RATES OF APPARENTLY ABNORMAL MMPI-2 PROFILES IN THE NORMAL POPULATION

by

Anthony P. Odland, M.S.

A Dissertation Presented to the School of Psychology
of Nova Southeastern University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

NOVA SOUTHEASTERN UNIVERSITY

2012

ACKNOWLEDGEMENTS

I extend my gratitude and thank my dissertation committee for their guidance, support, and time invested throughout this research project. The completion of this dissertation is the result of their meaningful contributions to my professional growth and development as a student scientist-practitioner.

I would also like to acknowledge NCS Pearson Inc. and Dr. Yossef Ben-Porath for providing the MMPI-2 interscale correlation data from the normative sample, and Dr. Marc Testa for providing a framework of SPSS syntax used to perform aspects of the resampling techniques.

I also gratefully thank my wife, Ashley, for her patience, love, support during my graduate school studies, and for her unyielding confidence in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

RATES OF APPARENTLY ABNORMAL MMPI-2 PROFILES IN THE NORMAL POPULATION

BY

Anthony P. Odland, M.S.

Nova Southeastern University

Abstract

Previous research suggests as more scores are interpreted, there is a coinciding increase in the chance significant scores will be obtained. Interpretation of the MMPI-2 can involve the analysis of as many as 98 or more separate scores, suggesting the measure has a strong proclivity for producing a high frequency of seemingly abnormal scores amongst normal healthy adults. In the current study the incidence of elevated MMPI-2 scores was simulated for the normal population using Monte Carlo methodology. Interscale correlations from the MMPI-2 restandardization sample were obtained to determine the percentage of the population with *N* or more seemingly abnormal scores. Simulations were conducted for all scales combined, and for the Clinical, Harris-Lingoes, Content, Content-Component, and Supplementary scales separately at varying *T*-score cutoffs. 36.8% of normal adults are expected exhibit at least one elevated score on the Clinical scales at 65*T*. The normal incidence of at least one seemingly abnormal score was 38.3% on the Content, and 55.1% on the Supplementary scales. When all scale groups are considered together, approximately 50% of the normal population has three or more significant scores, and at least seven seemingly meaningful scores are found for one out of five normal persons. These results imply that consideration of a large number of

MMPI-2 scales should be conducted with caution, and that high *T*-score cut-points may optimally increase confidence in the absence of corroborative test scores and extra test data.

CHAPTER I

Statement of the Problem

Accurate nosological classification is a primary concern in the field of clinical neuropsychology. Abnormality may be defined with varying inclusion criteria such as in terms of standard deviations from the population mean, percentile ranks, and confidence intervals. Parameters of atypicality are defined in order to facilitate the determination of whether a score or pattern of scores deviate significantly from a normative comparison standard (Lezak, Howieson, Loring, Hannay, & Fischer, 2004). Whether a performance is considered to be low, deviant, or abnormal reflects how an individual's quantified performance compares to an established cutoff score. In defining these criteria efforts are often made to strike a balance between specificity and sensitivity in order to maximize the clinical utility of a test. Although normative data are often provided within test manuals and present criterion defining abnormality, data are often not presented for the number of aberrant scores expected in a population. That is to say, manuals generally outline criteria corresponding to quality of individual scores or composites in reference to comparison sample, but do not furnish information about the meaning of a given frequency of scale elevations. This lack of base rate data may inadvertently require clinicians to speculate about the meaning of isolated abnormal scores (Palmer, Boone, Lesser, & Wohl, 1998).

Interpretation of a patient's performance should, in part, reflect how any abnormal scores compare in frequency to that expected to occur in the normative sample. A patient who scores significantly different from the mean on several subscales of a given test, for

example, may actually have the same number of aberrant scores as a meaningful percentage of the normative sample. Within a normative sample there may be a statistically high frequency of significantly high or low scores that are not necessarily clinically relevant (Brooks, Strauss, Sherman, Iverson, & Slick, 2009). Interpretations drawn from such data would vary from the interpretation that would otherwise be made for patients that obtain one or more scale scores that deviate significantly from the mean in the normative sample (e.g., occurring in the 1st percentile). When analyzing a pattern of scores a clinician must discern whether a patient's scores are atypical, and how many abnormal scores are necessary for there to be a departure in performance from the normative group (Ingraham & Aiken, 1996). Formulations that do consider a patient's frequency of elevated scores may lead to errors of interpretation. An obvious interpretive error is concluding impairment is present in a patient based on one or more significant scores when a condition does not exist.

The percentage of abnormal scores expected by chance within a given population generally increases as more scores are interpreted (Binder, Iverson, & Brooks, 2009; Ingraham & Aiken, 1996; Crawford, Garthwaite, & Gault, 2007). Generally speaking, a parallel can be drawn between the interpretation of multiple test scores and the increasing family-wise error rate associated with interpretation of an increasing number of statistical tests (Moran, 2003). A sort of psychometric paradox arises. As more scores are interpreted with the aim of performing a more thorough assessment, alpha is likely to inflate. In other words, there is an increase in statistically significant scores independent of clinical significance.

Over the last several years a productive line of research in this area stemmed from the work of Ingraham and Aiken (1996), and more recently Crawford and colleagues (2007). Crawford et al. (2007) outlined a generic method to determine the effects interpreting an increasing number scales has on the percentage of elevated scores observed in a given population. An advantage of this technique is base rates can be determined from interscale correlation tables easily accessible in most test materials.

An uninvestigated extension of these methodologies includes their application to the objective assessment of personality, an oftentimes important component of a neuropsychological battery. Therefore, the aim of the present study is to apply these methods to the Minnesota Multiphasic Personality Inventory – second edition (MMPI-2). Base rate data for the percentage of the population with significant scores expected in the normal population are not readily available. Moreover, the MMPI-2 lends itself to an examination of these base rates due to the large number of included scales (i.e., 118), its widespread use among neuropsychologists and clinical psychologists, and the potentially far reaching implications of its use diagnostically (Bow, Flens, & Gould, 2010; Rabin, Barr, & Burton, 2005)

CHAPTER II

Review of the Literature

Overview of the MMPI

The Minnesota Multiphasic Personality Inventory (MMPI) was developed across the 1930s and 1940s to provide a measure capable of sampling a varied array of behaviors. In part, the authors' goals were to create an all-inclusive inventory to replace the use multiple separate tests, that contained easily comprehensible items, and that provided characteristics that enabled its future development and refinement as personality research advanced. The 550 items comprising the measure were created by referencing the authors' personal clinical experiences, relevant texts, and other already available scales (Hathaway & McKinley, 1940a; McKinley & Hathaway, 1944). Items were constructed to elicit an answer of either "true" if an answer was mostly true or "false if an answer was not mostly true. An additional option of "cannot say" was also provided that the test taker could endorse if the item did not apply to them. Originally, the MMPI was administered as a set of cards that was sorted into 3 separate piles according to the test taker's responses (McKinley & Hathaway, 1940).

MMPI Test Construction

The original standardization sample included participants assumed to have no underlying illness whom were visiting family or friends at a university hospital that were aggregated into a normal comparison group. Pre-college high school graduates and skilled workers from local projects were also included as normal comparison groups. Additional participants included patients with different physical diseases from a

university hospital, all of whom did not have apparent psychiatric conditions, and outpatients from a university outpatient neuropsychiatric clinic (Hathaway & McKinley, 1940a).

For the construction of the MMPI scales a group of visitors seeing family at a university hospital were contrasted with a group of college students who were attending the University of Minnesota. Based on this contrast, for example, items were selected for Scale I (i.e., Hypochondriasis) if there was a percentage frequency difference in endorsement of at least two standard deviations between normal participants and those from the group that met criteria for Hypochondriasis. For Scale 1, 55 items were included (McKinley & Hathaway, 1940). Other Clinical scales developed through this type of methodology included the Depression scale (Hathaway & McKinley, 1942), Psychasthenia scale (McKinley & Hathaway, 1942), Hysteria scale, Hypomania scale, Psychopathic Deviate scale (McKinley & Hathaway, 1944), Paranoia scale, Schizophrenia scale, Masculinity-Femininity scale (Hathaway, 1956), and Social Introversion scale (Drake, 1946).

Efforts were also made to establish indicators of valid responding such as the F-scale (deviant responding), L-scale (defensiveness), a K-correction (indicator of subtle defensiveness), and a "Cannot say" score representing the frequency of items left unanswered (McKinley, Hathaway, & Meehl, 1948; Meehl & Hathaway, 1946). Raw scores obtained from the normal group were transformed into linear *T*-scores (Butcher et al., 2001).

Overview of the MMPI-2

Recognizing the need for a more current and improved standardized objective measure of personality, the University of Minnesota Press initiated a project to restandardize the original MMPI. During the restandardization project the item content of the MMPI changed from 550 to 567 items reflecting the addition of new content and the deletion of 16 duplicate items. A new more comprehensive normative sample of 2,600 people (1,462 females; 1,138 males) was obtained to more accurately reflect the United States' population and changes in its demographic composition (Butcher et al., 2001; Butcher, 1992). The Minnesota Multiphasic Personality Inventory – second edition (MMPI-2) restandardization sample, however, has an overrepresentation of both males and females with a college education or post-graduate education, and an underrepresentation of Hispanic and Asian-American subgroups. Raw scores for all scales except scales 0 and 5 (which had distinct distributions) were converted to uniform *T*-scores through a series of statistically normalizing steps, designed to approximate a composite of 16 individual distributions. This process involved deriving linear *T*-scores for each of the 16 distributions corresponding to percentile values, and then averaging them together (Butcher et al., 2001; Tellegen & Ben-Porath, 1992).

Since the inception of the MMPI-2 its utility as an objective assessment measure has been evinced by its wide use in clinical neuropsychology practice (Rabin et al., 2005) and by the voluminous number of related articles published in peer-reviewed articles, unpublished manuscripts, and presentations. The measure's use crosscuts a variety of clinical contexts, and has demonstrated the ability to provide contextually rich correlates

of human behavior. For example, the MMPI-2 is employed frequently in the assessment of emotional and neuropsychological symptoms such as memory and attention, is used to assess the possibility of malingering or symptom exaggeration, and is often used in the context of determining ability to return to work following the diagnosis or treatment of neurological conditions (Rabin et al., 2005). Review of the literature indicates the publication of thousands of articles associated with the MMPI and MMPI-2. As the MMPI-2 advanced from the first edition it has been investigated in a wide number of clinical populations that helped establish the utility of the Validity, Clinical, Harris-Lingoes, Content, Restructured Clinical, and Supplementary scales and subscales (Butcher et al., 2001).

Validity scales

The current edition of the MMPI has 9 indicators of valid responding including the "Cannot Say" score, Variable Response Inconsistency scale (VRIN), True Response Inconsistency scale (TRIN), Infrequency (F) scale, Back F (F_B) scale, Infrequency-Psychopathology (F_P) scale, Lie (L) scale, Correction (K) scale, and Superlative Self-Presentation (S) scale. These Validity scales are designed to detect underlying response styles and helps determine whether a test taker's item endorsement has compromised the validity of other scores (Butcher et al., 2001).

The Cannot Say score reflects the number of items on the MMPI-2 left unanswered. A raw score ≥30, for example, strongly indicates a profile is invalid due to a disproportionate number of unanswered items. If too many items are left unanswered it may affect a test taker's MMPI-2 profile by impacting the validity of other scales. The

VRIN and TRIN scales are not content specific and are used to determine general test taker response patterns that may be problematic for an accurate interpretation of the other scales. Each item included in the VRIN scale has a complementary item that is either similar or opposite in its content. If a certain number of these items are answered in an inconsistent manner it can suggest a haphazard testing approach. The TRIN scale is comprised entirely of item pairs that are opposite with respect to content. A large number of items endorsed false on the TRIN scale may suggest the test taker has a tendency to answer false to items on the test regardless of item content. A *T*-score ≥ 80 on either the VRIN or TRIN scales suggests the test taker's profile is invalid (Butcher et al., 2001).

Items included in the F-scale were selected on the basis of their infrequent endorsement in the normative sample. A high score on the F-scale is interpreted in combination with the VRIN and TRIN scales to determine profile validity. A high score may represent random or fixed responding, exaggeration of symptoms, or genuine psychopathology. The F_B-scale was developed to assess the back half of the MMPI-2 in the same way that the F-scale assesses the first 370 items. Significant differences (i.e., 30 *T*-score points) can indicate a change in response style from the first to the second half of the test. The Fp-scale uses the same rationale for construction as the F and F_B scales; however, its content was selected on the basis of endorsement infrequency in the normative sample and sample receiving services at inpatient psychiatric facility (Butcher et al., 2001).

The L, K, and S-scales were developed to assess defensive response styles. The L-scale is intended to determine test taker defensiveness in responding. The L-scale was

originally developed to detect more overt or unsophisticated defensive responding (Butcher, 2001; Graham, 2007). Unlike the L-scale, the K-scale was developed in an empirical fashion and is intended to detect less overt or more subtle attempts to deny psychological difficulties. A K-correction for defensiveness was developed from the items of the K-scale, and is used to adjust scores on some Clinical scales. The S-scale was constructed to determine whether item endorsement reflects a test taking attitude that is overly responsible and virtuous, such as when a test taker claims to be free of any psychological difficulties (Graham, 2006).

Clinical scales & Harris-Lingoes subscales

The ten Clinical scales are similar in structure to those included in the original MMPI with the exception of the previously mentioned changes in item content. The Clinical scales were constructed using a criterion keying approach. Different interpretations drawn from the Clinical scales, as outlined by Butcher et al. (2001), coincide with varying *T*-score ranges. A score of 45*T* is considered "low," 45-54*T* is "Average," 55-64*T* is "Moderate," 65-74*T* is "High," and ≥75*T* is considered "Very High." Specific descriptors for these ranges are outlined for each scale, and for two and three point code types. These characterizations reflect behavioral, cognitive, and emotional extra-test correlates determined from studies extending back to the original version of the MMPI (Graham, 2006). Seven of the Clinical scales were also deconstructed into the Harris-Lingoes subscales in an effort to separate the scales' content into more homogenous clusters. The Harris-Lingoes were developed with the aim to facilitate clinician understanding of the items so as to aid in the interpretative process

(Butcher, 2001). Coefficient alpha ranges from .34 on Scale 6 for males to .87 on Scale 7 for females. Test-retest reliabilities range from .54 on Scale 8 for females to .93 on Scale 0 for males (Graham, 2006).

Scale 1 (Hypochondriasis) generally represents specific and general complaints that are self-centered in focus, and that often involve a preoccupation with the body. Scale 2 (Depression) contains items generally relating to experiences a depressed individual may feel such as hopelessness, worry, and denial of happiness. Scale 2 is separated into five Harris-Lingoes subscales including Subjective Depression (D_1), Psychomotor Retardation (D_2), Physical Malfunctioning (D_3), Mental Dullness (D_4), and Brooding (D_5). Content on Scale 3 (Hysteria) includes items containing somatic complaints such as nausea, chest paint, headaches, and items that relate to the denial of psychological discomfort. Harris-Lingoes subscales for Scale 3 include Denial of Social Anxiety (Hy_1), Need for Affection (Hy_2), Lassitude-Malaise (Hy_3), Somatic complaints (Hy_4), and Inhibition of Aggression (Hy_5; Butcher et al., 2001).

Scale 4 (Psychopathic Deviate) includes topics such as family problems, delinquency, and difficulties with authority. Respective Harris-Lingoes subscales include Familial Discord (Pd_1), Authority Problems (Pd_2), Social Imperturbability (Pd_3), Social Alienation (Pd_4), and Self-Alienation (Pd_5). Scale 5 (Masculinity-Femininity) was constructed to include content pertaining to stereotypic gender roles, and does not have subscales. Scale 6 (Paranoia) contains content relating to topics such as suspiciousness, delusions of persecution, cynicism, and complaints against others. Subscales include Persecutory Ideas (Pa_1), Poignancy (Pa_2), and Naiveté (Pa_3). Scale 7 (Psychasthenia)

reflects content related to the modern day diagnosis of obsessive-compulsive disorder. Scale 8 (Schizophrenia) covers symptoms including disturbances of perception, bizarre thinking, social alienation, fears, as well as symptoms of neurological conditions (Gass & Russell, 1991). Harris-Lingoes subscales include Social Alienation (Sc_1); Emotional Alienation (Sc_2); Lack of Ego Mastery, Cognitive (Sc_3); Lack of Ego Mastery, Conative (Sc_4); Lack of Ego Mastery, Defective Inhibition (Sc_5); and Bizarre Sensory Experiences (Sc_6). Scale 0 (Social Introversion) has content dealing with social isolation, self-depreciation, and neurotic maladjustment. Coinciding subscales are Shyness/Self-Consciousness (Si_1), Social Avoidance (Sc_2), and Alienation-Self and Others (Sc_3; Butcher et al., 2001; Graham, 2006).

Content scales & Content-Component subscales

In contrast to the MMPI-2 Clinical scales which were developed using a criterion reference approach, the Content scales were constructed through a series of steps including use of deduction, theory, and consideration of underlying constructs verified statistically. Items were eliminated that shared variance with other scales so as to obtain scales that were statistically "clean" (Butcher, Graham, Williams, & Ben-Porath, 1990). Unlike the Harris-Lingoes subscales, the Content scales contain items from the entire body of the test (Graham, 2006). Unlike the Clinical scale items that were selected without regard to their face-validity but rather to maximize the dichotomy between two nosological categories, the Content scales are face-valid. Consistent with Content scales intended development they have a high degree of internal consistency (e.g., alpha from .72 to .86 for males), and only slightly lower test-retest reliability (Butcher, 1990). The

Content scales were developed using military personnel, college students, airline pilot applicants, inpatient psychiatric and chronic pain patients, and inpatient alcohol and drug abuser samples in addition to the MMPI-2 restandardization sample. Participants were excluded from the normative group if (1) they omitted more than 40 items, (2) obtained a raw F-scale score more than 25, (3) or had a raw F_B score more than 25 (Butcher et al., 1990). Similar to the Clinical scales, raw scores were transformed into uniform *T*-scores to facilitate comparisons between scales.

Fifteen separate Content scales were developed from the item pool including: Anxiety, Fears, Obsessiveness, Depression, Health Concerns, Bizarre Mentation, Anger, Cynicism, Antisocial Practices, Type A (E.g., hard-driving, irritable, and work-oriented), Low Self-Esteem, Social Discomfort, Family Problems, Work Interference, and Negative Treatment Indicators scales (Butcher et al., 1990). The construct, rational, and deduction based approach used to construct the Content scales is reflected in the scale labels. For example, items within the Anxiety scale reflect behavioral correlates in the spectrum of how anxiety is experienced.

Like the Clinical scales, the majority (12/15) of the Content scales were deconstructed into component subscales. These Content-Component subscales were developed using a rational and statistical approach using factor analysis and Cronbach's alpha.For example, the Depression scale is partitioned into Lack of Drive (DEP1), Dysphoria (DEP2), Self-Depreciation (DEP3), and Suicidal Ideation (DEP4). There are 27 Content-Component subscales in total. Their internal consistency is lower than that of their parent scales, ranging from .47 to .90.

Supplementary scales

Consistent with the vision set out by the original MMPI authors, the large item pool has allowed for the measure to be developed by others into literally hundreds of additional supplementary scales. Several of these scales were selected for inclusion when the MMPI-2 was developed based on validity, reliability, and their perceived clinical utility. These scales, collectively named the Supplementary scales represent an array of different domains intended to be used in combination with the Clinical and Validity scales (Graham, 2006). The Supplementary scales include 15 separate scales: Anxiety (A scale), Repression (R Scale), Ego Strength (Es), MacAndrew Alcoholism Scale-revised (MAC-R), Addiction Acknowledgement (AAS), Addiction Potential (APS), Marital Distress (MDS), Hostility (Ho), Overcontrolled hostility (O-H), Dominance (Do), Social Responsibility (Re), College Maladjustment (Mt), Masculine Gender Role (GM), Feminine Gender Role (GF), and Posttraumatic Stress Disorder-Keane (PK; Butcher et al., 2001).

MMPI-2 Interpretation

Prefacing a discussion on MMPI-2 interpretation, Graham (2006) asserts a profile of scores could not be developed capable of describing a test taker's characteristics to absolute certainty. Different behavioral, cognitive, and emotional extra-test correlates may apply to test takers in varied ways. Graham (2006) states the MMPI-2 should be used adjunct to other psychological tests, observational data, and interview. Even in the instance of using computerized interpretations, scores should be interpreted in the context of other available information. Consistent with these recommendations, Green (2000)

described interpretation of the MMPI-2 as a multistage process involving the incorporation of a depth of information.

Prior to interpreting the Clinical scales the Validity scales should be evaluated to help ensure a protocol reflects an accurate characterization of the test taker. The Validity scales also need to be interpreted with consideration to contextual and patient factors. For example, a high score on the F-Scale may reflect genuine psychopathology, or when interpreted simultaneously with the VRIN and TRIN scales can represent random responding or acquiescence (Butcher et al., 2001; Graham, 2006; Green, 2000). Graham (2006) recommended interpretive strategy answer questions regarding examinee test-taking attitude, level of adjustment, behavior, diagnosis, and implications for treatment.

Interpretation of the Clinical scales may include consideration of one, two, or three point code types. Code types refer to combinations of specific Clinical scales such as Scales 2 and 7 (i.e., 2/7), that are correlated with commonly occurring extra-test emotional and behavioral "correlates", and inform the psychologist which characteristics are most prominent in an individual. Three and two point code types are inclusive for combinations of various Clinical scales with the exception of Scales 0 and 5. Graham (2006) recommends interpreting only those code types where the lowest score is at least five *T*-score points greater than the next lowest Clinical scale, and that are also greater than a *T*-score = 59. He also states, however, that inferences about symptoms should not be made, or "…made with considerable caution…" unless a *T*-score $\geq$ 65 (Graham, 2007, p. 94).

Harris-Lingoes subscales are interpreted in concert with the Clinical scales, and should not be interpreted separately. Graham (2006) recommends the Harris-Lingoes subscales not be interpreted unless the coinciding parent scale is at least 65*T*. As the Clinical scales are heterogeneous in item content, the intent of the Harris-Lingoes subscale interpretation is to convey information about the specific clusters of items of endorsed. The Harris-Lingoes subscales may not add meaningful information to parent scales with very highly elevated (Graham, 2007; and Green, 2000) scores, however, since all or most of the subscales will also be elevated in such an instance.

Following interpretation of Clinical and Validity scales interpretive guidelines suggest the Content and Supplementary scales be interpreted. Similar to the Harris-Lingoes subscales, the Content-Component subscales are intended to be used to better understand elevated parent scales (i.e., Content scales). Guidelines recommend the Content-Component subscales not be interpreted unless their coinciding parent scale is significant at $T \geq 60$. The component subscales should not be interpreted independent of the Content scales (Graham, 2006). Similar to the majority of the previously mentioned scales, most low scores obtained on the Supplementary scales are not interpreted.

Throughout the interpretative process clinicians may also wish to consider the effect of specific patient demographics on observed scores (Green, 2000). In general, however, demographic characteristics such as age, education, and ethnicity have minimal influence on MMPI-2 scores, and the test has proved itself useful across various cultural and demographic groups (Butcher, et al., 2001).

Considerations for Interpretation of Multiple Tests

Normal Population and Abnormal Scores

Diagnostic accuracy is a central concern in the field of clinical neuropsychology. Neuropsychological tests and measures have historically been validated to distinguish between patients with and without central nervous system disorders (Benton, Sivan, Hamsher,Varney, & Spreen, 1994; Reitan & Wolfson, 1985), which oftentimes involves a balance between the rates of correct positive and correct negative diagnoses. Abnormality is commonly defined in terms of standard deviations from the population mean, and is used to determine if a given score deviates significantly from a normal standard of comparison (Lezak, Howieson, Loring, Hannay, & Fischer, 2004). Normative data are often provided in test manuals outlining criterion for abnormality in the form of standardized units from the mean, percentile ranks, or confidence intervals. Data that are presented typically reference the theoretical normal distribution as a comparison from which scores can be contrasted. Use of this theoretical distribution implicitly assumes, however, only one contrast is being made; and data are infrequently presented for the rates of elevated scores expected due to normal variation or measurement error in a population for instances where multiple test scores are interpreted together. While manuals generally outline information about individual scores or composites, they often do not furnish information about the normal co-occurrence of expected scale elevations.

If abnormality is defined by a test score falling one standard deviation from the mean on a characteristic that is normally distributed then about 84% of the population would be identified as normal and 16% as abnormal. When a given characteristic is

normally distributed one is able to estimate the percentage of a population that obtains a score for a test; however, when multiple measures are interpreted (which is almost always the case in neuropsychological assessment) the probability of obtaining more than one elevated score also increases (Binder, Iverson, & Brooks, 2009; Crawford, Garthwaite, & Gault, 2007; Ingraham & Aiken, 1996; and Palmer et al., 1998). Binder, Iverson, and Brooks (2009), for example, pointed out that a battery limited to scores obtained from the Wechsler Adult Intelligence Scale – III and Wechsler Memory Scale – III produces 37 subtest and index scores. This example draws attention to the increasing family-wise error rate that would be associated with a large battery of tests. This seemingly large aggregate of test scores may actually under predict the number of scores obtained and interpreted from a more inclusive test battery.

Crawford and colleagues (2007) pointed out that data for the percentage of a population expected to have one more abnormal score on several scales for many specific tests are not readily available. Moreover, because the number of abnormal scores in a population is likely to be greater when interpreting more than one scale (as opposed to the interpretation of a single score), one cannot reliably assume that with a cutoff score of one standard deviation from the mean that only 16% will be classified as deviant. Similar to the interpretation of multiple significance tests, type I error is likely to aggregate across tests and produce a family-wise error rate that is larger than that of its individual component tests. Consequent scores derived from the battery facilitate the assessment of a compendium of domains and lend to thoroughness, however, the coinciding large number of test scores accord with an increasing likelihood that there will be more

statistically elevated scores. This indicates some measures more comprehensive in scope, which rely on a large number of scores, may yield a correspondingly high percentage of elevated scores. Although few would argue that a single elevated score in a battery of tests represents a clinically significant departure, interpretation of (for example) two or three aberrant scores may be more ambiguous. That is to say, as the frequency of score elevations increases interpretation becomes increasingly complex and base rates for the occurrence of multiple elevations becomes critical. Practically speaking, without these base rates this type of ambiguity forces clinicians to surmise the significance of scores, such as whether they are indicative of psychopathology or are normally occurring among healthy test takers (i.e., people without significant psychological disturbance).

Heaton, Grant, & Matthews (1991), for example, demonstrated that when 40 measures from the expanded Halstead-Reitan Battery were considered together, the average participant from the normal standardization sample had four or more test scores considered significant using a cutoff score more than one standard deviation from the mean. When 26 Halstead-Reitan measures were examined in a larger standardization sample (Heaton, Miller, Taylor, & Grant, 2004), as many as 87% of the normal group produced one or more "impaired" score and 34% produced at least five.

Over the last several years there has been an increasing consideration for the number of abnormal scores in a patients' test performance (Brooks, et al., 2009). As mentioned previously, research has indicated deviant scores are common within normative groups. Previous analyses determined the occurrence of elevated scores coincides with the number of scales administered and average intercorrelations between

scales or tests. The percentage of the normal population with elevated scores generally increases when scales are weakly correlated and as additional scales are added. The occurrence of abnormal scores also varies with established cutoff values (more stringent cutoff values yield fewer significant scores), demographic characteristics, and the intelligence of those being tested (See Binder, Iverson, & Brooks, 2009; Brooks et al., 2008, 2009; Brooks & Iverson, 2010; Crawford et al., 2007; Diaz-Asper, Schretlen, & Pearlson, 2004). Higher levels of parental education are associated with lower percentages normal children deviant scores on tests of cognitive ability (Brooks, Sherman, & Iverson, 2010).

Crawford and colleagues (2007) used Monte Carlo simulation to estimate the number of low Index scores expected on the WAIS-III (Wechsler, 1997a) and WISC-IV (Wechsler, 2003), with abnormality defined at one standard deviation below the mean. At least one significant Index score was obtained by 34% of adults and 37% of children. Brooks, Iverson, Holdnack, & Feldman (2008) found that the majority of normal older adults in the Wechsler Memory Scale 3rd Edition (Wechsler, 1997b) standardization sample had at least one of eight subtest scores in the impaired range (one *SD* below the mean). Binder and colleagues (2009) reviewed the literature related to these base rates, and concluded that apparently abnormal test scores are common in normal populations across a variety of neuropsychological tests. More stringent definitions of clinical significance (e.g. 1.5 *SD*s from the mean) reduce but do not eliminate the incidence of significant scores in normal persons.

The Monte Carlo method demonstrated by Crawford and colleagues (2007) demonstrated a high level of accuracy for estimating the percentage of the normal population with *N* or more abnormal scores for batteries of tests from interscale correlation matrices available in test manuals (Schretlen, et al., 2008; Brooks & Iverson, 2010; Decker, Schneider, & Hale, 2012). The Monte Carlo methodology proved to be an efficient and effective resource for practicing neuropsychologists in helping delineate meaning from a battery of tests. Generation of these base rate data conveniently allows clinicians to avert interpretation errors associated with lack of access to base rate normative data. A noted limitation for cognitive ability tests, however, includes the over and under estimation of base rate data for those with low and high intelligence, respectively (Brooks & Iverson, 2010).

The Monte Carlo simulation recreates the distributions of test scores obtained by members of a standardization sample using scale variance and covariance contained in the interscale correlation matrix. A large sample (one million observations) of random normally distributed variates with a mean of zero and standard deviation of one are then subjected to the pattern of correlations from the restandardization sample. Generated scores can then be counted at a specific cut-point to determine the frequency of a population with scores falling beyond a specified criterion. Mathematically, this process involves (1) obtaining a Cholesky decomposition of an interscale correlation matrix, (2) generating one million random normally distributed vectors that have a specified number of scores with a mean of zero and *SD* of one consistent with the number of scales in the correlation matrix, (3) and then postmultiplying the scores by the Cholesky

decomposition matrix by each generated vector to produce one million normally distributed observations with scores reflecting the correlative pattern of the normative sample. A higher incidence of abnormal scores was shown to be associated with lower correlations among scales, increasing numbers of scales, and less conservative definitions of abnormality in terms of standard deviations from the normative mean.

Recent studies have reported psychometric data on the number of apparently abnormal scores expected in normal groups for various tests including the NEPSY-II (Brooks, Sherman, & Iverson, 2010), Children's Memory Scale (CMS; Brooks, Iverson, Sherman, & Holdnack, 2009), Neuropsychological Assessment Battery (NAB; Brooks, Iverson, & White, 2007), WAIS-III and WISC-IV (Crawford et al., 2007), Wechsler Memory Scale-III (WMS-III; Brooks, Iverson, Holdnack, & Feldman, 2008), Halstead Reitan Neuropsychological Test Battery (Axelrod & Wall, 2007), and for flexible batteries (Palmer et al., 1998; and Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008).

Brooks, Iverson, and White (2007) examined the potential for misclassification of mild cognitive (MCI) impairment or dementia in older adults on the NAB. The authors used the NAB standardization sample ($N = 742$), which was comprised of older adults ranging from 55 to 79 years of age ($M = 68.1$, $SD = 6.6$), and with a mean education of 13.5 years ($SD = 2.9$). The results showed that more than half of the standardization sample obtained one or more abnormal scores at one standard deviation from the mean, which is a sharp contrast to the 16% that would be expected using the binomial distribution. When cutoff criteria were set more stringently (i.e., 5^{th} percentile), as many as 30.8% of the standardization sample obtained one or more abnormal score consistent

with a diagnosis of MCI. The authors demonstrated increased intelligence to be associated with a lower prevalence of abnormal memory scores. For example, more than 80% of those with low average intelligence were found to have one or more low memory scores. They also noted that higher intelligence test scores were not necessarily commensurate with memory scores, and that low memory scores were common in those with above average intelligence (Brooks, Iverson, & White, 2007).

Older adults may also be misclassified as having MCI with the WMS-III when multiple scores (i.e., eight) are simultaneously interpreted. Using 550 adults from the WMS-III standardization sample ranging in age from 55 to 87 years (M = 72.8; SD = 9.0) Brooks, Iverson, Holdnack et al. (2008) found low memory scores to be common. At one standard deviation from the mean as many as 64.1% the sample had one or more abnormal scores, and more than one quarter had three or more abnormal scores. A significant effect was observed in the percentage of the sample with abnormal scores due to level of intelligence. For example, one half of those with an estimated full scale intelligence quotient of less than 80 had one or more significant scores at approximately 1.5 *SD*s from the mean. In comparison, 21.4% of those with a full scale intelligence quotient between 110 and 119 had one or more significant score.

Extending the above mentioned methodology to pediatric neuropsychology, Brooks and colleagues (2009, 2010) explored base rate data for the incidence of abnormal scores on the CMS and select subtests of the NEPSY-II. Results showed that as a larger number of scores are interpreted and as cutoff criteria are made less stringent there is an increased frequency of children with significant scores. When simultaneously evaluating

six indices from the CMS (i.e., Learning, Verbal Immediate, Verbal Delayed, Verbal Delayed Recognition, and Visual Delayed) using a sample of 1,000 children and adolescents it was found that the prevalence of abnormal memory scores were related to intelligence. Children and adolescents with below average intelligence (FSIQ between 70-89) were 7.1 times more likely to have one or more abnormal scores than those with above average intelligence (FSIQ ≥ 110). The percentage of the sample with one, two, three, or four or more abnormal scores decreased in accordance with increasing WISC-III full scale intelligence (Wechsler, 1993). As many as 37.6% of normal healthy children and adolescents were shown to present with one or more abnormal scores when abnormality was defined as one standard deviation from the mean. When cutoff criteria were tightened (i.e., two standard deviations from the mean) the frequency of children and adolescents with one or more significant index scores decreased to 20%. The authors also noted "slight" differences in the prevalence of abnormal scores across age bands. Specifically, there was an observed tendency for an increased prevalence of significant scores among lower age bands (Brooks, Iverson, Sherman et al., 2009).

The percentage of children with abnormal scores was also found to vary with the level of parental education. That is to say, there was an inverse relationship between parental education and the prevalence of abnormal scores. For the NEPSY-II the authors found that generally as parental education increases there is a subsequent decrease in the prevalence of abnormal scores regardless of age across different cutoff criteria. For example, when defining abnormality as a scaled score ≤ 8 on an abbreviated battery, 88.2% and 37.8% children between the ages of five and six have two or more significant

scores when parental education is ≤ 11 and ≥ 16 years, respectively. In comparison, when ignoring parental education, 55.2% of children between the ages of five and six will have two or more abnormal scores (Brooks et al., 2010).

When interpreting multiple scores, these results should encourage test interpretation that considers a patient's frequency of significant scores to the prevalence of elevated scores in a normal and healthy population. With the aim of gleaning a thorough portrait of a patient's abilities or characteristics, it is essential that clinicians integrate both the level and frequency of a patient's scores into their final analysis and interpretations. Formulations drawn from these methods are likely to increase the accuracy of subsequent conclusions and recommendations.

Significant scores are also common for neurologically healthy people on flexible neuropsychological batteries. Palmer and colleagues (1998) administered a flexible battery to 132 neurologically healthy older adults between the ages of 50 and 79 (*M*= 63.8, *SD*= 7.7) to determine the percentage of examinees with borderline and impaired scores. Participants were excluded on the basis of abnormal findings on a neurological/physical examination and interview. A percentage of the normal population obtained significant scores on each of the separate 26 measures in the battery. Looking across all 26 scores, 73.5% of the sample had one or more impaired scores, and almost half (i.e., 47.7%) had two or more impaired scores. 37.1% and 23.6% had one or more and at least two scores ≥ 2.0 SDs from the mean, respectively (Palmer et al., 1998). Similarly, on a battery comprised of 32 measures 66% of a sample of 130 normal healthy participants were found to have discrepancies of three or more standard deviations

between their highest and lowest scores. After excluding these participants' lowest and highest scores 27% still had discrepancies of three or more standard deviations between their next highest and lowest scores (Schretlen, Munro, Anthony, & Pearlson, 2003).

Schretlen and colleagues (2008) used a sample of 327 neurologically normal adults between the ages of 18 and 92 to compare the effectiveness of the Monte Carlo simulation method compared the binomial model. Results generally supported the utility of the Monte Carlo simulation method. Consistent with the previously mentioned studies, the prevalence of abnormal scores increased as more tests were interpreted (i.e., 10, 25, and 43 tests) and as cutoff criteria were made less stringent. Another important finding included an association between significant scores on tests of cognitive ability and demographic variables such as age, education, sex, race, and an estimation of premorbid intelligence.

Decker, Schneider, and Hale (2012) compared the binomial and Monte Carlo models' accuracy on a co-normed group of 14 tests (Woodcock – Johnson Tests of Cognitive Abilities – Third Edition; WJ III) by contrasting their performance with known base rates obtained through actual sampling. Results were compared across ages 6-8, 9-13, 14-19, 20-39, and 40+. Using regression-based model fit indices to contrast each methodology the authors determined that both techniques were significantly correlated with observed rates. However, the Monte Carlo method produced significantly larger correlations than the binomial model, and produced a better fit to the real base rates across all of the age ranges.

Taken in sum, these studies suggest neuropsychologists should consider several psychometric principles while gleaning meaning from a battery of tests. Regardless of the specific measure interpreted, statistically significant scores on a battery of tests is commonplace amongst normal healthy examinees. As the number of scores simultaneously interpreted increases there is likely to be a coinciding increase in the number of aberrant scores. Research also indicates demographic characteristics of an examinee influence the occurrence of significant scores on tests of cognitive ability. The defined parameters of significance impact the number of scores observed as being abnormal. Specifically, as criteria for significance is made more stringent the percentage of the population with abnormal scores decreases. Simultaneous interpretation of tests, measures, or scales that share little or no variance (as opposed to being strongly correlated) are more likely to result in a increased incidence of significant scores in normal healthy people (Axelrod & Wall, 2007; Binder, Iverson, & Brooks, 2009; Brooks, Iverson et al., 2009; Brooks et al., 2007; Brooks et al., 2010; Brooks, Strauss et al., 2009; Crawford et al., 2007; Palmer et al., 1998; Schretlen, Munro, Anthony, & Pearlson, 2003; and Schretlen et al., 2008).

Interpretation of Multiple MMPI-2 Scales

Use of multiple scales lends to the MMPI-2's comprehensiveness and allows clinician's to extract a broad number of patient descriptives. However, as suggested previously, past research suggests simultaneous consideration of several or more MMPI-2 scales will yield a high percentage of elevated scores (i.e., those scores significantly higher than the mean). Consistent with the suggested cutoff scores, elevated scores are

defined here as those scores within the upper 1st percentile for the Validity scales and scores falling within the upper 6.6th (i.e., 1.5 *SD*s from the mean) percentile for all other scales (Butcher et al., 2001). If previous research is correct in indicating the frequency of significant scores increases with the addition of each subsequent scale, then the MMPI-2 would seem particularly susceptible to being impacted by consideration of all of its scales. For example, interpretation of the Validity and Clinical scales yields 18 separate scores; with the addition of the Content and Supplementary Scales the number of scores more than doubles increasing to 48. Consistent with previous research, one expects that as additional scores are interpreted, there would be a general increase in the number apparently abnormal scores (Crawford et al., 2007; and Binder et al., 2009). In relation to interpretation of the MMPI-2 it is likely that there are statistical and clinical consequences of test interpretation that relies on a high frequency of data points. Specifically, it is highly plausible that a clinician utilizing many data points will have a correspondingly high number of significant scores independent of any psychopathology in the patient.

The MMPI-2 is the most frequently used test by neuropsychologists for the objective diagnosis of primary psychiatric disorders (Rabin, Barr, & Burton, 2005), psychiatric symptoms secondary to neurological disorders, and for the differential diagnosis of psychiatric versus neurological disorders (Lezak et al., 2004; and Gass, 2006). The original Clinical scales were constructed empirically to distinguish patients with known psychiatric diagnoses from normal individuals (Graham, 2006). A variety of additional scales may be interpreted including 15 Supplementary scales, 31 Harris-

Lingoes subscales, 15 Content scales, and 27 Content-Component subscales (Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001). A *T*-score of 65 or higher on a scale is considered a significant, on the basis of its ability to separate clinical groups from the normative sample (Butcher & Pope, 1992). This cutoff point would be expected to correctly characterize 92% of normal persons if a single Clinical scale was administered.

As previously discussed, practitioners may interpret any number of scores ranging from the 10 Clinical scales to all available 98 scales. Psychological or emotional difficulties may be hypothesized based upon even more lenient *T*-score elevations (e.g., *T*-score ≥ 60; Butcher et al., 2001). Data are currently unavailable for the incidence of what would otherwise appear to be clinically significant MMPI-2 score elevations in normal individuals when more than one scale is considered.

Monte Carlo Simulation

Monte Carlo methods, also referred to as resampling or simulation techniques encompass commonly employed methodologies used across various disciplines including the social sciences, chemistry, physics, engineering, and mathematics to approximate solutions through statistical sampling (Fishman, 1996; Lemieux, 2009; Madras, 2000; and Rubino & Tuffin, 2009). Monte Carlo methods invariably employ the use of computer systems in order to calculate large quantities of data; they are an efficient means for solving complex problems, and yield a tolerable degree of error (Fishman, 1996). In order for a simulation technique to be of value provided a given criterion, the properties of the target distribution must be understood. For example, if it is the aim of a researcher to draw 1,000,000 pseudorandom observations it is important to understand

the parameters of the population of interest, such as if the distribution of interest is normally distributed, and whether it has a negative or positive skew. Knowledge of the distribution of scores underlying an event enables one to set parameters for the simulation (e.g., $M = 0$, $SD = 1$; Gentle, 2003). Terminologically speaking, "random numbers" that are generated within specified parameters refer to those pseudorandom numbers generated from a uniform distribution, and "random variates" refer to random numbers derived from a uniform distribution that have been transformed into a related value. A noted strength of Monte Carlo methods is their flexibility and wide array of potential uses (Gentle, 2003).

One useful application of the Monte Carlo method involves generating 1,000,000 pseudorandom (N) uncorrelated normally distributed vectors each with k (as in a N x k matrix) scores that have a mean of zero and standard deviation of one, and then postmultiplying the vectors by the Cholesky decomposition (CD) of a targeted square matrix. Postmultiplication of the vectors by the CD of a k x k matrix acts to constrain the observations to the correlative patterns inherent to the normative interscale correlation matrix. In a sense, characteristics from the normative sample are mapped onto the scores from each of the 1,000,000 pseudorandom vectors.

For example, in attempting to simulate a response style of a normal healthy population on a given neuropsychological test, one could obtain a CD of a k x k interscale correlation matrix derived from a large normative sample, and then postmultiply the pseudorandom independent normal vectors by the CD. Consistent with the number of loops specified each iteration represents a certain observation or participant. Since the

values for each observation have a mean of zero ($SD = 1$) consistent with aforementioned parameters, they may be transformed into *T*-scores, scaled scores, or standard scores. If the desired product is a *T*-score, one simply multiplies each value by ten and adds 50 to the product (Crawford et al., 2007).

Purpose of the Study

In psychological and neuropsychological assessment multiple scales are often administered to patients, as in the instance of a flexible or fixed test battery. Such batteries often include several or more tests, each of which can be comprised of multiple subscales and scores. Use of multiple scores within a specific test is commonplace; and, practitioners may interpret any number of scores on the MMPI-2 that can range from the ten Clinical scales to more than 98 scales. As previously discussed, a *T*-score of 65 or higher is considered significant, because of its ability to separate clinical groups from the normative sample (Butcher & Pope, 1992). This cutoff point is expected to correctly characterize 93% of normal persons if a single score was interpreted. However, psychological or emotional difficulties may even be inferred from more moderate *T*-score elevations, which according to past research suggests results in a higher rate of significant scores (Ingraham & Aiken, 1996; and Butcher et al., 2001). Data are currently unavailable for the incidence of what would otherwise appear to be clinically significant MMPI-2 score elevations in normal individuals when more than one scale is considered. Significant scores become more common in the normal population as the number of test scores increase due to chance (Ingraham & Aiken, 1996).

The consequences interpreting a high number of scales could include mistakenly ascribing psychopathology in normal examinees from *N* or more significant scores. Patients may be given inappropriate treatment, denied civil competency, not permitted to live independently, make their own medical decisions, operate a motor vehicle, return to work, or manage their own finances. Potentially, the incorrect attribution of symptoms to psychological causes could also occur in the presence of neurological conditions such as epilepsy, chronic pain, mild cognitive impairment, head injury, and exposure to neurotoxins. The attribution of cognitive impairment to psychiatric disorders or emotional consequences to neurological disorders may occur without the presence of any real psychological impairment. A consequence could include an inappropriate referral or decision regarding the discharge of a patient in substance abuse treatment or chronic pain management treatment facilities. Faulty interpretation of the MMPI-2 may contribute toward incorrect decision making in child custody cases, career counseling, the prediction of violent behaviors, as well as in the instance of employment screening for high risk occupations, such as airline pilots, personnel at nuclear power facilities, graduate and professional students, firefighters, paramedics, law enforcement personnel, as well as seminary students (Graham, 2006; and Green, 2000).

In continuation of other practical efforts to make available psychometric data for the frequency of elevated scores in normal healthy people, it is the aim of this study to explore the occurrence of statistically significant scores using the MMPI-2 restandardization sample. Consistent with Binder and colleagues (2009) call for the examination of functional domains and tests with a proclivity for yielding a high

likelihood significant scores (e.g., high number of interpreted scales), the current study will examine the effects of simultaneous interpretation of commonly used combinations of Clinical, Harris-Lingoes, Content, Content-Component, and Supplementary scales. The expected rates of the normal population with significant scores will therefore be estimated for various groups of MMPI-2 scales at different cutoff scores using a Monte Carlo simulation technique (Schretlen et al., 2008).

Research Hypotheses

The main questions to be examined by the present study relate to the effects of interpretation of an increasing number of scales on the percentage of the population with significant scores. The primary research hypotheses include:

1. There will be a difference between males and females in the mean number of observed scores 1.5 standard deviations higher than the population mean for all scale groups combined. This exploratory hypothesis is based upon the observation that male and female scores on the MMPI-2 are scored using separate normative data because gender differences exist in the way items are typically answered (e.g. "I like mechanics magazines").
2. It is hypothesized that the frequency of one or more observed scores 1.5 standard deviations higher than the population mean as derived by Monte Carlo simulation will be significantly higher than expected for the theoretical normal distribution of scores for each scale group (Clinical, Content, Supplementary, Harris-Lingoes, Content-Component). More than 6.7% of the population is expected to obtain at

least one abnormal score due to the aggregation of Type I error across the multiple scales that comprise each scale group.

3. It is hypothesized that the frequency of one or more observed scores 1.5 standard deviations higher than the population mean as derived by Monte Carlo simulations will be significantly higher than expected for the theoretical normal distribution of scores for all scale groups combined (Clinical, Content, Supplementary, Harris-Lingoes, Content-Component). More than 6.7% of the population is expected to obtain at least one abnormal score due to the aggregation of Type I error across the 98 combined scales.

CHAPTER III

Methodology

Participants

This study used an archival data set provided by the University of Minnesota Press (2001) and Graham (2006). Data from the MMPI-2 restandardization sample were furnished that consisted of interscale correlation tables for the Clinical, Harris-Lingoes, Content, Content-Component, Supplementary, and Restructured Clinical (RC) scales and subscales. Interscale correlation data were available for all combinations of scales. The correlational data presented in the aforementioned materials were constructed from non-*k*-corrected linear *T*-scores (Butcher et al., 2001).

The restandardization sample was comprised of 2,600 (1,462 females; 1,138 males) participants. Included participants in the final sample were between the ages of 18 and 85 years ($M = 41.0$; $SD = 15.3$) and recruited from seven states including Minnesota, Virginia, California, Ohio, Pennsylvania, Washington, and North Carolina; participants were also recruited from military bases and a federal Indian reservation (Graham, 2006). Individuals were excluded from the restandardization sample for inadequate background information, excessive item omissions, or profile invalidity as evidenced by a high (i.e., raw score >20) F or F_B scale score. Demographically, 81.4% were Caucasian, 12.1% Black, 3.0% Native American, 2.8% Hispanic, and 0.7% were Asian; 26.9% had graduated college, 25.1% had some college, 24.6% were high school graduates, 18.5% received education beyond college, and 4.9% had less than a high school education (Butcher et al., 2001). Approximately 21% of females and 32% males held managerial or

professional occupations. Only 5% of females and 12% of males were considered to be laborers. More than 60% of males and females in the sample were married (Graham, 2006).

Assessment Measure

Minnesota Multiphasic Personality Inventory – Second edition

The Minnesota Multiphasic Personality Inventory – second edition (MMPI-2) is a self-report measure with 567 items. The MMPI-2 includes the Clinical, Validity, Content, Supplementary, Content-Component, and Harris-Lingoes subscales as well as numerous additional scales that have been introduced since the measure's inception in 1989. The final normative sample was composed of 2,600 (1,138 males and 1,462 Females) participants who were selected to match the 1990 U.S. Census data. The MMPI-2 was developed using a criterion-keyed approach in which scores were constructed by contrasting the performance of a reference group to a normal group. Uniform transformed *T*-scores allow for ease of test interpretation and for comparisons between subscales (Tellegen & Ben-Porath, 1992). The Validity and Clinical Scales are made up of eight and 10 scales, respectively. The Clinical subscales are deconstructed into 31 Harris-Lingoes and Si Subscales. There are 15 Content Scales and 20 Supplementary Scales. The Content Scales are also separated into the Content-Component subscales (Butcher et al., 2001). For the present study, data from the Clinical, Content, Supplementary, Content-Component, and Harris-Lingoes subscales were analyzed.

Procedure & Analyses

The general methodology outlined by Crawford and colleagues (2007) was used to analyze correlational data from the restandardization sample. This method involved (1) obtaining the Cholesky (alternatively spelled Choleski) decomposition of a *k* x *k* correlation matrix, (2) generating a random vector of independent, standard, normal variates (i.e., transformed random numbers from uniform distribution) that agrees with the number of tests or scales in a battery, (3) postmultiplying the lower triangular matrix of the Cholesky decomposition by the generated random vector, (4) and lastly steps 2 and 3 were looped or reiterated many times to simulate the scores of 1,000,000 cases. A generic program (i.e., PercentAbnormK.exe; Crawford et al., 2007) is available for calculating the percentage of a population with *N* or more significant scores using the previously outlined method. The program does not allow for inclusion of more than 20 scales and has limited options for selection of cutoff criteria.

Due to the different variants (e.g., number of included scales) of the above mentioned Monte Carlo method needed for the present study, a syntax grounded approach was determined to be most appropriate. Syntax based in the program PASW 18.0 (Predictive Analytics SoftWare, 2009) for Windows was adapted from Schretlen et al., (2008). See Appendix A for an example of the syntax used to resample distributions of scores. Following the methodology outlined by Crawford et al. (2007), for a *k* x *k* interscale correlation matrix (1) the syntax generates 1,000,000 random vectors of *k* independent standard normal variates with a mean of zero and a standard deviation of one; (2) computes the Cholesky decomposition of the *k* x *k* matrix; (3) and postmultiplies

each generated random vector by the Cholesky decomposition. The product of each "observation" is then computed into a new variable with a mean of zero and a standard deviation of one; (4) these values are multiplied by 10 and added to 50 - converting the values to *T*-scores; (5) the percentage of vectors or observations equal to or greater than a specified cutoff criteria for *N* – abnormal scores (e.g., significant as defined by a *T*-score ≥ 65; or, two abnormal *T*-scores ≥ 65) are then counted and displayed in separate frequency tables for each cutoff score with row values displayed from highest to lowest. The percentage of 1,000,000 observations with *N* or more significant scores from the 98 scales is then simply obtained by counting the cumulative percent of scores greater than the cut-score. This syntax can be easily manipulated depending on the specific question needing to be answered. For example, step (5) of the syntax was modified to count the number of observations for *N* – scores between a range of *T*-scores (e.g., 65 thru 74).

In the current study the Monte Carlo technique was used to determine the percentage of the population with significant scale elevations on the MMPI-2. In addition to the previously outlined hypotheses, a significant aim of the present study was to provide descriptive data for the impact of simultaneous interpretation of various combinations of scales on the percentage of the normal population with significant scores. Interscale correlational data from the restandardization sample (University of Minnesota Press, 2001) was input into the syntax for each of the separate scales and for combinations of scales. Combinations of scales input together included: Clinical and Harris-Lingoes scales; Content and Content-Component scales; Clinical, Content, and

Supplementary Scales; and Clinical, Harris-Lingoes, Content, Content-Component, and Supplementary scales.

Research Hypotheses

1. *There will be a difference between males and females in the mean number of observed scores 1.5 standard deviations higher than the population mean for all scale groups combined.* This hypothesis will be tested by a two tailed independent *t*-test between males and females on the mean number of scales above 65*T*. The alpha level for this analysis will be set at .05.
2. *It is hypothesized that the frequency of one or more observed scores 1.5 standard deviations higher than the population mean as derived by Monte Carlo simulation will be significantly higher than expected for the theoretical normal distribution of scores for each scale group (Clinical, Content, Supplementary, Harris-Lingoes, Content-Component).* This hypothesis will be tested by a series of *z*-tests for proportions (Glass & Hopkins, 1984) that compare observed frequencies of scores above and below 65*T* to those that occur in the normal distribution (6.7 percent and 93.3 percent respectively) for each scale group. The alpha level for each analysis will be set at .01 due to the number of statistical tests involved.
3. *It is hypothesized that the frequency of one or more observed scores 1.5 standard deviations higher than the population mean as derived by Monte Carlo simulation will be significantly higher than expected for the theoretical normal distribution of scores for all scale groups combined (Clinical, Content, Supplementary, Harris-Lingoes, Content-Component).* This hypothesis will be tested by a *z*-test

for proportions that compare observed frequencies of scores above and below 65*T* to those that occur in the normal distribution (6.7 percent and 93.3 percent respectively) for the scales in aggregate. The alpha level for the analysis will be set at .05.

CHAPTER IV

Results

"Operand Matrix is not Positive Definite"

Using the previously mentioned syntax, five of the 18 interscale correlation matrices (i.e., Content & Content-Component, all scales combined, and the Clinical & Harris-Lingoes subscales for females only) generated an error message stating "operand matrix is not positive definite for CHOL" (Predictive Analytics SoftWare, 2009), thereby preventing execution of the Cholesky factorization. Further inspection revealed these matrices each to be singular with no possible square root or inverse – two absolute requirements of a Cholesky factorization (Gentle, 2007). The largest of the interscale correlation matrices (98 scales) contained four negative eigenvalues for both males (-.003336, -.005437, -.005708, -.009135) and females (-.000627, -.001087, -.008614, -.010721; see Appendix D for uncorrected and corrected eigenvalues); combined Content and Content-Component matrices produced one negative eigenvalue for both males (-.00534089) and females (-.00352979); and a matrix comprised of the Clinical and Harris-Lingoes subscales produced one negative eigenvalue for females (-.00005249).

Spectral-decomposition is one viable methodology employed to for correct non-positive definiteness in real-world matrices, where for applied reasons correlations cannot be reinvented. Fundamentally, spectral-decomposition reconstructs acceptable matrices so as to be extremely similar to their original analogues, while at the same time retaining positive-semidefiniteness. See Rebonato & Jäckel (1999) for a more comprehensive procedural step-by-step methodology. Advantages to this technique included quick

implementation for large matrices, guaranteed production of a positive-semidefinite matrix, and resultant correlation matrices that were well defined and very closely approximated the original matrices (Rebonato & Jäckel, 1999). The above mentioned procedural description was written into PASW 18 so as to correct the matrices prior to attempting the Cholesky factorization (See Appendix A).

Negative eigenvalues were set to an extremely small number ($1\ X\ 10^{-8}$) in order to minimize any dissimilarity between the original interscale correlation matrix and the corrected matrix. Resultant corrected matrices were similar to their uncorrected counterparts, such that each corrected matrix was identical to its original form when rounded to the hundredths place – the format the data were provided (University of Minnesota Press, 2001). Although corrected and uncorrected versions of the matrices were identical at the hundredths place, minute and seemingly inconsequential differences thereafter were adequate to create positive-semidefiniteness and meet those assumptions necessary to carry out a Cholesky factorization. Corrected matrices allowed for further analysis and for Hypotheses I, II, and III to be evaluated.

Hypothesis I

Consistent with the predictions made in hypothesis I, an independent-samples *t*-test indicated females ($M = 4.858$, $SD = 7.362$) have more elevated scores than males ($M = 4.809$, $SD = 6.970$) when all 98 MMPI-2 scales were simultaneously considered at a *T*-score ≥ 65; $t(1,999,998) = 4.796$, $p = .0000016$, 95% Confidence Interval = 0.03, 0.07. A very small effect size ($d = 0.007$), however, suggested these differences were not clinically meaningful. Statistical significance was attributable to a large sample size ($N =$

2,000,000) and considerable degree of coinciding statistical power (0.999). See Appendices B - J for the percentages of the population separated by gender with elevated scores for different combinations of scales.

Hypotheses II & III

Consistent with hypotheses II and III, the frequency of observed scores 1.5 standard deviations above and below the mean as derived by Monte Carlo simulation was significantly higher than expected for the theoretical normal distribution when all scale groups were considered simultaneously (Clinical, Content, Supplementary, Harris-Lingoes, and Content-Component) and separately. All *z*-tests indicated substantial differences ranging from $z = 1203.61$, $p < .001$ to $z = 2942.10$, $p < .001$. Specific *z*-score values for each comparison above and below 1.5 standard deviations from the mean are presented in Table 1. Increasingly large differences in frequency between expected (normal distribution) and observed (Monte Carlo) percentages of elevated scores appeared to coincide with the quantities of scores that are simultaneously considered.

Table 1.

Hypotheses II & III: Comparison between the observed and expected percentages of the normal population to score above and below 1.5 standard deviations from the mean

MMPI-2 Scale	Observed Percentage	Theoretical Percentage[a]	*Z*-Statistic
Clinical Scales			
T-score ≥ 65	36.793	6.700	1203.614*
T-score ≤ 65	63.207	93.300	-1203.614*
Content Scales			
T-score ≥ 65	38.250	6.700	1261.889*
T-score ≤ 65	61.750	93.300	-1261.889*
Supplementary Scales			
T-score ≥ 65	55.108	6.700	1936.150*
T-score ≤ 65	44.892	93.300	-1936.150*
Harris-Lingoes Subscales			
T-score ≥ 65	71.321	6.700	2584.613*
T-score ≤ 65	28.679	93.300	-2584.613*
Content-Component Subscales			
T-score ≥ 65	60.863	6.700	2166.329*
T-score ≤ 65	39.137	93.300	-2166.329*
All Scales[b]			
T-score ≥ 65	80.259	6.700	2942.101*
T-score ≤ 65	19.741	93.300	-2942.101*

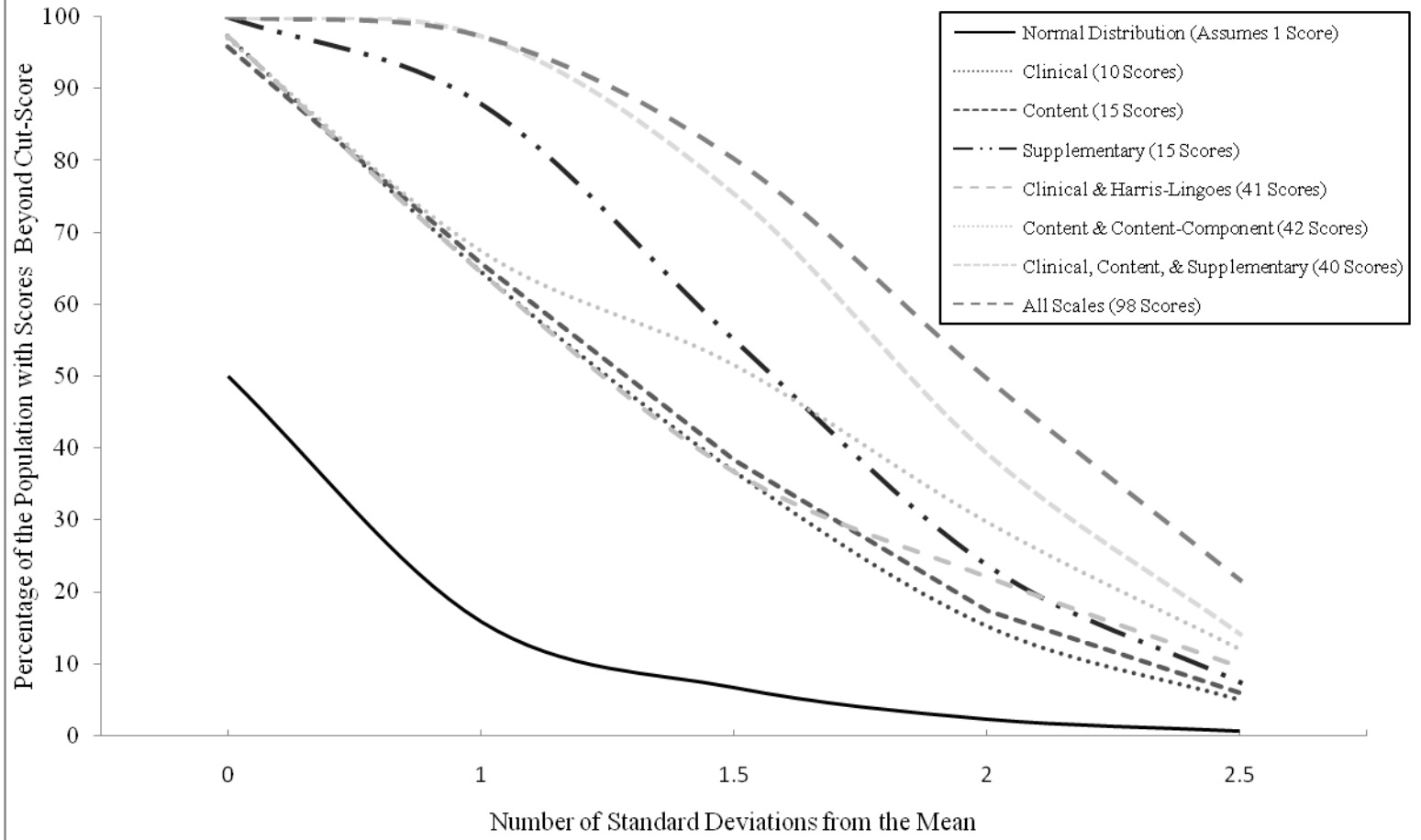
[a] "Theoretical Percent" represents the expected percentage of the population with one or more score above or below the specified cut-score for the normal distribution.

[b] "All Scales" includes Clinical, Content, Supplementary, Harris-Lingoes, and Content-Component Scales and subscales.

* $p < .001$.

See Figure 1 for a visual depiction of the percentage of the population (males and females combined) with one or more elevated score at 1.0, 1.5, 2.0, and 2.5 standard deviations from the mean for each scale group. The observed pattern suggested consideration of fewer scales at high cut-scores (i.e., 2.0 – 2.5 *SD*s) generally resulted in reductions to the percentage of the population with elevated scores. The theoretical normal distribution, which inherently assumes one score is being interpreted, predicted the fewest number of people with one or more elevated scores independent of the cut-score used.

Figure 1: *Percentage of the Population with One or More Elevations at 1.0 - 2.5 SDs from the Mean*



Elevation Qualities

Although specific patterns of elevated scores varied from males and females on certain scales (Butcher et al., 2001), total incidence of elevations observed here for families of scales were not clinically meaningful and thus averaged to facilitate interpretation (as previously mentioned, interested readers can reference Appendices B – J for results separated by gender).

Table 2.
Descriptive Statistics: Frequency of Scale Elevations for Different Combinations of Scales – Males and Females Combined

MMPI-2 Scale	Mean	Median	Mode	*SD*	Skew	Kurtosis	Range
Clinical Scales							
≥ 60*T*	1.587	1.000	0.000	1.805	1.343	1.424	0-10
≥ 65*T*	0.669	0.000	0.000	1.155	2.335	6.372	0-10
≥ 70*T*	0.228	0.000	0.000	0.635	3.892	19.727	0-9
≥ 75*T*	0.062	0.000	0.000	0.307	6.678	60.144	0-7
Content Scales							
≥ 60*T*	2.382	1.000	0.000	3.039	1.584	2.021	0-15
≥ 65*T*	1.004	0.000	0.000	1.911	2.803	9.122	0-15
≥ 70*T*	0.342	0.000	0.000	1.022	4.786	29.979	0-15
≥ 75*T*	0.094	0.000	0.000	0.475	8.405	100.256	0-15
Supplementary Scales							
≥ 60*T*	2.381	2.000	2.000	1.698	0.668	0.017	0-10
≥ 65*T*	1.002	1.000	0.000	1.245	1.578	2.754	0-8
≥ 70*T*	0.341	0.000	0.000	0.740	2.975	11.451	0-8
≥ 75*T*	0.093	0.000	0.000	0.375	5.524	41.476	0-8
Harris-Lingoes Subscales							
≥ 60*T*	4.921	4.000	3.000	3.868	1.308	1.688	0-25
≥ 65*T*	2.074	1.000	0.000	2.517	2.232	6.663	0-23
≥ 70*T*	0.706	0.000	0.000	1.371	3.636	19.739	0-21
≥ 75*T*	0.193	0.000	0.000	0.642	6.011	58.321	0-19
Content-Component Subscales							
≥ 60*T*	4.287	3.000	1.000	4.105	1.343	1.704	0-27
≥ 65*T*	1.806	1.000	0.000	2.503	2.284	6.781	0-24
≥ 70*T*	0.615	0.000	0.000	1.287	3.645	19.485	0-23
≥ 75*T*	0.168	0.000	0.000	0.578	5.772	52.622	0-18

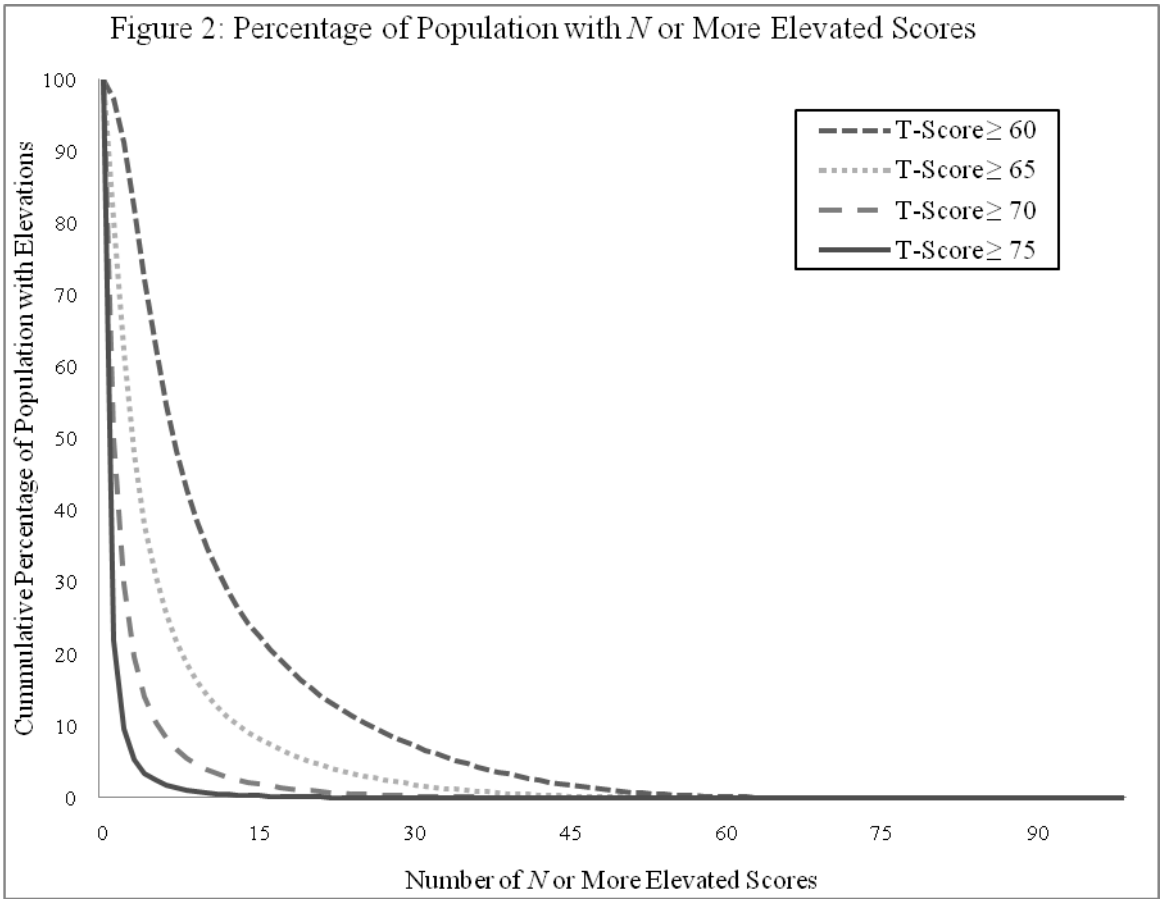
Descriptive Characteristics

Descriptive characteristics for the frequency of scale elevations are presented in Tables 2 and 3 for separate scale families and combinations of scale groups, respectively. For both separated (Table 2) and combined (Table 3) scale groups, the mean number of elevated scores increased as more scores are under consideration and cut-scores were relaxed. The mean number of elevated scores was greater than the median, which was consistent with the large skew and kurtosis values, and indicated that the frequency of elevated scores in the population was not normally distributed.

Table 3.
Descriptive Statistics: Frequency of Scale Elevations for Different Combinations of Scales – Male and Females Combined

MMPI-2 Scale	Mean	Median	Mode	*SD*	Skew	Kurtosis	Range
Clinical, Content, & Content-Component							
≥ 60*T*	6.346	4.000	3.000	5.524	1.574	2.097	0-33
≥ 65*T*	2.673	2.000	1.000	3.621	2.671	8.516	0-31
≥ 70*T*	0.909	0.000	0.000	1.974	4.496	27.299	0-29
≥ 75*T*	0.248	0.000	0.000	0.918	7.944	92.743	0-27
Clinical & Harris-Lingoes							
≥ 60*T*	2.774	1.000	0.000	3.889	2.174	5.484	0-33
≥ 65*T*	1.381	0.000	0.000	2.680	2.979	11.240	0-30
≥ 70*T*	0.556	0.000	0.000	4.505	4.505	27.916	0-26
≥ 75*T*	0.175	0.000	0.000	7.229	7.228	77.433	0-20
Content & Content-Component							
≥ 60*T*	5.197	3.000	0.000	6.713	1.750	2.940	0-42
≥ 65*T*	2.470	1.000	0.000	4.208	2.712	8.963	0-40
≥ 70*T*	0.898	0.000	0.000	2.187	4.359	26.181	0-36
≥ 75*T*	0.253	0.000	0.000	0.976	7.262	79.036	0-30
All Scales[a]							
≥ 60*T*	10.227	6.000	3.000	10.866	1.952	3.927	0-78
≥ 65*T*	4.834	2.000	0.000	7.169	2.930	10.532	0-76
≥ 70*T*	1.797	0.000	0.000	3.864	4.663	29.841	0-69
≥ 75*T*	0.522	0.000	0.000	1.752	7.946	94.914	0-50

[a] "All Scales" includes Clinical, Content, Supplementary, Harris-Lingoes, and Content-Component Scales and subscales.



For the present outcome, the frequency of elevated scales was determined by categorizing scores as either "elevated" (+1) or not elevated (+0) which prevented the occurrence of negative values, created a restriction of the range (floor effect), and resulted in a nonsymmetrical distribution of elevated scores. As *T*-score values became more conservative for each family of scales positive skew and kurtosis exacerbated, *SD*s constricted, and the distributions of scores drifted further from symmetry. Figure 2 illustrates these characteristics by presenting the percentage of the population with elevated scores at different cut-scores when all 98 MMPI-2 scales (Clinical, Harris-Lingoes, Content, Content-Component, Supplementary scales) were concurrently considered.

Percentage of the Normal Population with Elevated Scores

A primary aim of the current study was to present data for the occurrence of elevated scores in the normal population. Consistent with this intention data separated by gender are presented in Tables 4, 5, and 6 for the percentage of the normal population with *N* or more elevated scores (≥60*T*, ≥65*T*, ≥70*T*, and ≥75*T*) when different combinations of MMPI-2 scales were considered in tandem. The Harris-Lingoes and Content-Component subscales were not presented separately as interpretive practice dictates they be interpreted jointly with their respective parent scales (Graham, 2006).

Table 4 highlighted the percentage of the normal population with seemingly clinically significant scores for the Clinical, Content, and Supplementary scales. For example, when significance was defined at a traditional *T*-score of 65 as many as 36% of normal people had at least one elevated score. Referencing the theoretical normal distribution one expects this percentage to be nearer to 6.7%, which more closely reflected the observed percentage of the population (7.72) with at least three significant scores. This example highlighted the inverse relationship between cut-score and frequency of elevated scores in the normal population. This trend was consistently evident when comparing *T*-score cut-points within each family of scales. When analyzing rates of significance across the Supplementary scales, for instance, the percentage of the population with two or more elevated scores dropped by 61% when moving from a more relaxed *T*-score = 60 to a traditional cutoff of 65*T* (25.30%), and fell off even more drastically when moving to a more conservative 70*T* (6.62%) and 75*T* (1.31%).

Table 4.
Percentage of population with N or more abnormal scores at T-score values of 60, 65, 70, & 75: MMPI-2 Standardization Sample

		≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7
Clinical Scales	60*T*	64.552	38.832	23.996	14.841	8.751	4.669	2.138
	65T	***36.793***	***15.648***	***7.723***	***3.882***	***1.826***	***0.734***	***0.240***
	70*T*	15.647	4.446	1.719	0.679	0.246	0.075	0.016
	75*T*	4.976	0.899	0.272	0.084	0.023	0.004	0.001
Supplementary Scales	60*T*	87.955	65.034	41.533	23.792	12.210	5.284	1.766
	65T	***55.108***	***25.296***	***11.443***	***5.256***	***2.193***	***0.716***	***0.178***
	70*T*	23.709	6.620	2.408	0.962	0.321	0.075	0.015
	75*T*	7.393	1.313	0.418	0.143	0.035	0.006	0.001
Content Scales	60*T*	63.993	44.537	32.988	25.180	19.434	15.026	11.491
	65T	***38.250***	***20.900***	***13.291***	***9.023***	***6.281***	***4.389***	***3.039***
	70*T*	17.370	7.119	3.836	2.294	1.422	0.884	0.545
	75*T*	5.922	1.739	0.786	0.407	0.227	0.125	0.068

Results for the percentage of the normal population with elevated scores for combinations of scale groups (Clinical & Harris-Lingoes; Content & Content-Component; Clinical, Content, & Supplementary; Table 5) revealed the same relationship mentioned above. Similar to the data in Table 4, results offered in Table 5 underscored the impact analyzing more scales has on rates of significance in the normal population.

Consideration of the Clinical and Harris-Lingoes scales together yielded a nearly identical percentage of people with one or more elevated score (65*T*) as when interpreting the Clinical scales alone. This observation was a direct result of recommended guidelines for interpretation of the Harris-Lingoes subscales, which specifies corresponding parent scales have a minimum value of 65*T* (Graham, 2006). The addition of the Harris-Lingoes subscales therefore cannot increase the rate of the population with one or more elevated score at 65*T*, because its interpretation hinges upon its parent scale already being significant. Minute differences (at the hundredths place) at this cut-point represented statistical variation attributable to the process of generating 1,000,000 random normal variates, and not clinically meaningful dissimilarity.

Consideration of an increasing number of MMPI-2 scales and subscales enlarged the percentage of the population with elevated scores at the tail end of the distribution presented in Figure 1, suggesting that as more scores are interpreted there was a general observable increase in percentage of normal individuals with multiple elevations.

Table 5.
Percentage of population with N or more abnormal scores at T-score values of 60, 65, 70, & 75: MMPI-2 Standardization Sample

		≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7
Clinical & Harris-Lingoes [a]	60*T*	64.486	44.388	34.515	27.035	20.943	16.577	13.427
	65T	***36.754***	***27.489***	***19.171***	***13.056***	***9.680***	***7.477***	***5.814***
	70*T*	22.126	12.008	6.931	4.370	2.990	2.099	1.512
	75*T*	9.595	3.568	1.708	0.952	0.585	0.379	0.251
Content & Content-Component [a]	60*T*	67.446	61.482	50.631	42.100	37.475	32.521	28.556
	65T	***51.575***	***39.066***	***29.181***	***22.680***	***18.182***	***14.778***	***12.150***
	70*T*	29.727	17.791	11.197	7.637	5.498	4.065	3.089
	75*T*	12.031	5.409	2.743	1.629	1.039	0.696	0.481
Clinical, Content, & Supplementary	60*T*	97.304	89.502	77.098	62.941	49.955	39.716	32.317
	65T	***75.370***	***50.459***	***33.094***	***22.512***	***16.300***	***12.572***	***10.120***
	70*T*	39.242	17.595	9.464	5.960	4.211	3.175	2.480
	75*T*	14.114	4.339	2.063	1.234	0.844	0.611	0.456

[a] Harris-Lingoes & Content Component subscale elevations were only considered when the parent scale (i.e., Clinical & Content) from which the subscale was derived was ≥65*T* or ≥60*T*, respectively.

Data for the proportion of the normal population with *N* or more significant scores when all 98 scales were interpreted together are displayed in Table 6. Monte Carlo simulation indicated that more than twenty percent of normal people have as many as seven or more elevated scores at a traditional cut-off point of 65*T*. Using a more lenient definition of significance (60*T*) this rate more than doubled to 47.3%, which is similar

Table 6.
*Percentage of population with N-abnormal scores for 98 MMPI-2 Scales**

T-Score	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7
60	97.301	91.161	82.032	71.740	61.988	53.837	47.340
65	***80.259***	***61.854***	***47.362***	***37.172***	***30.127***	***25.114***	***21.376***
70	49.615	29.208	19.113	13.723	10.468	8.285	6.750
75	21.567	9.412	5.301	3.456	2.449	1.826	1.411

* Includes Clinical, Content, Supplementary, Harris-Lingoes, and Content-Component scales

to percentage of those expected to have one or more elevated scores at 70*T*. Almost all normal people could be expected to have at least one score significant at one standard deviation from the mean (97.3%). In contrast, the most conservative definition of significance (75*T*) resulted in the lowest rates of deviant scores.

CHAPTER V

Discussion

Clinical neuropsychologists use their expertise to make fine distinctions between normal and aberrant symptom presentations, discern diagnostic clarity from the "gray" area between psychological and neurological manifestations, and formulate pragmatic recommendations. The product of these efforts is in part anchored to the psychometric qualities of the tests and measures used in the neuropsychological battery. An instrument's value within the battery is measured by its ability to accurately identify intraindividual scatter across relevant domains, add incremental validity to other tests, and by the rate it correctly classifies scores as negative or positive for a specified characteristic. The latter quality oftentimes involves contrasting examinee performance with an appropriate comparison standard (Lezak et al., 2004). To facilitate this process scores can be placed in percentiles and transformed into standard scores (e.g., M=100, SD=10) to ascribe numerical weight to a performance relative to a comparison standard. With consideration for various factors, abnormality is then defined at a certain cut-point, such as one standard deviation above the mean.

If the theoretical normal distribution is referenced for comparison, for example, a cut-point one standard deviation from the mean would accurately categorize 84% of the normal population and misclassify the other 16%. This type of comparison, however, implicitly assumes one score is under consideration and does not account for scale interrelatedness (i.e., shared variance), and the consequences of interpreting more than one score, both of which impact the occurrence of significant scores (Crawford et al.,

2007; Ingraham & Aiken, 1996). The addition of multiple scales to the interpretive process is positively associated with the rate of significant scores and has been demonstrated to be related to considerable percentages of the normal population obtaining significant scores (Axelrod & Wall, 2007; Binder, Iverson, & Brooks, 2009; Brooks et al., 2007, 2008, 2009, 2010; Brooks, Iverson, Sherman et al., 2009; Crawford et al., 2007; Heaton et al., 2001; Heaton et al., 2004; Ingraham & Aiken, 1996; Palmer, Boone, Lesser, & Wohl, 1998, Schretlen et al., 2008). Additional pertinent factors (e.g., parental education, score cutoff, age, education, intelligence) have also been shown to correlated to the frequency of elevated scores in the normal population (Brooks et al., 2007, 2008, 2009, 2010; Brooks, Iverson, Sherman et al., 2009; Crawford et al., 2007; Decker et al., 2012; Schretlen et al., 2003, 2008). These studies emphasize the statistical commonplace of abnormal scores in nonclinical populations at commonly used defined cut-off points of significance, and punctuate the potential for misattributing one or more normally occurring elevations as basis for a clinical diagnosis.

Provision of base rates for the occurrence of deviant scores in nonclinical populations removes a degree of theoretical guesswork required in determining the *clinical* significance of a set of scores, and potentially increases diagnostic certainty and decision making. To date, however, empirical efforts to provide these data largely have focused on tests of intelligence and cognition, and have not considered the impact of multiple score interpretation on tests of personality and emotional functioning. Considering a majority of neuropsychologists work in some capacity with patients suspect of various psychological diagnoses, the absence of these data (up to this point)

represents a significant gap in the literature. In a large survey of 747 clinical neuropsychologists, over 60% of respondents indicated they frequently utilize objective personality tests, of which the MMPI-2 was most common, ranking ninth in overall use as a neuropsychological assessment instrument (Rabin, Barr, & Burton, 2005).

In a continuation of other practical efforts to make available base rates for the significant scores in normal people, this study was designed to investigate the commonality of elevated scores on the MMPI-2 in normal adults. This intention was consistent with Binder and colleagues' (2009) call for the examination of domains and tests with a proclivity for yielding a high likelihood of significant scores (i.e., high number of considered scales). The MMPI-2 lends itself to this type of examination due to the large number of included scales (i.e., 98), its widespread use amongst neuropsychologists and clinical psychologists, and the potentially far reaching diagnostic implications associated with its interpretation (Bow, Flens, & Gould, 2010; and Rabin, Barr, & Burton, 2005). The current investigation focused on the MMPI-2, and set out to explore gender differences in rates of significant scores, and whether the observed percentage elevated scores in the normal population differs from expected rates significance when multiple scales are simultaneously considered. An additional principal goal was to present data for the rates of significant scores for combinations of MMPI-2 scales at different cutoff scores in the normal population, which is consistent with Graham's (2006) observance that there is a general lack of statistical data available to improve classification accuracy.

Hypothesis I

The first hypothesis stated there would be a difference between males and females in the mean number of observed scores falling 1.5 standard deviations higher than the population mean when the Clinical, Harris-Lingoes, Content, Content-Component, and Supplementary scales are simultaneously considered. Despite separate normative data, male and females did not significantly differ in respect to sheer number of significant scores (Butcher et al., 2001). Although there are apparent differences between males and females in the normative sample (e.g., 3% males and 6% of females were in treatment for "mental health problems"; Graham, 2006) such as gender specific patterns of item endorsement and test-retest stability, they did not appear to outwardly translate into discrepant rates of abnormal scores in the normal population (Butcher et al., 2001).

The most feasible explanation for this finding resides in the MMPI-2 normative process and derivation of uniform *T*-scores. The unpublished MMPI®-2 scale intercorrelations provided by the University of Minnesota Press (University of Minnesota Press, 2001), correlations presented in Graham (2006), and tabled data in the *MMPI-2: Manual for Administration, Scoring and Interpretation - Revised Edition* (Butcher et al., 2001) all convey normative data for shared variance and covariance between MMPI-2 scale *T*-score values rather than raw scale scores. Individual variation in the normative sample was therefore weighted to a mean (*T*-score = 50) and standard deviation (*T*-score = 10) for males and females adjusting for any difference in raw scores. Gender differences are not outwardly perceptible when referencing *T*-score values; however,

evaluation of raw score equivalencies of same *T*-score values for males and females reveal incongruities.

For example, in order to reach a threshold of 65*T* on Scale 2 ("Depression") males need to positively endorse a minimum of 21-items whereas females require 26-items (Butcher et al., 2001). Variation in item endorsement and *T*-score equivalence between males and females is perhaps best exemplified by Scale 5 ("Masculinity-Femininity") which includes items reflecting stereotypical masculine and feminine characteristics (Hathaway, 1956). For example, male and female examinees may have the same score (51*T*), but a female would have to positively endorse eleven more items. A more extreme example on this scale is a raw score of ten which equates to 32*T* for males and 100*T* for females. Identical *T*-score values for males and females may, however, lead to different interpretative conclusions (Graham, 2006). More salient to the current investigation was the observation that inconsequential differences (Cohen's $d = 0.007$) between the proportion of males and females with elevated scores on the MMPI-2 allowed averaging of base rates without loss of information.

Hypotheses II & III

Hypotheses II and III stated that the frequency of one or more observed scores 1.5 standard deviations higher than the population mean, as derived by Monte Carlo simulation, would be significantly higher than expected for the theoretical normal distribution when all MMPI-2 scale groups are considered either separately or together (Clinical, Content, Supplementary, Harris-Lingoes, Content-Component). Both Hypotheses were substantiated by the current results. Findings indicated that differences

between the observed and theoretical percentages of the population with elevated scores (*T*-score ≥ 65) were not only significant but extensive (*z*-values ranging from 1203.61 to 2942.10). The magnitude of these findings is particularly strong with regard to the sizeable impact a large number of scores has on rates of apparent clinical significance in normal adults.

These results add to an expanding body of research (Axelrod & Wall, 2007; Binder, Iverson, & Brooks, 2009; Brooks et al., 2008, 2009; Brooks & Iverson, 2010; Brooks, Sherman, & Iverson, 2010; Crawford et al., 2007; Decker et al., 2012; Heaton et al., 2001; Heaton et al., 2004; Ingraham & Aiken, 1996; Moran, 2003; Palmer, Boone, Lesser, & Wohl, 1998; Schretlen et al., 2008) indicating the theoretical normal distribution consistently underestimates rates of practical significance in the instance of multiple score interpretation. The compelling rates of elevated scores accompanying the use of additional scales are akin to the accumulation of type I error. This analogy is helpful in understanding the ensuing family-wise error rate associated with 98 data points, and illuminates the divergence between expected rates of significance and the ballooning of rates observed (Moran, 2003).

Methodological Considerations

Restandardization Sample & Validity Scales

In the context of this study the Monte Carlo technique can best be described as an inferential technique because inferences are being made to population parameters from the characteristics of the MMPI-2 Restandardization Sample. In part, this methodology relies on the MMPI-2 Restandardization Sample being consistent with the general

population which it is supposed to represent. To the extent this assumption is met these data can be said to represent the percentage of the "normal" population with *N* or more significant scores. Efforts were made during the restandardization process to represent characteristics of the United States' Census data, however, "Hispanic and Asian-American" groups were underrepresented, and "Native Americans" and those with higher levels of education were overrepresented (Butcher et al., 2001, p. 3). However, these discrepancies may not in and of themselves translate into meaningful problems with the MMPI-2 standardization scores (Schinka & LaLone, 1997).

When trying to determine how "normal" the standardization sample was, there are a several important factors to consider, including the small percentage of the sample that was receiving psychological treatment (Butcher et al., 2001; Graham, 2006). Each of these factors has the potential to shape the backdrop in which the present results are interpreted. Unless the normative sample was uniquely free from psychopathology, in which case it would not be truly representative (Helmes & Reddon, 1993), and assuming examinee scores are interpreted as *T*-scores, these base rates either approximate or under predict rates of significance in nonclinical populations.

There was a small percentage (3% of males and 6% of females; Butcher et al., 2001; Graham, 2006) of participants receiving psychological services in the standardization sample. However, Monte Carlo simulations generated normal distributions of scores with consideration for shared variance between scales, and not from counting the number of significant scores present in the restandardization sample. The resultant high base rates of elevated scores were related to the accumulative effect of

multiple score interpretation (analogous to alpha inflation). Hypothetically, were this not the case, a relatively small number of participants with psychopathology would be highly unlikely to provide explanatory power for the very large percentages of people observed to have significant scores (Table 6).

Moreover, given the large restandardization sample ($N = 2600$; Butcher et al., 2001) it is unlikely that the approximate 5% ($N \approx 130$) of participants receiving psychological services significantly disrupted interscale correlation matrices to a degree that would overcome the impact the sheer number of MMPI-2 scores has on rates of significance. Even if the sample was significantly overrepresented by individuals with psychopathology then raw score totals would be greater for the mean (50*T*) than in a truly representative sample and average results would have been "built into" the standardized scores. Consideration for the slight positive skew of the uniform *T*-score distribution targeted in the MMPI-2 restandardization could to a degree impact the current results, but only in that there would be a larger percentage of the population with more extreme scores at the distribution's tail end, translating into even larger percentages of the "normal population" with significant scores (Butcher et al., 2001; Tellegen & Ben-Porath, 1992).

Exclusion of Validity Scales

The MMPI-2 Validity Scales were purposefully excluded from the Monte Carlo simulations because of variable cut-scores and complexities associated with configural interpretation (Butcher et al., 2001; Graham, 2006; Green, 2000). For example, True Response Inconsistency Scale (TRIN) item pairs are contrasted and evaluated for the consistency and then added to a constant which sets all TRIN *T*-scores ≥ 50. The absence

of *T*-scores < 50 suggests significant restriction of range in the distribution of scores. The "Cannot Say (?)" validity scale, on the other hand, counts the number of items left unanswered on the test form and uses *raw* scores for interpretation, not *T*-scores. The interpretation of other validity scales such as the *F* (Infrequency) Scale are predicated on configural interpretation involving simultaneous interpretation of multiple validity scales (e.g., VRIN, TRIN, *L*, and *K*), from which numerous specific patterns of validity scores are compared to yield a number of inferences (Butcher et al., 2001; Graham, 2006; Green, 2000).

Moreover, during the restandardization process an unspecified percentage of the 300 participants excluded from the final sample were precluded on the basis of "excessive" item omissions and "excessively" high *F* and F_B scores (Butcher et al., 2001, p. 3). Resampling the normative sample's performance on these scales would have yielded rates of the population after excluding persons with false positive validity elevations rather than actual false positive rates in the normal population.

Not Positive Definite Matrix

A potentially unanticipated consequence of considering a large number of MMPI-2 scales (i.e., 98 by 98, 41 by 41, and 42 by 42) was production of matrices lacking positive definiteness. A "not positive definite" matrix is one with a determinant (i.e., scalar function of a matrix) approximately zero ("singular"). Singular matrices cannot be factored (e.g., subjected to Cholesky decomposition), because inversion requires dividing by a matrix's determinant, which if equal to zero is undefined (Gentle, 2007). Non-singularity or positive definiteness is an absolute requirement of a Cholesky

decomposition, which as a critical component to the Monte Carlo simulation represented an initial stumbling block to resampling base rates of significant scores in the normal population.

There are numerous potential explanations for the occurrence of singular covariance matrices that are fundamental to understanding the nature of the data set and determining whether a matrix should and can be corrected. Several relevant hypotheses are discussed here. Oftentimes the most parsimonious explanation for a not positive definite matrix is human or typographical error in the data set that alters the correlation matrix making it nonsymmetrical and non-invertible. In this study correlation data were reviewed and there were no transcription errors (e.g., inappropriate decimal places, values, and signs).

Another potential explanation for resultant singularity was linear dependency among the variables in the correlation matrices. Linear dependency simply refers to a significant degree of shared variance between two or more variables to the extent they approach an almost perfect relationship. Psychological and neuropsychological tests and measures may be particularly prone to linear dependency by nature of interrelated underlying theoretical constructs (e.g., Lezak et al., 2004; Sattler, 2008; Strauss, Sherman, & Spreen, 2006). Several factors specific to the MMPI-2 suggested linear dependency as the root cause of singularity. First, the empirical keying approach used to construct the MMPI-2 and considerable item overlap between some scales contributes to the presence of high interscale correlations, which is in part consistent with the authors' original intent (Butcher et al., 2001; Hathaway & McKinley, 1940).

This issue is exacerbated by the proliferation of MMPI-2 scales and subscales, all of which draw from the same pool of 567 items. In evaluating the 98 scales used in the current study, for example, there were approximately 10 interscale correlation values between .90 - .95, and almost 30 correlations between .80 - .89 (University of Minnesota Press, 2001). This issue may be further compounded by 10 Clinical and 15 Content scales that are deconstructed into an additional 58 subscales (i.e., linear combinations of respective parent scales). For example, the Cynicism (CYN) scale and Misanthropic Beliefs (CYN1) component subscale correlate .95 in males and females from the normative sample (University of Minnesota Press, 2001).

Interscale correlation data provided by the University of Minnesota Press (2001) may have been further ill-conditioned as the product of rounding off values at the hundredths place, thereby attenuating already minimal differences between scale correlations. Evidence for this explanation is provided *ex post facto* by Spectral Decomposition. The "intuitively similar" matrices produced via Spectral Decomposition (98x98, 41x41, and 42x42) rounded off to the hundredths place are indistinguishable at every value point from the data provided by the University of Minnesota Press (2001; Rebonato & Jäckel, 1999, p. 13). In general terms non-positive definiteness can be thought of as being "cured" in this instance by correcting for the loss of information inherent to rounding and extending the number of values further beyond the decimal place.

Research Implications

Current findings replicated previous Monte Carlo simulation studies by demonstrating that the rates of apparently abnormal scores in the normal population increased as the number of scores under consideration increased and as definitions of significance become more relaxed (Binder, Iverson, & Brooks, 2009; Brooks et al., 2007, 2008, 2009, 2010; Brooks, Iverson, Sherman et al., 2009; Crawford et al., 2007; Decker et al., 2012; Schretlen et al., 2008). Resampling techniques appeared to be particularly impressive when used with the MMPI-2 Restandardization Sample in emphasizing pitfalls associated with referencing the theoretical normal distribution to extract clinical meaning from a set of scores. Specifically, this study showed that estimates for the proportion of the population with significant scores were *extremely* underestimated by the theoretical normal distribution (see Table 1.). For example, the normal distribution only predicted 8% of the actual 80.3% of the population with significant scores when 98 MMPI-2 scales were considered together at a cut-score of 65*T*, and thus misclassified over 70% of the population who actually had scale elevations.

The application of this technique to the MMPI-2 represented a unique contribution to the literature in a number of ways. Prior to this study, resampling techniques had not been applied to measures of personality and emotional functioning, or been attempted on psychological measures with a comparable number of data points. Spectral decomposition allowed for base rates to be resampled from very large correlation matrices (e.g., 98x98; Rebonato & Jäckel, 1999; University of Minnesota Press, 2001), and offered a unique vantage point to observe the consequence of

interpreting a large number scales with a significant degree of shared variance (Graham, 2006). Results from the Monte Carlo simulations suggested the strength of the interscale correlations was not compelling enough, at least at or below 98 scales, to noticeably lessen the effect that interpreting an increasing number of scores has on rates of elevated scores in the normal population. Although the strength of interscale correlations likely effected specific patterns of elevated scores, there was not a significantly obvious effect over and above the sheer number of scores interpreted (Crawford et al., 2007).

Rates of Significant Scores

The generic Monte Carlo technique used in this study accurately resampled distributions of scores for the normal population, which were used to count percentages of the population with *N* or more seemingly abnormal scores across multiple scales (Binder et al., 2009; Decker et al, 2012; Schretlen et al., 2008). Overall, statistically abnormal scores were common on the MMPI-2. This is consistent with idea there is a normal level of intraindividual cognitive and emotional variability in normal adults (Lezak et al., 2004; Sattler, 2008). These results confirm the majority of normal adults have at least one high score on MMPI-2 when multiple scales are considered together. When all MMPI-2 scores are considered simultaneously the number of significant scores ≥65 ranged from zero to 76 in normal people. Strikingly, approximately half of adults have at least three apparently abnormal scores, and 20% have seven or more. Moreover, the *MMPI-2 Manual* (Butcher et al., 2001) indicates that some psychological or emotional difficulties may be associated with *T*-scores from 55-64; however, these data demonstrate that at a *T*-score cutoff of 60, almost the entire normal population has at least

one seemingly pathological score, and as many as one half have at least seven (Table 6). Moderate symptoms will on average be picked up across three scales (Table 3).

Although these rates seem quite high, they actually underestimate the true frequency of interpretable scores. For example, interpretations regarding moderate symptoms can be drawn from *T*-scores ranging from 55-59, and for some scales interpretations can be made from low scores (Butcher et al., 2001; Graham, 2006). Depending on the interpretive source negative descriptive characteristics may be associated with low scores across some scales. Because the current simulation involved generating normally distributed scores, the incorporation of low scores into interpretation would effectively double the number significant scores for each observation by including scores on either tail of the distribution.

When evaluating configurations of profiles, for example, Gotts and Knudson (2005) associated low scores on Scale 0 (M =40T) with social phobia and low *Ho* scores with anger dyscontrol. In the context of a counseling practice, Watkins, Campbell, and Lynn (2000) described low Scale 1 scores (<40T) as potentially reflecting difficulties with closeness and denial of physical symptoms; Scale 3 scores (<41T) with trust issues in therapy, difficulties forming a therapeutic alliance, and unfriendliness; low scores (<40T) on Scale 5 as potentially being associated with inflexibility, recklessness, and being unsympathetic in males; low scores (<40T) on Scale 6 as being associated with being guarded, touchy, evasive, and stubborn; Scale 7 (<40T) with laziness; Scale 9 (<40T) with listlessness, lethargy, and symptoms of depression; and Scale 0 (<40T) with someone who is self-indulgent, impulsive, exhibitionist, and manipulative.

Although not all significant scores (low or high) are directly indicative of psychopathology or clinical diagnosis, unfavorable personality characteristics may be inferred from some scales (Graham, 2006; Green, 2011). The Supplementary scales (e.g., Ego Strength, Dominance, Social Responsibility, and Marital Distress) may fall within this category. In practice these findings may translate into a patient being mistakenly ascribed pathological characteristics, or even worse, a clinical diagnosis. These findings underscore the importance of exercising caution when interpreting a large number of MMPI-2 scales.

Potential costs of misdiagnosis may be significant and negatively impact patients' quality of life, by subjecting them to inappropriate forms of treatment, denial of competency, or psychological consequences associated with being misdiagnosed. The MMPI-2 is also used, for example, used to screen airline pilots, applicants to seminary school, and law enforcement personnel (Graham, 2006). A certain frequency (not interpretation specific scale elevations) of elevated MMPI-2 Clinical scales may also be used to make predictions regarding early treatment termination, functioning as a sort of "red flag" in the context of therapy (Minnix et al., 2005).

Two- & Three-Point Code Types

In the instance of analyzing two and three point code types on the Clinical scales rates of significance in the general population may be reduced by making interpretations when the scales making up a code type are at least five *T*-score points above the next lowest clinical scale (Graham, 2006; Graham, Timbrook, Ben-Porath, & Butcher, 1991). Munley and colleagues (1991, 2004) also highlighted the possible impact of measurement

error when interpreting two and three point code types by highlighting that well differentiated code types may be less susceptible to measurement error. Strongly differentiated code types at least five *T*-score points above the next highest scale that have strong external correlates may be less likely to be elevated at a significant rate in the normal population (Graham, 2006).

However, Graham (2006) sets out specific recommendations for interpreting code types with *T*-score values greater than 60. Graham (2006) stated "considerable caution" should be applied when making inferences about symptoms from code types that are not as high, but also suggests that inferences regarding personality characteristics do apply (p. 94). He posits an example of a 27 / 72 code type ($T \leq 65$), in which inferences would not be made regarding symptoms of anxiety, but with reference to characteristics of insecurity, passive-dependent relationships, and perfectionism. This example may be viewed as congruent with a normally occurring degree of variation in the general population; and is consistent with the observation that Clinical scale code types with lower elevations have greater temporal stability, and do not regress to the mean to the same extent as more significantly elevated code types (Munley, 1991). Meaning code types closer to the mean by virtue of their stability suggest more long standing characteristics. Such interpretations would be consistent with results from the current study. Although Monte Carlo simulations did not directly evaluate two and three point code types, approximately 40% and 25% (Table 4) of the normal population were demonstrated to have at least two or three or more elevations on the Clinical scales at 60*T*, respectively. However, in the case where descriptions of these characteristics extend

clinical weight, have authority of diagnosis, and affect examinee quality of life neuropsychologists should be prudent in narration. For example, at a *T*-score one *SD* beyond the mean 65% of normal people will have at least one significant score on the Clinical scales, and when all scores are considered together almost the entire population can be expected to have an elevation.

Finding a Balance

Current data indicated rates of significant scores in the normal population may be reduced when fewer scales are interpreted at higher cut scores. For example, consideration of the Clinical scales at 75*T* (Table 4) approximated the level of confidence assumed using the theoretical normal distribution. As additional scales were added to the Clinical scales (e.g., Content & Supplementary), however, the observed number of normal people obtaining at least one or more significantly elevated scores was comparable to rates expected one *SD* from the mean (Table 5). This phenomenon was similar to alpha inflation observed with when performing more than one *t*-test. For example, if 100 comparisons are made one expects five to be significant by chance at $p = .05$. This analogy is akin to considering 98 MMPI-2 scale scores. These results *strongly* encourage a prudent interpretive approach, particularly when descriptive accounts or diagnoses are being drawn to influence the trajectory of an examinee's career, shape treatment, impact decisions in criminal or civil court, or other instances where interpretive false positives could negatively affect an examinee's quality of life. As more scales were considered and interpretations became more liberal (i.e., *T*-scores between 55 - 64 or <55, Butcher et al., 2001) these findings gain even more ecological traction.

There are several general practices for reducing interpretive false positives in multiple score interpretation that could be used including (1) not over-interpreting isolated high or low scores, (2) knowing the impact of demographic variables on the frequency of elevated scores, (3) understanding base rates of a suspect condition, (4) minimizing the number of scores interpreted (when possible), (5) integrating patient history, behavioral observations, and data from other tests to corroborate observed scale elevations, (6) and use of other relevant diagnostic information such as clinically recognizable patterns of performance (Binder et al., 2009; Lezak et al., 2004; Schretlen et al., 2003).

In the context of minimizing MMPI-2 interpretive false positives interpretation should also consider other configural aspects of a profile. For example, some suggest the Content -Component subscales be interpreted when the parent scale is ≥60*T* and a score on one of the components scales is *at least* 10 *T*-score points higher the other component subscales (Graham, 2006). Graham et al. (1991) also recommended only interpreting two- and three-point codes when the scales were *at least* five *T*-score points higher than the next highest score. Moreover, there is a plethora of available data for using the MMPI-2 in different contexts and in different patient populations (e.g., Graham, 2006; Green, 2011; Gass, 2006) that can be accessed to increase the confidence of one's interpretive conclusions. Moreover, the current results portray potential error rates if all scores within each scale combination is considered. Current results might overestimate rates of significance if only specific scores are considered within a scale family, as in the instance of using an a priori approach to test interpretation. This type of approach would

be aided actively seeking out data to disconfirm a priori hypotheses, and by looking for patterns of consistency across MMPI-2 scales and with other measures (Palmer et al., 1998).

Another potential means of reducing interpretive false positive would be to eliminate highly redundant scales that do not contribute significantly to interpretation. As mentioned previously there are approximately 10 interscale correlation values between .90 - .95 and almost 30 correlations between .80 - .89 (University of Minnesota Press, 2001). The degree to which these scales are correlated indicates that they share a significant degree of variance and are measuring the same or very similar construct. When repeated tests are performed (i.e., consideration of multiple scores) the examiner is to an extent repeating the same test multiple times. For example, the Content scale Cynicism is correlated with its component scale (CYN1) at $r = .95$. Removing redundancy may in effect reduce the overall chance for committing a type I error of interpretation.

Equally important to not inappropriately misclassifying an examinee's symptoms as pathological, is the potential for an increase in false negatives at the expense of too stringent cutoff criteria. Using an overly conservative cutoff score could result in a clinician losing important information relevant to a patient's level of functioning and need for treatment. When evaluating a large number of data points Garcia (2004) referred to the potential for losing valuable information in ecological research at the cost of a high confidence as the "Bonferroni iron claw" (p. 657). The author highlighted the significant loss of power (finding significance when a difference exists) associated with using overly

conservative cut-scores. There is a balance that must be struck between false positives and negatives.

Recommendations for Future Research

An intriguing direction for future research efforts would be to determine how well clinicians trained in MMPI-2 interpretation could successfully classify *clinically meaningful* score profiles from profiles containing normal rates of *N* or more elevated scores in the normal population (such as demonstrated in this study) for various combinations of scales. For example, classification accuracy could be examined for simultaneous consideration of the Clinical Scales alone, the Clinical and Harris-Lingoes Scales together, Content Scales alone, Content and Content-Component Scales, Supplementary Scales, Clinical, Content, and Supplementary Scales, and for all scale groups combined. A median score profile could be generated by resampling distributions of scores. This average number of elevated scores could then be used as part of an internet survey. A participant pool could be obtained through membership directories of relevant professional organizations (e.g., APA Division 40, ABPP/ABCN listserv, National Academy of Neuropsychology, International Neuropsychological Society) and participants randomly assigned to interpret actual clinical profiles or the median profile generated by Monte Carlo simulation.

Results from the aforementioned hypothetical study might bridge the divide between the current results and clinical practice, providing useful information regarding how well MMPI-2 interpretive procedures are carried out. More importantly, perhaps,

such a study could determine whether specific *T*-score cut-points and consideration of different combinations of scales influences the formation of diagnostic conclusions.

Future research may aim to further investigate the effects of different demographic variables on rates of significance in normal adults on the MMPI-2 to determine if their impact on rates of significance is similar to other neuropsychological tests (Brooks et al., 2007, 2008, 2009, 2010; Brooks, Iverson, Sherman et al., 2009; Crawford et al., 2007; Decker et al., 2012; Schretlen et al., 2003, 2008). Some MMPI-2 scales such as Scale 2 ("Depression"), for example, have been shown to be impacted by age. African-Americans, Native Americans, and Hispanics score five to ten *T*-score points higher than Caucasians and Asian Americans on Scales 4 and 9 ("Psychopathic Deviate" & "Hypomania"), whereas scores tend to be slightly lower in older people (Graham, 2006). Higher levels of education are associated with scores five *T*-score points greater on Scale 5 ("Masculinity-Femininity"). Average *T*-score values on Scale 8 ("Schizophrenia") are frequently five points higher in ethnic minorities (Graham, 2006). Such an investigation would also be particularly interesting in light of counterintuitive findings found by Long, Graham, and Timbrook (1994), which suggested that although lower levels of education and income are associated with higher MMPI scores, their scores underpredicted problem behaviors. The sum-total of differences attributable to different demographic factors could have a cumulative impact on overall rates of significance in the normal population.

Future research might address base rates in the normal population associated with multiple score interpretation on the Minnesota Multiphasic Personality Inventory –

Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008). Characteristics of the MMPI-2-RF's construction may reduce the occurrence of clinical false positives. For example, a stated advantage of the MMPI-2-RF includes the use of relatively fewer scales (Green, 2011), which should in theory lead to reductions in the number of elevated scores (i.e., reduced "family-wise error" rate; Crawford et al., 2007). A stated strength of the MMPI-2-RF core Restructured Clinical scales over the MMPI-2 Clinical scales is the relative distinctiveness of each scale, due to the extraction of items loading onto a common factor labeled as "Demoralization" (Tellegen et al., 2003). Noteworthy is the likely effect of scale interrelatedness on the observed frequency of score elevations (Crawford et al., 2007). Comparatively, a weaker mean interscale correlation ($r \approx .20$) for the MMPI-2-RF scales should be at least partially associated with a proportionally higher number of normal adults with seemingly abnormal scores (Tellegen & Ben-Porath, 2003). Along this same line of thought, a contrastingly more significant degree of interscale overlap amongst the MMPI-2 scales may to an extent prevent modest increases in seemingly aberrant scores (Graham, 2007; Green, 2011).

The current study did not investigate the rates of specific high-point elevations, or two-point or three-point code types, which may be limiting to the extent the reader is interested in determining the commonality of specific profile configurations. Data for two- and three-point code types are already available for multiple different patient populations and have been extensively reviewed for the MMPI-2 (e.g., Arbisi, Ben-Porath, & McNulty, 2003; Archer, Griffin, & Aiduk, 1995; Graham, 2006; McGrath & Ingersoll, 1999a; McGrath & Ingersoll, 1999b; Munley et al., 2004; Sellbom, Graham, &

Schenk, 2005; Graham, Ben-Porath, McNulty, 1999), as well as code type data retained for interpretation from the original MMPI (see Graham, 2006). The current study determined that approximately 16% and 8% of the normal population is expected to have at least two or three significant scores ≥65*T* on the Clinical scales, respectively. The percentage of the normal population with two- or three-point code types, however, would be less than the rate of the population with two or three elevations, because all possible combinations of the 10 Clinical scales do not have two- or three-point external correlates. If all possible combinations of the 10 Clinical scales existed for two- and three-point code types 810 configurations would exist, however, external correlates only exist for 73 total combined code types (when 1/3 3/1 are counted separately; Graham, 2006). Use of two- and three-point code types may reduce the overall level of clinical false positives in the normal population, assuming code types are not considered in addition to their one-point elevations, in which circumstance the current Monte Carlo results may be an underestimate (Crawford et al., 2007).

A recent paper by Decker and colleagues (2012) suggested future studies evaluate the impact of skewness and kurtosis on the Monte Carlo's ability to accurately to produce base rates of *N* or more scores for non-normally distributed sets of scores, such as would occur when attempting to resample distributions of scores from a clinical sample. Such a study would require actual frequency data for a clinical sample. Correlation matrices could then be constructed from the known data set. Comparisons could then be made between actual base rates, Monte Carlo simulations using normally distributed random

variates, and Monte Carlo simulations that take into account skew and kurtosis (Schretlen et al., 2009; Decker et al., 2012).

Summary and Conclusions

The Minnesota Multiphasic Personality Inventory – second edition (MMPI-2) is one of the most widely used (Rabin, Barr, & Burton, 2005) instruments among practicing neuropsychologists; however, until now a critical aspect of its psychometric qualities had been largely unexplored: the effects of multiple score interpretation on base rates. Present findings provide the first meaningful data for the relationship between multiple score interpretation and the frequency of normal adults with significant MMPI-2 scale elevations. The magnitude at which these rates depart from the theoretical normal distribution (Hypotheses II & III) exemplify the importance of not over interpreting scores in clinical and non-clinical settings, as substantial proportions of the general population can be expected to score beyond traditional cut-points. This is not only true when considering all MMPI-2 scales in unison, but for consideration of separate distinct scale families. The nongendered base rates for percentages of the normal population with *N* or more significant scores created in this study were designed to be easily referenced (Hypothesis I; Tables 4, 5, & 6), and are strengthened by the large representative nature of the Restandardization Sample used for the Monte Carlo Simulations (Butcher et al., 2001).

In common practice these findings demonstrate the importance of understanding base rates of significance in the normal population. Pragmatically, clinical false positives may translate into real-world consequences for examinees, such as being mistakenly

ascribed psychopathology or experiencing reductions in overall quality of life. For example, at least a singular elevation is likely to occur in over 80% of normal adults, and over 20% can be expected to have *at least* seven significant scale elevations, all of which could be mistakenly ascribed clinical significance. As traditional notions of significance ($T \geq 65$) are relaxed the potential for clinical false positives increases in tandem. This application of the Monte Carlo method underscores the need for hypothesis driven interpretations of data that integrate extra-test information, minimize redundancy, and dispense with superfluous scales that do not add incremental validity (Graham, 2006). This is not to imply that elevated scores are to be sought out as they confirm *a priori* hypotheses, but rather that "disconfirmatory" strategies are applied. The latter issue becomes even more pronounced with the advent of the MMPI-2-RF and potential for interpreting 204 MMPI scales in total, all of which are derived from the same pool of 567 items (Green, 2011; Tellegen & Ben-Porath, 2008). Consideration for all 204 scales together is possible, but the significant overlap in content begs the question of whether such an approach increments novelty without unnecessarily inflating rates of significance in the normal population. Focused score interpretation and more stringent definitions of abnormality, as was demonstrated here and by others (Crawford et al., 2007), reduces the rate of elevated scores among normal adults by refining the numerical quality significance and decreasing the overall number of scores interpreted.

REFERENCES

Arbisi, P. A., Ben-Porath, Y. S., & McNulty, J. L. (2003). Empirical correlates of common MMPI-2 two-point codes in male psychiatric inpatients. *Assessment, 10*(3), 237-247.

Archer, R. P., Griffin, R., & Aiduk, R. (1995). MMPI-2 clinical correlates for ten common codes. *Journal of Personality Assessment, 65*(3), 391-407.

Axelrod, B. N., & Wall, J. R. (2007). Expectancy of impaired neuropsychological test scores in a non-clinical sample. *International Journal of Neuroscience, 117*, 1591–1602.

Benton, A. L., Sivan, A. B., Hamsher, K., Varney, N. R., & Spreen, O. (1994). *Contributions to neuropsychological assessment: A clinical manual, 2nd Edition*. New York, NY: Oxford University Press.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*, 31–46.

Bow, J. N., Flens, J. R., & Gould, J. W. (2010). MMPI-2 and MCMI-III in forensic evaluations: A survey of psychologists. *Journal of Forensic Psychology Practice, 10*, 37-52.

Brooks, B. L., & Iverson, G. L. (2010). Comparing actual to estimated base rates of "abnormal" scores on neuropsychological test batteries: Implications for interpretation. *Archives of Clinical Neuropsychology, 25*(1), 14-21.

Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society, 14,* 463-478.

Brooks, B. L., Iverson, G. L., Sherman, E. M. S., & Holdnack, J. A. (2009). Healthy children and adolescents obtain some low scores across a battery of memory tests. *Journal of the International Neuropsychological Society, 15*, 613–617.

Brooks, B. L., Iverson, G. L., & White, T. (2007). Substantial risk of "accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society, 13*, 490–500.

Brooks, B. L., Sherman, E. M. S., & Iverson, G. L. (2010). Healthy children get low scores too: Prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology, 25*, 182–190.

Brooks, B. L., Strauss, E., Sherman, E. M. S., Iverson, G. L., & Slick, D. J. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology*, *50*(3), 196–209.

Butcher, J. N. (1992). The research base, psychometric properties, and clinical uses of the MMPI-2 and MMPI-A. *Canadian Psychology, 33*(1), 61-86.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., & Dahlstrom, W. G. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory -2): Manual for*

administration, scoring, and interpretation (Revised ed.). Minneapolis, Minnesota: University of Minnesota Press.

Butcher, J. N, Graham, J. R., Williams, C. L., & Ben-Porath, Y. S. (1990). *Development and use of the MMPI-2 Content scales*. Minneapolis: University of Minnesota Press.

Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007) Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology, 21*(4), 419-430.

Decker, S. L., Schneider, W. J., & Hale, J. B. (2012). Estimating base rates of impairment in neuropsychological test batteries: A comparison of quantitative models. *Archives of Clinical Neuropsychology, 27*, 69-84.

Diaz-Asper, C. M., Schretlen, D. J., Pearlson, G. D., (2004). How well does IQ predict neuropsychological test performance in normal adults? *Journal of the Neuropsychological Society, 10*(1), 82-89.

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

Fishman, G. S. (1996). *Monte Carlo: Concepts, algorithms, and applications.* New York: Springer.

Gass, C. S. (2006). Use of the MMPI-2 in neuropsychological evaluations. In J.N. Butcher (Ed.), *MMPI-2: A practitioners guide* (p. 301-326). Washington, DC: American Psychological Association.

Gass, C. S., & Russell, E. W. (1991). MMPI profiles of closed head trauma patients: Impact of neurologic complaints. *Journal of Clinical Psychology, 47*(2), 253-260.

Gentle, J. E. (2007). *Matrix algebra: Theory, computations, and applications in statistics.* Springer.

Gentle, J. E. (2003). *Random number generation and Monte Carlo methods.* Springer.

Gotts, E. E., & Knudsen, T. E. (2005). *The clinical interpretation of the MMPI-2: A content cluster approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Graham, J. R. (2006). MMPI-2: *Assessing personality and psychopathology* (4th ed.). New York: Oxford University Press.

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

Green, R. L. (2000). *The MMPI-2: An interpretive manual* (2nd ed.). Boston: Allyn and Bacon.

Greene, R. L. (2011). *The MMPI-2/MMPI-2-RF: An interpretive manual* (3rd ed.). Boston, MA: Allyn & Bacon.

Hathaway, S. R. (1956). Scales 5 (Masculinity-Femininity), 6 (Paranoia), 8 (Schizophrenia). In G. S. Welsh & W. G. Dahlstrom (Eds.)*, Basic readings on the MMPI in psychology and medicine.* Minneapolis: University of Minnesota Press.

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

Hathaway, S. R., & McKinley, J. C. (1940b). A multiphasic personality schedule (Minnesota): III. The measurement of symptomatic depression. *The Journal of Psychology: Interdisciplinary and Applied, 14*, 73-84.

Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery: Demographic corrections, research findings, and clinical applications.* Odessa, Florida: Psychological Assessment Resources.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead–Reitan battery (norms, manual and computer program)*. Odessa, Florida: Psychological Assessment Resources.

Helmes, E., & Reddon, J. R. (1993). A perspective on developments in assessing psychopathology: A critical review of the MMPI and MMPI-2. *Psychological Bulletin, 113*(3), 453-471.

Ingraham, L. J., & Aiken, C. B. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology, 10*(1), 120-124.

Lemieux, C. (2009). *Monte Carlo and quasi-Monte Carlo sampling.* United States of America: Springer.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, J. H., & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). New York: NY, Oxford University Press.

Long, K. A., Graham, J. R., & Timbrook, R. E. (1994). Socioeconomic status and MMPI-2 interpretation. *Measurement and Evaluation in Counseling and Development, 27*, 158-177.

Madras, N. (Ed.) (2000). *Monte Carlo methods*. Providence, Rhode Island: American Mathematical Society

McGrath, R. E., & Ingersoll, J. (1999a). Writing a good cookbook: I. A review of MMPI high-point code system studies. *Journal of Personality Assessment, 73*(2), 149-178.

McGrath, R. E., & Ingersoll, J. (1999b). Writing a good cookbook: II. A synthesis of MMPI high-point code system study effect sizes. *Journal of Personality Assessment, 73*(2), 179-198.

McKinley, J. C., & Hathaway, S. R. (1940). A multiphasic personality schedule (Minnesota): II. A differential study of hypochondriasis. *The Journal of Psychology: Interdisciplinary and Applied, 10*, 255-268.

McKinley, J. C., & Hathaway, S. R. (1942). A multiphasic personality schedule (Minnesota): IV. Psychasthenia. *Journal of Applied Psychology, 26*(5), 614-624.

McKinley, J. C., & Hathaway, S. R. (1944). The Minnesota multiphasic personality inventory. V. Hysteria, hypomania and psychopathic deviate. *Journal of Applied Psychology, 28*(2), 153-174.

McKinley, J. C., & Hathaway, S. R. (1948). The Minnesota Multiphasic Personality Inventory: VI. The K scale. *Journal of Consulting Psychology, 12*(1), 20-31.

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

Minnix, J. A., Reitzel, L. R., Repper, K. A., Burns, A. B., Williams, F., Lima, E. N., Cukrowicz, K. C., Kirsch, L., & Joiner, T. E. (2005). Total number of MMPI-2 clinical scale elevations predicts premature termination after controlling for intake symptom severity and personality disorder diagnosis. *Personality and Individual Differences, 38*, 1745-1755.

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos, 100*(2), 403-405.

Munley, P. H. (1991). Confidence intervals for the MMPI-2. *Journal of Personality Assessment, 57*(1), 52-60.

Munley, P. H., Germain, J. M., Tovar-Murray, D., Borgman, A. L. (2004). MMPI-2 profile code types and measurement error. *Journal of Personality Assessment, 82*(2), 179-188.

Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1997). Base rates of "impaired" neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology, 3*(6), 503-511.

Predictive Analytics SoftWare (Version 18.0.0) [Computer software]. Chicago: SPSS Inc.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*, 33-65.

Rebonato, R., & Jäckel, P. (1999). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Quantitative Research Centre of the NatWest Group*.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation.* Tucson, Arizona: Neuropsychology Press.

Rubino, S., & Tuffin, B. (Eds.) (2009). *Rare event simulation using Monte Carlo methods*. United Kingdom: John Wiley & Sons Ltd.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). San Diego: Jerome M. Sattler, Publisher, Inc.

Schinka, J. A., & LaLone, L. (1997). MMPI-2 Norms: Comparisons with a census-matched subsample. *Psychological Assessment, 9*(3), 307-311.

Schretlen, D. J., Munro, C. A., Anthony, J. C., & Pearlson, G. D. (2003). Examining the range of normal intraindividual variability in neuropsychological test performance. *Journal of the International Neuropsychological Society, 9*, 864-870.

Schretlen, D. J., Testa, S. M., Winicki, J. M., Pearlson, G. D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society, 14,* 436–445.

Sellbom, M., Graham, J. R., & Schenk, P. W. (2005). Symptom correlates of MMPI-2 scales and code types in a private-practice setting. *Journal of Personality Assessment, 84*(2), 163-171.

Stein, L. A. R., Graham, J. R., Ben-Porath, Y. S., McNulty, J. (1999). Using the MMPI-2 to detect substance abuse in a mental health setting. *Psychological Assessment, 11*(1), 94-100.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2007). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). United States of America: Oxford University Press.

Tellegen, A., Ben-Porath, Y. S. (2008). *MMPI-2-RF technical manual.* Minneapolis: University of Minnesota Press.

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

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI-2 Restructured Clinical scales: Development, validation, and interpretation.* Minneapolis: University of Minnesota Press.

University of Minnesota Press (2001). MMPI®-2 Scale Intercorrelations—Normative Sample. Unpublished data. Reproduced by permission of the University of Minnesota Press on behalf of the Regents of the University of Minnesota, copyright-owner of the Minnesota Multiphasic Personality Inventory®-2 (MMPI®-2) 1989, 2001.

Watkins, C. E., & Campbell, V. L. (2000). *Testing and assessment in counseling practice.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale (3rd ed.).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale (3rd Ed.).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children (4th ed.).* San Antonio, TX: Psychological Corporation.

Appendix A

SPSS Syntax Used for Monte Carlo Procedure.

```
INPUT PROGRAM.
LOOP j=1 to 1000000.
COMPUTE X1 = rv.norm(0,1).
COMPUTE X2 = rv.norm(0,1).
COMPUTE X3 = rv.norm(0,1).
COMPUTE X4 = rv.norm(0,1).
COMPUTE X5 = rv.norm(0,1).
COMPUTEX… = rv.norm(0,1).
COMPUTE X98 =rv.norm(0,1).
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

matrix.
get x / variables=x1 to x98.

compute cor={INSERT CORRELATION MATRIX}.

/*******eigenvectors (E_vec) and eigenvalues (E_val) of correlation matrix********/.
call eigen (cor,E_vec,E_val).

/********find and set set negative eignevalues equal to a small value*********/.
compute E_valp=E_val.
compute FLAGS = E_val >= 0.
compute tflags=t(FLAGS).
compute  n=nrow(E_vec).
compute iflags=mdiag(FLAGS).
compute negindx=Trace(iflags) +1.

loop j=negindx to n.
+ do if (E_valp(j) < 0).
+ compute E_valp(j)= .00000001.
+ end if.
end loop.

/*********square root of adjusted eigenvalues)*********/.
compute lamb_adj=mdiag(sqrt(E_valp)).

/**********square each elements of eigenvectors********/.
compute E_vecsq=E_vec &** 2.

/**********compute the T scaling matrix and its diagonal ***********/.
compute ti=E_vecsq*E_valp.
compute T_matrx=mdiag(sqrt((ti &** -1.))).

/***********compute the adjusted correlation matrix*********/.
compute B=T_matrx*E_vec*lamb_adj.
compute cor_adj=B*t(B).
/*print cor_adj.
/*************Cholesky **********************/.
compute deter=det( cor_adj).
/*print deter / title "determinant of corr matrix" / format=f10.7 .
```

```
/*print sval( cor_adj) / title "singular value decomposition of corr".
/*print eval( cor_adj) / title "eigenvalues of input corr".

* In a symmetric matrix sval and eigenvalues are identical - choose 1 .

compute condnum=mmax(sval( cor_adj))/mmin(sval( cor_adj)).
/*print condnum / title "condition number of adjusted corr matrix" / format=f10.2 .
compute cho=chol(cor_adj).
/*print cho / title "cholesky factor of adjusted corr matrix" /format=f8.4.
compute chochek=t(cho)*cho.

/*print chochek / title "chol factor premult by its transpose " /format=f10.2.
compute newx=x*cho.
compute newx=newx*10 + 50.
save newx /outfile=* /variables= nr1 to nr98.
end matrix.

IF  (nr2 < 65) nr11=0.
EXECUTE.
IF  (nr2 < 65) nr12=0.
EXECUTE.
IF  (nr2 < 65) nr13=0.
EXECUTE.
IF  (nr2 < 65) nr14=0.
EXECUTE.
IF  (nr2 < 65) nr15=0.
EXECUTE.
IF  (nr3 < 65) nr16=0.
EXECUTE.
IF  (nr3 < 65) nr17=0.
EXECUTE.
IF  (nr3 < 65) nr18=0.
EXECUTE.
IF  (nr3 < 65) nr19=0.
EXECUTE.
IF  (nr3 < 65) nr20=0.
EXECUTE.
IF  (nr4 < 65) nr21=0.
EXECUTE.
IF  (nr4 < 65) nr22=0.
EXECUTE.
IF  (nr4 < 65) nr23=0.
EXECUTE.
IF  (nr4 < 65) nr24=0.
EXECUTE.
IF  (nr4 < 65) nr25=0.
EXECUTE.
IF  (nr6 < 65) nr26=0.
EXECUTE.
IF  (nr6 < 65) nr27=0.
EXECUTE.
IF  (nr6 < 65) nr28=0.
EXECUTE.
IF  (nr8 < 65) nr29=0.
EXECUTE.
IF  (nr8 < 65) nr30=0.
EXECUTE.
IF  (nr8 < 65) nr31=0.
EXECUTE.
```

```
IF  (nr8 < 65) nr32=0.
EXECUTE.
IF  (nr8 < 65) nr33=0.
EXECUTE.
IF  (nr8 < 65) nr34=0.
EXECUTE.
IF  (nr9 < 65) nr35=0.
EXECUTE.
IF  (nr9 < 65) nr36=0.
EXECUTE.
IF  (nr9 < 65) nr37=0.
EXECUTE.
IF  (nr9 < 65) nr38=0.
EXECUTE.
IF  (nr10 < 65) nr39=0.
EXECUTE.
IF  (nr10 < 65) nr40=0.
EXECUTE.
IF  (nr10 < 65) nr41=0.
EXECUTE.
IF  (nr43 < 60) nr57=0.
EXECUTE.
IF  (nr43 < 60) nr58=0.
EXECUTE.
IF  (nr45 < 60) nr59=0.
EXECUTE.
IF  (nr45 < 60) nr60=0.
EXECUTE.
IF  (nr45 < 60) nr61=0.
EXECUTE.
IF  (nr45 < 60) nr62=0.
EXECUTE.
IF  (nr46 < 60) nr63=0.
EXECUTE.
IF  (nr46 < 60) nr64=0.
EXECUTE.
IF  (nr46 < 60) nr65=0.
EXECUTE.
IF  (nr47 < 60) nr66=0.
EXECUTE.
IF  (nr47 < 60) nr67=0.
EXECUTE.
IF  (nr48 < 60) nr68=0.
EXECUTE.
IF  (nr48 < 60) nr69=0.
EXECUTE.
IF  (nr49 < 60) nr70=0.
EXECUTE.
IF  (nr49 < 60) nr71=0.
EXECUTE.
IF  (nr50 < 60) nr72=0.
EXECUTE.
IF  (nr50 < 60) nr73=0.
EXECUTE.
IF  (nr51 < 60) nr74=0.
EXECUTE.
IF  (nr51 < 60) nr75=0.
EXECUTE.
IF  (nr52 < 60) nr76=0.
```

```
EXECUTE.
IF  (nr52 < 60) nr77=0.
EXECUTE.
IF  (nr53 < 60) nr78=0.
EXECUTE.
IF  (nr53 < 60) nr79=0.
EXECUTE.
IF  (nr54 < 60) nr80=0.
EXECUTE.
IF  (nr54 < 60) nr81=0.
EXECUTE.
IF  (nr56 < 60) nr82=0.
EXECUTE.
IF  (nr56 < 60) nr83=0.
EXECUTE.

COUNT
  LT1.0sd98tests = nr1 nr2 nr3 nr4 nr5 nr6 nr7 nr8 nr9 nr10 nr11 nr12 nr13 nr14 nr15 nr16 nr17 nr18 nr19 nr20 nr21
nr22 nr23 nr24 nr25 nr26 nr27 nr28 nr29 nr30 nr31 nr32 nr33 nr34 nr35 nr36 nr37 nr38 nr39 nr40
nr41 nr42 nr43 nr44 nr45 nr46 nr47 nr48 nr49 nr50 nr51 nr52 nr53 nr54 nr55 nr56 nr57 nr58 nr59 nr60
nr61 nr62 nr63 nr64 nr65 nr66 nr67 nr68 nr69 nr70 nr71 nr72 nr73 nr74 nr75 nr76 nr77 nr78 nr79 nr80
nr81 nr82 nr83 nr84 nr85 nr86 nr87 nr88 nr89 nr90
nr91 nr92 nr93 nr94 nr95 nr96 nr97 nr98 (60 thru Highest)  .
VARIABLE LABELS LT1.0sd98tests 'Number of scores higher than 60 with 98 measures' .
EXECUTE .

COUNT
  LT1.5sd98tests = nr1 nr2 nr3 nr4 nr5 nr6 nr7 nr8 nr9 nr10 nr11 nr12 nr13  nr14 nr15 nr16 nr17 nr18 nr19 nr20 nr21
nr22 nr23 nr24 nr25 nr26 nr27 nr28 nr29 nr30 nr31 nr32 nr33 nr34 nr35 nr36 nr37 nr38 nr39 nr40
nr41 nr42 nr43 nr44 nr45 nr46 nr47 nr48 nr49 nr50 nr51 nr52 nr53 nr54 nr55 nr56 nr57 nr58 nr59 nr60
nr61 nr62 nr63 nr64 nr65 nr66 nr67 nr68 nr69 nr70 nr71 nr72 nr73 nr74 nr75 nr76 nr77 nr78 nr79 nr80
nr81 nr82 nr83 nr84 nr85 nr86 nr87 nr88 nr89 nr90
nr91 nr92 nr93 nr94 nr95 nr96 nr97 nr98 (65 thru Highest)  .
VARIABLE LABELS LT1.5sd98tests 'Number of scores higher than 65 with 98 measures' .
EXECUTE .

COUNT
  LT2.0sd98tests = nr1 nr2 nr3 nr4 nr5 nr6 nr7 nr8 nr9 nr10 nr11 nr12 nr13 nr14 nr15 nr16 nr17
nr18 nr19 nr20 nr21 nr22 nr23 nr24 nr25 nr26 nr27 nr28 nr29 nr30 nr31 nr32 nr33 nr34 nr35 nr36 nr37 nr38 nr39 nr40
nr41 nr42 nr43 nr44 nr45 nr46 nr47 nr48 nr49 nr50 nr51 nr52 nr53 nr54 nr55 nr56 nr57 nr58 nr59 nr60 nr61 nr62 nr63
nr64 nr65 nr66 nr67 nr68 nr69 nr70 nr71 nr72 nr73 nr74 nr75 nr76 nr77 nr78 nr79 nr80 nr81 nr82 nr83 nr84 nr85 nr86
nr87 nr88 nr89 nr90
nr91 nr92 nr93 nr94 nr95 nr96 nr97 nr98 (70 thru Highest)  .
VARIABLE LABELS LT2.0sd98tests 'Number of scores higher than 70 with 98 measures' .
EXECUTE .

COUNT
  LT2.5sd98tests =  nr1 nr2 nr3 nr4 nr5 nr6 nr7 nr8 nr9 nr10 nr11 nr12 nr13 nr14 nr15 nr16 nr17 nr18 nr19 nr20 nr21
nr22 nr23 nr24 nr25 nr26 nr27 nr28 nr29 nr30 nr31 nr32 nr33 nr34 nr35 nr36 nr37 nr38 nr39 nr40
nr41 nr42 nr43 nr44 nr45 nr46 nr47 nr48 nr49 nr50 nr51 nr52 nr53 nr54 nr55 nr56 nr57 nr58 nr59 nr60
nr61 nr62 nr63 nr64 nr65 nr66 nr67 nr68 nr69 nr70 nr71 nr72 nr73 nr74 nr75 nr76 nr77 nr78 nr79 nr80
nr81 nr82 nr83 nr84 nr85 nr86 nr87 nr88 nr89 nr90
nr91 nr92 nr93 nr94 nr95 nr96 nr97 nr98 (75 thru Highest)  .
VARIABLE LABELS LT2.5sd98tests 'Number of scores higher than 75 with 98 measures' .
EXECUTE .
FREQUENCIES
  VARIABLES= LT1.0sd98tests LT1.5sd98tests LT2.0sd98tests LT2.5sd98tests
  /FORMAT=DVALUE
  /ORDER=  ANALYSIS
```

Appendix B

MMPI-2 Clinical, Content, Supplementary, Harris-Lingoes, & Content-Component Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	97.291	97.310	80.358	80.159	49.980	49.249	21.861	21.273
2 or more	91.223	91.099	62.288	61.420	29.709	28.707	9.586	9.238
3 or more	82.253	81.811	48.009	46.714	19.545	18.681	5.373	5.229
4 or more	72.242	71.237	37.895	36.448	14.000	13.457	3.465	3.447
5 or more	62.785	61.191	30.792	29.461	10.599	10.336	2.441	2.457
6 or more	54.789	52.885	25.629	24.598	8.335	8.235	1.802	1.850
7 or more	48.293	46.386	21.761	20.990	6.735	6.764	1.388	1.434
8 or more	42.970	41.295	18.724	18.208	5.557	5.648	1.099	1.150
9 or more	38.577	37.136	16.277	15.977	4.678	4.795	.878	.936
10 or more	34.821	33.685	14.283	14.164	3.966	4.118	.719	.774
11 or more	31.660	30.710	12.667	12.663	3.401	3.565	.591	.647
12 or more	28.893	28.151	11.302	11.371	2.930	3.120	.489	.547
13 or more	26.460	25.937	10.114	10.273	2.546	2.732	.408	.466
14 or more	24.327	23.973	9.097	9.330	2.225	2.409	.341	.398
15 or more	22.448	22.210	8.196	8.485	1.943	2.133	.290	.338
16 or more	20.725	20.628	7.414	7.735	1.702	1.883	.240	.292
17 or more	19.171	19.184	6.717	7.072	1.492	1.666	.206	.247
18 or more	17.753	17.883	6.095	6.473	1.311	1.477	.174	.216
19 or more	16.461	16.669	5.547	5.922	1.157	1.316	.144	.183
20 or more	15.280	15.569	5.040	5.427	1.012	1.166	.121	.156
21 or more	14.159	14.547	4.583	4.970	.886	1.038	.103	.131
22 or more	13.169	13.603	4.169	4.551	.775	.917	.086	.110
23 or more	12.231	12.714	3.784	4.176	.683	.809	.074	.094
24 or more	11.370	11.894	3.426	3.821	.597	.716	.063	.081
25 or more	10.572	11.114	3.107	3.496	.519	.636	.053	.070
26 or more	9.807	10.394	2.818	3.195	.449	.563	.045	.059
27 or more	9.086	9.702	2.544	2.913	.389	.499	.038	.048
28 or more	8.426	9.056	2.295	2.648	.341	.440	.032	.040
29 or more	7.807	8.430	2.065	2.407	.298	.385	.027	.034
30 or more	7.225	7.843	1.858	2.190	.260	.337	.021	.029
31 or more	6.672	7.290	1.667	1.982	.223	.296	.017	.024
32 or more	6.158	6.770	1.495	1.793	.188	.253	.013	.021
33 or more	5.669	6.287	1.327	1.619	.158	.218	.011	.017
34 or more	5.204	5.817	1.177	1.450	.133	.188	.009	.015
35 or more	4.766	5.384	1.044	1.298	.114	.161	.007	.012
36 or more	4.362	4.956	.924	1.154	.096	.138	.005	.010
37 or more	3.977	4.557	.808	1.033	.082	.117	.004	.008
38 or more	3.625	4.184	.708	.914	.066	.098	.003	.006
39 or more	3.300	3.835	.619	.804	.055	.084	.002	.005
40 or more	2.997	3.513	.538	.711	.046	.071	.002	.003
41 or more	2.706	3.212	.468	.627	.039	.060	.002	.003
42 or more	2.429	2.918	.403	.551	.033	.048	.001	.002
43 or more	2.181	2.641	.343	.482	.027	.040	.001	.002
44 or more	1.950	2.390	.293	.417	.021	.034	.000	.001
45 or more	1.740	2.156	.249	.363	.019	.028	.000	.001
46 or more	1.539	1.930	.213	.306	.014	.023	.000	.001
47 or more	1.354	1.721	.180	.261	.011	.018	.000	.000
48 or more	1.191	1.531	.152	.220	.009	.015	.000	.000
49 or more	1.043	1.345	.125	.185	.006	.012	.000	.000
50 or more	.904	1.189	.105	.157	.005	.010	.000	.000

Appendix C

MMPI-2 Content & Content-Component Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	68.147	66.745	52.163	50.987	30.101	29.352	12.186	11.876
2 or more	61.965	60.998	39.464	38.668	17.943	17.639	5.453	5.364
3 or more	51.074	50.188	29.419	28.943	11.248	11.146	2.733	2.753
4 or more	43.363	42.636	22.818	22.542	7.657	7.616	1.609	1.648
5 or more	37.724	37.225	18.245	18.118	5.471	5.525	1.014	1.064
6 or more	32.690	32.351	14.779	14.777	4.018	4.111	.667	.725
7 or more	28.657	28.455	12.102	12.198	3.031	3.147	.455	.507
8 or more	25.185	25.153	9.981	10.162	2.307	2.427	.323	.361
9 or more	22.150	22.249	8.297	8.497	1.777	1.890	.232	.260
10 or more	19.550	19.722	6.892	7.133	1.369	1.492	.165	.190
11 or more	17.252	17.512	5.748	6.012	1.063	1.180	.119	.140
12 or more	15.214	15.537	4.795	5.071	.831	.936	.086	.105
13 or more	13.414	13.772	4.005	4.275	.646	.746	.060	.074
14 or more	11.819	12.199	3.340	3.604	.502	.577	.044	.051
15 or more	10.391	10.796	2.774	3.021	.384	.454	.032	.036
16 or more	9.088	9.516	2.283	2.527	.298	.356	.024	.025
17 or more	7.934	8.358	1.880	2.094	.229	.277	.015	.017
18 or more	6.907	7.330	1.535	1.735	.176	.211	.011	.012
19 or more	5.979	6.411	1.257	1.435	.135	.160	.008	.009
20 or more	5.170	5.571	1.023	1.175	.102	.119	.006	.006
21 or more	4.438	4.822	.828	.962	.077	.088	.004	.004
22 or more	3.782	4.154	.658	.777	.057	.064	.003	.003
23 or more	3.210	3.562	.518	.626	.041	.045	.002	.002
24 or more	2.701	3.015	.403	.488	.029	.033	.001	.002
25 or more	2.244	2.522	.308	.386	.020	.023	.001	.001
26 or more	1.852	2.105	.235	.292	.014	.017	.001	.001
27 or more	1.501	1.727	.179	.223	.009	.010	.000	.000
28 or more	1.201	1.390	.127	.166	.006	.008	.000	.000
29 or more	.951	1.113	.092	.119	.004	.005	.000	.000
30 or more	.739	.873	.063	.084	.002	.003	.000	.000
31 or more	.553	.671	.042	.057	.001	.002	.000	.000
32 or more	.412	.507	.027	.037	.001	.001	.000	.000
33 or more	.294	.365	.017	.022	.000	.001	.000	.000
34 or more	.202	.259	.010	.013	.000	.000	.000	.000
35 or more	.132	.173	.005	.008	.000	.000	.000	.000
36 or more	.082	.109	.003	.004	.000	.000	.000	.000
37 or more	.048	.065	.001	.002	.000	.000	.000	.000
38 or more	.025	.035	.001	.001	.000	.000	.000	.000
39 or more	.012	.016	.000	.000	.000	.000	.000	.000
40 or more	.004	.007	.000	.000	.000	.000	.000	.000

Appendix D

MMPI-2 Clinical & Harris-Lingoes Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	64.851	64.120	37.092	36.416	22.455	21.796	9.771	9.419
2 or more	45.173	43.602	28.071	26.906	12.225	11.790	3.582	3.554
3 or more	35.162	33.867	19.544	18.797	6.924	6.937	1.661	1.755
4 or more	27.448	26.622	13.174	12.937	4.269	4.470	.894	1.010
5 or more	21.165	20.720	9.605	9.754	2.837	3.142	.528	.642
6 or more	16.622	16.532	7.302	7.652	1.937	2.261	.331	.427
7 or more	13.349	13.504	5.570	6.058	1.364	1.660	.208	.293
8 or more	10.845	11.197	4.240	4.790	.959	1.222	.134	.202
9 or more	8.850	9.350	3.220	3.781	.681	.901	.086	.136
10 or more	7.178	7.789	2.480	2.994	.482	.675	.054	.094
11 or more	5.769	6.450	1.892	2.383	.338	.501	.035	.064
12 or more	4.583	5.284	1.444	1.883	.236	.369	.022	.044
13 or more	3.654	4.332	1.098	1.489	.159	.269	.013	.029
14 or more	2.913	3.554	.812	1.154	.106	.193	.008	.020
15 or more	2.300	2.917	.598	.885	.073	.134	.004	.014
16 or more	1.806	2.350	.426	.675	.047	.093	.003	.010
17 or more	1.389	1.869	.307	.502	.028	.064	.001	.006
18 or more	1.044	1.462	.212	.368	.018	.042	.001	.004
19 or more	.770	1.122	.143	.263	.011	.026	.000	.002
20 or more	.563	.854	.093	.187	.006	.016	.000	.001
21 or more	.399	.633	.060	.127	.003	.010	.000	.000
22 or more	.272	.457	.038	.083	.002	.006	.000	.000
23 or more	.184	.316	.023	.050	.001	.004	.000	.000
24 or more	.119	.219	.013	.029	.000	.002	.000	.000
25 or more	.072	.142	.007	.017	.000	.001	.000	.000
26 or more	.044	.087	.003	.010	.000	.000	.000	.000
27 or more	.022	.049	.001	.004	.000	.000	.000	.000
28 or more	.011	.027	.000	.002	.000	.000	.000	.000
29 or more	.006	.014	.000	.001	.000	.000	.000	.000
30 or more	.002	.006	.000	.001	.000	.000	.000	.000

Appendix E

MMPI-2 Clinical, Content, & Supplementary Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	97.290	97.317	75.359	75.381	39.347	39.137	14.223	14.004
2 or more	89.532	89.471	50.771	50.146	17.907	17.283	4.418	4.259
3 or more	77.301	76.895	33.573	32.615	9.613	9.315	2.086	2.040
4 or more	63.402	62.479	22.985	22.039	6.006	5.914	1.225	1.242
5 or more	50.706	49.204	16.606	15.993	4.193	4.228	.823	.865
6 or more	40.496	38.936	12.709	12.434	3.134	3.215	.591	.631
7 or more	32.959	31.674	10.135	10.105	2.419	2.540	.435	.476
8 or more	27.447	26.656	8.300	8.430	1.909	2.038	.322	.363
9 or more	23.381	22.996	6.900	7.122	1.516	1.652	.240	.279
10 or more	20.207	20.144	5.771	6.057	1.207	1.338	.178	.211
11 or more	17.629	17.768	4.839	5.151	.955	1.084	.131	.156
12 or more	15.445	15.722	4.036	4.362	.748	.869	.094	.116
13 or more	13.514	13.917	3.360	3.680	.577	.686	.065	.085
14 or more	11.828	12.293	2.767	3.079	.439	.532	.046	.060
15 or more	10.258	10.802	2.256	2.553	.331	.411	.029	.043
16 or more	8.847	9.422	1.811	2.086	.242	.304	.019	.029
17 or more	7.518	8.133	1.427	1.676	.170	.224	.012	.020
18 or more	6.313	6.920	1.099	1.306	.117	.157	.007	.013
19 or more	5.212	5.809	.820	.999	.076	.106	.004	.007
20 or more	4.219	4.772	.592	.740	.049	.073	.002	.004
21 or more	3.322	3.825	.412	.529	.029	.045	.002	.002
22 or more	2.537	2.985	.272	.364	.017	.027	.001	.001
23 or more	1.866	2.233	.177	.238	.010	.015	.000	.000
24 or more	1.308	1.604	.105	.147	.004	.007	.000	.000
25 or more	.877	1.097	.060	.084	.002	.004	.000	.000
26 or more	.548	.694	.031	.045	.001	.002	.000	.000
27 or more	.312	.404	.015	.020	.000	.001	.000	.000
28 or more	.165	.207	.006	.009	.000	.000	.000	.000
29 or more	.074	.093	.002	.002	.000	.000	.000	.000
30 or more	.029	.033	.001	.001	.000	.000	.000	.000

Appendix F

MMPI-2 Clinical Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	64.916	64.188	37.152	36.434	15.839	15.454	5.044	4.907
2 or more	39.400	38.264	15.855	15.441	4.445	4.446	.882	.915
3 or more	24.208	23.784	7.656	7.790	1.655	1.782	.246	.298
4 or more	14.696	14.986	3.702	4.061	.615	.743	.066	.101
5 or more	8.437	9.065	1.666	1.985	.203	.288	.014	.031
6 or more	4.345	4.993	.641	.827	.057	.092	.002	.006
7 or more	1.924	2.351	.200	.279	.012	.019	.000	.001
8 or more	.682	.832	.050	.064	.001	.003	.000	.000
9 or more	.154	.178	.007	.008	.000	.000	.000	.000
10 or more	.014	.018	.000	.000	.000	.000	.000	.000

Appendix G

MMPI-2 Content Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	64.604	63.382	38.697	37.802	17.595	17.144	6.003	5.841
2 or more	45.003	44.070	21.056	20.735	7.164	7.074	1.737	1.740
3 or more	33.233	32.743	13.329	13.253	3.808	3.863	.772	.799
4 or more	25.247	25.113	8.967	9.078	2.255	2.333	.391	.423
5 or more	19.398	19.470	6.169	6.392	1.379	1.465	.214	.239
6 or more	14.901	15.151	4.270	4.507	.836	.931	.115	.135
7 or more	11.289	11.693	2.926	3.151	.512	.577	.061	.074
8 or more	8.403	8.848	1.967	2.155	.307	.355	.033	.042
9 or more	6.112	6.516	1.266	1.424	.173	.202	.015	.022
10 or more	4.253	4.618	.781	.895	.098	.114	.007	.009
11 or more	2.795	3.090	.456	.524	.045	.060	.003	.004
12 or more	1.672	1.899	.227	.276	.019	.027	.001	.002
13 or more	.865	1.022	.100	.127	.007	.011	.000	.001
14 or more	.348	.421	.033	.040	.002	.003	.000	.000
15 or more	.082	.101	.004	.006	.000	.000	.000	.000

Appendix H

MMPI-2 Supplementary Scales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	87.835	88.075	54.974	55.241	23.638	23.780	7.402	7.384
2 or more	65.123	64.944	25.449	25.143	6.696	6.543	1.328	1.297
3 or more	41.629	41.437	11.469	11.417	2.397	2.418	.404	.432
4 or more	23.709	23.875	5.170	5.342	.925	.998	.133	.153
5 or more	12.078	12.341	2.159	2.226	.301	.341	.033	.037
6 or more	5.269	5.298	.712	.720	.074	.076	.007	.004
7 or more	1.808	1.723	.184	.171	.016	.013	.001	.000
8 or more	.399	.349	.030	.025	.002	.001	.000	.000
9 or more	.016	.012	.000	.000	.000	.000	.000	.000
10 or more	.000	.000	.000	.000	.000	.000	.000	.000

Appendix I

MMPI-2 Harris-Lingoes Subscales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	95.036	94.745	71.861	70.780	37.097	36.028	13.093	12.669
2 or more	84.550	83.593	46.594	45.220	15.535	15.125	3.513	3.500
3 or more	70.900	69.407	29.547	28.560	7.484	7.506	1.344	1.433
4 or more	56.559	55.027	18.743	18.355	3.975	4.179	.610	.707
5 or more	43.858	42.649	12.329	12.358	2.363	2.596	.320	.404
6 or more	33.578	32.739	8.420	8.704	1.478	1.696	.178	.244
7 or more	25.788	25.379	5.933	6.306	.953	1.141	.101	.147
8 or more	19.983	19.984	4.263	4.659	.619	.783	.058	.091
9 or more	15.634	15.977	3.060	3.448	.398	.535	.032	.056
10 or more	12.196	12.774	2.173	2.563	.249	.360	.019	.033
11 or more	9.454	10.130	1.535	1.876	.154	.238	.011	.019
12 or more	7.228	7.985	1.046	1.350	.091	.154	.006	.011
13 or more	5.440	6.197	.701	.946	.053	.095	.003	.007
14 or more	4.026	4.710	.451	.646	.030	.056	.002	.004
15 or more	2.892	3.477	.276	.423	.016	.033	.001	.002
16 or more	1.993	2.494	.166	.264	.009	.017	.000	.001
17 or more	1.315	1.709	.088	.157	.004	.009	.000	.000
18 or more	.803	1.098	.042	.084	.001	.004	.000	.000
19 or more	.459	.656	.018	.040	.001	.002	.000	.000
20 or more	.226	.342	.007	.016	.000	.001	.000	.000

Appendix J

MMPI-2 Content-Component Subscales: Percentage of Normal Population with N or More Abnormal Scores at 1.0 to 2.5 Standard Deviations from the Mean

	$T \geq 60$		$T \geq 65$		$T \geq 70$		$T \geq 75$	
	Males	Females	Males	Females	Males	Females	Males	Females
1 or more	86.058	85.466	61.249	60.476	31.708	31.287	11.606	11.402
2 or more	70.555	69.682	38.341	37.819	13.468	13.456	3.046	3.069
3 or more	57.030	56.244	25.146	24.903	6.799	6.868	1.130	1.183
4 or more	45.957	45.340	17.035	17.048	3.750	3.861	.485	.546
5 or more	36.906	36.532	11.777	11.921	2.173	2.301	.236	.267
6 or more	29.545	29.480	8.205	8.447	1.300	1.415	.115	.139
7 or more	23.616	23.741	5.734	6.023	.795	.878	.061	.076
8 or more	18.783	19.014	4.030	4.299	.483	.554	.033	.042
9 or more	14.829	15.180	2.822	3.060	.295	.345	.018	.026
10 or more	11.599	12.023	1.977	2.178	.178	.219	.009	.013
11 or more	8.957	9.443	1.355	1.532	.105	.139	.005	.008
12 or more	6.875	7.322	.924	1.062	.060	.085	.003	.004
13 or more	5.196	5.609	.619	.725	.035	.050	.002	.003
14 or more	3.864	4.219	.400	.485	.020	.032	.001	.001
15 or more	2.806	3.127	.259	.319	.012	.020	.000	.001
16 or more	1.999	2.268	.161	.206	.007	.012	.000	.000
17 or more	1.377	1.598	.098	.129	.004	.007	.000	.000
18 or more	.923	1.101	.058	.077	.002	.004	.000	.000
19 or more	.587	.713	.033	.044	.001	.002	.000	.000
20 or more	.360	.443	.017	.025	.001	.001	.000	.000
21 or more	.205	.259	.008	.012	.000	.001	.000	.000
22 or more	.108	.140	.003	.006	.000	.000	.000	.000
23 or more	.050	.070	.001	.003	.000	.000	.000	.000
24 or more	.018	.030	.001	.001	.000	.000	.000	.000
25 or more	.007	.011	.000	.000	.000	.000	.000	.000

Appendix K

Comparison of Eigenvalues from Input Correlation Matrices

MALES		FEMALES	
Matrix Eigenvalues	Adjusted Matrix Eigenvalues	Matrix Eigenvalues	Adjusted Matrix Eigenvalues
33.837991	33.8299032	35.983368	35.97575694
9.4984047	9.4947321	8.7003271	8.69811364
6.3230118	6.3201723	5.9089819	5.90554622
3.9255857	3.92467574	4.0534187	4.05220827
2.9292043	2.92805401	3.2863017	3.2855038
2.8580835	2.85683352	2.8250575	2.82423596
2.4946612	2.49412982	2.4594476	2.45914899
2.3485904	2.34833767	2.2037161	2.20328407
2.0254533	2.02483279	1.8977207	1.89736141
1.7380866	1.73769269	1.6424664	1.64211362
1.5167226	1.51636901	1.3483786	1.34824646
1.4687224	1.4684399	1.2747054	1.27455176
1.262636	1.26249721	1.1507457	1.15057526
1.1441568	1.14404304	1.1177204	1.1176216
1.0100942	1.01000994	1.0330716	1.03292063
0.9949264	0.99482064	0.9816989	0.98161303
0.9141845	0.91398979	0.9421142	0.94204085
0.8568754	0.85678816	0.8952697	0.89512983
0.8414993	0.84136003	0.8797369	0.87962297
0.8213113	0.82122301	0.8155568	0.81548902
0.7965826	0.79650821	0.770283	0.7702154
0.7456114	0.74553372	0.7387974	0.73874246
0.7264607	0.72639778	0.7238647	0.72381935
0.6926428	0.69258799	0.7006733	0.70058624
0.6774284	0.67736865	0.6657724	0.66571562
0.6592314	0.65918159	0.6287403	0.62869019
0.640391	0.64035285	0.6194985	0.61947355
0.6240362	0.62398373	0.596341	0.59629022
0.6194199	0.61936183	0.5819651	0.58193511
0.5650337	0.56497767	0.5661645	0.56612228
0.5500784	0.5500384	0.5447493	0.54469459
0.54818	0.54811983	0.5294406	0.52937663
0.5208719	0.52079993	0.5009509	0.50087807
0.4892397	0.48919657	0.4826836	0.48262985
0.4758145	0.47576096	0.4636091	0.46351081

0.4711809	0.47113852	0.4570747	0.45700622
0.4501857	0.45015218	0.4437577	0.44370701
0.4397454	0.43969636	0.4113795	0.41128575
0.4132221	0.41317187	0.3958735	0.3958305
0.3977316	0.39768949	0.3806784	0.38064801
0.3861934	0.38611658	0.3721828	0.37213521
0.3737375	0.37369433	0.3521678	0.35213
0.3524893	0.35243848	0.3405509	0.34050129
0.3380114	0.33798458	0.3321675	0.33213573
0.3363016	0.33625748	0.3201713	0.32010644
0.3157593	0.31572185	0.3082749	0.30823284
0.3063908	0.30635992	0.3021719	0.30213425
0.2850842	0.2850589	0.2793171	0.27926449
0.2702225	0.27018224	0.2723902	0.27235979
0.2672869	0.26725391	0.2558262	0.25579071
0.2489605	0.24892516	0.2532326	0.25316561
0.2368055	0.23678745	0.2353543	0.23533129
0.2313227	0.23129609	0.2289972	0.22897225
0.2269954	0.22696742	0.2177673	0.21772964
0.2247251	0.22469067	0.2117899	0.2117543
0.2030105	0.20297813	0.2031598	0.20313308
0.1953891	0.19536489	0.1990291	0.19898789
0.1918771	0.19185415	0.1832164	0.18318857
0.1821499	0.182107	0.1798496	0.17978002
0.1775504	0.17752025	0.1709492	0.17094075
0.1763447	0.17632108	0.1678504	0.16783612
0.1660287	0.16597891	0.1621049	0.16206981
0.1604951	0.16044503	0.15478	0.15475779
0.1509213	0.15088754	0.1433358	0.14331698
0.1410727	0.14105423	0.1326775	0.13265817
0.1341015	0.13406129	0.1258558	0.12579096
0.1231905	0.1231758	0.1185252	0.11851388
0.1072303	0.10719141	0.1041189	0.10410104
0.1014509	0.10142537	0.093705	0.0936822
0.0951134	0.09510416	0.0912357	0.09122204
0.087752	0.08772439	0.0893301	0.08932131
0.0844196	0.0844035	0.0739821	0.07397126
0.0762531	0.07623741	0.0705148	0.07049765
0.0741677	0.07415099	0.0630791	0.0630685
0.0685617	0.0685425	0.0600058	0.05999948
0.0642009	0.06418816	0.0580435	0.05803252

0.0574509	0.05743828	0.0539992	0.05399087
0.0560613	0.05604805	0.0521331	0.05212357
0.0552097	0.05519892	0.0443997	0.04439446
0.0501461	0.05013384	0.0414662	0.04146066
0.0482692	0.04826189	0.03784	0.03783695
0.0423604	0.04234907	0.0368949	0.03688909
0.0349547	0.0349498	0.032874	0.03286882
0.0338453	0.03384013	0.0321914	0.03218553
0.0304256	0.03042089	0.0285753	0.02856713
0.0286137	0.02860398	0.0258561	0.0258495
0.0248755	0.0248602	0.02207	0.02206436
0.0205855	0.02058124	0.0214602	0.02145384
0.0183099	0.01830647	0.0207941	0.0207888
0.0142737	0.0142709	0.0132285	0.01322473
0.0122687	0.01226418	0.0115943	0.01159198
0.0098852	0.00988091	0.0085888	0.00858401
0.0076727	0.00766749	0.0065476	0.0065458
0.0035524	0.00354983	0.0007245	0.00072383
-0.003336	***0.00000001****	***-0.000627***	***0.00000001****
-0.005437	***0.00000001****	***-0.001087***	***0.00000001****
-0.005708	***0.00000001****	***-0.008614***	***0.00000001****
-0.009135	***0.00000001****	***-0.010721***	***0.00000001****

* Negative eigenvalues were set to 0.00000001 to create positive-definite correlation matrices.