol problem treatment units, military personnel, and eventually applicants in industrial settings who applied for highly responsible or stressful positions, such as airline pilots, police officers, or nuclear power plant operators (Butcher & Williams, 2000). Counseling psychologists and community counselors used the test extensively (Watkins, Campbell, & McGregor, 1988), and clinicians who might seem unlikely to value the MMPI reported it to be an important assessment instrument (Graham, 2006). The use of the MMPI was even endorsed by members of the Society for Personality Assessment, a group typically associated with projective techniques, as well as by members of the American Association for Behavior who additionally indicated that it was important for professional clinics to be skilled in the use of the MMPI (Piotrowski, Shery, & Keller, 1985; Piotrowski, & Keller, 1984).
CHAPTER 2

LITERATURE REVIEW

Concerns about the MMPI

Despite its reputation as a sound psychological assessment procedure and its ability to provide reliable evaluations consistently across various administrations, by the 1960s, a number of new nonclinical settings into which it had been introduced (e.g., employment screening, admission to academic programs and military induction), began to develop some concerns about some of the MMPI items dealing with sexual adjustment, bodily functions, and religious matters (Butcher et al., 2001). During this time Hathaway (1965) even indicated that the MMPI could be criticized for “its perpetuation of the Kraeplin-derived diagnostic nosology” (p. 462), which emphasized that all psychiatric diseases are caused by biological and genetic disorders. These problems became more evident with the changes in psychiatric diagnosis, particularly with the transformations in the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 1952, 1968, 1980, 1987). In addition to this, some began to question the appropriateness of the original normative sample regarding contemporary use in clinical settings (Greene, 1987). The MMPI was standardized on friends and family members in the university hospitals at the University of Minnesota in the late 1930’s. The sample was one of convenience, and little effort had been made to ensure that it was representative of the U.S. population (Graham, 2006). Although it is not known for sure, it is probable that few if any, minority individuals were represented in the original normative sample (Greene, 1987). Standardized participants came primarily from one geographic area around Minneapolis, Minnesota, almost all of them were white, and the typical person was about 35 years of age, married, residing in a small town or rural area, working in a skilled or semiskilled trade (or married to a man of this occupational level), and having about 8 years of formal education (Dahlstrom, Welsh, & Dahlstrom, 1972). In addition to this, as American culture changed,
concerns were also expressed about sexist wording, outmoded idiomatic expressions, and references to increasingly unfamiliar literary material and recreational activities (Butcher et al., 2001). By the 1970s, others noted problems with the original MMPI concluding that many of the items in the inventory were out of date or objectionable and recommended that the instrument be revised by deleting obsolete items and broadening the item pool to include more contemporary themes (Butcher & Williams, 2001). For example, few items concerned suicide attempts, use of drugs other than alcohol, and treatment-related behaviors (Graham, 2006). Although many additional scales were developed using the original MMPI item pool, the success of these scales often was limited by the inadequacy of the item pool (Graham, 2006). Although these concerns created many questions regarding the continued professional endorsement of the MMPI, perhaps the greatest threat to its popularity came from the growing evidence that people were endorsing some of the items in substantially different ways, making the need for contemporary national norms apparent (Colligan, Osborne, Swenson, & Offord, 1983; Butcher et al., 2001).

In response to many of these concerns, Hathaway and McKinley articulated plans to collect data from a substantially larger number of individuals for their nonpatient samples. However, funding limitations in the period of the economic depression forced them to settle for less than a third of their goals (Butcher et al., 2001). Although the sample that they were able to gather at that time matched the Minnesota population of the 1930’s in terms of age range, educational level, and socioeconomic background, few Black, Native American, or other minority members were recruited by their sampling procedures (Butcher et al., 2001). Although most researchers and users of the MMPI at that time began to recognize the need for test revision and restandardization, the enormity of the task and the remaining unavailability of funds delayed revision plans for years (Graham, 2006). Finally in 1982, Beverly Kaemmer, MMPI manager at the University of Minnesota Press, appointed a committee to undertake the restandardization of the MMPI. James N. Butcher (University of Minnesota) and W. Grant Dahlstrom (University of North Carolina) began the work, joined later
that year by John R. Graham (Kent State University) and in 1986 by Auke Tellegen (University of Minnesota) (Butcher & Williams, 2001). Funds to support the revision were provided by the University of Minnesota Press while the test distributor, National Computer Systems (now Pearson Assessments), provided support in the form of test materials, forms, and scanning and scoring data (Graham, 2006). These efforts initiated a nationwide sampling program to remedy many of the limitations identified with the original test norms (Butcher et al., 2001).

Restandardization

From the start of the restandardization project, it was determined that every effort would be made to maintain continuity between the original MMPI and its revision (Graham, 2006). This decision was popularly endorsed because new criterion groups and item derivation procedures were not going to be used on the standard Validity and Clinical Scales (Greene, 2000). Although major revisions of the existing validity and clinical scales were not part of the restandardization project, it was hoped that the project would produce data that later could lead to improvements in the basic scales (Graham, 2006; Greene, 2000). In the first year of the project, the committee decided to develop two separate experimental booklets, one for adults (Form AX) and one for adolescents (Form TX), for use in the data collection (Butcher & Williams, 2001). In addition to having a primary goal of collecting a contemporary normative sample that would be more representative of the general population than had been true for Hathaway’s original sample, efforts would be made to improve the MMPI item pool by rewriting some of the items that would expand the content dimensions of the item pool (Graham, 2006). Items measuring new content (e.g., suicidal behavior, treatment readiness, Type A behaviors, and problematic alcohol and other drug use) were added to both experimental booklets and developmentally relevant items were added to the appropriate booklets (e.g., work adjustment items were added to Form AX and school adjustment items to Form TX (Butcher & Williams, 2001).
Preparing the experimental booklet (Form AX) involved several simultaneous processes (Graham, 2006). To maintain continuity between the original and revised forms of the MMPI, a decision was made to include all 550 unique items in the Form AX (Graham, 2006). Of the 550 original items, they reworded 141 to eliminate outdated and sexist language and to make these items more easily understood (Greene, 2000). Rewording these items did not change the correlations of the items with the total scale score in most cases (Ben-Porath & Butcher, 1989). In addition to rewriting dated, difficult and sexist items, the restandardization project deleted 16 repeated items originally only included to facilitate machine scoring (Meyer & Weaver, 2007). Another major change in the restandardization project involved adding new items (Graham, 2006). After seeking recommendations from experts in personality measurement and clinical assessment concerning content dimensions that should be added to the pool, 154 provisional items were generated, bringing the item total to 704 which was used to collect the normative data for the MMPI-2 (Graham, 2006; Butcher et., 2001; Greene, 2000). Although some of the new items were an alternate version of existing items introduced to determine whether they would constitute improvement over the original scales, most of these additions were designed to provide better coverage of topics and areas of concern than did the original item pool: family functioning, eating disorders, substance abuse, readiness for treatment or rehabilitation, and interference with performance at work (Butcher et al., 2001).

Preparation of the National Norms

The next step of the restandardization project involved obtaining a large normative group that was broadly representative of the U.S. population (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). Collateral forms were created to gather biographical and supplementary information about the sample of adults whose responses would be used to establish the new test norms. Census data from 1980 were used to guide participant solicitation (Graham, 2006). Individuals between the ages of 18 and 90 were then contacted through a variety of methods, most by direct mail from directories and advertising lists, and at one
site, Chapel Hill, North Carolina, some participants were solicited by advertisements and special appeals, as well as by follow-up contact with persons listed in catchment area rolls (Butcher et al., 2001). To ensure geographic representativeness, seven testing sites (Minnesota, Ohio, North Carolina, Washington, Pennsylvania, Virginia, and California) were selected (Graham, 2006). In addition, individuals were added proportionately to the sample from groups of individuals tested on a federal Indian reservation (Tacoma, Washington area) and on four military bases, since these people would not have been contacted through the other methods (Butcher et al., 2001).

After an initial trial period, it was decided that individuals would be paid $15 for their participation, and couples that participated together would be paid $40 (Graham, 2006). All subjects were administrated the 704-item experimental Form AX of the MMPI, a biographical questionnaire, and a questionnaire assessing significant life events in the past 6 months called the Dyadic Adjustment Questionnaire (Filsinger, 1983), (Butcher & Williams, 2001). Heterosexual couples (N = 800) additionally completed two other forms, describing the nature and length of their relationships and were asked to rate each other on 110 characteristics, using a revised version of the Katz Adjustment Scale (Katz, 1968). This information provided validity descriptors for the MMPI-2 scales with nonclinical samples (Butcher & Williams, 2001).

Using these procedures over 2,900 individuals were initially tested for inclusion in the restandardization sample (Butcher et al., 2001). After eliminating persons because of test invalidity or incompleteness of other forms, in the late 1980s a final sample of 2,600 community participants (1,138 men and 1,462 women) was constituted, which included 841 couples (Graham, 2006). The most common reason for dropping an individual from the sample was an unacceptable MMPI protocol: excessive item omissions (40 or more of the 704 items) or an excessively high score (20 or more) on either the F scale or Back F (Fb), derived from the latter part of the AX test booklet. Additional sources of problematic records were incomplete or missing biographical or recent life-events forms, and omitted birth dates or gender identification (Butcher et al., 2006)
In addition to including individuals residing in seven different states, the normative sample for the MMPI-2 varied significantly from the original normative sample on a number of demographic variables. Graham (2006) has reported that the racial composition of the MMPI-2 sample was as follows: Caucasian, 81%; African-American, 12%; Hispanic, 3%; Native American, 3%; and Asian American, 1%. These percentages were designed to reflect the 1990 national census parameters on ethnicity (Greene, 2000). Although there was appropriate representation of blacks and whites, Hispanic and Asian American subgroups were underrepresented in the normative sample (Butcher et al., 2001). Alternatively, Native Americans were somewhat overrepresented in the normative sample (Butcher et al., 2001). Although distinctions have been made between the terms race and ethnicity, much of the literature examining the validity of the MMPI tests across racial groups has used these terms interchangeably. As a result it should be noted that all subsequent reference to either of these terms should be interpreted similarly, in that race and ethnicity are used interchangeably.

MMPI-2 participants ranged in age from 18 to 85 years (M = 41.04; SD = 15.29) and in formal education from 3 to 20+ years (M = 14.72; SD = 2.60) (Graham, 2006). Although individuals in the normative sample for the MMPI-2 are more representative of the United States as a whole, similar comparisons between census data and the normative sample for age and education reveal that the most disparate feature of the community sample in the comparison with the U.S. Census data is in their educational background (Butcher et al., 2001). In the original MMPI, educational factors were considered to influence, to some extent, performance on two MMPI basic scales: K and Mf, thus requiring that the new MMPI-2 national census parameters accurately be utilized in the restandardization (Butcher & Williams, 2001). In the normative sample, there is an excess of adult men and women with college and post graduate education and an under-representation of those who completed high school or who did not obtain a high school diploma (Butcher et al., 2001). The potential impact of this higher level of education in the MMPI-2 normative sample has been a focus of
ongoing concern (Helmes & Reddon, 1993); however, research involving a census matched subsample within the MMPI-2 restandardization sample found only one difference that exceeded 3 T score points between these two samples on the Standard Validity and Clinical scales, Content scales, and Supplementary scales (Schinka, Lalone & Greene, 1998; Greene, 2000). These studies determined that although a relatively small amount of variance in the clinical scales was found to be associated with demographic variables such as gender, aside from clinical scale 5, gender did not account for any significant amount of variance in scale elevation. Other studies comparing mean scores by gender have suggested that other than clinical scale 5, some minimal to moderate effects of gender on the clinical scales have been identified (Caldwell, 1997). These findings suggested that women consistently endorse more items on clinical scales 1 (Hypochondriasis), 2 (Depression), 3 (Hysteria), 7 (Psychasthenia) and 0 (Social Introversion), while men consistently endorse more items on clinical scale 9 (Hypomania). Butcher et al. (1989) has also reported that clinical scale 4 (Psychopathic Deviate) raw score means tend to be higher for men than women. Greene (2000) has reported that although gender differences when found tend to be small, previous research indicates that women may be more willing to report negative symptoms and behaviors than men. These findings may demonstrate that although some demographic variables may have a subtle effect on MMPI-2 scales and items, there remains no simple, direct significant relationship between any single demographic variable and scores on a given MMPI-2 scale (Greene, 2000). Although the MMPI-2 normative sample has an average education level closer to the contemporary subjects taking the test (15.0 years for men and 14.4 for women) than those used in developing the MMPI, all educational levels except the lowest used in creating the normative sample for the original MMPI, produce mean profiles that match the full MMPI-2 normative sample. These findings indicate that for most people the new norms for the MMPI-2 can be applied and compared with the MMPI without any needed adjustment (Butcher & Williams, 2001).
Additional demographic variables indicate that most men (61.6%) and women (61.2%) in the new sample were married and that approximately 32% of the men and 21% of the women had professional or managerial positions, and approximately 12% of the men and 5% of the women were laborers (Graham, 2006). Social class factors, except in the very low socioeconomic level on two of the standard scales, have not proven to be of much importance in interpreting the MMPI profiles – at least not sufficiently important to recommend the development of special norms for various socioeconomic groups (Butcher & Williams, 2001). As a result, no additional norms were created for the MMPI-2 sample which reported a median family income of $30-35,000 for men and $25-30,000 for women (Graham, 2006). In summary, although the normative sample for the MMPI-2 is more representative of the general population than was Hathaway’s original sample, researchers like Schinka and LaLone (1997) have concluded that the small demographic differences (including education and gender) that exist between the restandardization sample and the 1990 census data are not clinically significant.

**Finalizing MMPI-2 Item Selection**

Although research suggested that demographic variability accounts for little variation in MMPI-2 profile differences, to provide data necessary for making decisions, such as which items from the Form AX booklet would be included in the final revised booklet, data were collected from a variety of additional groups. These individuals included psychiatric patients, alcoholics, chronic pain patients, marital counseling clients, college students, and job applicants (Graham, 2006). When finalizing the items to be included on the MMPI-2, the Restandardization Committee deleted 77 items from the original MMPI in addition to 13 items deleted from the standard Validity and Clinical scales and the 16 repeated items (Greene, 2000). Despite these modifications, only relatively small changes occurred in the items that contribute to the 3 Validity Scales and the 10 basic Clinical Scales, which can be scored if the client completes the first 370 items (Meyer & Weaver, 2007). These scales as previously mentioned were all kept intact in the MMPI-2 to ensure continuity.
between the two versions of the instrument for clinical and research purposes (Butcher, Graham, & Ben-Porath, 1995). Consequently, most special and research scales that have been developed on the MMPI are still capable of being scored unless the scale has an emphasis on religious content or the items are drawn predominantly from the last 150 items on the original MMPI (Greene, 2000). This was fortunate, as it left the essence and great bulk of all the research concerning these scales directly applicable to the MMPI-2 (Butcher, 2006; Ben-Porath, 1989). As a result, the final version of the revised MMPI (MMPI-2) includes 567 items from the Form AX booklet (Graham, 2006). Although the rationale for including and dropping items from Form AX that resulted in the 567 items on the MMPI has not been made explicit (Greene, 2000), Graham (2006) has offered several criteria employed in deciding which items were to be included in the final booklet. First, all items entering into the standard validity and clinical scales were provisionally included, as were items needed to score supplementary scales judged to be important, while other items were maintained because they would be included in new scales developed from the item pool (Graham, 2006). While other items were deleted due to objectionable content, the MMPI-2 remained relatively similar in most ways to the original MMPI (Graham, 2006). As such, although the MMPI was revised in 1989 to update its norms and to expand its measurement scope by adding new items and developing new scales, a large portion of the original MMPI is contained in the MMPI-2 (Butcher, Graham, & Ben-Porath, 1995).

Due to the similarity between the two instruments, investigators who have collected data on the original version can convert MMPI scores to MMPI-2 scores by deleting the 13 items that were dropped from the original scales using the modified raw scores to derive T scores from MMPI-2 norms (Butcher, Graham, & Ben-Porath, 1995). Before developing T-score conversions however, the potential impact of item changes and/or modifications was evaluated by examining item-endorsement shifts, by computing item-scale correlations (corrected for item membership) on the scales in question, and by a special test-retest study by Ben-Porath and Butcher (1989) comparing responses to the
original and the rewritten MMPI items. They found generally, that the consistency of responses of the group administered both original and rewritten items did not differ significantly from that of those who were administered the original items twice (Butcher et al., 2001). For those exceptional items whose response patterning did change significantly, no appreciable differences were found when comparing their contribution to the clinical and supplementary scales with that of the original (Butcher et al., 2001; Butcher, Graham, & Ben-Porath, 1995).

Although these modifications did not alter any of the psychometric properties of the MMPI-2, present-day subjects, including individuals from the new normative sample, tend to endorse more items in the pathological direction, thus producing higher mean scores [approximately 5 T-score points on every scale except Scales L (Lie), 1 (Hypochondriasis), and 0 (Social Introversion)] than the original MMPI normative sample (Butcher & Williams, 2001). Thus, the transition to the MMPI-2 norms meant that the new profiles are slightly less elevated when compared to the original MMPI norms (Greene, 2000). These findings are believed to occur because a somewhat different set of instructions are used today (Butcher & Williams, 2001). Butcher and Williams (2001) have additionally explained that originally item omissions were allowed, even encouraged, but that in current practice, test administrators tend to encourage completing all of the items, consequently, the original MMPI norms are inaccurate for today’s test usage. Despite the differences that exist between the normative groups for the MMPI and the MMPI-2, the new norms, based on contemporary instructions, should allow for more accurate assessment (Graham, 2006; Butcher & Williams, 2001). Subsequent research has resulted in the MMPI-2 replacing the MMPI in virtually all settings (Graham, 2006; Butcher & Williams, 2001).

**Derivation of Standardized Scores**

The development of the new normative standards has altered the MMPI profile (Ward, 1991). Despite the large number of modifications made to the test, the most important changes incorporated into the MMPI-2 are new normative data and an altered method for computing T scores from raw scores for eight of
the standard clinical scales (1, 2, 3, 4, 6, 7, 8, and 9) (Butcher et al., 1989). Linear T scores were retained on Scales 5 and 0 because these two scales were derived in a different manner than the other clinical scales and the distribution of raw scores was less skewed (Butcher et al., 1989). As previously stated new norms were necessary because the original MMPI normative sample is not representative of the U.S. population (Graham, 2006) and because the MMPI normative values are known to deviate importantly from data obtained in contemporary samples (Pancoast & Archer, 1989). It is well known that raw scores on psychological scales are usually not directly interpretable, suggesting that some form of standardization was needed in using the new MMPI-2 normative sample (Butcher et al., 2001). One familiar form is the linear T-score transformation (Butcher et al., 2001). The original MMPI norms were developed using a linear T-score transformation (Hathaway & McKinley, 1940). This process involved establishing T-score distributions that were assigned a mean of 50 and a standard deviation of 10 (Butcher & Williams, 2000). If in addition, the distributions of the scales in question have the same overall shape, then the same linear T score on different scales will also have the same percentile value (Butcher et al., 2001). The fact that the same linear T score value could have different percentile values for different scales was not a psychometrically desirable feature (Butcher et al., 2000). As a result, the original linear T score transformation approach was followed during the restandardization, with an important modification that solved the problem of nonequivalency of percentile values across scales that occurred with these scores (Butcher & Williams, 2000). This modification has now been corrected, and with a minimum change in the original linear T-score distributions, by the derivation of uniform T (UT) scores (Tellegen & Ben-Porath, 1992). To arrive at this composite target distribution, linear T scores were first derived for each of the 16 raw score distributions, using the formula \( T = 50 + \frac{10(X - M)}{SD} \), where \( X \) is the raw score, and \( M \) and \( SD \) are the mean and standard deviation of the raw scores (Butcher et al., 2001). For the eight basic clinical scales involved, regression methods were then developed to transform the raw scores directly into uniform T scores that would conform as
closely as possible to the composite standard or average T score (Butcher et al., 2001). These composite T scores were then used to create uniform T scores for each of the eight clinical scales while maintaining the underlying positive skew in the distribution (Greene, 2000). This new uniform T-score procedure was originally designed to improve the percentile comparability among the eight basic clinical scales while minimizing the departure from the original linear MMPI T scores (Edwards, Morrison, & Weissman, 1993). By transforming each of the 16 MMPI-2 clinical scale distributions (8 male, 8 female) and the 15 MMPI-2 content scales in the normative sample, every uniform T score now has an equivalent percentile across these scales making the distribution of T scores more similar across the clinical scales than would be found using the linear T scores (Edwards, Morrison & Weissman, 1993; Greene, 2000; Butcher et al., 1989). Today uniform T scores are also used for the Personality Psychopathology Five (PSY-5) scales, and the Restructured Clinical (RC) scales while Linear T scores are used for other MMPI-2 supplementary scales (Graham, 2006).

**Establishing Pathology**

In interpreting original MMPI profiles, clinicians followed the strategy of considering a T score of 70 the point at which an elevation was clinically significant because this cutoff was thought to fall at a percentile rank of 95 for each MMPI scale (Butcher & Williams, 2000). However, being that differences in T scores between the MMPI and the MMPI-2 vary a bit by scale and by level of scores on a scale, by on average about 5 T score points, the MMPI-2 manual (Butcher et al., 1989) recommends that MMPI-2 T scores between 65 and 69 are now considered potentially significant in interpreting MMPI-2 results (Keller & Butcher, 1991; Graham, 2006). Consequently, a T score of 65 or greater was chosen to demarcate the “clinical range” on the MMP-2 (Butcher et al., 1989). Each scale now has approximately 4% of the scores above T = 69, and about 8% of the scores are above T = 64 meaning that a T score of 65 falls uniformly at the 92nd percentile for the eight clinical scales discussed and the MMPI-2 content scales (Edwards, Morrison, & Weissman, 1993; Butcher & Williams, 2000).
New Validity Measures

Additional MMPI-2 scales have also been developed since the test was revised including some new validity measures (e.g., Variable Response Inconsistency [VRIN], True Response Inconsistency [TRIN], Back Infrequency [Fb], Infrequency-Psychopathology [Fp], and Superlative Self-Presentation [S]) (Butcher, 2006). The VRIN and TRIN scales were designed to assess if the test taker has responded to the items randomly or without consideration of their content (Graham, 2006). Sometimes persons respond to the MMPI-2 items with the intention of portraying themselves as having more problems and symptoms than they really have. The Fb and Fp scales were designed to detect for this type of overreporting (Graham, 2006). The S scale on the other hand was designed to assess the tendency of some persons to present themselves on the MMPI-2 as highly virtuous, responsible individuals, who are free of psychological problems, have few or no moral flaws, and get along extremely well with others (Graham, 2006). Clinical syndrome measures – such as the content scales and addiction measures ACK (Alcohol Admission Scale) and PRO (Proclivity) – have also come to be widely used (Butcher, 2006).

Cross Cultural Use of the MMPI-2

Today, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is understood to be a broad band test designed to assess a number of the major patterns of personality and psychological disorders (Butcher et al., 2001). With the MMPI revision in the 1980’s, the MMPI-2 remains the most widely used clinical instrument in personality assessment (Butcher, 2006). In fact, survey data not only indicates that the MMPI-2 is the most widely used psychological test in the United States but that it is commonly used in other countries around the world (Graham, 2006). In comparison to other commonly used psychological instruments, data indicates that the MMPI/MMPI-2 is the personality test most frequently used by clinical psychologists (86%), second in use only to the WAIS-R (94%; Camara, Nathan, & Puente, 2000). More than 14,000 books and articles have been published on the MMPI and MMPI-2 instruments (Butcher, Atlis, & Hahn, 2003), and a number of current text books are devoted to introducing the
MMPI-2 to students and practitioners interested in updating their base for information (Graham, 2006; Butcher, 2006; Butcher & Williams, 2000; Greene, 2000, Freidman, Lewak, Nichols, & Webb, 2001). Psychologists who use the MMPI-2 in their practice do so with the confidence that they have more than a half-century of research to support their decisions and recommendations (Graham, Ben-Porath & McNulty, 1999) and that this empirical foundation was built through thousands of investigations over the earlier years of the MMPI’s life (Graham, Ben-Porath & McNulty, 1999).

**Current Overview of the MMPI-2**

Many of the changes introduced by the MMPI-2 included the development of several new scales designed to assist the clinician in interpreting the standard Validity and Clinical scales (Greene, 2000). The MMPI-2 and its numerous scales now represent the restandardization of the MMPI that marks the advent of a new era of clinical usage and research of this venerable inventory (Greene, 2000). A complete list of the MMPI-2 validity scales include the following: Variable Response Inconsistency (VRIN) Scale which measures whether the client is inconsistent in his or her responses, True Response Inconsistency (TRIN) Scale which measures the degree and tendency that a participant may have to answering “true” or “false”, Infrequency (F, Fb, and Fp) Scales which were all designed to detect unusual or atypical ways of answering the test items, the FBS Scale which was designed to measure the extent to which one may be malingering or falsely endorsing pathology, the L (Lie) Scale which is composed of items intended to identify persons who are deliberately trying to avoid answering the MMPI frankly and honestly (Dalhstrom at al., 1972), the K (Correction) Scale which was designed to identify persons who displayed significant psychopathology yet had profiles within the normal range, and the Superlative (S) Scale which was designed to assess persons who present themselves in a superlative manner that is encountered frequently in individuals who are being screened in personnel settings (Green, 2000).

The MMPI-2 clinical scales include Scales 1 through 8. Although each scale is believed to measure a specific component of one's personality, some
intercorrelation between scales has been identified (Tellegen et al., 2003). Generally speaking Scale 1, also identified as the Hypochondriasis (Hs) scale, is expected to measure a patient’s level of concern about their own health (Butcher et al., 2001). Elevated scores on this scale are expected to identify a variety of somatic complaints with little or no organic basis (Butcher et al., 2001). Scale 2: Depression (D) is expected to measure symptomatic depression, which is a general attitude characterized by poor morale, lack of hope in the future, and a general dissatisfaction with one’s own status (Hathaway & McKinley, 1942). Scale 3: Hysteria (Hy) was originally developed by Hathaway and McKinley (1940) with the intention of providing an objective measure of a complex clinical phenomenon referred to as conversion hysteria, which today we call conversion disorder (Butcher & Williams, 2000). Scale 4: Psychopathic Deviate (Pd) measures a number of concerns relating to one’s general social maladjustment and the absence of strongly pleasant experiences. This scale also taps concerns about family and authority figures, as well as self and social alienation and boredom (Green, 2000). Scale 5: Masculinity-Femininity (Mf) is different from the other standard scales in several ways (Butcher & Williams, 2000). The construct underlying its development is not a clinical syndrome and scores on this scale do not seem to be related to symptoms or problems for nonclinical persons (Graham, 2006; Long & Graham, 1991). This being said, the scale was originally developed by Hathaway and McKinley (1940) to identify homosexuality in men. Scale 6: Paranoia (Pa) assesses the behavior pattern of suspiciousness, mistrust, delusional beliefs, excessive interpersonal sensitivity, rigid thinking, and externalization of blame commonly found in paranoid disorders (Butcher & Williams, 2000). Scale 7: Psychasthenia (Pt) is expected to assess the neurotic syndrome of psychasthenia, which is characterized by the person’s inability to resist specific actions or thoughts regardless of their maladaptive nature (Green, 2000). Scale 8: Schizophrenia (Sc) assesses a wide array of content areas, including bizarre thought processes and peculiar perceptions, social alienation, poor familial relationships, difficulties in concentration and impulse control, lack of deep interests, disturbing questions of self-worth and self-identity, and sexual
difficulties (Green, 2000). Scale 9: Mania (Ma) was developed to assess elevated mood, accelerated speech and motor activity, irritability, flight of ideas, and brief periods of depression (Graham, 2006). Scale 0: Social Introversion (Si) much like Scale 5 was not based on a psychiatric syndrome. It is used to assess the social introversion-extroversion dimension with high scores reflecting social introversion (Butcher et al., 2001).

The K-Corrected profile was developed to determine the proportion of K that, when added to the raw score on the clinical scale, would maximize the discrimination between the normative groups and the criterion group (Greene, 2000). McKinley, Hathaway, and Meehl (1948) determined that the discriminations could be improved on five of the clinical scales by the addition of the proportion of K. Thus Clinical Scales 1 (Hs), 4 (Pd), 7 (Pt), 8, (Sc) and 9 (Ma) can be corrected by the addition of a whole raw score or fraction of a raw score of K (Greene, 2000). Raw scores thus are needed in evaluating MMPI-2 data, which if needed can later easily be converted visually to T scores because the T scores are printed at the side of each profile sheet (Graham, 2006).

In addition to its utilization in the construction of the standard validity and clinical scales, the original MMPI item pool was used to develop numerous other scales by variously recombining the 566 items using item-analytic, factor-analytic, and intuitive procedures (Graham, 2006). These additions included the Harris-Lingoes subscales, the Content scales and the Supplementary scales all of which were designed to assist in clarifying the clinical interpretation of MMPI-2 profile elevation created by clinical scale heterogeneity (Graham, 2006). The Harris-Lingoes subscales were designed to provide an analysis of systematic content endorsement of subgroups of items within the standard clinical scales (Graham, 2006). Harris and Lingoes (1968) constructed subscales for 6 of the 10 standard clinical scales (scales 2, 3, 4, 6, 8, and 9). They did not develop subscales for scales 1 or 7 because they considered them homogeneous in content (Graham, 2006). Following a combined rational/empirical procedure, fashioned after the approach used in developing the original MMPI content scales, Butcher, Graham, Williams, and Ben-Porath (1990) constructed a set of
15 content scales for the MMPI-2. These scales were specifically designed to assess content areas introduced in the MMPI-2 and to assess more broadly, via items added in the revision process, areas covered by the original content scales (Butcher et al., 2001). Evidence of the validity of the content scales has been reported by Archer, Aiduk, Griffin, and Elkins (1996) and a comprehensive study of the MMPI-2 content scales’ empirical correlates was reported by Graham et al., (1990) who identified empirical correlates for the scales using the using the Symptom Checklist-90-Revised (SCL-90-R) (Derogatis, 1977) as a criterion. The SCL-90-R is a 90-item self-report inventory designed to reflect the psychological symptom patterns of psychiatric and medical patients (Derogatis, 1983). The Supplementary scales were designed to assess a wide variety of psychological constructs adding to the interpretation of the validity and clinical scales, augmenting the coverage of clinical scale problems and disorders (Butcher et al., 2001). Included in the Supplementary scales are the Weiner Subtle-Obvious subscales. Wiener (1948) differentiated between MMPI items that were easy to detect as indicating emotional disturbance and items that were relatively difficult to detect as indicating emotional disturbance. The former were labeled as obvious items and the latter as subtle items (Graham, 2006). Wiener (1948) rationally developed Obvious and Subtle subscales for scales 2, 3, 4, 6, 8, and 9 of the original MMPI, hypothesizing that test takers who were trying to fake bad on the MMPI would endorse many of the obvious and few of the subtle items in the clinical scales.

**Problems Associated with the MMPI-2**

In contrast to many of the changes introduced by the MMPI-2, the well known Clinical scales were transferred virtually unchanged from the original MMPI to the MMPI-2 (Tellegen et al., 2003). This was done out of a desire, appropriate at the time of the restandardization, to maintain substantial continuity between the MMPI and the MMPI-2, particularly with respect to the Clinical scales with which the MMPI has long been identified (Tellegen et al., 2003, Graham, 2006). Indeed, because of their empirical foundations, the Clinical scales have continued to make a major contribution to clinical interpretation...
The development of these empirically keyed clinical scales, which were designed to facilitate psychiatric diagnoses have, not gone without certain problems (Forbey & Ben-Porath, 2007). These include the high intercorrelations that resulted in large part from the empirical method of their construction, which primarily emphasized convergent validity while largely ignoring discriminant validity (Sellbom & Ben-Porath, 2005). This method essentially consisted of selecting items that differentiated patients with a specifically targeted diagnosis from nonpatients (Sellbom & Ben-Porath, 2005).

Dissatisfied with the transparency of face-valid measures, Hathaway and McKinley (1940) initially relied on empirical scale development for their development of the MMPI scales. Irrespective of content and theory, scale items were selected solely based on their discriminability between criterion groups (Rogers, Sewell, Harrison, & Jordan, 2006; Sellbom, Ben-Porath & Graham, 2006). As a result, any item that statistically differentiated a specific patient group from the normal group was added to a clinical scale regardless of its content or whether it was also scored on another scale, resulting in a lack of discriminant validity due to considerable item overlap and correlated measurement errors across scales (Sellbom, Ben-Porath & Graham, 2006; Sellbom & Ben-Porath, 2005). More importantly, no assumptions were made about the veridicality of the clients’ self-ascriptions regarding either their personality characteristics or past experiences (Roger et al., 2006). As a result, in addition to identifying phenomena unique to a particular diagnostic group, each clinical scale also tends to measure characteristics common to patients across syndromes (Sellbom & Ben-Porath, 2005). Moreover, as some items were found to differentiate groups merely by virtue of comparing a psychiatric group to normals; other items that differentiate a person with emotional distress from a person without emotional distress were also included on each of the clinical scales (Sellbom, Graham, & Schenk, 2006).

**Interscale Correlation**

Despite attempts at clarifying MMPI-2 interpretation, item and construct overlap continue to contaminate the accuracy of MMPI-2 clinical scale
interpretation due to excessively high interscale correlations (Rogers & Sewell, 2006). High clinical-scale intercorrelations have been attributed to the empirical keying approach used by Hathaway and McKinley (1943) to distinguish psychiatric populations from so-called normals (Tellegen et al., 2003), resulting in significant item overlap among the scales. In the MMPI-2, the average number of overlapping items per pair of the 10 standard clinical scales is 6.4 items (Greene, 2000), with notably higher overlap in certain scale pairs: 1-3 (20 items overlapping), 7-8 (17), 2-7 (13), 2-3 (13), 6-8 (13), 4-0 (11), 8-9 (11), 1-2 (10), 3-4 (10), and 4-8 (10) (Simms, Casillas, Clark, Watson, & Doebbeling, 2005). As a result a pair of scales such as Scale 7, Psychasthenia, a putative “neurotic” scale, and Scale 8, Schizophrenia, a putative “psychotic” scale, might have 75% or more of their variance in common (Nichols, 2006). Although not entirely surprising considering that less than 20% of Scale 8 items describe psychotic phenomena, and a similar proportion of its items overlap with Scale 7, the extent of their covariation routinely confounds the interpretation of these two scales (Nichols, 2006; Greene, 2000). Redundancy of this magnitude is problematic for a number of reasons (Simms et al., 2005). First, item overlap results in artifactualy high intercorrelations among scales and decreases their distinctiveness and discriminant validity (Simms et al., 2005). Item overlap also poses significant challenges to understanding the factor structure of the MMPI/MMPI-2 scales and renders the meaning of the individual scale scores unclear, especially when endorsement of common items elevates many scales simultaneously (Forbey & Ben-Porath, Simms et al., 2005; Greene, 2000; Waller, 1999). This undesirable shortcoming may result in multiple elevations on scales that would not be expected to be elevated based on what is known about the test takers presenting problems (Forbey & Ben-Porath, 2007). These excessive correlations sometimes result in floating profiles where many or all of the clinical scale scores exceed the threshold for clinical significance (i.e., $T > 64$) (Wallace & Liljequist, 2005). Floating profiles are challenging to interpret because most of all of the scales are significantly elevated, making it difficult to determine which scale(s) to focus on or which scale(s) actually represent the origin(s) of the
distress reported by the client (Wallace & Liljequist, 2005; Sellbom & Ben-Porath, 2007). For example, substance abusers often present with elevations on clinical Scale 4, a measure of acting out proclivities, but also on Scales 2, 6, 7, and 8, indicators of emotional and thought disturbance (Forbey & Ben-Porath, 2007; Sellbom & Ben-Porath, 2005). Such multiple (and anomalous) elevations often prove to be difficult for clinicians to interpret and require the utilization of the Harris-Lingoes subscales, in addition to the code-type reference materials and other supplementary sources, adding many unnecessary time consuming steps to the interpretative process (Forbey & Ben-Porath, 2007; Sellbom, Graham, & Schenk, 2006).

MMPI-2 researchers have developed a number of techniques to address the challenge to differential assessment posed by the substantial clinical scale intercorrelations (Sellbom & Ben-Porath, 2005). There have been several methods employed over the years to deal with this problem, some of which included an early effort to use code types to control for some of the variance (i.e., general maladjustment) common to all of the Clinical scales (Sellbom, Graham, & Schenk, 2006). Little research however, has demonstrated that code types in fact add unique information above and beyond what can already be extracted from the Clinical scales comprising a particular code type (Sellbom, Graham, & Schenk, 2006). Additional research involving the Harris-Lingoes and Content scales have been more successful in clarifying elevations on the MMPI-2 thus facilitating Clinical scale interpretation (Sellbom, Graham, & Schenk, 2006; Sellbom & Ben-Porath, 2005). In fact, content interpretation as assessed through the Harris-Lingoes and Content scales of the MMPI-2 has been widely accepted as an adjunct to traditional empirical scale interpretation (Butcher & Williams, 2000). These assumptions have generally supported the use of these content oriented scales validity to augment the traditional Clinical scales however; the level of incremental validity is sometimes very modest (Rogers et al., 2006). Archer and colleagues (1996) found the average change in $R^2$ was 2.8% for male inpatients and 2.9% for female inpatients when using the Symptom Checklist-90-Revised (Derogatis, 1977) as a criterion. By relying on these techniques,
however MMPI-2 users are able to disentangle overlapping variance across the clinical scales and identify which of the core elements of the scales to emphasize in an interpretation (Sellbom & Ben-Porath, 2005).

Individuals like Tellegen and colleagues (2003) believed that the clinical scales did contain theoretically meaningful and clinically useful concepts. Therefore these authors proposed a solution to decrease the high intercorrelations among the clinical scales by filtering out a common core feature of each: demoralization (Wallace & Liljequist, 2005). Tellegen et al. (2003) defined demoralization as “a broad affectively colored dimension represented to some degree in each of the clinical scales” (p. 1). Tellegen et al. (2003) hypothesized that by removing the demoralization variance from each scale, a more distinctive and useful scale could be created. Demoralization was considered clinically significant but was not believed to be the specific core component of any single clinical scale and instead was believed to be responsible for a major portion of the variance that consistently inflated the correlations between the clinical scales (Wallace & Liljequist, 2005). After removing the general distress factor of demoralization, Tellegen and colleagues then sought to employ a complex set of steps for refining unidimensional scales (Rogers & Sewell, 2006). These steps began with the goal of identifying and measuring a set of core constructs whose roots are in the clinical scales but that (a) demonstrate greater distinctiveness, (b) are more homogeneous, and (c) disentangle nonspecific demoralization variance from substantive construct variance (Simms, 2006).

**Heterogeneity**

Another problem associated with the development and subsequent interpretation of the MMPI-2s clinical scales concerns the issue of heterogeneity of item content. This concern is exacerbated by the inclusion of “subtle items” which have marginal discriminant validity (Tellegen et al., 2003). Although these items did not contribute to the construct or criterion related validity of the MMPI clinical scales, the Weiner Subtle-Obvious subscales were maintained in the MMPI-2 because some persons believe that they are useful in detecting certain
response sets that invalidate profiles (Graham, 2006). Although originally expected to attenuate the convergent validity of the clinical scales, subsequent research has indicated that many of these items are not valid indicators of clinical scale constructs but rather behave, at least collectively, closer to a randomly composed set of items than valid scales when correlated with conceptually relevant constructs (Ben-Porath & Butcher, 1990). Given the polythetic nature of the MMPI-2 clinical scales, the inclusion of these seemingly irrelevant items only further complicates MMPI-2 interpretation (Rogers & Sewell, 2006).

In response to the long standing concerns regarding the conceptual overlap and heterogeneity among the clinical scales of the MMPI-2, one solution would be to abandon these scales and replace them with a set of differently developed measures (such as are found in other recently published clinical inventories) (Simms, 2006; Tellegen, 2003). No effort would then be made to preserve the distinctive and empirically supported assessment and perspective that the clinical scales have afforded (Tellegen et al., 2003).

**A Solution**

These goals and potential solutions then led to the adoption of a flexible test construction strategy that permitted Tellegen et al. to identify and measure the meaningful and distinct components of each standard clinical scale (Simms et al., 2005). In doing so they described four stages of test construction: (a) creation of a Demoralization scale that “measures a broad, emotionally colored variable that underlies much of the variance common to the MMPI-2 Clinical Scales” (Tellegen et al., 2003, p. 11), (b) identification of distinct “core” components for each clinical scale, (c) development of “seed scales” to tap each core component using items from the standard clinical scales, and (d) derivation of a final set of scales composed of items from the entire MMPI-2 item pool (Simms et al., 2005). These new scales developed using the MMPI-2 item pool would be known as the Restructured Clinical (RC) scales (Graham, 2006).

**RC Scales Development**

As stated, the construction of the RC scales proceeded in four steps, of which the first was to identify and embody a general dimension, Demoralization
(Nichols, 2006). Tellegen et al. (2003) began this process using the original 321 items that form the clinical scales. It was anticipated that removal of the general demoralization factor from the clinical scales would result in a set of restructured scales that would be less intercorrelated and have greater discriminant validity than the original clinical scales (Graham, 2006). The factor of demoralization was first conceptualized through Watson and Tellegen’s (1985) model of self-reported affect and was viewed as corresponding to the pleasantness versus unpleasantness (PU) axis between the orthogonal dimensions of negative affect (NA) and positive affect (PA). Tellegen (1985) theorized that anxiety is in part a product of experiencing high NA, whereas individuals low on PA would be more prone to depression. In developing the RC scales Tellegen et al. (2003) argued that the nonspecific component of the MMPI-2 clinical scales could be modeled as the highest level of affect hierarchy, a pleasantness-unpleasantness dimension, which reflects general hedonic valence. By extracting this variance as an independent scale they hoped to reduce the artificial covariation among, and therefore highlight the distinctiveness of, the remaining clinically meaningful dimensions of the clinical scales (Simms et al., 2005). Because scale 2 and scale 7 are known to be related to anxiety, depression and other emotional distress, these scales were thought to contain items that would assess the demoralization dimension (Graham, 2006). Using four clinical samples made up of men and women in residential substance abuse treatment and men and women receiving inpatient psychiatric care, and the MMPI-2 normative sample, factor analyses of items in scale 2 and scale 7 yielded a set of items marking the Demoralization factor (Graham, 2006; Wallace & Liljequist, 2005). This factor was further correlated with the remaining items in the MMPI-2 item pool that were not included in Clinical scales 2 and 7, resulting in a 24-item Demoralization (RCd) scale (Tellegen, 2003).

The next step in the RC scale development involved identifying the distinct core components of each of the MMPI-2 clinical scales after removal of common demoralization variance (Forbey & Ben-Porath, 2007). This step involved conducting separate factor analyses of the items for each clinical scale
augmented with the earlier selected set of demoralization items (Sellbom & Ben-Porath, 2005). More specifically, for each clinical scale, they conducted principal-components analyses of the scales items plus the Demoralization component (which emerged as the first principal component in all analyses) and a component with content that appeared psychologically coherent and distinctive from all other core components (Simms et al., 2005). For most of the clinical scales, two factors emerged: one for demoralization and one for a core component (Sellbom & Graham, 2006). Of the eight clinical scales for which they developed RC scales, the core component appeared as the second component for five scales – Scales 1, 2, 7, 8, and 9 – and the third component for Scales 3, 4, and 6 (Tellegen et al., 2003).

The third step in constructing the RC scales involved identifying items with high loadings on the factor representing the core of each scale to serve as a “seed” scale for each RC scale (Graham, 2006). This step was designed to enhance the discriminant validity of the restructured scales by removing from the set corresponding to each core component those items that correlated substantially (or overlapped) with measures of the other core constructs (Forbey & Ben-Porath, 2007). These seed scales were refined by various means to reduce overlap and increase internal consistency (Nichols, 2006). These analyses yielded a set of 12 maximally independent “seed” scales, enhancing the discriminant validity of the final RC scales (Forbey & Ben-Porath, 2006; Graham, 2006).

Each of the seed scales (Demoralization plus one for each original clinical scale except for scale 5, which yielded two seed scales representing aesthetic/literary interests and mechanical/physical interests) were then correlated with all other MMPI-2 items in the four clinical samples (Graham, 2006). Each item was added to a seed scale if it correlated above a minimum value with that seed (the “convergence criterion”; Tellegen et al., 2003, p. 19) and below a maximum value with all of the remaining seed scales (the “discriminant criterion”; Tellegen et al., 2003, p. 19); these values were allowed to differ for each scale (Nichols, 2006). A final series of ad hoc adjustments were
then made to these augmented seed scales (including, in some cases, the relaxation of these criteria) to optimize scale content, increase internal consistency, or to increase desired relationships with selected external criteria (Nichols, 2006). In addition to RCd, these analyses yielded eight RC scales whose names reflect the core components assessed by each: Somatic Complaints (RC1), Low Positive Emotions (RC2), Cynicism (RC3), Antisocial Behavior (RC4), Ideas of Persecution (RC6), Dysfunctional Negative Emotions (RC7), Aberrant Experiences (RC8), and Hypomanic Activation (RC9) (Simms et al., 2005). Although all 10 clinical scales were subjected to this procedure, only 8 of the RC scales were ultimately used (Sellbom, Graham, & Schenk, 2006). Notably, seven of these scales represent face valid derivatives of the content of the original clinical scales, but the restructuring markedly changed the content of Scale 3 by including only items tapping cynical views of the world, because the somatic items that originally composed a significant portion of Scale 3 were redundant with the core component of RC1 (Simms et al., 2005). These complex psychometric and statistical procedures culminated in the final RC Scales (Graham, 2006; Nichols, 2006). Figure 1 can be used to visually correlate each RC scale with its clinical scale derivative:
<table>
<thead>
<tr>
<th></th>
<th>RCd</th>
<th>1 (Hs)</th>
<th>2 (D)</th>
<th>3 (Hy)</th>
<th>4 (Pd)</th>
<th>6 (Pa)</th>
<th>7 (Pt)</th>
<th>8 (Sc)</th>
<th>9 (Ma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Visual Correlation of Each RC Scale with its Clinical Scale Derivative Scale
Improved Psychometric Properties

Data from several samples were analyzed to examine the psychometric performance of the RC scales and to conduct comparisons with the clinical scales (Tellegen et al., 2003). This process utilized several samples and procedures, including the MMPI-2 normative sample (1462 women and 1138 men; Butcher et al., 1989, 2001), a sample of 410 male and 601 female adults seeking outpatient psychotherapy (Graham, Ben-Porath, & McNulty, 1999), and an inpatient psychiatric sample consisting of a total of 1951 male and 501 female adults (Arbisi, Ben-Porath, & McNulty, 2003). As one would expect, given the methodology for constructing the RC scales, intercorrelations of the RC scales are lower than for the clinical scales, providing the opportunity for the scales to have greater discriminant validity than the clinical scales (Graham, 2006). When comparing the RC scales and the original clinical scales from the MMPI-2 normative sample, Tellegen et al. (2003) found the RC scales to be equal or superior to the standard clinical scales. Median coefficient alphas ranged from .76 (for the normative sample of men and women) to .86 (male psychiatric patients) for the RC scales, whereas the comparable median alphas generally were lower for the standard clinical scales, ranging from .59 (for the normative sample of men) to .79 (female psychiatric patients) (Simms et al., 2005). The RC scales also yielded higher temporal stability (median $r_s = .81$ and .86 for men and women, respectively) in a portion of the normative sample over a short interval ($M = 9$ days) compared with the standard clinical scales (median $r_s = .74$ and .68 for men and women respectively) (Simms et al., 2005).

Although Hathaway and McKinley did not consider internal consistency when they constructed the clinical scales, internal consistency was one of the considerations in the development of the RC scales (Graham, 2006). Therefore once measured, it was not surprising that the RC scales had higher internal consistency values than the clinical scales (Tellegen et al., 2003; Graham, 2006). Although at this time the only test-retest reliability data available for the RC scales are for 82 men and 111 women from the MMPI-2 normative sample who completed the test twice with a retest interval of approximately one week, this
data indicates the stability of the RC scales over this short time period is quite acceptable and in most cases greater than for the corresponding clinical scales (Graham, 2006; Wallace & Liljequist, 2005).

The next set of analyses was designed to explore the internal validity of the RC scales through a series of within-MMPI-2 correlational analyses (Tellegen, 2003). This was accomplished by examining the intercorrelations and their correlations of each RC scale with their clinical scale counterparts (Forbey & Ben-Porath, 2007). Tellegen et al. (2003) demonstrated that owing to increased variance, the correlations are generally higher for the clinical samples than for the normative ones. In addition, subsequent findings indicate a substantial drop in the correlations between RCd and the RC scales relative to the correlations between RCd and their clinical scale counterparts (Tellegen et al., 2003). These findings indicate that the RC scales, are substantially less saturated with Demoralization than are the clinical scales (Tellegen et al., 2003). Further, the majority of RC scales were found to be meaningfully associated with their clinical scale counterparts, as reflected by moderate to high correlations (Forbey & Ben-Porath, 2007). Overall, the internal validity analyses indicate that the RC scales resemble their clinical scale counterparts, that they are substantially less correlated with Demoralization than are the clinical scales, and are generally (although not uniformly) less correlated among themselves, suggesting that the RC scales are positioned well to demonstrate improved discriminant validity in comparison with the clinical scales (Tellegen et al., 2003). Moreover, to the extent that Demoralization saturation attenuates at least some of the convergent validity of the clinical scales, the RC scales are expected to show improvement in this respect as well (Tellegen et al., 2003).

Tellegen et al. (2003) examined the external validity of the RC scales by comparing their empirical correlates with those of the clinical scales. Utilizing clinicians ratings for the outpatient sample, and record review data for the inpatients, Tellegen and colleagues reported considerable increased discriminant validity and comparable to improved convergent validity for the RC scales compared to their clinical scale counterparts (Forbey & Ben-Porath, 2007).
Convergent validity was appraised by comparing correlations for each RC scale and its clinical scale counterpart with their non-corresponding extratest criterion variable scales (Tellegen et al., 2003). These correlations generally supported the convergent validity of the RC scales (Graham, 2006). Tellegen et al. (2003) additionally concluded that as intended, the RC scales predict as well as or better than their clinical scale counterparts the variables that are linked conceptually to the core constructs of these scales. These include the findings that the RC scales have comparable to improved internal consistencies over the clinical scales as well as comparable to improved convergent and improved discriminant validity in outpatient and inpatient samples (Tellegen et al., 2003). Moreover, several preliminary reports have reported the RC scales’ improved psychometric characteristics (Sellbom & Graham, 2006). Sellbom and Ben-Porath (2005) found the RC scales to correspond well to Tellegen’s (1982) model of personality as conceptualized through the Multidimensional Personality Questionnaire. Sellbom, Ben-Porath, & Graham (2006) examined correlations between RC scales and clinical scales and external criterion measures for 813 male and female clients at a university clinic and found that the RC scales had equal to higher correlations with conceptually relevant therapist’s ratings of various criteria than did their clinical scale counterparts. Forbey and Ben-Porath (2003) conducted a study of the RC scales with a male substance abuse disorder treatment sample and found general support for increased internal consistency over the clinical scales, decreased intercorrelations among the RC scales, and equal to improved convergent and discriminant validity over their clinical scale counterparts. Forbey, Ben-Porath, and Tellegen (2004) compared the external validity of the RC scales and the content scales using both outpatient and inpatient samples finding that both the RC scales and their content scale counterparts were strongly correlated with conceptually related criterion measures. In summary, data currently suggest that when compared with the standard clinical scales, the RC scales (a) converged well with their parent scales, (b) were more distinctive, (c) yielded lower relations with RCd (Demoralization), and (d) generally demonstrated improved discriminant validity and equivalent or superior convergent validity.
when correlated with variables conceptually related to the core components of the scales (Tellegen et al., 2003).

Thus, the initial reliability and validity data are promising, but replication and extension of these findings in independent samples are necessary (Simms et al., 2005). Despite the RC scales’ ability to eliminate item overlap, reducing interscale correlations as well as improving upon internal and external validity measures when compared to their clinical scale counterparts, the true test appears to be whether the individual RC scales can provide reliable differential interpretations as measured against valid external criteria (Rogers & Sewell, 2006).

Need for Validity Studies across Racial Groups

Although several studies have been identified which confirm these initial findings concerning the convergent and discriminant validity of the RC scales for inpatient and outpatient samples, none to date have evaluated the convergent or discriminant validity of the RC scales across racial groups (Sellbom, Ben-Porath & Graham, 2006; Sellbom and Ben-Porath, 2005; Forbey, Ben-Porath, and Tellegen 2004). As previously stated the RC scales were developed to provide a purer measure of the core constructs associated with the clinical scales while reducing the extent to which the scales are influenced by general maladjustment or demoralization (Graham, 2006). When researching differential interpretations of the RC scales one would expect that they would have strong correlations with conceptually relevant extratest measures (convergent validity) and would be less correlated than their corresponding clinical scales with non relevant criterion measures (discriminant validity) (Graham, 2006).

The MMPI scales were developed and normed on almost exclusively Caucasian samples (Graham, 2006). Most of the research to date has compared African-American and Caucasian groups, mainly because African-Americans have composed the largest minority group (Arbisi, McNulty, & Ben-Porath, 2002; Graham, 2006). However, as the composition of the U.S. population changes, it becomes more important to study other minority groups as well (Graham, 2006; Timbrook, & Graham, 1994; Greene, 1987). Researchers have stated that similar
studies are needed with different minority groups, such as American Indians, Hispanics, and Asian Americans (McNulty, Graham, Ben-Porath, & Stein, 1997). Native Hawaiians and Other Pacific Islanders (NHOPI) should be included in such requests (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). This is especially apparent after acknowledging the rapid population growth of NHOPI people in the United States (Harris, & Jones, 2005; Greico, 2001; Brittingham, & de la Cruz, 2005).

The MMPI-2 provides an opportunity for comparison across racial groups, however it has often have been cited as deficient when used with minority groups (Robin et. al., 2003). Since the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) was published more than 60 years ago, its usage with minority ethnic groups has been widely studied and reviewed (Pace, et al., 2006; Velasquez, et al., 2000; Timbrook & Graham, 1994; Dalhstrom & Gynther, 1986). The possibility that this instrument might be susceptible to potentially biasing race effects has been the subject of research for several decades (Ben-Porath, Shondrick, & Stafford, 1995). It is important to investigate potential biases and limitations of the MMPI-2 as an accurate and reliable psychological test (Ben-Porath, Shondrick, & Stafford, 1995). Although research has suggested ethnic differences on the MMPI tests, many of these studies have indicated an inconsistent pattern of score differences between minority and White groups (Timbrook & Graham, 1994). The original MMPI did not include ethnic minorities in its standardization sample leading some to contend that the test is biased against ethnic minority groups (Hall, Bansal & Lopez, 1999). Although ethnic minority persons were included in the standardization sample of the MMPI-2 (Butcher et al., 1989) the sample used did not include NHOPI people. Much of the research evaluating potential ethnic group differences on the MMPI-2 since the restandardization has involved European Americans, African Americans and Latino Americans. Hall, Bansal and Lopez (1999) conducted a meta-analysis of 37 studies over the last 31 years concerning the use of the MMPI and the MMPI-2 with African American, Caucasian and Latino men and women. Their findings revealed that although some ethnic group differences
appear to exist, most of the MMPI/MMPI-2 differences among European Americans, African Americans and Latino Americans are trivial. For example, although some early studies suggested that African Americans tended to score higher (approximately five T-score points) than Caucasians on scales F, 8, and 9, (Ball, 1960) later studies comparing these two groups found that these differences were small or even non-existent when groups were matched for age, education and other demographic characteristics (Dahlstrom, Lacher & Dahlstrom, 1986). Campos (1989) conducted a meta-analysis of 16 studies that compared Hispanics and Caucasians and concluded that the only consistent finding was that Hispanics scored higher (approximately four T-score points) than Caucasians on the L scale. Graham (2006) has additionally suggested that although some differences have been identified, many of the differences reported between these two groups have not been consistent across studies and that most of these studies have not controlled for differences between Hispanic and Caucasian groups on variables such as gender, age, education, socioeconomic status or level of acculturation. Greene’s review (1987) identified only seven studies that compared MMPI scores of American Indians and Caucasians. Although Greene reported to have identified a clear pattern that outpatient Native Americans tend to score higher on most of the clinical scales than their White counterparts, Graham, (2006) suggested that when examining all comparisons made across the clinical groups evaluated, there was no clear pattern to these differences across the studies. Although few research studies are available concerning MMPI and MMPI-2 scores of Asian Americans, in some studies Asian Americans obtained higher scores than Caucasians on many of the MMPI scales. For example, Tsushima and Onorato (1982) examined MMPI racial group comparisons between Asian American and Caucasian psychiatric samples concluding that Asian Americans scored higher than their Caucasian counterparts on clinical scales 1, 2, 4, 6, 7, 8, and 0. Subsequent data regarding racial group comparisons between Asian Americans and Caucasians using data from the MMPI-2 normative sample suggests that Asian Americans did not score consistently higher than the Caucasians on any of the MMPI-2 scales. Although
additional research is needed concerning MMPI-2 differences between these two groups, Graham (2006) has suggested that because very few Asian Americans were included in the MMPI-2 normative sample, it is not appropriate to reach conclusions about differences between these two groups on the MMPI-2. Although several studies have included minority populations, there remains no published research evaluating ethnic differences on the MMPI-2 involving NHOPI people. Because of the comparisons that have been made between NHOPI and other minority group’s mental health outcomes, it could be inferred that similar findings are likely to be identified through MMPI-2 racial group comparisons involving NHOPI’s and Caucasians (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003; U.S. Department of Health and Human Services, 1998).

Although several studies have identified some racial group differences, these assessments have often been criticized as presenting erroneous representations of racial minority’s mental health (Pace, et al., 2006; Butcher, Cheung, & Lim, 2003; Hoffman, Dana & Bolton, 1985). These erroneous representations are particularly striking given that minority differences have been investigated for nearly 40 years on the MMPI and over 10 years on the MMPI-2 (Robin et. at., 2003). This is of special concern because effective psychopharmacological and psychotherapeutic treatment planning and continuity of care are predicated in part, on accurate evaluation of symptom severity and diagnosis (Malgady, 1996). Continuity regarding uniform assessment is especially problematic given the more rapid increases in ethnically diverse populations in this country relative to the growth rate of the non-Hispanic White population (Malgady, 1996).

**Population Trends**

Between 1980 and 1990, United States population trends indicated a 107.8% increase in Asian/Pacific Islander populations (Barringer, Gardner, & Levin, 1993). By the year 2050, non-Hispanic Whites are projected to be 50% of the population (Nagayama-Hall, 2005). Although the non-Hispanic White population is projected to grow only 7% over the next 45 years, projected growth
is 71% for African Americans, 213% for Asian American and Pacific Islanders, and 188% for Hispanic Americans (U.S. Census Bureau, 2000a). In areas of the country, such as Hawaii and Florida, the population distribution is expected to be even more diverse. Hawaii is already perhaps the most ethno culturally diverse state in the United States where no single ethnic or racial group holds the majority (Oliveira et al., 2006). Hawaii’s indigenous people, Native Hawaiians, represent one of the five major ethnicities and number approximately 200,000 in a population of 1.2 million (Andrade, et al., 2006).

**Native Hawaiian and Other Pacific Islanders**

The term “Native Hawaiian and Other Pacific Islander” refers to people having origins in any of the original peoples of Hawaii, Guam, Samoa, Tonga, Fiji, Tahiti, New Zealand or other Pacific Islands (Greico, 2001). The Office of Management and Budget (OMB) revised Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting, to include two critically important changes affecting NHOPI populations. These changes include (1) separating “Asians” from “Native Hawaiians and Other Pacific Islanders,” and (2) allowing respondents completing household surveys, administrative forms and records, and other data collections (for example, the U.S. Census) to designate more than one racial or ethnic category (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003; Ross, 2000).

Today the population distribution of Hawaii is as follows: 21.1% Native Hawaiians, 20.3% Japanese, 23.7% Caucasians, 16.8% Filipinos, 5.8% Chinese, and 12.4% various other ethnic groups (e.g., Samoans, Tongans, Maoris, Tahitians, Fijians, Koreans, South East Asians, and Micronesians; Hawaii State Department of Health, 2003). In addition, the people of Hawaii are diverse in their acculturation status as subsequently explained, and include Native people (e.g., Hawaiians), immigrants (e.g., Filipinos and Koreans), and refugees (e.g., Vietnamese and Cambodians) (Hawaii Primary Care Association, 2004). Thus, various ethnic identities and degrees of acculturation (e.g., bicultural, assimilated) exist within and between ethnic groups and across generations
within ethnic groups, both of which add to the level of complexity when considering health care status and service delivery systems (Oliveira et al., 2006).

In 2007, the United States population was 296.41 million (U.S. Census Bureau, 2007). This is an increase from the Census 2000 which showed that the United States population was 281.4 million on April, 2000 (Grieco, 2001). Of that total, 874,000 reported as NHOPI (Greico, 2001). When compared to other minority groups, Native Hawaiians and Other Pacific Islanders represent a group in the United States larger than those reporting to be Arab or to have Arab ancestries (Brittingham, & de la Cruz, 2005). In addition to this, between 1990 and 2000 the NHOPI population increased by 509,000, or by 140 percent, a growth rate nearly double that of those reporting to have Arab ancestries from the same period of time (Harris, & Jones, 2005; Greico, 2001; Brittingham, & de la Cruz, 2005).

Despite this rapid growth, NHOPIs have struggled to distinguish themselves from the broader categorization of Asian Americans and Pacific Islanders (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). Historically, their cultural dignity as diverse and indigenous peoples has been ignored, making NHOPI communities invisible within American culture (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). The Asian American and Pacific Islander (AA/PI) category has been used as a social and political convenience because the use of the term allows researchers, service providers, and policymakers to easily describe and discuss groups who seemingly share similar backgrounds (U.S. Department of Health and Human Services, 2001). Unfortunately, this classification masks the social, cultural, and psychological variations that exist among AA/PI ethnic groups and constrains analyses of the interethnic differences in mental illness, help-seeking, and service use (U.S. Department of Health and Human Services, 2001). The conclusions drawn from analyses using AA/PIs as a single racial category may be substantively different than ones made when specific AA/PI ethnic groups are examined (Uehara, Takeuchi, & Smuckler, 1994).
NHOPIs as a Unique Racial Group

The recognition of NHOPIs as a unique racial group is the first step in revealing the community’s true social and economic condition (Pacific American Research Center, 2006). Of these people living in the United States, 80 percent of this population lives in just 10 states (Harris & Jones, 2000; Greico, 2001). These states include Hawaii, California, Washington, Texas, New York, Florida, Utah, Nevada, Oregon and Arizona (Greico, 2001). Although 71 percent of the total NHOP population resides in the western United States, states such as Florida, Texas and New York have recorded to have experienced a significant increase in the NHOP population (U.S. Department of Health and Human Services, 2006; Greico, 2001). In fact, as of 2000 Florida had the fifth largest Native Hawaiian population in the country and the seventh largest NHOP population (U.S. Census Bureau, 2000a).

It is also significant to note that 33 percent of this group is under the age of 18 (U.S. Department of Health and Human Services, 2006). In 2000, NHOPs had a median age of 28 years, compared with 35 years for the total U.S. population (Harris, & Jones, 2005; U.S. Census Bureau, 1993). Although in 2000, about 89 percent of the total population and 80 percent of NHOPs were born native to the United States, it is also significant to note that forty-four percent of the NHOP foreign born arrived in the United States between 1990 and 2000 (Harris, & Jones, 2005). Thirty percent of NHOPs arrived in the United States between 1980 and 1989 and 18 percent between 1970 and 1979 (Harris, & Jones, 2005). This significant trend of population growth can easily be seen in a number of states where the NHOP total inclusive population percentage change has increased by as much as 440% since 1990 (U.S. Census Bureau, 2000b).

Professional knowledge of the mental health needs for NHOP people remains somewhat limited (U.S. Department of Health and Human Services, 1999). The paucity of health data obtained for these populations is a major barrier to effective health care for NHOP people (Palafox, & Kaanoi, 2000). Accumulating evidence suggests that, contrary to public belief, NHOP Americans are experiencing significant mental health problems (U.S. Department
of Health and Human Services, 1998). Although data on NHOPI health disparities are limited, a growing body of evidence suggests that, compared to the general U.S. population, NHOPI communities are at risk for poorer health outcomes (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). It is believed that early estimates based on utilization rates among clinical samples seriously underestimated the actual need in the general population (U.S. Department of Health and Human Services, 1998). This is because NHOPI Americans tend to underutilize mental health services. There is convergent evidence that NHOPI Americans underutilize mental health services regardless of service type, based on reports that compare this group’s service use to their proportion in the general population (U.S. Department of Health and Human Services, 1998).

**Mental Health and NHOPIs**

While overall prevalence rates of diagnosable mental illnesses among NHOPIs may at times appear similar to those of the white population, when symptom scales are used, NHOPIs show higher levels of depression symptoms than do white Americans (U.S. Department of Health and Human Services, 1999). When compared to the suicide rate of white Americans (12.8 per 100,000 per year), Native Hawaiian adolescents have a higher risk of suicide than other adolescents in Hawaii (U.S. Department of Health and Human Services, 2001; U.S. Department of Health and Human Services, 1999). Several studies also report that NHOPI Americans exhibit other more severe disturbances compared to non NHOPI Americans. This finding suggests that NHOPI Americans may tend to delay seeking help, and consequently come to the attention of the mental health system at a point of acute breakdown and crisis (U.S. Department of Health and Human Services, 1998). Further studies show that NHOPI Americans are more likely to drop out after initial contact or terminate prematurely from mainstream service settings (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003; U.S. Department of Health and Human Services, 1998). This reality may best be understood when considering that the traditional medical model of care-by itself-is not well received in most
NHOPI communities (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). A major challenge to securing additional funding for NHOPI communities is the severe paucity of data on the mental health status of NHOPI populations (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003; Harris, & Jones, 2005; U.S. Department of Health and Human Services, 1998). Without question, the lack of data generally translates into a lack of funding.

As data begin to document health disparities, jurisdictions may find themselves better positioned to compete for limited resources for capacity building, prevention, cultural assessment and mental health treatment (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). Studies have linked such underutilization to the shame, stigma, and other cultural factors that influence symptom expression and conceptions of illness (Harris, & Jones, 2005; U.S. Department of Health and Human Services, 1998). These potential disparities are complicated by the realities of economic hardship, poverty, joblessness, and underemployment in NHOPI communities; limited access to primary health care and specialty medical care services an underutilization of services (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). This underutilization of services creates a major challenge in securing additional funding for NHOPI communities due to the limited availability of research on the mental health condition of NHOPI people. Without question, this lack of data generally translates into a lack of funding. As data begins to document health disparities, jurisdictions may find themselves better positioned to compete for limited resources for capacity building, prevention, cultural assessment and mental health treatment (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). As previously stated, the recognition, mental health assessment, and cultural examination of NHOPIs as a unique racial group is the first step in revealing the community’s true social and economic condition (Pacific American Research Center, 2006).
Racial Group Comparisons Using the RC Scales

Some clinicians have questioned whether the MMPI-2 is appropriate for use with ethnic minorities such as NHOPI’s that are likely to be very different from the original normative sample (Graham, 1990). Comparing population means from different racial backgrounds is important in determining the applicability and limits of psychological theories and tests (Ben-Porath, 1990; Okazaki & Sue, 1995). Researchers, however, have emphasized that a simple comparison of means across racial groups is not sufficient to indicate whether scores for one group are biased (Timbrook & Graham, 1994). Perhaps more importantly in making racial comparisons is whether the test is an equally valid predictor of the behaviors and psychological characteristics it is intended to measure (Ben-Porath, Shondrick, & Stafford, 1995). In attempting to examine potentially biased MMPI-2 results Greene (1999) has suggested in addition to the simple comparison of means across racial groups that one would first need to account for group differences on demographic variables. Greene (1999) suggested that some measured mean scale differences could be related to such factors as gender, age, education, and socioeconomic status. Prior research involving large sample sizes have suggested that these variables may be related to MMPI scale scores (Greene, 1999). Because many studies have not accounted for these demographic variables when evaluating racial group comparisons, one cannot be sure that some scale differences found in earlier studies cannot be attributed to differences on important demographic variables (Ben-Porath, Shondrick, & Stafford, 1995). Although a limited number of demographic differences such as gender have been found for the MMPI-2 Clinical scales, current studies have yet to establish similar findings regarding the MMPI-2 RC scales.

Additional approaches to the assessment of significant mean scale score differences and potential test bias employ reference to population differences in how test scores relate to an external criterion-related measure (Malgady, 1996). McNulty, Graham, Ben-Porath, and Stein (1997) have suggested that studies examining the relation between MMPI-2 scores and extratest criteria are needed
to determine whether mean score differences between racial groups indicate test bias or reflect actual differences in symptomology and personality characteristics that are deserving of treatment consideration. If the relationships between MMPI-2 scale scores and relevant extratest criteria differ between racial groups, racial group membership can be said to function as a moderator variable, reflecting evidence of test bias (McNulty et al., 1997). If racial group membership were found to affect MMPI-2 scale scores when potential confounding factors are analyzed, the critical issue would then become whether the differences are of sufficient magnitude to affect clinical interpretation (Greene et al., 2003). If such differences were found, perhaps the interpretation of these data might be useful in helping others to provide a fuller understanding of behavior in local cultural contexts.

Consequently, understanding these potential differences would prove to be highly important to professionals in attempting to interpret and design treatment interventions relevant to specific cultural contexts such as that of NHOPi populations. Relatively little research has been conducted with NHOPi populations concerning potential ethnic differences on the MMPI-2. Research specifying and examining this potential appears to be much needed. Without this type of information regarding NHOPi populations, data may be misanalyzed or misleading and may not adequately inform research, program planning, funding, and project implementation (President’s Advisory Commission on Asian Americans and Pacific Islanders, 2003). Specifically, the newly constructed RC scales should be evaluated.

**Statement of Problem and Research Purpose**

It is important to evaluate the extent to which MMPI-2 test scores are related to relevant extratest measures (Graham & Lilly, 1984). One can use criterion-related validity to explore ethnic differences in the accuracy with which an RC scale predicts extratest characteristics (Timbrook & Graham, 1994). In other words, differential prediction of conceptually relevant extratest data across racial groups would be a demonstration of test bias in a particular MMPI-2 scale (Arbisi, Ben-Porath, & McNulty, 2002). Doing so would prove useful in
establishing reliable differential interpretations as measured against valid external criteria (Rogers & Sewell, 2006).

These processes require the analysis of both slope and intercept bias. McNulty, Graham, Ben-Porath, and Stein (1997) have suggested that studies examining the relation between MMPI-2 scores and extratest criteria will help to determine whether mean score differences between racial groups indicate test bias or reflect actual differences in symptomatology and personality characteristics that are deserving of treatment consideration. If racial group membership were found to affect MMPI-2 scale scores when potential confounding factors are examined, the critical issue would then become whether the differences are of sufficient magnitude to affect clinical interpretation (Greene et. al., 2003). If such differences were found, perhaps the interpretation of these data might be useful in helping others to provide a fuller understanding of behavior in local cultural contexts.

When evaluating for mean racial group differences on MMPI-2 scales, bias has often been attributed to the over prediction of pathology. The under prediction of pathology has not received the same level of attention (Lopez, 1989). Although there is a limited amount of research for systematic MMPI-2 based over-prediction in minority populations, the possibility of under prediction of psychopathology remains relatively unknown. In order to assess the relationship between MMPI-2 RC scale scores and relevant extratest criteria, the strength of the association between the predictor and criterion variable (slope bias) in NHOPIs and Caucasians as well as both systematic over- and under prediction (intercept bias) need to be assessed (Arbisi et al., 2002).

Being that research of this nature is needed regarding ethnic minority groups such as NHOPIs, it appears necessary to identify an up-to-date, comprehensive set of empirical correlates to guide clinicians in the interpretation of NHOPI MMPI-2 profiles using the RC scales (Graham, Ben-Porath & McNulty, 1999). Therefore the purpose of the present study includes analyzing the comparative validity of the MMPI-2 RC scales scores in a sample of NHOPI and Caucasian mental health center outpatients. This was accomplished by looking
for significant slope and/or intercept differences that were present in the MMPI-2 RC scales between NHOPI and Caucasian populations. If the relationships between MMPI-2 RC scale scores and relevant extratest criteria differed between racial groups, racial group membership could be said to function as a moderator variable, reflecting test bias (McNulty et al., 1997).

**Research Questions**

Given the statement of the problem and the purpose of the present study the following research questions were evaluated as related to the use and applicability of the MMPI-2 with NHOPI populations:

1a. Do NHOPI people tend to score significantly different than Caucasian people on the following nine MMPI-2 RC scales: Demoralization (Rcd), Somatic Complaints (RC1), Low Positive Emotions (RC2), Cynicism (RC3), Antisocial Behavior (RC4), Ideas of Persecution (RC6), Dysfunctional Negative Emotions (RC7), Aberrant Experiences (RC8), and Hypomanic Activation (RC9)?

1b. How are the demographic variables gender, education and age related to mean RC scale differences?

2. Is there evidence of test bias using the MMPI-2 RC scale scores?
CHAPTER 3

METHODOLOGY

This section provides the methods and procedures for this study. The presentation of the chapter has been divided into the following sections: statement of hypotheses, participants, instrumentation, procedure and data analysis. The following hypotheses were made:

Statement of the Hypotheses

1a. It is hypothesized that the NHOPI sample will score significantly higher on the MMPI-2 RC scales RCd (dem) Demoralization, RC3 (cyn) Cynicism and RC9 (hpm) Hypomanic Activation than the Caucasian sample. In contrast it is hypothesized that the Caucasian group will score significantly higher on the MMPI-2 RC scale RC2 (ipe) Low Positive Emotions than the NHOPI sample.

Most of the current research evaluating mental health trends of NHOPI people has been collected by the U.S. Department of Health and Human Services (2001; 1999). Their findings have consistently suggested that when symptom scales are used, NHOPIs show higher levels of depression symptoms than do white Americans and that NHOPI Americans tend to exhibit other more severe disturbances compared to non NHOPI Americans (U.S. Department of Health and Human Services, 2001; 1999). A more recent study involving the comparison of NHOPI and Caucasian MMPI-2 RC scale mean scores has suggested alternative findings (Kehoe, Li, Dao & Peterson, in press). This research identified significant differences between NHOPI and Caucasian mental health center outpatients on several MMPI-2 RC scales. These findings have helped to establish a hypothesis regarding racial group differences for the current study. It has been recognized that ethnic differences research is needed in determining the applicability and limits of tests like the MMPI-2 (Ben-Porath, 1990). Although a general lack of substantial research exists, involving NHOPI populations, the current hypothesis is based on prior research involving a similar outpatient population.
1b. It is hypothesized that the MMPI-2 RC scale scores will not significantly differ by gender, education or age. Although previous findings not including NHOPi populations have suggested minimal effects of gender on clinical scales 1, 2, 3, 4, 7, 9 and 0 (Caldwell, 1997; Butcher et al., 1989) there remains no simple, direct significant relationship between any single demographic variable and scores on a given MMPI-2 clinical scale (Greene, 2000). Although each RC scale is composed of the core component of its clinical scale derivative the lack of current research assessing for demographic differences on the RC scales using ethnic minorities helps to justify the reasoning for the current hypothesis.

2. It is hypothesized that the current research will not find any consistent evidence for racial test bias in the MMPI-2 RC scales. Although no research has yet been published regarding racial test bias in the MMPI-2 involving NHOPi populations, previous studies comparing African American and Caucasian populations have failed to find any consistent evidence for racial test bias using several MMPI-2 scales (McNulty et al., 1997; Timbrook & Graham, 1994). Arbisi et al. (2002) was able to identify some evidence for potentially significant under prediction of psychopathology in African Americans, although they were unable to find any consistent bias that results in clinically significant errors in MMPI-2 based predictions for African Americans. Although it is possible that the current research may suggest evidence for racial test bias in the MMPI-2 RC scales for NHOPi people, the lack of such evidence for test bias when researching other minority groups creates reason for the current hypothesis. This research is intended to assist in determining the validity of MMPI-2 RC scales for NHOPi populations.

**Participants**

Data were collected at a large mental health clinic and diagnostic center serving Honolulu County and the island of O'ahu located in Hawaii. As is the case in most community mental health centers, a variety of treatment programs are available to clients being served by this clinic, including individual or group outpatient counseling or therapy. All of the participants in this study were referred to the clinic and diagnostic center through the Hawaii Department of Health and
the Adult Mental Health Division of Hawaii. All of the participants in this study were clinically diagnosed by a licensed psychologist within the Hawaii Department of Health as needing Intensive Case Management mental health services. Persons seeking services at the mental health center between May 1, 2007 and May 1, 2008 of Caucasian or Native Hawaiian or Other Pacific Islander (NHOPI) descent were potential participants in the study. MMPI-2’s are administered to all willing and eligible individuals undergoing routine evaluations at this facility unless they are seen as being unable to complete psychological testing because of language difficulties, severe disturbance or retardation, or lack of cooperation. A data set of completed MMPI-2’s administered at this mental health facility was used for the present study. All Caucasian and NHOPI clients who produced valid MMPI-2 profiles from this data set were included in the present study, which resulted in a potential n of 321. 232 valid MMPI-2 profiles from this set were used for the present study. The age range of the participants for this study ranged from eighteen to seventy five years of age. The norms set for the MMPI-2 are for ages 18-80. Participants were selected by using a list of NHOPI and Caucasian samples from the North Shore Mental Health (NSMH) total population. NSMH serves an approximately equal number of people from each of these two groups. The criteria for inclusion in the present study were as follows and reflect the criteria established by McNulty et. al., (1997) in their study evaluating the comparative validity of MMPI-2 scores of African American and Caucasian Mental Health Center Clients:

1. Having 30 or fewer items omitted.
2. Variable Response Inconsistency [VRIN] scale T score equal to or less than 80.
3. True Response Inconsistency [TRIN] raw score greater than 5 and less than 13.
4. $F$ raw score less than 28 for men and 30 for women, and $F$-back [Fb] raw score less than 24 for men and 25 for women (Arbisi et al., 2002; McNulty et al., 1997).
Of the 321 individuals initially identified as having completed the MMPI-2 and a general biographical form, 89 were identified as having produced an invalid MMPI-2. Of those who produced an invalid MMPI-2 profile, 11 were due to an elevated F; 4 were due to an elevated Fb; 25 were due to an elevated VRIN; 16 were due to an elevated TRIN and 33 due to a combination of F, Fb, VRIN or TRIN elevations and were excluded from further analysis. The NHOPI sample was more likely to produce an invalid profile than the Caucasian sample: \( \chi^2(1,N=321) = 9.767, p < .05 \). The VRIN and TRIN scales were designed to assess for content-nonresponsiveness while the F and Fb scales were designed to assess for content responsiveness (Graham, 2006). As previously described, the VRIN scale provides an indication of a tendency to respond inconsistently to MMPI-2 items, while the TRIN scale identifies persons who respond inconsistently to items by giving true responses to items indiscriminately (acquiescence) or by giving false responses to items indiscriminately (nonacquiescence) (Butcher et. al., 2001). The F scale on the other hand was developed to detect deviant or atypical ways of responding to test items (Graham, 2006). In the MMPI-2 normative sample, scores on the F scale are related to ethnicity, with African-Americans, Native Americans and Hispanics scoring on average three to five T-score points higher on the F scale than Caucasians (Graham, 2006). Although NHOPI’s were not included in the MMPI-2 normative sample, their identification and experience as an ethnic minority may have contributed to their elevated F scale scores. High scores on the F scale are also associated with elevated scores on clinical scales 6 and 8 (Graham, 2006). Although the RC scale scores were not analyzed for those who produced invalid profiles, for those included in this study the NHOPI group did score significantly higher than the Caucasian group on RC8 which was derived from clinical scale 8. Although not as commonly identified in this study, an elevated Fb scale score could indicate that the test taker responded to items in the second half of the MMPI-2 in an invalid manner. As with the F scale, persons who respond randomly to MMPI-2 items or who “fake bad” in responding to many of the items also will produce very high scores on the Fb scale (Graham, 2006). The final
sample included 126 (67 men and 59 women) Caucasian and 106 (48 men and 58 women) Native Hawaiian or Other Pacific Islander mental health center outpatients.

Gender, age, educational level, marital status and primary language spoken were determined for both the NHOPI and Caucasian participants. Data regarding the participants' demographic characteristics are presented in Table 1. The only significant difference was found between Caucasian and NHOPI participants for years of education. The Caucasian population was more likely to have participated in a greater number of years of education than the NHOPI sample. There were no differences between races in age, gender, marital status or primary language spoken.
Table 1: *Demographic Characteristics of the Sample*

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Caucasian (n = 126)</th>
<th>Native Hawaiian and Other Pacific Islander (n = 106)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>67</td>
<td>48</td>
</tr>
<tr>
<td>Female</td>
<td>59</td>
<td>58</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>45.1</td>
<td>42.4</td>
</tr>
<tr>
<td>SD</td>
<td>12.3</td>
<td>12.8</td>
</tr>
<tr>
<td><strong>Years of Education</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>13.1</td>
<td>12.2</td>
</tr>
<tr>
<td>SD</td>
<td>2.5</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Marital Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Married</td>
<td>62</td>
<td>55</td>
</tr>
<tr>
<td>Married</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Widowed</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Divorced</td>
<td>41</td>
<td>24</td>
</tr>
<tr>
<td>Separated</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td><strong>Primary Language Spoken</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>126</td>
<td>106</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*p < .05.*
Instruments

The Minnesota Multiphasic Personality Inventory – 2 (MMPI-2)

The MMPI-2, a revised and updated version of the original MMPI, is a 567-item personality inventory (McNulty et al., 1997). In addition to the 10 standard clinical scales, the MMPI-2 includes a variety of validity, and restructured clinical scales (Butcher & Williams, 2000; Tellegen et al., 2003). It is a broad-band test designed to assess a number of the major patterns of personality and psychological disorders (Butcher et al., 1989). The internal consistencies and test-retest reliabilities of the MMPI-2 scales are acceptable and are reported in the test manual (Butcher et al., 1989) along with some validity information. The clinical scales test-retest coefficients for a one week interval range from .70 to .93 for a sample of 82 men and from .54 to .92 for a sample of 111 women in the MMPI-2 normative sample (Butcher et al., 1989). For the RC scales, utilizing the same normative data, Tellegen et al. (2003) reported internal consistencies ranging from .63 to .87 (median = .74) for men, and from .54 to .80 (median = .76) for women. Test-retest correlations for men ranged from .76 to .91 (median = .81) for men and from .54 to .90 (median = .86) for women. In addition to the RC scales tending to have largely comparable or greater internal consistencies than their clinical scale counterparts, the test-retest reliabilities of the RC scales and the clinical scales are comparable exceeding .70 with the exception of RC6, which probably reflects the restricted variance of RC6 in the normative sample (Tellegen et al., 2003). The present study used the following nine MMPI-2 RC scales: Demoralization (RCd), Somatic Complaints (RC1), Low Positive Emotions (RC2), Cynicism (RC3), Antisocial Behavior (RC4), Ideas of Persecution (RC6), Dysfunctional Negative Emotions (RC7), Aberrant Experiences (RC8), and Hypomanic Activation (RC9).

Patient Description Form (PDF)

The patient description form (PFD) was developed by Graham, Ben-Porath, and McNulty (1999) for their outpatient study. Its 187 items are rated on a 5-point Likert scale (1 = not at all to 5 = very high) by therapists who are asked to describe the personality and symptom characteristics of their clients. The items
represent a comprehensive list of MMPI and MMPI-2 extratest descriptors. The list was compiled by extracting all of the profile descriptors contained in these sources and eliminating those that were redundant or inconsistent. Using a combined empirical-rational approach, 25 scales were developed to assess the major content dimensions of the PDF. Each scale was designed specifically to assess symptoms and problems that outpatients are likely to have and that should be related to some MMPI-2 scales. Internal consistencies range from .69 to .92 (Graham, Ben-Porath, & McNulty, 1999) indicating that the scales are internally consistent. Of the 25 scales originally developed 15 were included in the present study. Those 15 PDF scales are (followed by coefficient alpha based on the original larger sample for men and women, respectively) Somatic Symptoms (.92, .91), Depressed (.86, .87), Insecure (.87, .91), Introverted (.84, .86), Achievement-Oriented (.84, .87), Agitated (.69, .75) Antisocial (.90, .87), Aggressive (.82, .77), Angry/Resentful (.92, .90), Suspicious (.83, .79), Anxious (.87, .87), Obsessive-Compulsive (.83, .85), Pessimistic (.73, .72), Psychotic Symptoms (.80, .87), and Narcissistic (.92, .93). Each of these 15 scales was predicted to associate with one or more of the RC scales used in this study. PDF scales not used in this study were identified to have little conceptual relevance or association with the RC scales used in this study.

*General Biographical Form (GBF)*

The GBF was designed to be completed on the basis of a personal interview with the client. The GBF form includes demographic data such as age, education, ethnicity, gender and the primary language spoken by the client. All persons completing the GBF were licensed professionals (e.g., social workers, psychologists) who have had considerable experience in conducting intake interviews and who were trained specifically to complete the GBF used in this study. The GBF was used to determine gender which was used as a categorical variable. All other demographic information was used to characterize the sample as provided in Table 1.
Procedure

All potential subjects participating in this study completed the MMPI-2 within 7 days of their initial intake session as part of their clinical outpatient evaluation. Participants were asked to complete the booklet form of the MMPI-2 and each assessment was administered individually. The administration of the MMPI-2 was conducted at a large mental health clinic and diagnostic center serving Honolulu County and the island of O'ahu located in Hawaii. All MMPI-2 testing was administered individually by a psychologist or a trained psychology assistant, with subjects receiving the standard MMPI-2 administration instructions. Once completed all MMPI-2’s were scored by computer through Pearson Assessments. The general biographical information form was additionally completed by a trained NSMH Intensive Case Manager prior to the assessment and administration of the MMPI-2. None of the participants in this study were identified by name and an informed consent form was included in a statement promising confidentiality.

After at least three treatment appointments the PDF was completed by a trained Intensive Case Manager. Before completing the PDF, Intensive Case Managers did not have access to any participants MMPI-2 results. The trained psychology assistants or Intensive Case Managers included eighty-six individuals who held a masters of higher level degree in psychology or a related field, a masters degree in social work and/or were licensed as a clinical social worker or clinical substance abuse counselor. This procedure and provided methodology is supported by Dr. John Graham, Dr. Yossef Ben-Porath and Dr. John McNulty who are widely accepted as three of the most prominent MMPI-2 researchers in the world. Their study using a different outpatient population is considered the most procedurally sound comprehensive study of MMPI-2 correlates published to date.

Data Analyses

Research Question 1.

a.) In an effort to identify mean score differences, MMPI-2 RC scale scores will be evaluated by computing mean score differences for race, gender,
and age. This will be accomplished through a three way factorial MANOVA. Although significant differences in mean scale scores alone cannot address the issue of bias (McNulty et al., 1997), this data will be used to provide valuable information concerning the relative level of functioning of the observed groups at the time of the assessment. (McNulty et al., 1997). This analysis will further help to clarify if the differences found can be attributed to racial group membership, gender, and/or age.

b.) Because preliminary analyses identified a racial group difference by education, the previous analysis will be run again using education as a covariate. This process will help to explain the effect education has on any previously identified differences in RC scale scores.

Research Question 2.
Detection of test bias will be accomplished by the analysis of the prediction errors. This analysis will include a hierarchical regression analytic strategy based on the procedures used in the Arbisi et al. (2002) study.

Selection of Predictor and Criterion Variables

Relevant criteria will be identified for each predictor variable. In this study the predictor variables will be the nine MMPI-2 Restructured Clinical scale scores.

The criterion variables will include fifteen Patient Description Form (PDF) scales. These scales have been identified in the instruments section. Regression analyses will be conducted between an MMPI-2 scale and a PDF variable if, a priori, the scale and PDF variable were deemed to be conceptually related (Arbisi et al., 2002). Previously reported descriptors for each of the MMPI-2 RC scales will be used to pair each scale with one or more PDF scales judged to have relevant content (McNulty et al., 1997). Pairings that have significant correlations will be included in the analyses. MMPI-2 and PDF scales that can not be suitably matched will not be included in the correlational analyses. In addition, if the correlation between an MMPI-2 RC scale/PDF pair is statistically significant for at least one of the two study groups (NHOPI or Caucasian), the pair will be included for further analysis. Table 2 includes the predicted associations between the RC
scales and PDF scales; however, actual selections may be different, based on the results of the correlational analyses.
Table 2: *Predicted associations between Restructured Clinical scales and Patient Description Form scales*

<table>
<thead>
<tr>
<th>RC Scale</th>
<th>Patient Description Form Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCd (Demoralization)</td>
<td>Insecure, Depressed, Anxious</td>
</tr>
<tr>
<td>RC1 (Somatic Complaints)</td>
<td>Somatic Symptoms, Depressed</td>
</tr>
<tr>
<td>RC2 (Low Positive Emotions)</td>
<td>Depressed, Achievement Oriented, Introverted, Pessimistic</td>
</tr>
<tr>
<td>RC3 (Cynicism)</td>
<td>Suspicious, Somatic Symptoms</td>
</tr>
<tr>
<td>RC4 (Antisocial Behavior)</td>
<td>Antisocial, Aggressive, Angry/Resentment</td>
</tr>
<tr>
<td>RC6 (Ideas of Persecution)</td>
<td>Suspicious</td>
</tr>
<tr>
<td>RC7 (Dysfunctional Negative Emotions)</td>
<td>Anxious, Obsessive-Compulsive, Insecure</td>
</tr>
<tr>
<td>RC8 (Aberrant Experiences)</td>
<td>Psychotic Symptoms, Anxious</td>
</tr>
<tr>
<td>RC9 (Hypomanic Activation)</td>
<td>Antisocial, Aggressive, Narcissistic</td>
</tr>
</tbody>
</table>
Slope and Intercept Bias

There are two ways that the relation between a predictor variable and criterion variable can be systematically biased (Arbisi et al., 2002). The first is a systematic difference in the slope of the regression line between the predictor variable (MMPI-2 RC scale) and the criterion variable for NHOPIs and Caucasians (Nunnally & Berstein, 1994). A significant difference between NHOPIs and Caucasians in the magnitude of correlation coefficients between a particular MMPI-2 RC scale and a criterion variable would indicate differential slope or slope bias.

Another way that bias can be evaluated is when the predictor variable (MMPI-2 RC scale) under- or over predicts the criterion variable for either NHOPIs or Caucasians. This type of interpreted bias is referred to as intercept bias and can be identified even when the regression coefficients between the predictor variable (MMPI-2 RC scale) and the criterion variable are the same for both races.

One way of identifying differential prediction including slope and intercept factors, is by using a step-down hierarchical multiple regression procedure (Lautenschlager & Mendoza, 1986). This type of regression was applied to the evaluation of differential predictive bias by race by Arbisi et al. (2002) and Rotundo and Sackett (1999) and was used in this study. This process helped to determine any racially differential relationship between MMPI-2 RC scale scores and relevant Patient Description Form extratest criteria. The procedures used to test for bias can include up to four separate analyses. Although all comparisons were examined for racial bias, slope and intercept bias were only evaluated if racial bias is first identified. In other words, testing for slope and intercept bias were only completed if the previous regression model for racial bias identified a significant increment in $R^2$. The four tests potentially administered first included a test for racial bias, which if needed was followed by a test for slope bias, which if needed was followed by a test for a combination of both slope and intercept bias, which if needed was followed by a test for intercept bias alone. This process began by first testing for the presence of racial bias by making a series of
comparisons between a regression model that includes only the predictor variable (MMPI-2 RC scale) with a model that includes the predictor variable (MMPI-2 RC scale), suspected moderator variable (racial group), and the cross product of the predictor and the moderator variable (full model). The full model in this study includes the MMPI-2 RC scale, the racial background variable, and the interaction term. This first test for racial bias can visually be understood through the following comparative sequencing:

Step 1 - MMPI-2 RC scale → PDF Scale
Step 2 - MMPI-2 RC scale, race, RC scale x race (interaction) → PDF Scale.

A significant increase in $R^2$ obtained by using the full model instead of the model containing the predictor variable alone demonstrates the presence of bias (Arbisi et al., 2002). If this analysis was significant, additional analyses needed to be conducted to identify whether slope differences, intercept differences or both were present. Potential slope differences if needed were next evaluated by conducting a comparison of the full model to a model containing the predictor variable (MMPI-2 RC scale) and race alone. This second test can visually be understood through the following comparative sequencing:

Step 1 - MMPI-2 RC scale, race → PDF Scale
Step 2 - MMPI-2 RC scale, race, RC scale x race (interaction) → PDF Scale.

If a significant increase in $R^2$ was obtained then this indicates the presence of slope bias and a further test was needed to be performed to detect intercept bias (Arbisi et al., 2002). Intercept bias was tested by comparing the full model to a model that includes the predictor variable (MMPI-2 RC scale) and the cross product of race and the MMPI-2 RC scale (interaction term). This third test can visually be understood through the following comparative sequencing:

Step 1 – MMPI-2 RC scale, RC scale x race (interaction) → PDF Scale
Step 2 - MMPI-2 RC scale, race, RC scale x race (interaction) → PDF Scale.

If this comparison results in a significant increase in $R^2$, then intercept bias was also present; if there is no significant increase in $R^2$, then the bias identified was solely due to differences in slope (Arbisi et al., 2002). If the omnibus test for bias was significant but no slope bias was identified, then a separate test for intercept bias...
bias alone was needed. To test for intercept differences when no slope differences are present a model containing the predictor variable (MMPI-2 RC scale) is compared to a model containing the predictor variable and race. This fourth test can visually be understood through the following comparative sequencing:

Step 1 – MMPI-2 RC scale $\rightarrow$ PDF Scale
Step 2 - MMPI-2 RC scale, race $\rightarrow$ PDF Scale.

A significant increase in $R^2$ would indicate presence of intercept bias. An indication of slope differences would suggest a difference in the validity of the MMPI-2 RC scales for NHOPIs and Caucasians; while intercept differences would indicate that an MMPI-2 RC scale is predicting related items differently for NHOPIs and Caucasians.

Comparisons across race, using the described step-down hierarchical multiple regression procedure, was used to indicate the presence of prediction bias. Evidence of statistically significant prediction bias for the identified set of criterion variables is illustrated in Table 6.

The full model includes the MMPI-2 scale, the racial background variable and the interaction term. The Prediction bias column compares the MMPI-2 RC scale alone with the full model. The Slope bias column compares the MMPI-2 RC scale plus racial background with the full model. If there is evidence of slope bias, then the Intercept bias column compares the MMPI-2 RC scale plus interaction term with the full model. If there is no evidence for slope bias, then the Intercept bias column compares the MMPI-2 RC scale alone with the MMPI-RC scale plus racial background variable. Test of slope and intercept bias will be conducted only if evidence of predictive bias is indicated.
CHAPTER 4

RESULTS

Research Question 1a: Individual mean scores were computed for race, gender and age for each of the MMPI-2 Restructured Clinical Scales. These demographic variables were analyzed through a three way MANOVA to examine overall mean scale differences across the MMPI-2 RC scale scores. Prior to running this analysis, age was dichotomized by splitting this variable into two groups. These groups were created by identifying the median age (or 50% point) and placing those that fell below this point into group 1 while placing those that fell above this point into group 2. The results of this analysis are summarized in Table 3, and the three-way means are presented in Table 4. The overall MANOVA was significant $F(9, 216) = 42.44, p < .00$. There were significant main effects for Race $F(9,216) = 4.35, p < .00$, Gender $F(9,216) = 3.74, p < .00$, and Age $F(9,216) = 2.48, p < .00$. The two and three way interactions were all non-significant ($p > .05$ for all tests).
Table 3: Individual Means for Racial Group, Gender and Age on the MMPI-2 RC Scales

<table>
<thead>
<tr>
<th>RC Scale</th>
<th>Race</th>
<th>Gender</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Caucasian (n = 126)</td>
<td>NHOPI (n = 106)</td>
<td>Male (n = 115)</td>
</tr>
<tr>
<td>RCd</td>
<td>13.95(.58)</td>
<td>12.00 (.63)</td>
<td>5.24 (.02)*</td>
</tr>
<tr>
<td>RC1</td>
<td>10.61 (.51)</td>
<td>10.03 (.56)</td>
<td>.60 (.44)</td>
</tr>
<tr>
<td>RC2</td>
<td>8.32 (.34)</td>
<td>5.96 (.37)</td>
<td>21.83 (.00)*</td>
</tr>
<tr>
<td>RC3</td>
<td>8.26 (.32)</td>
<td>9.67 (.35)</td>
<td>8.83 (.00)*</td>
</tr>
<tr>
<td>RC4</td>
<td>10.11 (.39)</td>
<td>9.79 (.43)</td>
<td>.31 (.58)</td>
</tr>
<tr>
<td>RC6</td>
<td>4.15 (.33)</td>
<td>4.96 (.36)</td>
<td>2.75 (.10)</td>
</tr>
<tr>
<td>RC7</td>
<td>11.39 (.55)</td>
<td>11.46 (.59)</td>
<td>.01 (.93)</td>
</tr>
<tr>
<td>RC8</td>
<td>6.52 (.38)</td>
<td>7.67 (.41)</td>
<td>4.26 (.04)*</td>
</tr>
<tr>
<td>RC9</td>
<td>13.44 (.48)</td>
<td>15.28 (.52)</td>
<td>6.81 (.01)*</td>
</tr>
</tbody>
</table>

* p < .05.
These results suggest a number of interesting findings. When assessing for racial group mean differences, five scales studied were identified to be of significance. They include Restructured Clinical Scales, RCd (Demoralization), RC2 (Low Positive Emotions), RC3 (Cynicism), RC8 (Aberrant Experiences) and RC9 (Hypomanic Activation). The Caucasian group was identified to have scored significantly higher than the NHOPI group on scale RCd and RC2 whereas the NHOPI scored significantly higher than the Caucasian group on scale RC3, RC8 and RC9. These results suggest that the Caucasian population sampled was more likely to have endorsed items that are indicative of the generalized unhappiness associated with depression and/or anxiety, whereas the NHOPI sample appeared to be more likely to endorse items that are consistent with hypomanic activation, bizarre experiences and an assertion that other people are untrustworthy, untruthful, and likely to only look out for themselves (Graham, 2006).

When assessing for age mean differences two scales were observed to have been of significance. These were scales RC4 and RC9. For each of these scales the younger group scored significantly higher than the older group. These findings suggest that those in the younger group may be more likely to experience difficulty conforming to social norms and expectations. Consistent with the identified racial group differences, the younger group was more likely to endorse items related to hypomanic activation. There was no significant interaction between race and age for either of these scales.

No differences in gender were demonstrated and there were no gender by race interactions. These findings suggest that gender and age had less effect on how a participant would respond to MMPI-2 questions than race.

**Research Question 1b:**

After controlling for education using a MANCOVA, the overall effect was significant $F(9,215) = 23.80, p < .00$. There were significant main effects for Race $F(9,215) = 4.34, p < .00$, Gender $F(9,215) = 3.58, p < .00$, and Age $F(9,215) = 2.29, p < .018$. The between-subjects effects for the RC scales remained significant for four of the five racial group differences. These include
RCd ($F = 7.55$, $p = .01$), RC2 ($F = 22.57$, $p = .00$), RC3 ($F = 5.29$, $p = .02$), and RC9 ($F = 4.64$, $p = .03$). However, there was no longer a significant racial group effect for RC scale 8 (Aberrant Experiences) $F = 2.14$, $p = .15$. The results of this analysis are summarized in Table 4.
Table 4
Three-way Means for Racial Group, Gender, and Age on the MMPI-2 RC Scales

<table>
<thead>
<tr>
<th></th>
<th>Caucasian (n = 126)</th>
<th></th>
<th></th>
<th>NHOPI (n = 106)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (n = 67)</td>
<td>Female (n = 59)</td>
<td></td>
<td>Male (n = 48)</td>
<td>Female (n = 58)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean(SD)</td>
<td>Mean(SD)</td>
<td></td>
<td>Mean(SD)</td>
<td>Mean(SD)</td>
<td></td>
</tr>
<tr>
<td>0-44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 and up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 27)</td>
<td>(n = 25)</td>
<td></td>
<td>(n = 26)</td>
<td>(n = 22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean(SD)</td>
<td>Mean(SD)</td>
<td></td>
<td>Mean(SD)</td>
<td>Mean(SD)</td>
<td></td>
</tr>
<tr>
<td>Rc Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCd</td>
<td>13.70(7.14)</td>
<td>12.78(6.69)</td>
<td>13.72(7.16)</td>
<td>15.62(5.34)</td>
<td>12.46(6.38)</td>
<td>11.82(6.33)</td>
</tr>
<tr>
<td>RC1</td>
<td>9.22(5.08)</td>
<td>10.10(6.10)</td>
<td>10.44(7.14)</td>
<td>12.68(5.30)</td>
<td>9.42(5.05)</td>
<td>10.32(4.52)</td>
</tr>
<tr>
<td>RC2</td>
<td>8.59(3.71)</td>
<td>8.35(3.77)</td>
<td>7.44(4.51)</td>
<td>8.91(3.96)</td>
<td>6.38(3.65)</td>
<td>6.95(3.70)</td>
</tr>
<tr>
<td>RC3</td>
<td>7.78(3.30)</td>
<td>9.52(3.93)</td>
<td>7.72(4.38)</td>
<td>8.00(3.92)</td>
<td>9.62(3.35)</td>
<td>8.95(3.57)</td>
</tr>
<tr>
<td>RC4</td>
<td>11.67(4.00)</td>
<td>9.98(4.64)</td>
<td>10.32(4.91)</td>
<td>8.47(4.22)</td>
<td>9.92(4.34)</td>
<td>10.23(4.06)</td>
</tr>
<tr>
<td>RC6</td>
<td>4.93(3.82)</td>
<td>4.43(3.91)</td>
<td>3.80(3.40)</td>
<td>3.44(2.23)</td>
<td>5.04(4.23)</td>
<td>4.23(2.69)</td>
</tr>
<tr>
<td>RC7</td>
<td>11.67(5.85)</td>
<td>10.68(6.05)</td>
<td>11.76(6.33)</td>
<td>11.47(6.46)</td>
<td>11.38(5.54)</td>
<td>10.64(5.79)</td>
</tr>
<tr>
<td>RC8</td>
<td>7.44(4.56)</td>
<td>6.45(4.42)</td>
<td>6.28(4.22)</td>
<td>5.91(3.68)</td>
<td>7.54(4.48)</td>
<td>8.36(3.47)</td>
</tr>
<tr>
<td>RC9</td>
<td>14.89(5.65)</td>
<td>13.88(5.34)</td>
<td>13.28(5.39)</td>
<td>11.71(4.95)</td>
<td>16.15(5.38)</td>
<td>14.95(5.71)</td>
</tr>
</tbody>
</table>
Research Question 2: The results of the correlational analyses are shown in Table 5. Conceptually relevant criteria were identified for Restructured Clinical Scales RCd, RC1, RC2, RC4, RC7, RC8, and RC9. No conceptually relevant criteria were identified for RC scales RC3 or RC6. As a result these scales were omitted from the regression analyses.
Table 5: Correlations between MMPI-2 RC Scale Scores and Therapist Ratings for Caucasian and Native Hawaiian and Other Pacific Islanders

<table>
<thead>
<tr>
<th>Scale Pairings</th>
<th>Caucasian (n = 126)</th>
<th>NHOPI (n = 106)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insecure</td>
<td>.48**</td>
<td>.29**</td>
</tr>
<tr>
<td>Depressed</td>
<td>.43**</td>
<td>.38**</td>
</tr>
<tr>
<td>Anxious</td>
<td>.29**</td>
<td>.32**</td>
</tr>
<tr>
<td>RC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somatic Symptoms</td>
<td>.28**</td>
<td>.37**</td>
</tr>
<tr>
<td>Depressed</td>
<td>.33**</td>
<td>.28**</td>
</tr>
<tr>
<td>RC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressed</td>
<td>.37**</td>
<td>.26**</td>
</tr>
<tr>
<td>Achievement</td>
<td>-.31**</td>
<td>-.26**</td>
</tr>
<tr>
<td>Introverted</td>
<td>.29**</td>
<td>.14</td>
</tr>
<tr>
<td>Pessimistic</td>
<td>.43**</td>
<td>.08</td>
</tr>
<tr>
<td>RC3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspicious</td>
<td>-.06</td>
<td>.09</td>
</tr>
<tr>
<td>Somatic Symptoms</td>
<td>.04</td>
<td>-.07</td>
</tr>
<tr>
<td>RC4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antisocial</td>
<td>.34**</td>
<td>.38**</td>
</tr>
<tr>
<td>Aggressive</td>
<td>.18*</td>
<td>.35**</td>
</tr>
<tr>
<td>Angry/Resentment</td>
<td>.08</td>
<td>.42**</td>
</tr>
<tr>
<td>RC6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspicious</td>
<td>.11</td>
<td>.09</td>
</tr>
<tr>
<td>RC7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxious</td>
<td>.18*</td>
<td>.26**</td>
</tr>
<tr>
<td>Obsessive-Compulsive</td>
<td>.03</td>
<td>.07</td>
</tr>
<tr>
<td>RC8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychotic Symptoms</td>
<td>.26**</td>
<td>.29**</td>
</tr>
<tr>
<td>Anxious</td>
<td>.16</td>
<td>.24*</td>
</tr>
<tr>
<td>RC9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antisocial</td>
<td>.24**</td>
<td>.21*</td>
</tr>
<tr>
<td>Aggressive</td>
<td>.05</td>
<td>.24*</td>
</tr>
<tr>
<td>Narcissistic</td>
<td>.09</td>
<td>.20*</td>
</tr>
</tbody>
</table>

** p < .01.
* p < .05.
Of the 19 regression analyses (See Table 6), two revealed statistically significant slope bias, however the effect sizes were small ranging from .024 to .031. In addition, nine demonstrated statistically significant intercept bias. Five of these effect sizes ranged from $R^2 = .02$ to 0.04. According to Cohen (1988), a small effect size ($R^2$) for hierarchical multiple regression is .02, a medium effect size is .13 and a large effect size is .25. The magnitude of the overall prediction bias effect sizes ($R^2$) ranged from .00 to .04.
Table 6: Regression Analyses of Patient Description Form Variables on Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Restructured Clinical Scales, Racial Background, and Scale x Racial Background Interaction Term (n = 232)

<table>
<thead>
<tr>
<th>MMPI-2 RC scale predictor/criterion variable</th>
<th>b</th>
<th>Prediction</th>
<th>Slope</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IV</td>
<td>IV x Race</td>
<td>R²</td>
<td>bias R²D</td>
</tr>
<tr>
<td>RCd Insecure</td>
<td>.081**</td>
<td>.170</td>
<td>-.021</td>
<td>.177***</td>
</tr>
<tr>
<td>RCd Depressed</td>
<td>.062**</td>
<td>-.170</td>
<td>-.006</td>
<td>.197***</td>
</tr>
<tr>
<td>RCd Anxious</td>
<td>.037</td>
<td>-.165</td>
<td>.001</td>
<td>.105***</td>
</tr>
<tr>
<td>RC1 Depressed</td>
<td>.036</td>
<td>-.353</td>
<td>.004</td>
<td>.115***</td>
</tr>
<tr>
<td>RC1 Somatic Symptoms</td>
<td>.054</td>
<td>-.336</td>
<td>.006</td>
<td>.140***</td>
</tr>
<tr>
<td>RC2 Depressed</td>
<td>.103**</td>
<td>-.012</td>
<td>-.022</td>
<td>.142***</td>
</tr>
<tr>
<td>RC2 Achievement Oriented</td>
<td>-.044</td>
<td>.011</td>
<td>-.008</td>
<td>.082***</td>
</tr>
<tr>
<td>RC2 Introverted</td>
<td>.096**</td>
<td>.399</td>
<td>-.029</td>
<td>.054**</td>
</tr>
<tr>
<td>RC2 Pessimistic</td>
<td>.197***</td>
<td>.551**</td>
<td>-.086</td>
<td>.112***</td>
</tr>
<tr>
<td>RC4 Antisocial</td>
<td>.054</td>
<td>-.224</td>
<td>.012</td>
<td>.130***</td>
</tr>
<tr>
<td>RC4 Aggressive</td>
<td>-.014</td>
<td>-.370</td>
<td>.048*</td>
<td>.084***</td>
</tr>
<tr>
<td>RC4 Angry/Resentment</td>
<td>-.064</td>
<td>-</td>
<td>.081**</td>
<td>.090***</td>
</tr>
<tr>
<td>RC7 Anxious</td>
<td>.018</td>
<td>-.323</td>
<td>.008</td>
<td>.062**</td>
</tr>
<tr>
<td>RC7 Insecure</td>
<td>.058**</td>
<td>-.029</td>
<td>-.015</td>
<td>.087***</td>
</tr>
<tr>
<td>RC8 Psychotic Symptoms</td>
<td>.051</td>
<td>-.140</td>
<td>.010</td>
<td>.076***</td>
</tr>
<tr>
<td>RC8 Anxious</td>
<td>.018</td>
<td>-.367*</td>
<td>.014</td>
<td>.055**</td>
</tr>
<tr>
<td>RC9 Antisocial</td>
<td>.043</td>
<td>-.142</td>
<td>-.003</td>
<td>.054**</td>
</tr>
<tr>
<td>RC9 Aggressive</td>
<td>-.030</td>
<td>-.495</td>
<td>.038*</td>
<td>.034**</td>
</tr>
<tr>
<td>RC9 Narcissistic</td>
<td>-.004</td>
<td>-.692*</td>
<td>.019</td>
<td>.056**</td>
</tr>
</tbody>
</table>

*p <.05.  **p <.01.  ***p <.001.
Restructured Clinical Scale d

Of the 3 analyses performed on the RCd scale one showed a statistically significant intercept difference. These findings indicate an underprediction of depression for NHOPI’s although this effect size was small ($R^2 \Delta = .020$). No slope bias was evidenced for the RCd scale.

Restructured Clinical Scale 1

Both of the analyses for the RC1 scale indicated statistically significant intercept bias. These results evidenced an underprediction of depression and somatic symptoms for NHOPI’s, although only depression demonstrated a small effect size. No slope bias was demonstrated for the RC1 scale.

Restructured Clinical Scale 2

Of the four analyses performed on the RC2 scale only one of the items showed statistically differential prediction. Pessimism evidenced both statistically significant slope and intercept bias indicating an overprediction for NHOPI’s when using a common regression equation. The findings for slope bias evidenced a small effect size ($R^2 \Delta = .024$).

Restructured Clinical Scale 4

One of the three analyses performed on the RC4 scale showed statistically differential prediction. Of these items, angry/resentment evidenced statistically significant slope and intercept bias indicating an underprediction for NHOPI’s when using a common regression equation. However, the findings of the effect sizes for this statistically significant slope and intercept bias were small.

Restructured Clinical Scale 7

Both of the analyses for the RC7 scale indicated statistically significant intercept bias. These results indicate an underprediction of anxiety and insecurity for NHOPI’s. No slope bias was demonstrated for the RC7 scale.

Restructured Clinical Scale 8

The RC8 scale also demonstrated statistically differential prediction. Of the two analyses performed, anxiety was the only item to have evidenced statistically significant intercept bias. These results indicate an underprediction
for NHOPI’s, although this effect was small ($R^2 \Delta = .024$). No slope bias was
demonstrated for the RC8 scale.

*Restructured Clinical Scale 9*

For the RC9 scale, one of the three analyses revealed statistically
significant intercept bias. These findings suggest an underprediction of
narcissism for NHOPI’s although it was a small effect size ($R^2 \Delta = .044$). No
slope bias was evidenced for the RC9 scale.

In sum, of the 19 scale-criterion comparisons, nine demonstrated
statistically significant intercept bias. Scales RCd, RC1, RC2, RC4, RC7, RC8
and RC9 evidenced intercept bias. Of the nine incidents of intercept bias, none
exceeded a small effect size. Overprediction of psychopathology for NHOPI’s
was found on only one scale-criterion prediction: Scale RC2 and pessimism.
CHAPTER 5

DISCUSSION

A number of mean score differences were observed across the demographic variables of race, gender and age on several of the MMPI-2 Restructured Clinical Scales. Although these observed differences cannot address the issue of bias (McNulty, Graham, Ben-Porath, & Stein, 1997), they do provide valuable information concerning the relative level of functioning of the two racial groups at the time the assessments were made, assuming the scales are unbiased (Graham, 2006). Although previous studies have demonstrated setting specific differences in mean profile scores between minority racial groups and Caucasian psychiatric patients (Greene, 1987, Butcher et al., 1983, Tsushima & Onorato, Lawson et al., 1982), the present study was the first to examine these differences when comparing MMPI-2 RC scale scores of Caucasian and Native Hawaiian and Other Pacific Islander populations. When assessing for racial group differences on the MMPI-2 RC scales, five scales were found to be statistically different across racial groups. These significant differences include Restructured Clinical Scales, RCd (Demoralization), RC2 (Low Positive Emotions), RC3 (Cynicism), RC8 (Aberrant Experiences) and RC9 (Hypomanic Activation). The Caucasian group was identified to have scored significantly higher than the NHOPI group on scale RCd and RC2 whereas the NHOPI scored significantly higher than the Caucasian group on scale RC3, RC8 and RC9. Although these differences may say little about whether the MMPI-2 is biased (Greene, 1987), these scores may reflect actual differences in symptomatology and personality characteristics that are deserving of treatment consideration (McNulty et al., 1997). In general, these findings suggest that the Caucasian sample studied may have lower self esteem, express a lack of interest in things, display feelings of apathy and be more likely to report feeling discouraged and demoralized than the NHOPI sample. Conversely these findings also suggest that the NHOPI sample may be more likely to see other people as untrustworthy or uncaring, more likely
to report hallucinations, delusions, and bizarre sensory experiences, and be more likely to report a variety of characteristics consistent with hypomanic activation than their Caucasian counterparts (Butcher & Williams, 2000).

More specifically, the Caucasian groups elevated RCd scale suggests that they may be more pessimistic and feel more overwhelmed and incapable of coping with their current life circumstances than the NHOPI group sampled. The Caucasian group’s higher RC2 scale additionally suggests that this group may be less happy, more demoralized and have less energy to deal effectively with the demands of their lives than the NHOPI group sampled (Graham, 2006). The NHOPI sample however, endorsed more items consistent with an unstable mood, psychomotor excitement, and flight of ideas while viewing other people as potentially more exploitive than their Caucasian counterparts (Tellegen, 2003). These findings aside from the observed difference on RCd are consistent with the hypotheses that the NHOPI group would score significantly higher than the Caucasian group on MMPI-2 RC scales RC3 and RC9. These results are also consistent with the hypothesis that the Caucasian group would score significantly higher than the NHOPI group on scale RC2.

These findings appear to contradict prior assumptions that when symptom scales such as the MMPI-2 are used, NHOPI’s show higher levels of depression and demoralization symptoms than do white Americans (U.S. Department of Health and Human Services, 1999).

Greene (2000) has suggested that some of the mean scale differences identified in MMPI-2 research could simply be related to such factors as gender, age and education; because past research has suggested that these variables may be related to MMPI scale scores. Because many similar studies have not accounted for these demographic variables, one cannot be sure if some differences found in earlier studies can be attributed to differences on these or other important demographic variables (Timbrook & Graham, 1997). The current research was partially completed in an attempt to determine the extent or influence that the demographic variables of race, gender, age, and education might have on identified group MMPI-2 RC scale differences.
When collectively examining the influence of demographic variables such as race, gender and age across these two groups it was determined that gender and age appear to have little influence on group mean MMPI-2 RC scale scores, with the exception of younger participants tending to more likely endorse items consistent with at risk behavior as seen in their higher RC4 scale. More specifically, these results suggest that the younger group may be more likely to have histories of difficulties with the law while also being at a higher risk for substance abuse than the older group in this study. Graham (2006) has suggested that scores on MMPI-2 Clinical Scale 4 tend to be related to age, with younger persons scoring slightly higher than older persons. In fact in the MMPI-2 normative samples, Caucasians and Asian Americans scored somewhat lower on scale 4 (5-10 T-score points) than African Americans, Native Americans, and Hispanics (Graham, 2006; Butcher et. al., 2001). These findings did not significantly interact with any racial group differences.

Because preliminary analyses identified a racial group difference by education, this variable was controlled for through a MANCOVA. It was found that, with the exception of RC scale RC8, the racial group differences were not confounded by education.

Although previous studies have not found consistent evidence for racial test bias in the MMPI-2 (McNulty et al., 1997; Timbrook & Graham, 1994), the present study was the first to examine differential prediction or the possibility of test bias in the MMPI-2 Restructured Clinical scale scores between Caucasian and Native Hawaiian and Other Pacific Islander’s. One of the major goals of this study was to evaluate the extent to which MMPI-2 test scores are related to relevant extratest measures (Graham & Lilly, 1984). The Patient Description Form was specifically developed for the purposes of extracting extratest information drawn from clinical case management. Each of the scales from this instrument was designed specifically to assess symptoms and problems that outpatients are likely to have. These scales were intended to correlate with a number of MMPI-2 scales. Pairings that were identified to be significantly correlated were included in the analyses while those that were not were dropped.
from further analyses. Racial group differences were evaluated through criterion related validity so as to determine the accuracy with which an RC scale predicts Patient Description Form extratest characteristics (Timbrook & Graham, 1994). Differential prediction of conceptually relevant extratest data across racial groups demonstrated some evidence of test bias in a number of MMPI-2 scales.

Test bias was evaluated in this study by comparing across race the relation between MMPI-2 RC scale scores and conceptually related therapist rating scales. Although significant differences in mean scale scores were identified in five of the MMPI-2 RC scales, those differences were generally reflected in the conceptually related Patient Description Form scales. For example, the Caucasian sample was more likely to have been depressed according to PDF data, and this difference was reflected in the Caucasian groups significantly higher mean scale scores on RCd and RC2. This correlation was further evaluated using a step-down hierarchical multiple regression procedure to compare races. The indication of prediction bias was observed in 9 of 19 comparisons. These results were typically due to intercept differences between Caucasians and Native Hawaiian and Other Pacific Islander’s. When bias was present, it was often in the direction of underprediction of psychopathology in Native Hawaiian and Other Pacific Islander’s. Underprediction of psychopathology can be related to the identification of a false negative. Underprediction therefore creates the potential to overlook the reality of an observed difference in endorsed psychopathology between NHOPI and Caucasian people when in fact there may be a difference. Assuming underprediction of psychopathology in NHOPI’s therefore could contribute to the failure to provide mental health services when in fact treatment may be necessary. The potential failure to provide treatment when needed could contribute the observed underutilization of mental health services and overall mental health outcome measures of NHOPI people. The results however, do present some evidence for the overprediction of problems in Native Hawaiian and Other Pacific Islander’s by the RC2 scale. These results could be compared to the identification of a false positive, concluding that NHOPI’s experience
greater levels of Low Positive Emotions than Caucasians, when in fact this may not be true. Providing mental health treatment based on the measured overprediction of psychopathology can potentially place mental health professionals at risk for proving a service when such treatment is not necessary. Despite these findings, when the incremental change in $R^2$ was examined, the effect sizes consistently fell in the small range and would likely have no real impact on the interpretation of the MMPI-2 RC scales. In summary, the results of the present study suggest that the MMPI-2 can be used with NHOPI people with little fear of bias.

Although mean score differences on five of the MMPI-2 RC scales were identified in this research, an obvious pattern of similarity between mean scale scores and extratest correlations was evident. These results are similar to those by Arbisi et al. (2002) who identified evidence of differential prediction, particularly intercept bias, in the MMPI-2 clinical and content scales. They determined that racial group membership was not a substantive moderator between MMPI-2 scales scores and clinically relevant phenomena because the effect size of race on prediction was uniformly small. It is suggested that the MMPI-2 RC scales not be interpreted differently for Caucasian and Native Hawaiian and Other Pacific Islander’s as the magnitude of the predictor error does not warrant separate interpretive guidelines. Unfortunately, because no extratest criteria were significantly correlated with the RC scales RC3 and RC6, these scales could not be evaluated for potential bias. RC3 tends to measure an avowal of excessive trust of others. High scorers on this scale see other people as untrustworthy, uncaring, and exploitive while low scorers are likely to be naïve and gullible (Graham, 2006). RC6 is a measure of persecutory thinking. High scorers tend to feel targeted, controlled or victimized by outside forces while others may experience delusions, hallucinations and other symptoms of schizophrenia or delusional disorders (Graham, 2006). Due to the lack of research regarding racial test bias in the MMPI-2 involving NHOPI populations as well as other racial minority groups it was hypothesized that this study would not find any consistent evidence for racial test bias in any of the MMPI-2 RC scales.
including RC3 and RC6. In assessing for racial test bias associated with these MMPI-2 RC scales, future research should identify extratest criteria that correlate with RC3 and RC6.

Additional research is needed in a number of areas. To begin, the current study sample did not contain a sufficient number of participants for the correlational analyses to be conducted separately by gender. Although the correlational analyses suggest that racial group membership did not function as a moderator variable for the scales analyzed, it is important that the assertions made here are true for both genders (McNulty et al., 1997). Future research with similar populations should additionally attempt to examine a greater number of NHOPI’s and Caucasians to be studied so as to improve upon the verifications identified in the current study. Secondly, although the current research was helpful in identifying scale mean differences that appear to exist between the two groups studied, because only the Restructured Clinical scales were examined, it is unknown whether similar findings might be found with the remaining content scales (Butcher et al., 1990), the supplementary scales, the Harris-Lingoes subscales (Timbrook & Graham, 1994) or the PSY-5 scales (Harkness, McNulty & Ben-Porath, 1995).

An additional limitation identified in this study is the potential for intake criterion unreliability. For example, it is possible that the assignment of a clinical descriptor by the clinician or case manager may in some way be biased. Although the Patient Description Form was completed by eighty-six trained and licensed professionals, the inter-rater reliability of their ratings on this instrument remains unknown. Because reliability checks could not be measured in this setting, systematic bias may have been introduced serving as a potential source of error (Arbisi et al., 2002; Graham et al., 1999). The reliability of the other measures used in this study have however been demonstrated. In addition, the small correlations between the MMPI-2 RC scales and the extratest criteria created a lower probability that test bias would be identified. Future research should attempt to establish higher correlations between the predictor and criterion variables.
Finally, an additional matter concerns whether the present study and its subsequent results could be replicated in other settings and/or geographical locations. Due to the fact that this population had been assigned to an outpatient community mental health setting in Hawaii, future research is needed to determine if these findings apply to other psychiatric inpatient, medical, or forensic populations. In addition, it is unclear as to whether the current findings could be generalized to similar Caucasian and NHOPI mental health center populations outside of Hawaii. Because only Caucasian and NHOPI populations were examined in this study, it is also unclear as to whether similar studies with other minority groups such as Asian Americans, Native Americans and Hispanic populations would produce similar findings. It is anticipated that future research will examine these issues using the strategies implemented in this study. In closing, it is recommended that additional research intended to replicate the methodology of the present study include ethnically diverse populations whose primary language is not English. Details concerning the current study site, demographic characteristics of the client population, and information on the study participants MMPI-2 results were provided here to facilitate comparisons with future research in this area (McNulty et al., 1997).
APPENDIX

IRB/IACUC APPROVAL LETTER AND INFORMED CONSENT FORM

Florida State UNIVERSITY
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8633 · FAX (850) 644-4392

APPROVAL MEMORANDUM

Date: 4/25/2007

To: Brian Kehoe
2403 Hartfield Road, #402
Tallahassee, Florida 32303

Dept.: EDUCATIONAL PSYCHOLOGY AND LEARNING SYSTEMS

From: Thomas L. Jacobson, Chair

Re: Use of Human Subjects in Research
An Examination of the MMPI-2 in Native Hawaiian and Other Pacific Islander and Caucasian Mental Health Center Clients

The forms that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Human Subjects Committee at its meeting on 4/11/2007. Your project was approved by the Committee.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals which may be required.

If the project has not been completed by 4/9/2008 you must request renewed approval for continuation of the project.

You are advised that any change in protocol in this project must be approved by resubmission of the project to the Committee for approval. The principal investigator must promptly report, in writing, any unexpected problems causing risks to research subjects or others.

By copy of this memorandum, the chairman of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department and should review protocols of such investigations as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Protection from Research Risks. The Assurance Number is IRB00000446.

cc: Huijun Li
HSC No. 2007.287
Florida State University  
Department of Educational Psychology and Learning Systems  
307 Stone Building  
Tallahassee, FL 32306  
A Comparison of MMPI-2 Validity in Native Hawaiian and Other Pacific Islander and Caucasian Mental Health Center Clients  

CONSENT FORM FOR RESEARCH  

The following information is provided for you to decide whether you wish to participate in the present study. You may refuse to sign this form and not participate in this study. You should be aware that even if you agree to participate, you are free to withdraw at any time. As such, there is no coercion and no consequence now or ever for declining to participate. In addition, if you do withdraw from this study, it will not affect your relationship with North Shore Mental Health, the services it may provide to you, or Florida State University.

I HAVE BEEN INFORMED THAT:

1. Brian Kehoe, who is doctoral student at Florida State University, has requested my participation in a research study at North Shore Mental Health.

2. The purpose of this research is to evaluate the concurrent validity of the Minnesota Multiphasic Personality Inventory 2nd Edition (MMPI-2) concerning multicultural populations.

3. As a participant in this study I will be asked to complete the MMPI-2. This inventory is currently among the most widely used and researched objective personality inventories in the country. In completing this inventory I will be asked to sort 567 statements into one of three categories: “true,” “false,” or “cannot say.” Sorting these statements into these categories may exceed two hours. My responses to these statements will be scored on a number of scales designed to assess several categories of personality. My consent to participate in this study will additionally involve the completion of the patient description form (PDF). The PDF is a 187 item form which will be completed by a psychologist or trained psychology assistant. If permitted, they will use the PDF to rate the degree to which I possess each of the identified personality and symptom characteristics listed on the PDF. I understand that the researcher will be allowed to review and gather my existing demographic data from my NSMH file. As previously stated, if I chose not participate or at any time withdraw from participating in this study, this decision will in no way affect the relationship or services that I would ordinarily receive from North Shore Mental Health or Florida State University.

4. There are a limited amount risks or discomforts if I agree to participate in this study. Some physical discomfort or fatigue may be part of the study and confidentiality will adhere to Hawaii and Federal HIPPA law. In the case that I believe that I will lose
privacy by being involved in this study, master’s level case managers, licensed psychologists, and licensed psychiatrists will be available to assist and counsel me. Licensed psychiatrists and psychologists will also be available for any emotional complication I may experience by participating in this research. I will additionally be informed that participation is voluntary and that I can ask questions or drop-out anytime during this process without any negative consequence.

5. If requested, I understand that the researcher will inform me of my MMPI-2 results and interpretation. It is my understanding that the researcher has received extensive training with the MMPI-2 administration and interpretation and that any feedback provided to me will be supervised by a licensed clinical psychologist. This information is especially expected to benefit the AMHD Intensive Case Management population of Hawaii.

6. The results of this research study may be published but my name or identity will not be revealed. The following steps will be taken so as to ensure and maintain the confidentiality of my records. All information obtained during the course of this study will remain confidential, to the extent allowed by law. Federal Health Insurance Portability and Accountability Act (HIPAA) law guidelines will be adhered to and all test measurement information will be stored under locked files at North Shore Mental Health. Data will be stored at NSMH for a minimum of seven years or as directed under Hawaii State and Federal HIPPA law. Data will not be destroyed unless recommended by the AMHD as it should prove helpful in assisting in the improvement of case management to the general AMHD population in Hawaii.

Confidentiality will additionally be maintained as the researcher will destroy any master list containing identifying information after he has assigned subject codes. This separate data set will be created 'on site' and the researcher will only take a coded data set away from the NSMH site for analysis. This confidential data set will be evaluated at Florida State University.

7. I will not be paid for my participation however after completing the MMPI-2 I will be given a McDonalds food coupon for my participation in the amount of $5.00.

8. Any questions I have concerning the research study or my participation in it, before or after my consent, will be answered by a participating NSMH service provider, Brian Kehoe or Dr. Huijun Li, 307 Stone Building, Tallahassee, FL 32306 (850-228-1100 or 850-644-4592). NSMH can be reached by phone at 808-638-8700.

9. If I have questions about my rights as a subject/participant in this research, or if I feel I have been placed at risk, I can contact the Chair of the Human Subjects Committee, Institutional Review Board, through the Florida State University Office of the Vice President for Research, at (850) 644-8633.

10. I have read the above informed consent form. I understand that I may withdraw my consent and discontinue participation at any time without penalty or loss of benefits to
which I may otherwise be entitled. In signing this consent form, I am not waiving any legal claims, rights or remedies. A copy of this consent form will be offered to me.

Subject's Signature ___________________________________________ (Date)

__________________________________________

Other Signature (if appropriate) ________________________________ (Date)

__________________________________________
REFERENCES


measurements yearbook (pp. 104-107). Rutgers, NJ: Rutgers University Press.


Evans, C., & McConnell, T. R. (1941). A new measure of introversion-


Tellegen, A., & Ben-Porath, Y. S. (1992). The new uniform T-scores for the


of the Surgeon General. Rockville, MD: US Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services.


BIOGRAPHICAL SKETCH

Brian Kehoe

Brian Kehoe was born in Mesa, Arizona and spent his early years in Northern California. He later moved to the North Shore of Oahu, Hawaii where he graduated from Kahuku High School in 1995. Later that year, Brian began his undergraduate studies at Brigham Young University, in Provo, Utah. In 1996, Brian moved to Chicago, Illinois where he served a voluntary two year full time religiously oriented mission for The Church of Jesus Christ of Latter-day Saints. He resumed his undergraduate studies, graduating from Brigham Young University in 2001 with a B.S. in Psychology and Psychology Teaching with a minor in Special Education: Mild/Moderate Disabilities. Brian also received a Masters degree from the University of Hawai’i at Manoa in Counseling Psychology in 2004. He then moved to Tallahassee, Florida where he received his Ph.D. from The Florida State University’s combined doctoral program in Counseling Psychology and School Psychology. Brian has worked as a clinician, psychology extern, psychology resident and psychology intern in numerous settings including Tallahassee Memorial Hospital in Tallahassee, Florida, North Shore Mental Health in Oahu, Hawaii, Michael E. DeBakey VA Medical Center in Houston, Texas and Tripler Army Medical Center in Honolulu, Hawaii. Brian is an avid soccer player and enjoys staying active. He loves living close to the ocean and plans to remain in Hawaii.