

THE MMPI SUBTLE SUBSCALES:

A TEST OF VALIDITY

By

JONI LYNN HOLLRAH

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1987

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1990

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1995

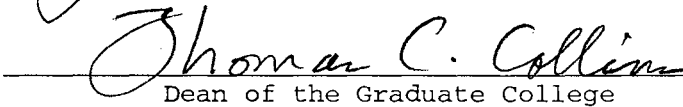
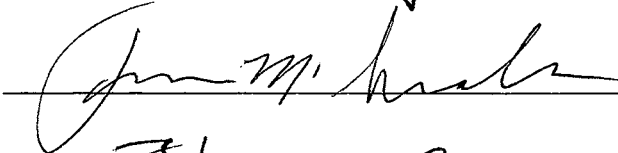
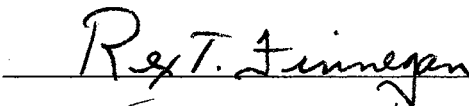
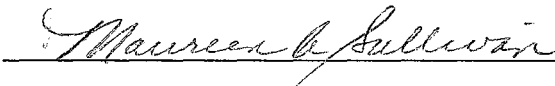
THE MMPI SUBTLE SUBSCALES:

A TEST OF VALIDITY

Dissertation Approved:



Dissertation Adviser



Dean of the Graduate College

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my major adviser, Dr. Bob Schlottmann, for his available and caring research guidance, his professional mentorship, and for a respectful, amiable, and special friendship. My appreciation also goes out to committee members Drs. Rex Finnegan, Ken Sandvold, Jim Seals, and Maureen Sullivan for their support and dedication in the process of my dissertation preparation and completion. Thanks to Dee Huggins for her professional assistance and good-natured support in the clinic during my data collection. I would also like to thank Dr. Jim Price for his intelligent statistics advice and for his support and belief in my abilities.

I express my caring and appreciation to Anne Scott, my colleague, research partner, and friend, for her faithful support and insightful, searching conversation. Thanks to my research colleague Dave Brunetti for believing in me and serving as a collegial trailblazer, and to Jolene Gordon and Maria Ast for their treasured personal and academic support. I also wish to extend my gratitude to Drs. Bob Weber and Don Fromme, and Molly Tovar, for believing in my personal and professional capacities and giving me the support I needed to initially pursue my career.

I want to express my special love and appreciation for my husband, Matt, for working with me in committed concert with love and wisdom in creating and balancing our careers and married life. I thank my aunt, (Dr.) Edie Long, who has provided me with a familial female and professional heroine with whom to identify, in conjunction with enduring unconditional emotional and spiritual support. I also want to thank my grandparents, Ed and Lucy Long, and my parents, Jim and Fran Mihura, for

securing faith in me during my arduous yet rewarding quest in pursuing my doctoral degree.

Finally, I would like to thank the Department of Psychology and the Ford Foundation for giving me the opportunity and support to make my professional career possible.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. REVIEW OF THE LITERATURE	4
Personality Test Construction:	
The Inclusion of Subtle Items	4
The Validity of the MMPI Subtle Items	9
Resistance to Faking/Paradoxical Effect	11
Convergent and Discriminative Validity	15
III. STATEMENT OF THE PROBLEM	26
IV. METHOD	29
Participants	29
Instrument	30
Procedure	30
V. RESULTS	33
VI. DISCUSSION	46
BIBLIOGRAPHY	57
APPENDIX A--VOLUNTEER FORM	62
APPENDIX B--INFORMED CONSENT FORMS	64

LIST OF TABLES

Table	Page
I. Random Assignment to Groups (Full-Scale): Mean T-Score Between-Subjects Univariate Analyses.34
II. Random Assignment to Groups (Subtle Subscales): Mean T-Score Between-Subjects Univariate Analyses.35
III. Random Assignment to Groups (Obvious Subscales): Mean T-Score Between-Subjects Univariate Analyses.36
IV. Effects of Fake-Good Instructions: Mean T-Score Within Subjects Univariate Analyses.38
V. Effects of Fake-Bad Instructions: Mean T-Score Within Subjects Univariate Analyses.39
VI. Paradoxical Effect (Fake-Good): Mean T-Score Within-Subjects Univariate Analyses.41
VII. Paradoxical Effect (Fake-Bad): Mean T-Score Within-Subjects Univariate Analyses.42
VIII. Predictive Validity of Fake-Good MMPI Scores: Mean T-Score Within-Subjects Correlations.44
IX. Predictive Validity of Fake-Bad MMPI Scores: Mean T-Score Within-Subjects Correlations.45

CHAPTER I

INTRODUCTION

The development of objective personality inventories has enhanced the efficiency and reliability of personality assessment. Inclusion of consistent scoring systems and use of normative standardization samples in objective personality tests has facilitated communication between psychologists about the personality variables of individuals. The self-report nature of the tests provides a sample of behavior for assessment purposes that broadens the scope of information when added to clinical interviews.

Various objective personality tests, employing different methods of test construction, have been developed and used in the past several decades. One of the most widely used objective personality tests (Lubin, Larsen, & Matarazzo, 1984) is the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1942). The MMPI is comprised of ten clinical scales and several validity scales which combine to form a profile that can provide valuable information about the personality characteristics of the respondent. The original MMPI has since been revised and was published in 1989 as the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer).

The MMPI was constructed using an empirical method whereby the decision to place items on particular scales was based on their statistical ability to discriminate between clinical and normal groups. This method of test construction led to the empirical inclusion of many items whose manifest content is not obviously related to psychopathology (i.e., *subtle items*). MMPI items have since been systematically rated into categories of subtle and obvious items (e.g., Wiener & Harmon, 1946). The subtle items hypothetically should provide the means to unobtrusively measure the characteristics of their scales. Ideally, if

a respondent were attempting to distort responses on the MMPI, these subtle items should still assess the desired scale dimension.

Although no MMPI interpretive manuals advise using the subtle subscale scores as unobtrusive measures of their scales' characteristics, this procedure is occasionally carried out in the clinical practice of psychological assessment. Especially when a respondent's traditional validity scales indicate an invalid profile due to a distorted response set, clinicians can be tempted to use the subtle subscale scores to provide a glimpse of the respondent's "true" scale score. Research has yet to validate this procedure, however, which will be a primary focus of the present study.

Indirect support for the subtle items' validity as unobtrusive measures, is found in their resistance to faking attempts. In other words, they are not usually manipulatable in the desired direction of distortion by the respondent. In fact, attempts to distort responses often lead to a *paradoxical effect* of the subtle items (e.g., Harvey & Sippelle, 1976; Wales & Seeman, 1968) whereby the obvious subscale scores change markedly in the expected direction and the subtle subscale scores show a minimal change in the opposite, unintended direction.

Although their resistance to faking offers support for the subtlety of these items' relationship to psychopathology, researchers have questioned the validity of the subtle items as measures of their scale's characteristics (Jackson, 1971). Research has not offered much consistent evidence for the validity of the subtle items except in their ability to aid in the detection of distorted profiles, especially fake-good profiles in psychiatric settings (e.g., Grow, McVaugh, & Eno, 1980; Kelly & Greene, 1989; Woychyshyn, McElheran, & Romney, 1992). Furthermore, these detected distorted profiles are typically discarded as "invalid" profiles. Therefore, although the subtle items may help detect invalid profiles, their ability to measure the characteristics of their scale has not yet been verified. There is even some evidence that

the subtle items may simply add random noise to their scales (Weed, Ben-Porath, & Butcher, 1990). It has been suggested that the empirical method used in developing the MMPI selected subtle items by sampling error because the scales were not cross-validated (e.g., Graham, 1990).

The purpose of the present study is to investigate the validity of the MMPI subtle items. Specifically, the goal is to determine if, when a distorted response set is used when taking the MMPI, the subtle subscale scores can predict honestly-taken, full-scale scores. In addition to investigating the validity of the subtle items, the present study may help determine whether the MMPI has any clinical usefulness when deviant test-taking attitudes would normally render the test invalid.

CHAPTER II

REVIEW OF THE LITERATURE

Personality Test Construction:

The Inclusion of Subtle Items

The prototype for personality inventories was first developed during World War I for use in screening draftees for symptoms that signaled unfitness for war. This prototype was The Woodworth Personal Data Sheet (Woodworth, 1920), which included a single measure of "neuroticism" based on 116 heterogeneous items. Personality inventories that were subsequently developed (e.g., the Bernreuter Personality Inventory, 1935) included additional scales with better item homogeneity which were constructed by a rational selection of items. The rational method of test construction is accomplished by rationally selecting items for scales based on a theory underlying the particular scale's construct. In all of the personality inventories in the 1920's and 1930's, the content of the items were "face-valid," which means they had a clear, obvious relationship to the intended constructs or symptoms to be assessed. In other words, the respondent could easily identify the constructs being measured by the item to which he or she was responding.

It soon became apparent that these early personality inventories resulted in an excess of misclassifications and, therefore, were not adequate measures of the intended constructs (e.g., Hathaway, 1965). Part of the problem with the inventories seemed to be with the face-valid items. Given any conscious or unconscious motivation to deceive, respondents could easily manipulate their scores because of the test items' obvious relationship to the constructs or symptoms assessed by their scales. Researchers concluded that, in order to remedy this problem, test items were needed without an obvious relationship to the characteristics measured on each scale. One attempt to determine such items was made by using the empirical method of item selection.

The MMPI (Hathaway & McKinley, 1942) was the first personality inventory with clinical scales constructed entirely by the empirical method. Initially, a large pool of statements was administered in a true-false format to normal samples (i.e., those not seeking psychiatric or psychological services) and psychiatric samples. Items' selection for scales was based on their *statistical* ability to discriminate between the normal and psychiatric groups. As a result, some items were selected empirically whose manifest content bore no obvious relationship to the characteristics measured by their scales. These items were thereafter labeled subtle items (e.g., Wiener & Harmon, 1946).

The MMPI initially included eight clinical scales and then added two more before the final product was published by Hathaway and McKinley in 1942. The original eight clinical scale descriptors used for the MMPI were based on the current nosology of psychopathology of the 1930's. These scales were: Hypochondriasis (*Hs*), Depression (*D*), Hysteria (*Hy*), Psychopathic Deviate (*Pd*), Paranoia (*Pa*), Psychasthenia (*Pt*), Schizophrenia (*Sc*), and Hypomania (*Ma*). Later, two additional clinical scales were developed using slightly different approaches. The Masculinity-Femininity (*Mf*) scale was developed with the initial intent to discriminate between males with homosexual and heterosexual orientations. However, because of difficulty obtaining an adequate sample of homosexual males, items were first chosen based on their frequency of endorsement by males and females. A scale refinement was later made by comparing a criterion group of men scoring high on the Terman and Miles masculinity-femininity test (Terman & Miles, 1936) and the *Mf* scale was subsequently included in the inventory. The tenth MMPI clinical scale (Social Introversion; *Si*) was developed by Drake (1946) using an empirical method that discriminated between introversion and extroversion in a female college sample. The *Si* scale has since been used with both male and female populations.

In 1945, Meehl published an article on structured personality tests that advocated the empirical method of item selection on the MMPI that results in subtle items. He suggested that rationally assigning items to scales resulted in complications, which were: (a) the possibility of conscious or unconscious distortion of responses and (b) that all verbal statements have the potential for different interpretations of meaning. The point he made was that, yes, people with particular personality traits do tend to say certain things about themselves. However, one cannot always take these statements at face-value. For example, the items "I am sure I get a raw deal from life (answered *true*)" and "I have been quite independent and free from family rule (answered *false*)" on the *Pd* scale are endorsed more often by those in the Psychopathic group than normals. However, one certainly should not assume that these responses are in fact true more often for psychopaths than for normals. Therefore, a method of test construction that differentiates groups based on an empirical selection of items to scales is more appropriate than one that uses a rational method of selection.

Although the outcome of the empirical method of test construction used on the MMPI was the inclusion of some items with a non-obvious relationship to psychopathology, the classification of *subtle* and *obvious* items did not come until later. Wiener and Harmon (1946; Wiener, 1948) personally rated MMPI items according to their subtle/obvious (S/O) relationship to general psychopathology. To determine which items were to be scored as subtle or obvious, Wiener employed several criteria. Initially, all items on the *F* scale were labeled as obvious. The remaining items were judged as subtle or obvious by Wiener and Harmon. This method of selection resulted in a sufficient number of subtle items on the following five scales: *D*, *Hy*, *Pd*, *Pa*, and *Ma*. *T*-scores exist for females and males (e.g., Greene, 1991).

Wiener's S/O item classifications has since been criticized for not including the inter-rater reliability results, only using two raters, the lack of information on the validity of their classifications (e.g., Dubinsky, Gamble, & Rogers, 1985) and their ratings of general psychopathology instead of scale-specific ratings (Ward, 1986). Additional attempts have been made to establish S/O item pools using different methods of classification. Four S/O rating studies which include the methodology for their S/O classifications have been published. It is important to delineate some distinguishing characteristics of the different methodologies which were employed. These distinctions have relevance to the particular definition of what is meant by "subtle" items.

In some studies (Wales & Seeman, 1969) the S/O distinction is made by statistical assignment (see Dahlstrom, Welsh, & Dahlstrom, 1972). Items endorsed predominantly in opposite directions (true/false) by normal and psychiatric groups are defined as obvious and have been labeled X items. Those items endorsed frequently in the same direction by both groups, but by a larger percentage of the psychiatric group, are defined as subtle and labeled 0 (or Zero) items.

In another instance, Duff (1965) determined the S/O distinction on the *Hy*, *Pd*, and *Sc* scales by using the ratings of psychology graduate students. Additional items from other clinical scales and items not on any clinical scales were also included as foils. Items were labeled obvious if judges could correctly identify the corresponding scale and the pathological response (true or false). Cutoff percentages of correct identification for classification of items were: (a) 10 percent or less = subtle, (b) 50 percent or greater = obvious, and (c) between 10 and 50 percent = intermediate subtle. Only *Hy* and *Pd* acquired enough subtle items to be of use (45% and 46%, respectively, compared to 6% of *Sc* items).

Christian, Burkhart, and Gynther (1978) gathered S/O ratings for all the MMPI items using undergraduate students as raters. Items were rated according to how indicative they were of a "psychological problem." A five-point scale was used with a range of *very obvious* (5) to *very subtle* (1), with *neither obvious or subtle* (3) employed as the mid-point. The mid-point classification has been labeled "neutral," and these Christian et al. subtle, neutral, and obvious ratings have been used in many studies.

However, Hryckowian and Gynther (1988) investigated the neutral classification and found that the phrase "neither obvious or subtle" may not have been interpreted in the manner originally intended by Christian et al. (1978). Focusing on the *D* scale, participants rated different definitions of the mid-point "neither subtle or obvious" and 58% interpreted it to mean "neither subtle or obvious *and therefore totally unrelated to depression.*" Therefore, many of the Christian et al. neutral items may more closely fit the subtle classification, which reduces reliability and validity in studies using these S/O classifications.

Of the studies that devised unique methods of S/O classifications, only Duff's (1965) defined the S/O ratings as the items' ability to measure the characteristics assessed by their particular scale. Duff instructed doctoral students in psychology to judge items for the particular scale to which they belonged, along with their keyed pathological response. The Wiener (1948) and Christian et al. (1978) S/O item classifications were judged with respect to general psychopathology, not the dimension of psychopathology measured by the respective scale. Wiener's study used clinicians as raters, and the Christian et al. study used naive raters who would more closely approximate the knowledge of most respondents. The *X* and *O* S/O items were not rated by judges, but labeled subtle or obvious because of their differential endorsement rate by normal and psychiatric groups.

Therefore, the term "subtle" may be used in different contexts across studies. One major difference of these ratings is whether subtle means "not obvious in its relationship to the characteristics measured by the scale" or "not obvious in its relationship to general psychopathology."

The most popular S/O ratings in the research literature are the Christian et al. (1978), Wiener-Harmon (1946, Wiener, 1948), and X and O (e.g., Wales & Seeman, 1969) S/O ratings. Other studies (McCall, 1958; Snyder & Graham, 1984) used less sophisticated or unpublished methods, and, like Duff's (1965), have only been used in their respective articles. The Wiener-Harmon S/O ratings are popularly used in clinical settings and in the current scoring systems of the MMPI (e.g., National Computer Systems).

In summary, early personality inventories included items based on their rational, obvious relationship to psychopathology. These items were easily manipulated by respondents and resulted in an excess of misclassifications. Test items whose manifest content was not obviously related to their scales' characteristics were determined by the empirical method of item selection used on the MMPI. These items have since been labeled subtle items and many researchers have systematically made this subtle/obvious distinction by slightly different methods.

The Validity of the MMPI Subtle Items

In 1971, Jackson offered a reply to Meehl's 1945 article on the advantages of the empirical method of test construction that results in subtle items. One of his main objections to the empirical method is that the selection of items for scales is based on the post-hoc statistical ability of items to discriminate between groups. He states that items should be based on relevant theoretical definitions of the trait measured by their scale, else the test does not have construct validity. His contention is not against the inclusion of subtle items,

but rather against a method for inclusion that is post-hoc. Jackson proposed that a more clinically useful method for item selection is to devise item pools for scales that include some items with a subtle relationship to the trait that nevertheless have a sound theoretical basis.

Jackson extended a challenge to the current empirical method of test construction, suggesting that this method "be pitted against two hours of work by a couple of good item writers...an introductory class of psychology students..." (pp. 237-238). He advocated a method of test construction that used a rational method of item selection, with an internal consistency approach to determine each scale's construct validity. Ashton and Goldberg (1973) responded to this challenge by comparing scales written by psychology students to scales measuring the same constructs on an empirically constructed test (i.e., the California Psychological Inventory, CPI; Gough, 1969). Their results showed that the three scales (Sociability, Achievement, and Dominance) written by psychology students were superior in reliability and validity over the scales which measured the same three constructs on the CPI.

Although Ashton and Goldberg's study was conducted with college students and more "normal" personality measures, it did extend a challenge to the empirical method of test construction. Most studies on MMPI subtle item validity began around this time, and many have been unfavorable. Some researchers propose that the subtle items do not assess the characteristics measured by their respective scales. For example, Jackson (1971) suggests that their inclusion is merely due to sampling error. He supported this hypothesis with Wiener's (1948) finding that the Wiener-Harmon subtle subscales correlated negatively with their respective scale's obvious subscales. However, this negative correlation was found only with scales *D* and *Hy*. Wiener also noted that the obvious items have a positively skewed distribution and the subtle items have a relatively normal distribution. Relatively few normals

responded to the obvious items in a scorable (pathological) direction, producing a skewed distribution. It has been suggested that the subtle items produce a normal distribution because they are more idiosyncratic and are unrelated to the dimension measured by each scale. Graham (1990) speculates that the items were only included because Hathaway and McKinley (1942) failed to cross-validate their item analysis.

Resistance to Faking/Paradoxical Effect

Many studies have supported the subtle items' resistance to faking (e.g., Wales & Seeman, 1968; 1969; 1972). Although resistance to faking is a desirable quality of the subtle items, it has been suggested that the subtle items are inappropriately keyed and are instead a general measure of non-pathological adjustment (e.g., Wales & Seeman, 1969). Some support for this is the fact that the direction for keying of 65 of the 110 Wiener-Harmon subtle items was opposite of the direction expected by Hathaway and McKinley (Wiener, 1948). Further justification is based on the paradoxical effect of the subtle items. The paradox occurs when MMPI respondents are instructed to fake. Their obvious subscale scores change considerably in the intended direction (i.e., fake-good or fake-bad), while the subtle subscale scores change in the opposite direction. For example, if a respondent were attempting to respond in a pathological manner, their obvious subscale scores would increase, but their subtle subscale scores would decrease. In these studies, the magnitude of change in the subtle subscale scores is also considerably less than that of the obvious subscale scores.

Three Wales and Seeman (1968, 1969, 1972) studies found this paradoxical effect using the statistical assignment definition of item subtlety (X and O items). College students, psychiatric inpatients, and nurses, took the MMPI once under standard instructions and once under either fake-good or ideal-self instructions. The paradoxical effect occurred with all three populations. Anthony (1971) found the

paradoxical effect using the Wiener-Harmon S/O subscales with military outpatients under honest and fake-bad instructions. Rice, Arnold, and Tate (1983) found the paradoxical effect for forensic psychiatric patients when faking bad, but not when faking good.

Two other studies (Gloye & Zimmerman, 1967; Hiner, Ogren, & Baxter, 1969) demonstrated the paradoxical effect using the Wiener-Harmon S/O ratings with college students under standard and ideal-self instructions. Burkhart, Christian, and Gynther (1978) and Dannenbaum and Lanyon (1993) used the Christian et al. S/O ratings, and Harvey and Sipprelle (1976) and Timbrook, Graham, Keiller and Watts (1993) used the Wiener-Harmon S/O ratings. Using college students, the first three of these four studies found the paradoxical effect under both fake-good and fake-bad instructional sets. Timbrook et al. did not find the paradoxical effect for the fake-bad condition.

Investigating specific scales, Peterson, Clark, and Bennett (1989) found the paradoxical effect for college students only for scale *Ma* in the fake-good group and for scales *D* and *Hy* in the fake-bad group. However, when using psychiatric outpatients, Vesprani and Seeman (1974) found the paradoxical effect for all five scales when using the *X* and *O* items and fake-good ("ideal-self") instructions.

Timbrook et al. (1993) propose that the paradoxical effect under fake-good instructions is accounted for by a simple nay-saying bias. In the fake-good condition, subtle items keyed false were answered false more frequently, and subtle items keyed true were answered true less frequently, compared to the standard condition. In other words, there was a tendency to respond false to the subtle items when faking good, regardless of their keyed direction. No corresponding bias occurred for the obvious items. These researchers suggest that the subtle subscale score increase under fake-good conditions occurs because of a tendency to answer false, or a nay-saying bias, for subtle items. This is not a general nay-saying bias when faking good, but limited to the subtle

items. How this occurs in the context of test taking is not explained. Perhaps when the psychopathological content of an item is vague or ambiguous, if faking good, then the bias would be to answer false.

In comparison, Dannenbaum and Lanyon (1993) suggest that the paradoxical effect occurs because respondents are answering items according to the misleading content, or face-validity, of these items when faking. *Deceptive* subtle items, those with misleading face-validity, were defined as the Christian et al. subtle items whose mean psychopathology rating in the keyed (pathological) direction fell in the low subtle range, and the mean rating in the non-keyed direction was higher. In other words, subtle items whose non-keyed direction was interpreted as more pathological than the keyed direction were considered deceptive or as having misleading face-validity. Seventy-three of the 100 Christian et al. subtle items qualified as deceptive; those not meeting these criteria were labeled *other* subtle. Their results showed that respondents attempting to fake-bad endorsed less deceptive subtle items and more other subtle items than those in the standard and fake-good groups. The standard and fake-good groups endorsed more deceptive subtle items than other subtle items.

In accordance with Wiener's (1948) observation, most of the subtle items were determined as having misleading face-validity; not only are they subtle in their general relationship to psychopathology, but they are interpreted as more pathological in their non-keyed direction. The previous results suggest that the paradoxical effect occurs because respondents attempt to fake in the desired direction, but are misled by manifest content or face-validity. Although the Timbrook et al. (1993) results suggest that the fake-good paradoxical effect is due to a nay-saying bias, perhaps the majority of the subtle items' non-pathological face-validity is in the false direction, a possibility which needs to be investigated by future research.

A study by Burkhart, Gynther, and Christian (1978) showed a tendency for psychologically-minded respondents to endorse more subtle items. Using college students and the Christian et al S/O items, high scorers on the California Psychological Inventory (CPI; Gough, 1975) Psychological Mindedness scale (Py) endorsed more MMPI subtle items than low scorers on the Py. However, this relationship did not occur under fake-good or fake-bad conditions.

Greene (1991) also reported that the MMPI-2 normative group had higher scores on the subtle scales compared to the original MMPI sample (except for scale D). In fact, the higher subtle subscale scores are the main reason the MMPI-2 normative sample achieved higher full-scale scores than the original MMPI sample; excepting Ma-O, the obvious subscale scores were virtually identical in these two groups. He suggested that the higher subtle subscale scores were due to the higher educational and occupational level of the MMPI-2 sample.

Although not conclusive, the findings of these last three studies (Burkhart et al., 1978; Dannenbaum & Lanyon, 1993; Greene, 1991) are consistent with the Wales and Seeman (1969) assertion that the subtle items may be a measure of social perceptiveness. However, if answering purely on the grounds of social desirability, a high subtle subscale score may be a sign of faking good (as suggested by the paradoxical effect).

In conclusion, the subtle items have an abundance of support for their resistance to faking and instead show a paradoxical effect under faking instructions. The paradoxical effect itself does not discredit the subtle items as measures of their scales' characteristics. The fact that the subtle subscale scores change opposite of the intended direction under faking attempts is completely consistent with the rationale to include subtle items on a personality test. This paradoxical change could help offset intended distortion in the easily manipulated obvious items. Although this correction function of the

subtle items would be beneficial, this benefit should not come at the cost of reducing scale validity.

Convergent and Discriminative Validity

Another approach used to determine if these subtle items are related to the personality dimensions measured on their respective scales is to assess their relationship to some criterion measured by their specific scale's characteristics. The following is a review of studies investigating the convergent and discriminative validity of the various classification types of S/O subscales. For the purposes of this paper, and because of the problems with the neutral classification (Hryckowian & Gynther, 1988), only the subtle and obvious subscale results will be presented for those studies using the Christian et al. S/O ratings.

One method evaluating the validity of the subtle items assesses their relationship to some external measure of their scales' characteristics. This method determines their convergent validity. Gynther, Burkhart, and Hovanitz (1979) administered the MMPI to college students along with a social nonconformity questionnaire (SNC; developed by the authors) in an attempt to determine the validity of the *Pd* subscales. The SNC consists of items assessing the occurrence of acts that break societal norms, laws, and regulations. For males, both *Pd-O* and *Pd-S* (*Pd-Full* was not analyzed) were related to the SNC. However, for females, only *Pd-O* was related to the SNC. In an analysis using semi-partial and multiple correlations, *Pd-O* was the strongest predictor for both males and females. For males only, *Pd-S* contributed a significant amount of unique variance to the prediction of scores on the SNC.

Worthington and Schlottmann (1986) also used the Christian et al. S/O ratings for the *Pd* scale and the SNC as the criterion. Male college students responded to the MMPI items under honest, fake-good (i.e., as

if "applying for a very desirable job"), and fake-bad (i.e., appearing "maladjusted") instructions. Under honest conditions, *Pd-S*, and not *Pd-O*, correlated with the SNC. In the faking conditions, all scales (*Pd-Full*, *Pd-O* and *Pd-S*) were successfully manipulated; curiously, *Pd-S* was not resistant to faking. None of the *Pd* scales provided any predictive ability under faking conditions.

Their study does not support the *Pd-S* subscale resistance to manipulation nor the paradoxical effect as has been shown for the subtle subscales as a whole (e.g., Burkhardt et al., 1978). Also, these results do not support the *Pd-S* ability to predict a criterion under faking conditions, which suggests the *Pd* Christian et al. subtle items are not effective unobtrusive measures of their scale. Due to the fact that participants were instructed to fake according to general psychopathology, these results also suggest that the Christian et al. ratings for scale *Pd* may not be so subtle. However, the fact that *Pd-S* predicted the criterion under *honest* conditions suggests it is related to the characteristics measured by its scale.

Snyter and Graham (1984) investigated the convergent validity of the *Pd* and *Ma S/O* subscales using male and female college students. The *S/O* scale and the corresponding external measures were: *Pd* and the SNC, and *Ma* and the Activity-level Biographical Questionnaire (ABQ; Hovanitz & Gynther, 1980) and the four factors of Zuckerman's Sensation Seeking Scale (SSS; 1977): Disinhibition, Thrill and Adventure Seeking, Experience Seeking, and Boredom Susceptibility. One major purpose of this study was to investigate the validity of *S/O* items derived by scale-specific ratings instead of those made for general psychopathology. The scale-specific ratings were taken from an unpublished study (Harper, cited in Snyter & Graham) and methodological details were not included. Using these ratings, *Pd-Full*, *Pd-O*, but not *Pd-S*, correlated with the SNC. *Ma-Full* correlated only with the ABQ,

Ma-S correlated with Thrill and Adventure Seeking and Boredom Susceptibility, and Ma-O correlated with none of its criteria.

The predictive validity of the Ma scale was also investigated by Hovanitz and Gynther (1980) using a male college student sample and the Christian et al. S/O ratings. The external measures were: the ABQ, the Porteus Maze Test (PMT; Porteus, 1965) as a performance measure of hypomania, and the four factors of Zuckerman's SSS. Ma-Full, Ma-O, and Ma-S each correlated with two of the six measures: Ma-Full with Experience Seeking and the Disinhibition Scale, Ma-O with Experience Seeking and the PMT, and Ma-S with Thrill and Adventure Seeking and the ABQ. Using canonical correlation analysis, Ma-O was a better definition of the variate than Ma-S. However, Gynther and Burkhart (1983) performed an additional item analyses of this data and noted that although Ma-O emerged as the best predictor, four of the five most predictive items were from the Ma-S scale.

McCall (1958) personally rated *D* scale items into categories of "face valid," "congruent," and "irrelevant." Face valid items obviously describe depressive symptomatology, congruent items apparently fit with theories of depression, and irrelevant items do not appear to be related to depression. Therefore, these ratings might fall into the following categories: (a) face valid--obvious to expert and naive raters, (b) congruent--somewhat obvious to expert raters, but subtle to naive raters, and (c) irrelevant--subtle to expert and naive raters. Of these three groups, only the face valid items significantly discriminated samples with and without depression. Therefore, this study does not support the discriminative validity of McCall's type of subtle *D* items.

Burkhart, Gynther, and Fromuth (1980) investigated the convergent validity of the *D* Christian et al. S/O subscales using three different criteria. College students completed the MMPI, the Depression Scale on the Profile of Mood States (POMS-D; McNair, Lorr, & Droppleman, 1971), the Beck Depression Inventory (BDI; Beck, 1967), and an abbreviated

Pleasant Events Schedule (PES; MacPhillamy & Lewinsohn, 1974). *D-Full* and *D-O* were positively correlated with the BDI, POMS-D, and PES, and *D-S* was negatively correlated with the BDI and POMS-D and not related to the PES.

Burgess, Campbell, and Zylberberg (1984) similarly found no convergent validity for *D-S*. Male cardiac patients were administered the MMPI, BDI, and the Hamilton Rating Scale for Depression (HRS; Hamilton, 1960). The HRS is designed to be rated by two clinical psychologists for reliability; however, there was no information in their study concerning the number of raters or reliability of the ratings. *D-Full* and *D-O* correlated positively with the BDI and HRS, and *D-S* was negatively related to the BDI and HRS. As before, *D-O* emerged with a higher correlation with criteria than *D-Full*. The authors suggest that the subtle items are invalid due to a lack of theoretical relevance.

Turner and Romano (1984) investigated the validity of scale *D* using the Wiener-Harmon S/O ratings and a chronic pain sample with and without depression. For convergent validity, they found that the BDI, the Zung Self-Rating Depression Scale (Zung, 1965), and diagnosis of major depression (made blind to the self-report results) correlated with *D-Full* and *D-O*, but not *D-S*. This study did not find a negative correlation for the *D-S* scale as has earlier studies investigating the *D* scale. For discriminative validity, these researchers found *D-Full* to have the best overall classification rate (75%), with sensitivity (ability to detect truly depressed patients) and specificity (ability to detect truly non-depressed patients) also both found to be 75%. Overall classification rates for *D-S* and *D-O* were very similar to each other (65% and 63%, respectively). *D-S* had very good specificity (86%), but very poor sensitivity (17%); *D-O* had good sensitivity (75%) and poor specificity (57%). Therefore, although the convergent validity of *D-S* was not supported, the finding that *D-Full* was a better classification

tool than *D-O* lends support to the inclusion of the *D-S* items on their scale.

Nelson and Cicchetti (1991) also investigated the convergent and discriminative validity of the *D S/O* subscales, using the Wiener-Harmon ratings and psychiatric outpatients with and without major depression or dysthymia. They used the BDI and the Depression scale on the Symptoms Check List-90-Revised (Derogatis, 1983) as criteria. As in all the previous *D S/O* studies, they found *D-Full* and *D-O*, but not *D-S* to correlate with these measures. Although the correlation of *D-O* with the criteria was greater than *D-Full*, statistical differences were not calculated. In terms of classification rates for psychiatric versus normal groups, *D-Full* had an overall hit rate of 77%, with a sensitivity of 78% and specificity of 75%. These rates were not calculated for *D-O* and *D-S* to determine whether the inclusion of *D-S* items improved classification rates as in Turner and Romano (1984).

Hovanitz, Gynther, and Marks (1983) investigated the convergent validity of the *Pa S/O* subscales using the Christian et al. *S/O* ratings and a male college student sample. The external measures were Mehrabian's Stimulus Screening Questionnaire (SSQ; 1977), Rotter's Interpersonal Trust Scale (IPT; 1967), the Einstellung Test (Luchins, 1951), a paranoia questionnaire (PAQ), and an unobtrusive measure of ideas of reference (i.e., answers that participants "personalized;" labeled PERS). The last two measures were developed by the authors. The following subscales and measures were significantly correlated in the expected direction: *Pa-Full* with the PAQ and Mehrabian's SSQ, *Pa-O* with the IPT, PAQ, and Mehrabian's SSQ, and *Pa-S* with PERS. The Einstellung Test was not related to any *Pa* subscales and the IPT correlated in the wrong direction with *Pa-S*. These *Pa* scale results are similar to the Hovanitz and Gynther (1980) results with the *Ma* scale in that various *S/O* subscales are related to different external measures.

Hovanitz et al. (1983) suggest possible problems and also important factors to note about many S/O studies. First, there is a problem with the criteria used for prediction, especially for the *Ma* and *Pa* scales. The predictive validity of the subtle and obvious subscales appears to be dependent on which criteria are used for prediction. Secondly, they suggest that studies using college students samples be replicated with a psychiatric population because of the low scores obtained on the MMPI with a normal population. And lastly, they point out the problems with the Christian et al. (1978) S/O ratings in that the subtle-obvious distinctions are rated for general psychopathology and not for specific scale characteristics.

Duff (1965) investigated the discriminative validity of the *Hy* and *Pd* S/O subscales using the scale-specific ratings of advanced psychology graduate students and an item discrimination index similar to that used in the original validation of the MMPI. The groups used for discrimination purposes were a normal group, taken from samples used in standardizing the MMPI, and two psychiatric inpatient groups, with diagnoses corresponding to scales *Hy* and *Pd*. These psychiatric groups were, respectively: a) conversion hysteria; psychoneurosis, hysteria; or psychoneurosis, mixed, and b) psychopathic personality. About 40% of the subtle items discriminated between normal and psychiatric groups, compared to over 90% of the obvious items. Duff concluded that manifest content is an important factor and that the more subtle items should be eliminated from the MMPI. This study was not designed to determine if the subtle items could discriminate, but to determine whether more subtle items or more obvious items could discriminate.

Hovanitz, Gynther, and Green (1985) investigated the discriminative validity of the *Pa* and *Ma* S/O subscales with the Christian et al. S/O ratings. They administered the MMPI and several of the previously cited questionnaires to a male college student sample. The scales and corresponding measures were: (1) the *Pa* scale with

Rotter's IPT, Mehrabian's SSQ, the Einstellung Test, the PERS, and the PAQ, and (2) the *Ma* scale with the PMT, ABQ, and the four factors of Zuckerman's SSS. The full *Pa* and *Ma* scales discriminated between these external measures better than the subtle and obvious *Pa* and *Ma* subscales did. Furthermore, although the obvious subscales correlated with their appropriate external measures more highly than the subtle subscales correlated with their external measures, the obvious subscales proved to be the worst discriminators. Therefore, concerning the discriminative validity of *Pa* and *Ma* S/O subscales, this study found more support for the subtle subscales than the obvious subscales, but found the greatest support for the full scales.

Wrobel and Lachar (1982) investigated multiple MMPI S/O subscales. They used two psychiatric samples and the Wiener-Harmon S/O ratings. For one sample, they took criteria from a symptom checklist completed by one rater; they obtained criteria for the second sample by extracting symptoms from the patients' medical charts (two raters, 91% agreement rate reported). Both of these rating data were reduced by an iterative principal-axis factor analysis and the resulting four factors were labeled *Discomfort*, *Somatization*, *Sociopathy*, and *Reality Distortion*. The authors did not clearly state which factor they intended to correspond to which scale. From their comments and discussion, apparently the factors and corresponding scales were: *Discomfort* and *D*, *Somatization* and *Hy*, *Reality Distortion* and *Pa*, and *Sociopathy* and *Ma*.

The following are the four factors and the MMPI scales with which they correlated: *Discomfort* (*D*-Full, *D*-O, *Hy*-Full, *Hy*-O, *Pd*-Full, *Pd*-O, *Pa*-Full, *Pa*-O, & *Ma*-O), *Somatization* (*Hy*-Full & *Hy*-O), *Reality Distortion* (*Pa*-Full & *Pa*-O), and *Sociopathy* (*Ma*-Full, *Ma*-O, & *Ma*-S). The results of this study support the convergent validity of all the obvious subscales. Of the subtle subscales, only *Ma*-S showed convergent validity. However, the content of three factors appear to be less than optimal scale descriptors (i.e., *Discomfort*, *Reality Distortion*, and

Sociopathy). The Discomfort factor proved to be a particularly poor discriminator, correlating with all five obvious subscales. The factor corresponding to scale *Hy*, Somatization, appears to be the most appropriate. Wrobel and Lachar concluded that the subtle items do not possess content or empirical validity, whereas the obvious items do.

With these inconsistent validity findings, the question arises as to whether the actual removal of the MMPI subtle items might in fact increase the validity of the instrument. A study by Hovanitz and Jordan-Brown (1986) investigated this possibility and also the subtle subscales' discriminant and convergent validity. Another purpose of their study was to use the Christian et al. S/O ratings with a psychiatric population and include criteria intended to be more relevant to MMPI validity. The S/O subscales investigated were *D*, *Pd*, *Pa*, and *Ma*. The MMPI was administered to psychiatric inpatients and the criteria used were: a mental status measure (i.e., the Brief Psychiatric Rating Scale; BPRS, Overall & Gorham, 1962), diagnosis (made blind to MMPI results), psychotropic medication, and four questionnaires. These same four questionnaires were those used in many of the previous S/O validity studies using college student samples. These questionnaires and their related MMPI scales were: 1) the BDI and scale *D*, 2) Zuckerman's SSS and scale *Ma*, 3) Rotter's IPT and scale *Pa*, and 4) the SNC and scale *Pd*.

They found that only the full scales and obvious subscales were significantly related to all of the scale-appropriate questionnaires. The relationships obtained with the scale-appropriate questionnaires for *D-S* and *Pa-S* were opposite of the predicted direction. However, all of the obvious subscales correlated at least as strongly with questionnaires other than the one intended for their scale. *Ma-O* actually correlated more strongly with all the questionnaires intended for other scales. *Pd-O* correlated more strongly with the questionnaires for scale *Pa* and *D* than the one for its scale. And the SNC used in

previous *Pd* convergent validity studies, correlated more strongly with the *Ma* scale than with *Pd*. For the mental status exam, *D-Full*, *D-O*, *Pa-Full*, *Pa-O*, and *Ma-S* correlated with the scale-appropriate measure. *Pd-Full* correlated with the *Ma* measure and *Ma-O* correlated with the *Pa* measure.

The *D* and *Ma S/O* subscales were also analyzed for their relationship with diagnostic criteria. *D-Full*, *Ma-Full*, and both *D-S* and *Ma-S* successfully differentiated diagnostic classifications. Neither *D-O* nor *Ma-O* discriminated successfully. The *D*, *Pa*, and *Ma S/O* subscales, analyzed for differentiating among classes of drugs prescribed to patients, produced success with *D-Full*, *D-S*, and *Ma-Full*. The obvious subscales (*D-O*, *Pa-O*, and *Ma-O*) did not differentiate between any of the diagnostic and drug classifications. In all cases, the full scales and subtle subscales discriminated better than the obvious subscales.

The data were also analyzed for the effects of removing the subtle items from their scales. For the mental status exam and questionnaires, the removal of the *D* subtle items slightly improved prediction, whereas the removal of the *Pd*, *Pa*, and *Ma* subtle items had no effect. For diagnosis, only the *D*, *Pa*, and *Ma* scales were analyzed; removal of the subtle items resulted in a loss of prediction for *D* and *Ma*, and had no effect for *Pa*. For medication, only scale *D* and *Ma* were analyzed; removal of the *D* subtle items again resulted in a loss of prediction, and had no effect for *Ma*.

Therefore, subtle item removal only improved scale validity for *D* when predicting the questionnaires and mental status exam. Conversely, when criteria were the more expedient measures of diagnosis and medication, removing the *D* subtle items reduced scale validity. Removing the *Ma* subtle items also reduced scale validity when the criterion was diagnosis. For all other scales and criteria, removing

the subtle items did not effect their scale's predictive validity. Therefore, this study offers support to the inclusion of the subtle items when using the appropriate population and clinically valid measures.

A study by Weed, Ben-Porath, and Butcher (1990) used a unique method to address the contention that the subtle items were included by sampling error (e.g., Jackson, 1971). They used the Wiener-Harmon S/O ratings and had married couples (in marital therapy and from the MMPI-2 normative sample) complete spousal rating forms. Therefore, the external measure was based on ratings made by each MMPI respondent's spouse. The researchers selected items from these forms to rationally correspond to characteristics of the five S/O scales. Addressing the sampling error contention, random numbers were added to each obvious subscale score to obtain pseudo full-scale scores. The resulting pseudo full-scales were compared with the regular full scales in the strength of their relationship to items from the spousal ratings forms.

Analyzing convergent validity, a significantly greater number of the spousal ratings items correlated with the appropriate obvious subscales than with the subtle subscales. Unfortunately, the results of individual S/O subscales were not reported, which prevents determining which subtle subscales obtained correlations. Addressing the sampling error hypothesis, regular full-scale correlations with the spousal rating form were no better or worse than those with the pseudo full-scales. In other words, including the subtle items with the obvious items was no different than adding a random variable to the obvious items.

This study supports the contention that the subtle items were included by sampling error, although closer inspection of their methodology may make this conclusion more tenuous. As seen earlier, the S/O subscales perform differently regarding convergent and discriminative validity. The authors reported that a few correlations

were obtained between subtle subscales and their external measure although they did not indicate which of the five subtle subscales obtained these correlations. More spousal rating items correlated with the appropriate obvious subscales; however, no analysis was performed on discriminative validity, which has been shown to be important to interpretation. Also, no validity or reliability was reported for their external measures.

Important to the issue of content, items were chosen rationally from the spousal rating form based on their obvious relationship to the five MMPI scales which could make the items more prone to correlate with the obvious subscales. The most notable problem was that, although the scales were declared better predictors without the subtle items, no statistical significance was reported for the obvious or full-scale scores as predictors. The authors state that an "arbitrary" validity coefficient of .10 was chosen as a cutoff, with no report of statistical significance (.16 was the highest mean correlation reported). Thus, using the spousal rating form items as criteria is very questionable.

In summary, the validity of the MMPI subtle items has been supported in some studies and not in others. In terms of convergent and discriminative validity, many studies found some support for the subtle items. Of all the subtle subscales, *Ma-S* has shown the best validity. However, the results from Weed and colleagues (1990) suggest that keeping the subtle items on their scales simply adds random noise, although other studies have found that they actually add to the validity of particular scales (e.g., Hovanitz & Jordan-Brown, 1986).

CHAPTER III

STATEMENT OF THE PROBLEM

To date, many different approaches have been used to determine the validity of the subtle items which were included on the MMPI by the empirical method of test construction (Hathaway & McKinley, 1942). Investigating the appropriateness of the subtle items for their scales has been a way to analyze the effectiveness of the empirical method that placed them on those scales. The most popular method of investigating the validity of the MMPI subtle items has been to use external measures of their scales' characteristics as criteria. Studies that have attempted to use this method have produced equivocal results.

Some difficulty exists in finding suitable external criteria for the characteristics measured by each of the MMPI scales. Most studies have used self-report, obvious content questionnaires as criteria (e.g., the Beck Depression Inventory). Some of the criteria do not even correlate significantly with the relevant full-scale score (e.g., Rotter's Interpersonal Trust Scale and the Einstellung Test with the Pa scale; Hovanitz et al., 1983). Although the obvious subscales have shown the best success in correlating with criteria, their discriminative validity is often lacking when using obvious self-report questionnaires (e.g., Hovanitz & Jordan-Brown, 1986). In this regard, the success of the obvious subscales may be exaggerated due to the consistent self-report of obvious, psychopathological symptomatology across measures. Also, the correlations of the S/O subscales with scale-appropriate questionnaires vary depending on the particular questionnaire used. For example, Hovanitz and Gynther (1980) found that the Ma-O and Ma-S subscales were significantly related to different questionnaires.

Most of the studies investigating the subtle items have used non-psychiatric college students as their population, a procedure which

poses some problems in the interpretation of their results. Statistically, non-psychiatric MMPI T-scores do not show much variation, as there are fewer higher scores reflective of more severe pathology. Using this population, therefore, results in a "floor effect," or a restricted range of T-scores (e.g., Peterson et al., 1989). Research on testing which uses non-psychiatric college students is also different in a variety of ways from that which uses outpatients or inpatients. For example, differences include the respondents' expectation of the clinical use of the test and motivation to respond to pathological items. The subtle items may measure some area of personality specific to their scale that is present, but not dysfunctional, in normals (e.g., psychological defenses). Importantly, the MMPI was constructed by a method that distinguished normals from those with specific psychopathology. Therefore, using protocols from normals in research studies would attenuate reliability and validity of the results. A more ecologically valid and statistically interpretable approach is to use the psychiatric population for which the test was constructed.

Instead of using a non-psychiatric subject sample recruited from college classes, the present study has used student and non-student outpatients receiving treatment in a university-based clinic as participants. Secondly, instead of using external criteria with questionable relationships to the MMPI, the present study has employed the respondent's honestly-taken MMPI full-scale T-scores as criteria, using as predictors their subtle subscale T-scores when they take the test under faking conditions (i.e., fake-good or fake-bad). This method is employed as a different way to approach the problem of finding a suitable external criterion to measure the characteristics represented on each of the subtle scales.

Concerning the major purpose of this study, obtaining subtle subscale scores under faking conditions enables a test of the validity of the subtle items as unobtrusive measures of their scales

characteristics. Along these lines, preliminary analyses are conducted to see whether the subtle items prove nonmanipulable, and perhaps show what has been called the "paradoxical effect." Individual scales were analyzed, as most studies have limited their analyses to overall S/O changes. Due to the subtle items' purported resistance to faking, if they truly measure the characteristics addressed by their respective scales, then respondents' subtle subscale scores under faking conditions should still be related to their honestly-taken, full-scale scores. However, because of the conflicting results of studies investigating the validity of the subtle subscales and the problem with choosing external criteria, no predictions will be made as to the relationship of the subtle scales to the full scales.

Concerning the experimental manipulation, or the effects of faking instructions, some predictions can be made. For the fake-good condition, it is predicted that the means for the *T*-scores on scales *F*, *Hy*, *D*, *Hs*, *Pd*, *Pa*, *Pt*, *Sc*, *Ma*, and *Si*, and the *F* - *K* index (Gough, 1950), will be lower in the fake-good condition than in the honest condition. The means for the *T*-scores on scales *L* and *K* are predicted to be higher in the fake-good condition than in the honest condition. No prediction is made for scale *Mf* because of a lack of information as to the effect of faking-good on this scale. In the fake-bad condition, it is predicted that the means for the *T*-scores on scales *F*, *Hy*, *D*, *Hs*, *Pd*, *Pa*, *Pt*, *Sc*, *Ma*, and *Si*, and the *F* - *K* index, will be higher in the fake-bad condition than the honest condition. The means for the *T*-scores on scale *K* are predicted to be lower in the fake-bad condition than in the honest condition. No predictions are made for scales *L* and *Mf* in the fake-bad condition because of a lack of information as to the effect of faking-bad on the *Mf* scale and the floor effect on *L*.

CHAPTER IV

METHOD

Participants

The 40 participants were 14 male and 26 female psychiatric outpatients (ages 20-58; $M = 29.55$; $SD = 9.73$), from a university-based training clinic in the Southwest United States. Participants were randomly assigned to one of two conditions (Honest/Fake-Good, HFG, or Honest/Fake-Bad, HFB), with 20 participants per group. Every participant first took the MMPI under the Honest set of instructions so that their test results could be used for diagnostic and treatment purposes. A complete sequence in the order of instructions in each condition (i.e., H-FG-FB; H-FB-FG) was not used in order to maximize participation rates. In regard to counterbalancing, many previous studies have shown no significant order effects (e.g., Gough, 1950; Rice et al., 1983; Wales & Seeman, 1968). Initial volunteers were informed that they would need to complete two MMPI questionnaires (Form-R, the first 399 items) to be included in the study. Incentive to participate in the study was provided by a \$10.00 cash payment, given at the end of the two MMPI administrations, and a free scoring of their honestly-taken MMPI. The testing feedback was given either by the participant's therapist or an experimenter (see Procedure section for details).

The HFG condition contained 12 males and 8 females (ages 20-50; $M = 27.55$ years, $SD = 7.66$); there were 19 self-identified Caucasians and 1 Hispanic. Fourteen out of 20 were enrolled at least part-time as college students. The HFB condition contained 2 males and 18 females (ages 20-58; $M = 31.70$ years; $SD = 11.23$); there were 19 self-identified Caucasians and 1 non-identified ethnicity (choosing the label of "Other"). Thirteen out of 20 were enrolled at least part time as college students. A 20-year-old, non-student, Caucasian male who did not comply with instructions on the Fake-Good protocol was excluded from

the analyses. The random assignment to groups resulted in an unbalanced number of males and females between groups. However, the analyses for the major purpose of this study are within-subjects.

All participating clients met the four following criteria for inclusion in the present study: (1) They were not involved in any legal matters that might increase their motivation to distort their responses on their honestly-taken MMPI. (2) They were at least 18-years-old. (3) They demonstrated the ability to read and understand the items on the MMPI (determined by reading aloud and responding to the first five items). (4) Their honestly-taken MMPIs obtained a Cannot Say score < 30 and T -scores of $L < 65$, $K < 70$, and $F < 100$ on the validity scales.

Instrument

Minnesota Multiphasic Personality Inventory (MMPI). The first 399 items of the MMPI Form-R (Hathaway & McKinley, 1942) were individually administered to participants using the pencil and paper questionnaire format. Scores were calculated for the 13 standard MMPI scale T -scores (L , F , K , Hs , D , Hy , Pd , Mf , Pa , Pt , Sc , Ma , & Si), the $F - K$ index (Gough, 1950), the Cannot Say score, and the five Wiener-Harmon (1946) S/O subscale T -scores (on scales D , Hy , Pd , Pa , and Ma).

Procedure

Initially, clinic therapists were given information about the study, and thereafter had the choice of informing their clients of the study. This procedure was implemented because of the confidentiality restraints of identifying and approaching clients for research purposes. In addition, therapists were asked to screen out any clients who might have legal motivations to fake (e.g., disability pending, child custody cases). Interested clients were subsequently given a "Volunteer Information" form (see Appendix A) to review upon leaving their therapy session. This form briefly describes the basic experiment,

confidentiality, their compensation, and the procedures for feedback on their test results. The form allows them to indicate their decision to participate in the study and whether they wished to be contacted by the experimenter. The client returned the form in a sealed envelope to the clinic secretary who put it in an experimenter's file.

Upon a participant's arrival for testing, he or she was given an informed consent form to read and sign. Dependent on their decision whether or not to use the MMPI in treatment, the consent form was one of two types (see Appendix B). After the participant indicated he or she was finished reading the consent form, the basic points of the form, which includes a general description of their part in the research, were then reviewed with the participant. They then read the "Honest" (standard) MMPI instructions, heard a brief summary of these instructions by the experimenter, and read aloud and completed the first five items with the experimenter present. After completing the Honestly-taken MMPI, they were randomly assigned one of the two sets of faking instructions (Fake-Bad or Fake-Good). After participants read these faking instructions, the experimenter inquired into their understanding of the instructions before they began the second test.

The Honest instructions were the standard instructions used when normally administering the MMPI. The Fake-Bad condition was preceded by the following instructions:

This inventory consists of numbered statements. I want you to read each statement and imagine a situation in which you are trying to fool a psychologist. More specifically, imagine a situation in which it would be to your advantage to appear as if you were mentally disturbed. Examples of such a situation could be: applying for rehabilitative services, trying to qualify for disability benefits, or trying to beat a legal charge on grounds of insanity. In short, I want you to take this test and deliberately FAKE BAD so that your deception could not be detected

by a professional psychologist.

The Fake-Good condition was preceded by the following instructions:

This inventory consists of numbered statements. I want you to read each statement and imagine a situation in which you are trying to fool a psychologist. More specifically, imagine a situation in which it would be to your advantage to appear as if you were completely normal and sane. Examples of such a situation could be: trying to secure an early release from prison, trying to secure a release from a mental hospital, or applying for a good job. In short, I want you to take this test and deliberately FAKE GOOD so that your deception could not be detected by a professional psychologist.

After completing the two MMPIs, the participants were debriefed on the nature of the study and paid for their participation. Participants were offered the opportunity to receive feedback on their test results (Honest MMPI) and all chose this option. The options for feedback modality included having feedback provided by either the therapist, the experimenter, or the therapist and experimenter, depending on the decisions made by the therapist and his or her supervisor. Concerning the actual usage of modalities, only the therapist alone and experimenter with the therapist present options were chosen; none of the feedback sessions were provided by only the experimenter. The decision to have the feedback provided by the therapist alone or the experimenter with the therapist present, was made according to the expertise of the therapist with the MMPI.

CHAPTER V

RESULTS

A preliminary analysis was conducted to determine if there were any pre-existing differences in the honestly-taken MMPI profile *T*-scores between the two groups that were not controlled for by the random assignment to groups. (An alpha level of .05 was used for all subsequent statistical tests.) Three separate MANOVAs (using Wilks' Lambda statistic) were conducted with Group (HFG & HFB) as the independent variable. The dependent variables on each of the three MANOVAs were the *T*-scores for (a) the 13 traditional validity and clinical full scales (*L*, *F*, *K*, *Hs*, *D*, *Hy*, *Pd*, *Mf*, *Pa*, *Pt*, *Sc*, *Ma*, & *Si*) and the *F* - *K* index (exceptional use of raw score), (b) the five subtle subscales, and (c) the five obvious subscales. The five S/O subscales are derived from scales *D*, *Hy*, *Pd*, *Pa*, and *Ma*.

Analyses showed no overall significant difference between the *T*-scores and *F* - *K* raw score for the HFB and HFG groups in the honest condition on any of the three MANOVAs: for the 4 validity and 10 clinical full-scale scores, $F(14, 25) = 1.36$, *ns*, for the subtle subscale *T*-scores, $F(5, 34) = 2.00$, *ns*, and for the obvious subscale *T*-scores, $F(5, 34) = 1.83$, *ns* (see Tables I, II, and III). These results indicate that there were no pre-existing differences in the standard MMPI profile *T*-scores between faking groups and that random assignment was effective.

Analyses were conducted to determine if the experimental manipulations (i.e., instructional sets to fake) were effective. Two one-way within-subjects MANOVAs were conducted with Condition as the independent variable (Honest vs. Fake-Good; Honest vs. Fake-Bad) and the 13 *T*-scores from the validity and clinical scales (*L*, *F*, *K*, *Hs*, *D*, *Hy*, *Pd*, *Mf*, *Pa*, *Pt*, *Sc*, *Ma*, & *Si*) and the raw score for the *F* - *K* index as

TABLE I

RANDOM ASSIGNMENT TO GROUPS (FULL-SCALE): MEAN T-SCORE
BETWEEN-SUBJECTS UNIVARIATE ANALYSES

	Honest (FG)	Honest (FB)	
Full Scale	MEAN (SD)	MEAN (SD)	F
<i>L</i>	46.80 (5.35)	45.00 (4.95)	1.22
<i>F</i>	66.90 (10.42)	66.05 (12.34)	0.06
<i>K</i>	48.45 (9.84)	50.25 (7.52)	0.42
<i>F - K¹</i>	-1.25 (9.00)	-2.50 (9.03)	0.19
<i>Hs</i>	61.10 (13.22)	62.65 (15.77)	0.11
<i>D</i>	75.55 (11.82)	78.05 (13.45)	0.39
<i>Hy</i>	64.90 (11.01)	66.30 (11.62)	0.15
<i>Pd</i>	73.10 (13.82)	76.35 (10.72)	0.69
<i>Sc</i>	78.05 (15.97)	75.05 (14.59)	0.38
<i>Ma</i>	62.20 (11.75)	55.70 (9.72)	3.63
<i>Si</i>	60.45 (11.95)	66.55 (12.68)	2.45

Note. For all comparisons, $df = 1, 38$; There were no statistically significant differences
¹raw score

TABLE II

RANDOM ASSIGNMENT TO GROUPS (SUBTLE SUBSCALES): MEAN T-SCORE
BETWEEN-SUBJECTS UNIVARIATE ANALYSES

	Honest (FG)	Honest (FB)	
Subtle Subscale	MEAN (SD)	MEAN (SD)	F
D-S	51.65 (10.96)	51.50 (8.67)	0.00
Hy-S	51.60 (9.97)	51.30 (10.64)	0.01
Pd-S	58.75 (9.78)	61.90 (10.83)	0.93
Pa-S	57.50 (9.81)	64.25 (7.99)	5.70
Ma-S	50.85 (9.24)	47.10 (9.69)	1.57

Note. For all comparisons, $df = 1, 38$; there were no statistically significant differences.

TABLE III

RANDOM ASSIGNMENT TO GROUPS (OBVIOUS SUBSCALES): MEAN T-SCORE
BETWEEN-SUBJECTS UNIVARIATE ANALYSES

	Honest (FG)	Honest (FB)	
Obvious Subscale	MEAN (SD)	MEAN (SD)	<i>F</i>
<i>D-O</i>	71.20 (12.61)	75.45 (13.93)	1.02
<i>Hy-O</i>	65.85 (12.03)	67.15 (14.35)	0.10
<i>Pd-O</i>	71.40 (16.85)	73.30 (13.65)	0.15
<i>Pa-O</i>	66.60 (14.08)	62.40 (10.12)	1.17
<i>Ma-O</i>	65.10 (12.50)	60.80 (9.37)	1.51

Note. For all comparisons, $df = 1, 38$; there were no statistically significant differences.

the independent variables in each condition. Therefore, the validity scores and clinical scale *T*-scores were compared for differences between (a) the fake-bad condition and its honest condition and (b) the fake-good condition and its honest condition.

The overall results of the experimental manipulations are as follows and are shown per scale in Tables IV and V. Results for the HFG group revealed a significant overall difference between conditions, $F(14, 6) = 16.20$, $p < .005$. Individual scale score differences were significant in the predicted direction for all scales except *Ma*, which was in the predicted direction but did not reach statistical significance. Therefore, overall differences suggest that the general Fake-Good manipulation was successful. Results for the HFB group showed an overall significant difference between conditions, $F(14, 6) = 39.82$, $p < .0001$. All individual scale score differences were significant in the predicted direction except for *D* and *Si*. The fake-bad *T*-scores for these scales was slightly higher than the honest *T*-scores, but did not reach significance level; the honestly-taken *T*-scores for scale *D* were already fairly high ($M = 78.05$; $SD = 13.45$).

Another purpose of preliminary analyses was to compare the obvious and subtle subscale *T*-scores from the five relevant clinical scales (*D*, *Hy*, *Pd*, *Pa*, & *Ma*) between each faking condition and their respective honest condition. The purpose of these comparisons is to attempt to replicate the paradoxical effect shown in the literature (e.g., Burkhart, Christian, & Gynther, 1978) which has previously used a normal population (i.e., college students not in treatment) to determine if it generalizes to an outpatient population using the Wiener-Harmon S/O items.

Both obvious and subtle subscale *T*-scores were compared between the honest and fake-good conditions and the honest and fake-bad conditions. No predictions were made given the sparse research on the paradoxical effect with outpatients and for individual scales. These comparisons

TABLE IV

EFFECTS OF FAKE-GOOD INSTRUCTIONS: MEAN T-SCORE
WITHIN-SUBJECTS UNIVARIATE ANALYSES

	Honest Condition	Fake-Good Condition	
Full Scale	MEAN (SD)	MEAN (SD)	F
<i>L</i>	46.80 (5.35)	68.30 (12.62)	49.86***
<i>F</i>	66.90 (10.42)	53.15 (9.74)	22.79***
<i>K</i>	48.45 (9.84)	66.45 (4.67)	61.75***
<i>F - K</i> ¹	-16.95 (6.21)	-1.25 (9.00)	50.24***
<i>Hs</i>	61.10 (13.22)	50.75 (6.08)	10.11***
<i>D</i>	75.55 (11.82)	50.00 (6.93)	75.06***
<i>Hy</i>	64.90 (11.01)	55.85 (5.72)	9.83**
<i>Pd</i>	73.10 (13.82)	57.25 (12.42)	13.64**
<i>Mf</i> [†]	60.60 (16.76)	58.55 (7.21)	0.42
<i>Pa</i>	69.60 (9.09)	56.15 (5.80)	31.29***
<i>Pt</i>	73.05 (12.50)	53.85 (4.45)	45.67***
<i>Sc</i>	78.05 (15.97)	56.35 (6.41)	31.90***
<i>Ma</i>	62.20 (11.75)	60.90 (8.76)	0.26
<i>Si</i>	60.45 (11.95)	41.30 (5.56)	38.03***

Note. For all comparisons, *df* = 1, 19.

¹raw score

** *p* < .01, *** *p* < .001

TABLE V

EFFECTS OF FAKE-BAD INSTRUCTIONS: MEAN T-SCORE
WITHIN-SUBJECTS UNIVARIATE ANALYSES

	Honest Condition	Fake-Bad Condition	
Obvious Subscale	MEAN (SD)	MEAN (SD)	F
<i>L</i> †	45.00 (4.95)	46.40 (8.10)	0.42
<i>F</i>	66.05 (12.34)	109.20 (3.58)	254.27***
<i>K</i>	50.25 (7.52)	41.55 (5.61)	15.78***
<i>F - K</i>	41.90 (11.41)	-2.50 (9.03)	227.09***
<i>Hs</i>	62.65 (15.77)	89.25 (16.26)	23.72***
<i>D</i>	78.05 (13.45)	83.20 (10.65)	1.85
<i>Hy</i>	66.30 (11.62)	78.05 (11.78)	7.75*
<i>Pd</i>	76.35 (10.72)	98.10 (7.72)	40.80***
<i>Mf</i> †	47.10 (12.79)	69.15 (10.71)	36.14***
<i>Pa</i>	69.25 (8.66)	108.30 (9.43)	174.28***
<i>Pt</i>	72.20 (11.02)	87.50 (8.36)	18.89***
<i>Sc</i>	75.05 (14.59)	116.10 (6.50)	153.60***
<i>Ma</i>	55.70 (9.72)	92.45 (8.80)	214.22***
<i>Si</i>	66.55 (12.68)	71.95 (11.79)	1.57

Note. For all comparisons, $df = 1, 19$.

†raw score

* $p < .05$, *** $p < .001$

were made using a MANOVA with condition (Honest vs. Faking) as the independent variable and the S/O subscale *T*-scores as the dependent measures.

Analyses confirmed the paradoxical effect for both groups (Honest/Fake-Good and Honest/Fake-Bad) with the respective overall effects: $F(5, 15) = 22.27, p < .0001$ and $F(5, 15) = 98.77, p < .0001$. However, the paradoxical effect was only significant for particular scales, as shown in Tables VI and VII. There were no significant differences between the Honest conditions and their respective Faking conditions for the *T*-scores on subscales *Pd-S* and *Pa-S*. In other words, these subscales proved resistant to faking, but did not show the paradoxical effect. For scale *Ma-S*, the paradoxical effect occurred in the Fake-Good condition, but, in the Fake-Bad condition, *Ma-S* was successfully manipulated, or not resistant to faking. In other words, whether faking good or bad, the *Ma-S* *T*-score increased. Subscales *D-S* and *Hy-S* showed the paradoxical effect for both Fake-Good and Fake-Bad groups. In contrast, all obvious subscales were successfully manipulated in the instructed direction (i.e., fake-good or fake-bad).

The following analyses address the basic question of the present study. For each of the five clinical scales with subtle and obvious subscales, zero-order correlations were calculated between the full-scale *T*-score obtained when participants responded honestly and the full-scale, subtle subscale, and obvious subscale *T*-score obtained under each of the faking conditions. For the *Pd* and *Ma* scales, the full-scale *T*-scores were computed with and without the *K* correction given the lack of cross-validation studies for using *K* corrections in general, and some evidence that it may actually reduce reliability in college settings (see Greene, 1991).

Statistical tests of significance for zero-order correlations were computed first with a liberal alpha level, that is, not controlling for number of comparisons. This approach was necessary to be able to

TABLE VI

PARADOXICAL EFFECT (FAKE-GOOD): MEAN T-SCORE
WITHIN-SUBJECTS UNIVARIATE ANALYSES

Scale	Subtle Subscales			Obvious Subscales		
	Honest	Fake-Good	<i>F</i>	Honest	Fake-Good	<i>F</i>
	MEAN (<i>SD</i>)	MEAN (<i>SD</i>)		MEAN (<i>SD</i>)	MEAN (<i>SD</i>)	
<i>D</i>	51.65 (10.96)	58.25 (6.29)	8.72**	71.20 (12.61)	41.65 (5.29)	94.03***
<i>Hy</i>	51.60 (9.97)	64.15 (6.22)	40.74***	65.85 (12.03)	38.25 (9.01)	74.40***
<i>Pd</i>	58.75 (9.78)	55.65 (10.38)	0.89	71.40 (16.85)	43.05 (11.19)	37.38***
<i>Pa</i>	57.50 (9.81)	61.55 (6.92)	3.07	66.60 (14.08)	45.60 (7.03)	37.56***
<i>Ma</i>	50.85 (9.24)	62.40 (9.03)	17.72***	65.10 (12.51)	46.95 (6.66)	54.18***

Note: *df* = 1,19 for all comparisons; ** *p* < .01, *** *p* < .001

TABLE VII

PARADOXICAL EFFECT (FAKE-BAD): MEAN T-SCORE
WITHIN-SUBJECTS UNIVARIATE ANALYSES

Scale	Subtle Subscales		<i>F</i>	Obvious Subscales		<i>F</i>
	Honest	Fake-Bad		Honest	Fake-Bad	
	MEAN (<i>SD</i>)	MEAN (<i>SD</i>)		MEAN (<i>SD</i>)	MEAN (<i>SD</i>)	
<i>D</i>	51.50 (8.67)	31.05 (8.32)	72.87***	75.45 (13.93)	92.05 (11.31)	15.65***
<i>Hy</i>	51.30 (10.64)	40.00 (7.33)	14.60**	67.15 (14.35)	91.10 (13.30)	28.68***
<i>Pd</i>	61.90 (10.83)	62.90 (8.82)	0.08	73.30 (13.65)	103.85 (8.25)	57.40***
<i>Pa</i>	64.25 (7.99)	59.25 (7.07)	4.22	62.40 (10.12)	115.60 (7.69)	433.11***
<i>Ma</i>	47.10 (9.69)	66.80 (10.48)	42.25***	60.80 (9.37)	95.25 (5.46)	281.70***

Note: *df* = 1,19 for all comparisons; ** *p* < .01, *** *p* < .001

compare the results of the present study with other S/O studies as none have controlled for number of comparisons. Without controlling for number of comparisons, the *T*-scores from the Fake-Good and Fake-Bad *Ma*-O subscales achieved significant positive correlations with the *T*-scores on the respective Honestly-taken *Ma* full-scale (with and without the *K* correction). The Fake-Bad *Pd*-O subscale *T*-scores had a significant negative correlation with its Honestly-taken *Ma* full-scale *T*-scores (with and without the *K* correction). No other full, subtle, or obvious *T*-scores from the faking conditions achieved significant correlations with their honest full-scale *T*-scores (see Tables VIII and IX). No significant correlations were achieved after controlling alpha for number of comparisons per scale (3). These tests of significance were conducted using $r_{\text{critical}} (.05/2, 18)$, where # of comparisons (= 3) - 1 = 2 and $N (= 20) - 2 = 18$.

TABLE VIII

PREDICTIVE VALIDITY OF FAKE-GOOD MMPI SCORES:
MEAN T-SCORE WITHIN-SUBJECTS CORRELATIONS

Honest	Fake-Good	<i>r</i>
Scale <i>D</i>	<i>D</i> -Full	.08
	<i>D</i> -O	.16
	<i>D</i> -S	-.15
Scale <i>Hy</i>	<i>Hy</i> -Full	-.10
	<i>Hy</i> -O	-.02
	<i>Hy</i> -S	-.07
Scale <i>Pd</i>	<i>Pd</i> -Full	-.07
	<i>Pd</i> -O	.03
	<i>Pd</i> -S	.00
Scale <i>PdNK</i>	<i>PdNK</i> -Full	-.01
	<i>Pd</i> -O	.00
	<i>Pd</i> -S	-.04
Scale <i>Pa</i>	<i>Pa</i> -Full	.00
	<i>Pa</i> -O	.01
	<i>Pa</i> -S	-.03
Scale <i>Ma</i>	<i>Ma</i> -Full	.42
	<i>Ma</i> -O	.55*
	<i>Ma</i> -S	.17
Scale <i>MaNK</i>	<i>MaNK</i> -Full	.43
	<i>Ma</i> -O	.52*
	<i>Ma</i> -S	.20

^aNK denotes that scale is without *K* correction

* Before controlling for alpha, $p < .05$

TABLE IX

PREDICTIVE VALIDITY OF FAKE-BAD MMPI SCORES:
MEAN T-SCORE WITHIN-SUBJECTS CORRELATIONS

Honest	Fake-Bad	<i>r</i>
Scale <i>D</i>	<i>D</i> -Full	.03
	<i>D</i> -O	-.13
	<i>D</i> -S	.32
Scale <i>Hy</i>	<i>Hy</i> -Full	-.30
	<i>Hy</i> -O	-.17
	<i>Hy</i> -S	-.36
Scale <i>Pd</i>	<i>Pd</i> -Full	-.35
	<i>Pd</i> -O	-.50*
	<i>Pd</i> -S	-.03
Scale <i>PdNK</i> ^a	<i>PdNK</i> -Full	-.27
	<i>Pd</i> -O	-.47*
	<i>Pd</i> -S	.03
Scale <i>Pa</i>	<i>Pa</i> -Full	.07
	<i>Pa</i> -O	-.01
	<i>Pa</i> -S	-.15
Scale <i>Ma</i>	<i>Ma</i> -Full	.27
	<i>Ma</i> -O	.46*
	<i>Ma</i> -S	.05
Scale <i>MaNK</i>	<i>MaNK</i> -Full	.31
	<i>Ma</i> -O	.45*
	<i>Ma</i> -S	.10

^aNK denotes that scale is without *K* correction

* Before controlling for alpha, $p < .05$

CHAPTER VI

DISCUSSION

These results do not offer support to the validity of the MMPI subtle items as unobtrusive measures of their scales. Specifically, it was sought to determine whether, under faking conditions, the subtle subscale scores would predict the full-scale scores obtained under standard instructions. This method of using the MMPI scores as the criteria to be predicted was employed as an alternative way to measure the validity of the subtle items in comparison to using external criteria (e.g., BDI to predict *D* scale scores; Turner & Romano, 1984). Given the problems discussed in the literature with finding appropriate criteria (e.g., Ward, 1986), a major purpose of the present design was to use the MMPI scores themselves as the criteria to be predicted by the subtle subscale scores. As a result, using the Wiener-Harmon S/O ratings and a psychiatric outpatient population, the respondents' subtle subscale scores obtained under both fake-good and fake-bad conditions were not significantly correlated with their full-scale scores obtained under standard instructions.

The present study's general non-support of the validity of the subtle items joins several of those in the literature that uniformly found no support for their validity (e.g., Burkhart et al., 1980; Burgess et al., 1984; Duff, 1965; McCall, 1958; Nelson & Cicchetti, 1991; Weed et al., 1992). However, many other studies have found at least some construct validity for the subtle subscales themselves and/or their inclusion in their respective scales (Gynther & Burkhart, 1983; Gynther et al., 1979; Hovanitz et al., 1983; Hovanitz et al., 1985; Hovanitz & Gynther, 1980; Hovanitz & Jordan-Brown, 1986; Snyder & Graham, 1984; Turner & Romano, 1984; Wrobel & Lachar, 1982). In fact, an abundance of support has been found for the validity of *Ma-S* which was not confirmed in the present study.

Although the validity of the subtle items as unobtrusive measures of their scales was not supported, the fact that the faked subtle subscales did not correlate with the regular full scales is not a clear indicator of their general invalidity as most of the faked obvious subscales also did not predict their regular full-scale scores. The single scale on which subjects were faking whose score positively correlated with its honestly-taken full-scale score was *Ma-O*, both in the fake-good and fake-bad condition, and with and without the *K* correction. The only other faked scale score to achieve a significant correlation with its honestly-taken full-scale score was in the negative direction (i.e., *Pd-O* in the fake-bad condition, with and without the *K* correction). These significant correlations were achieved only when using very liberal error rates where the number of comparisons was not taken into consideration when controlling for error. The purpose of this decision was based on the ability to compare the findings of the present study to previous MMPI S/O studies, all of which have not controlled for number of comparisons. When controlling for number of comparisons within each scale, all of these significant correlations dropped out.

Additionally, the present study does not address whether the inclusion of the subtle items on their respective scales improves or hinders their full scale's validity. Therefore, these results do not indicate that the subtle items should be removed from their scales. A more direct approach to address their inclusion on scales would be to compare the construct validity of the full scales with and without the subtle items. Both support and non-support for the inclusion of the subtle items has been shown in the literature (e.g., Burgess et al., 1984; Hovanitz & Jordan-Brown, 1986). It has been speculated that the subtle items may serve as some type of correction factor for their scales (e.g., Burkhart & Gynther, 1983) and therefore may be beneficial to their scales regardless of their criterion validity. For example, the subtle items might help correct for false positives similar to the

empirically derived C_k correction factor on the *Hs* scale (McKinley & Hathaway, 1940). However, the present study does not address the subtle items' ability to correct for false positives or false negatives and future studies investigating this question will need to use external criteria other than the MMPI itself.

The present results supported the subtle items' resistance to manipulation which has been widespread in the literature (e.g., Burkhart et al., 1978; Vesprani & Seeman, 1974) for all subtle subscales except *Ma-S* in the fake-bad condition. Peterson et al. (1989) is the only published study that also analyzed the change status of individual Wiener-Harmon S/O subscales under fake-bad instructions. Using college students, their study similarly found that *Ma-S* was not resistant to manipulation in the fake-bad condition, although all other subtle subscale were resistant to faking. In the present study, although the majority of the subtle subscales were resistant to faking, none of the subtle subscales scores obtained under faking instructions were found to be correlated with their honestly-taken full-scale scores. In other words, although the subtle subscales may be resistant to faking, the present findings do not support the subtle subscale scores obtained under faking instructions as accurate measures of their full-scale scores taken under standard instructions.

One of the most direct clinical applications of these results is that they address the feasibility of using the Wiener-Harmon subtle subscale scores to predict a respondent's "true" score when it is suspected he or she is faking good or bad. The Wiener-Harmon S/O subscales are popularly used in clinical settings and are calculated when using NCS as an automated scoring system. Given that the MMPI subtle items can be understood as unobtrusive measures of their scales' characteristics, one might infer that the Wiener-Harmon subtle subscale scores could be used as a guide to what the person's true score would have been if not faking. However, the results from the present study do

not support the alternative use of the five Wiener-Harmon subtle subscale scores as predictive of a person's score when that person is thought to be consciously distorting their responses.

As a special phenomenon that has occurred in studies addressing the subtle items' resistance to faking, the present study also investigated the paradoxical effect of the subtle items, which has vast support in the literature (e.g., Wales & Seeman 1968; 1969; 1972). Individual scales were analyzed for the paradoxical effect, which has not yet been addressed using the Wiener-Harmon S/O items and a psychiatric outpatient population. The paradoxical effect was found in the fake-good group for *D-S*, *Hy-S*, and *Ma-S*, and in the fake-bad group for *D-S* and *Hy-S*; in the fake-bad group, *Ma-S* showed an effect in the opposite direction--i.e., was manipulated successfully. For *Pd-S* and *Pa-S*, no paradoxical effect was found in either the fake-good or fake-bad group.

These paradoxical effect results are very similar to those found by the previously mentioned Peterson et al. (1989) study. The only difference is that their study did not find the paradoxical effect for *D-S* and *Hy-S* in the fake-good group although a nonsignificant change occurred in the paradoxical direction. Concerning this nonsignificant change, their study described a floor effect that was found in fake-good conditions when using college students where the scores were already very low and had a restricted range. Therefore, their failure to find the statistical difference for these two subtle subscales that was found in the present study may have been limited by the floor effect in the fake-good condition that occurs with a normal population. Nevertheless, the comparability of the paradoxical results between these two studies is remarkably consistent.

The only other study analyzing for the paradoxical effect in individual scales (Vesprani & Seeman, 1974) used "ideal self" instructions with the *X* and *O* items and psychiatric outpatients and found the paradoxical effect for all *O* subscales (i.e., subtle

subscales). Overall, the paradoxical effect has shown the most consistent support with the X and O items and all attempts that failed to find this effect in some capacity have occurred with the Wiener-Harmon S/O items. Although not explored by the present study, this phenomenon may be explained by the statistical definition of the X and O items. The O (subtle) items are those endorsed by a majority of normals, but by yet even more of the diagnostic groups; the X (obvious) items are those endorsed by a minority of the normals and significantly more of the diagnostic group. Therefore, although subtle items classified as such by ratings (e.g., Christian et al. and Wiener-Harmon S/O ratings) may have face-validity that appears misleadingly normal (e.g., Dannenbaum & Lanyon, 1993), the X and O items may be more likely to show a paradoxical effect by their statistical and scale-specific method of classification.

In critique of the method used in the present study, using the MMPI scores as criteria is not completely free of the problems encountered when using criteria consisting largely of obvious self-report items (e.g., the BDI) as most of the MMPI items are obvious. Therefore, the present study runs into similar problems as other studies using obvious items. The fact that many studies have found correlations between face-valid self-report measures and MMPI full-scale scores, but no such correlations with the subtle subscales (e.g., Nelson & Cicchetti, 1991), may reflect a general tendency to report obvious psychopathological symptomatology or be confounded with social desirability. This latter possibility is also suggested by the high correlation found between a social desirability measure and the MMPI obvious items (Christian et al., 1978).

Future studies might use criteria similar to the more clinically expedient measures of diagnosis and medication employed by Hovanitz and Jordan-Brown (1986). Alternatively, external criteria might consist of independent ratings of respondents by clinicians based on the

characteristics measured by the MMPI scales. Similar procedures were used by Wrobel and Lachar (1982) and Weed et al. (1990), but both lacked reliability and validity of their measures and in the latter, ratings were made by lay people. Therefore, criteria might consist of independent ratings by clinicians on measures with proven external validity. However, the feasibility of this latter method is questionable given that the rating clinician must have a well-informed understanding of the person being rated. In an outpatient setting, there is likely to be only one such clinician, the person's therapist, thereby eliminating the possibility of inter-rater reliability. An alternative is to use an inpatient setting where many mental health workers might have first-hand knowledge of the patients and provide such ratings.

Two unpublished studies (Brantley, 1991; Taylor, 1990) have encountered difficulties when using a method compatible with the latter suggestion. These researchers used U.S. Veterans with and without motivation to fake (i.e., applying for disability) and component scales on the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) as the external criteria. The interpretation of their S/O results was impeded by inadequate inter-rater reliability and the failure for three of the four raters' BPRS scores to correlate with any of the MMPI full, subtle, or obvious scores in either the standard or faking condition.

There is some debate in the literature over the use of S/O ratings which are made for general psychopathology, a classification to which the Wiener-Harmon S/O items in the present study belong. For example, Ward (1986) asserts that subtle ratings based on a general degree of psychopathology instead of the characteristics of the individual scales violates commonsense notions of subtlety. Holden and Jackson (1979) state that subtlety should be defined as "the respondent's ability to relate an item to its actual, keyed scale" (p. 461). Therefore, the failure to find predictive ability for the Wiener-Harmon subtle subscale

scores in the present study does not necessarily imply that an item with a subtle relationship to its scale tends to be a poor scale item. What the specific findings indicate is that when respondents take the MMPI with general instructions to fake good or bad, the scale items which are considered by psychologists to have a non-obvious relationship to psychopathology are not related to the respondents' regular full-scale scores.

Therefore, the specific type of Wiener-Harmon subtle items do not appear to be good unobtrusive measures of their scales' characteristics. Whether items rated as subtle for their particular scale are good unobtrusive measures is less clear. Currently, there are no published scale-specific ratings for S/O items. As previously discussed, Snyder and Graham (1984) used unpublished scale-specific ratings and found support for *Ma*-S but not *Pd*-S. Interestingly, studies using the Christian et al. S/O items rated for general psychopathology (e.g., Gynther et al., 1979; Hovanitz & Gynther, 1980), conversely have supported *Pd*-S and not *Ma*-S. Hovanitz and Gynther found that many of the Christian et al. *Ma* subtle items were on the Harris and Lingoes' (1955) *Imperturbability* scale, a rationally defined scale. This suggests that, if scale-specific S/O ratings are used for *Ma* instead of those for general psychopathology, many *Ma* items that are rated as subtle for general psychopathology might be rated as obvious for the hypomania construct (and vice versa). Scale-specific subtle items' potential to unobtrusively measure the characteristics of their scale needs to be assessed using these particular items and should not be deduced from the general psychopathology ratings of the Wiener-Harmon subtle items.

In critique of the future use of the Wiener-Harmon S/O items, a decision to use S/O ratings for general psychopathology or for specific scales should depend on the situation. Ratings for general psychopathology by naive raters seem appropriate when assessing the S/O

subscales' ability to detect overreporting or underreporting of psychopathology in general. Therefore, studies investigating the S/O subscales as a general validity index should use these ratings. Scale-specific ratings by naive raters are appropriate in research addressing the subtle items' ability to unobtrusively measure the characteristics of their scales, or their potential to serve as a correction function for their full scale. Scale-specific ratings by expert raters may intuitively provide researchers with the most likely candidates of subtle items that do not belong on their scales. Of course, a more direct way of determining invalid MMPI items is with an item analysis of all items or a cross-validation of the MMPI.

Nevertheless, most of the subtle items studies have been conducted with the Christian et al. S/O items. Only two predictive validity studies have been published using the Wiener-Harmon S/O items (Weed et al., 1990; Wrobel & Lachar, 1982) with which to directly compare the results of the present study. Weed et al. found some ability of the Wiener-Harmon subtle subscale scores to predict their criteria although they did not state which scales. However, none of the subtle subscales were found to be better predictors than a random variable (although limitations of this study have been previously discussed). Wrobel and Lachar found predictive validity for *Ma-S*, which is the subtle subscale that has obtained the most support for its validity but was not predictive in the present study.

However, one limitation to comparing the present results to the previously cited Wiener-Harmon S/O studies is the use of faking instructions. The method of using fake good and bad subtle subscale scores to predict honestly-taken full-scale scores may not generalize to the ability of the subtle subscale scores under standard instructions to predict scale-related criteria. In support of using faked subtle subscale scores, a large number of studies have shown that the subtle items are usually resistant to faking in the desired direction, except

for the Wiener-Harmon *Ma-S* subscale. Therefore, for the results of the faked subtle subscale scores to be compared to the ability of the subtle subscale scores under standard instructions to predict criteria, one must assume that the scores would not significantly change under faking instructions. However, the fact that the faked scores often do change in the opposite of the desired direction (i.e., the paradoxical effect) indicates that the faked subtle subscale scores are not stable across instructional sets. Therefore, the present results more clearly address the relationship of the subtle subscales taken under faking instructions to their honestly-taken full-scale scores, rather than the validity of the subtle items under normal conditions.

The generalizability of the type of faking or distorted response sets to which the present study applies also is limited to the general faking instructions used. The faking instructions used in the present study were non-specific fake-good and fake-bad instructions and included a number of examples which the respondents might use as a mental set (e.g., fake good--applying for a good job, etc.). In everyday usage of the MMPI, the context for the potential faked profiles will probably be known (e.g., psychotherapy, occupational, insanity plea, etc.) and the results based on non-specific faking instructions can only be a guideline at best. The general faking instructions also do not give any information as to what happens when respondents are faking on specific scale characteristics (e.g., depression).

Additionally, the conscious distortion used in the present study may obtain different results than when respondents are unconsciously motivated to distort their responses, a faking set that would be difficult to experimentally manipulate or detect as a subject variable. The closest one might come to finding this population is to determine particular settings (e.g., applying for a job) where one is likely to find distorted responses. However, the unconscious nature of any distortion seems impossible to predict a priori. Even the ability to

determine the direction of the distortion may elude researchers. For example, a masochistic or self-defeating person might be unconsciously motivated to look worse when applying for a job.

The present study used the original MMPI instead of its revised edition, the MMPI-2. However, it may be safe to generalize these MMPI findings to the MMPI-2. Ben-Porath and Butcher (1989) give evidence that the psychometric properties of the MMPI-2 are not significantly different than those of the MMPI. Furthermore, only three of the thirteen items dropped from the original MMPI were on the five S/O scales and all three were on scale D. For the Wiener-Harmon S/O items, two qualify as subtle and one as obvious. However, because the MMPI-2 normative sample achieved higher scores on the Wiener-Harmon subtle subscales than the original MMPI sample (Greene, 1991), which also accounts for the overall increase in the normative samples' higher full-scale scores on the MMPI-2, new validity studies could be performed with these new norms.

Concerning other design limitations of the present study, the low number of subjects does not lend much power to find a significant statistical relationship if indeed one does exist. Additionally, although random assignment to groups (fake-bad or fake-good) was employed, there was a disproportionately large number of females in the fake-bad group which could further limit generalizability.

In summary, the main findings of the present study do not support the ability of the Wiener-Harmon subtle items, when answered under faking instructions, to predict standard full-scale scores. These results do not indicate, however, whether the MMPI full-scale scores would be better off without the subtle items. They more directly indicate that respondents' Wiener-Harmon subtle subscale scores obtained under conscious distortion are not related to what their full-scale scores would be when taken honestly. Suggestions for future studies

have been described in detail, with specific emphasis on the need for scale-specific S/O ratings.

BIBLIOGRAPHY

- Anthony, N. (1971). Comparison of clients' standard, exaggerated, and matching MMPI profiles. *Journal of Consulting and Clinical Psychology*, 36, 100-103.
- Ashton, S. G., & Goldberg, L. R. (1973). In response to Jackson's challenge: The comparative validity of personality scales constructed by the external (empirical) strategy and scales developed intuitively by experts, novices, and laymen. *Journal of Research in Personality*, 7, 1-20.
- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. New York: Harper & Row.
- Ben-Porath, Y. S., & Butcher, J. H. (1989). Psychometric stability of rewritten MMPI items. *Journal of Personality Assessment*, 53, 645-653.
- Bernreuter, R. G. (1935). The theory and construction of the personality inventory. *Journal of Social Psychology*, 4, 387-405. (Inventory published by Stanford University Press, Stanford, CA, 1931, rev. 1935.)
- Brantley, D. E. (1991). *The predictive validity of MMPI subtle and obvious items under honest and exaggerated response conditions in a clinical population*. Unpublished doctoral dissertation, Oklahoma State University, Stillwater.
- Burgess, P. M., Campbell, I. M., & Zylberberg, A. (1984). Face validity vs. item subtlety in the MMPI D scale. *Journal of Clinical Psychology*, 40, 499-504.
- Burkhart, B. R., Christian, W. L., & Gynther, M. D. (1978). Item subtlety and faking on the MMPI: A paradoxical relationship. *Journal of Personality Assessment*, 42, 76-80.
- Burkhart, B. R., Gynther, M. D., & Christian, W. L. (1978). Psychological mindedness, intelligence, and item subtlety endorsement patterns on the MMPI. *Journal of Clinical Psychology*, 34, 76-79.
- Burkhart, B. R., Gynther, M. D., & Fromuth, M. E. (1980). The relative predictive validity of subtle vs. obvious items on the MMPI depression scale. *Journal of Clinical Psychology*, 36, 748-751.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Manual for administration and scoring: MMPI*. Minneapolis: University of Minnesota Press.
- Christian, W. L., Burkhart, B. R., & Gynther, M. D. (1978). Subtle-obvious ratings of MMPI items: New interest in an old concept. *Journal of Consulting and Clinical Psychology*, 46, 1178-1186.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook: Volume I: Clinical interpretation*. (rev. ed.) University of Minnesota, Minneapolis.

- Dannenbaum, S. E., & Lanyon, R. I. (1993). The use of subtle items in detecting deception. *Journal of Personality Assessment*, 61, 501-510.
- Derogatis, L. (1983). *SCL-90-R: Administration, scoring, and procedures manual-II*. Clinical Psychometric Research (2nd ed.). Towson, MD: Author.
- Drake, L. E. (1946). A social I. E. scale for the MMPI. *Journal of Applied Psychology*, 30, 51-54.
- Dubinsky, S., Gamble, D. J., & Rogers, M. L. (1985). A literature review of subtle-obvious items on the MMPI. *Journal of Personality Assessment*, 49, 62-68.
- Duff, F. L. (1965). Item subtlety in personality inventory scales. *Journal of Consulting Psychology*, 29, 565-570.
- Gough, H. G. (1950). The F minus K dissimulation index for the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology*, 14, 408-413.
- Gloye, E. E., & Zimmerman, I. L. (1967). MMPI item changes by college students under ideal-self response set. *Journal of Projective Techniques and Personality Assessment*, 31, 63-69.
- Gough, H. G. (1969). *California Psychological Inventory* (revised manual). Palo Alto, CA: Consulting Psychologists Press.
- Gough, H. G. (1975). *Manual for the California Personality Inventory* (rev. ed.). Palo Alto, CA: Consulting Psychologists Press.
- Graham, J. R. (1990). *MMPI-2: Assessing personality and psychopathology*. New York: Oxford University Press.
- Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretative manual*. Needham Heights, MA: Allyn and Bacon.
- Grow, R., McVaugh, W., & Eno, T. D. (1980). Faking and the MMPI. *Journal of Clinical Psychology*, 36, 910-917.
- Gynther, M. D., & Burkhart, B. R. (1983). Are subtle MMPI items expendable? In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (pp. 115-132). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Gynther, M. D., Burkhart, B. R., & Hovanitz, C. (1979). Do face-valid items have more predictive validity than subtle items? The case of the MMPI Pd scale. *Journal of Consulting and Clinical Psychology*, 47, 295-300.
- Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56-62.
- Harris, R. E., & Lingo, J. C. (1955). *Subscales for the MMPI: An aid to interpretation* [Mimeographed materials]. Department of Psychiatry, University of California.
- Harvey, M. A., & Sippelle, C. N. (1976). Demand characteristic effects on the subtle and obvious subscales of the MMPI. *Journal of Personality Assessment*, 40, 539-544.

- Hathaway, S. R. (1965). Personality inventories. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. 451-476). New York: McGraw-Hill.
- Hathaway, S. R., & McKinley, J. C. (1942). *Minnesota Multiphasic Personality Inventory*. Minneapolis, Minn.: University of Minnesota Press.
- Hiner, D. L., Ogren, D. J., & Baxter, J. C. (1969). Ideal-self reporting on the MMPI. *Journal of Projective Techniques and Personality Assessment*, 33, 389-396.
- Holden, R. R., & Jackson, D. N. (1979). Item subtlety and face validity in personality assessment. *Journal of Consulting and Clinical Psychology*, 47, 459-468.
- Hovanitz, C. A., & Gynther, M. D. (1980). The prediction of impulsive behavior: Comparative validities of obvious vs. subtle MMPI hypomania (Ma) items. *Journal of Clinical Psychology*, 36, 422-427.
- Hovanitz, C. A., Gynther, M. D., & Green, S. B. (1985). Discriminant validity of subtle and obvious items: The MMPI Pa and Ma scales. *Journal of Clinical Psychology*, 41, 42-44.
- Hovanitz, C. A., Gynther, M. D., & Marks, P. A. (1983). The prediction of paranoid behavior: Comparative validities of obvious vs. subtle MMPI paranoia (Pa) items. *Journal of Clinical Psychology*, 39, 407-411.
- Hovanitz, C. A., & Jordan-Brown, L. (1986). The validity of MMPI subtle and obvious items in psychiatric patients. *Journal of Clinical Psychology*, 1, 100-108.
- Hryckowian, M. J., & Gynther, M. D. (1988). MMPI item subtlety: Another look. *Journal of Clinical Psychology*, 44, 148-152.
- Jackson, D. N. (1971). The dynamics of structured personality tests: 1971. *Psychological Review*, 78, 229-248.
- Kelly, D. B., & Greene, R. L. (1989). Detection of faking good on the MMPI in a psychiatric inpatient population. *Psychological Reports*, 65, 747-750.
- Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist*, 39, 451-454.
- Luchins, A. S. (1951). On recent usage of the Einstellung-effect as a test of rigidity. *Journal of Consulting Psychology*, 15, 89-94.
- MacPhillamy, D. J., & Lewinsohn, P. M. (1974). Depression as a function of levels of desired and obtained pleasure. *Journal of Abnormal Psychology*, 83, 651-657.
- McCall, R. J. (1958). Face validity in the D scale of the MMPI. *Journal of Clinical Psychology*, 14, 77-80.
- McKinley, J. C., & Hathaway, S. R. (1940). A multiphasic personality schedule (Minnesota): II. A differential study of hypochondriasis. *Journal of Psychology*, 10, 255-268.

- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology*, 1, 296-303.
- Mehrabian, A. (1977). A questionnaire measure of individual differences in stimulus screening and associated differences in arousability. *Environmental Psychology and Nonverbal Behavior*, 1, 89-103.
- Nelson, L. D., & Cicchetti, D. (1991). Validity of the MMPI depression scale for outpatients. *Psychological Assessment*, 3, 55-59.
- Overall, J. E., & Gorham, D. R. (1962). The Brief Psychiatric Scale. *Psychological Reports*, 10, 799-812.
- Peterson, G. W., Clark, D. A., & Bennett, B. (1989). The utility of MMPI subtle, obvious scales for detecting fake good and fake bad response sets. *Journal of Clinical Psychology*, 45, 575-583.
- Porteus, S. D. (1965). *Porteus maze test: Fifty years' application*. Palo Alto, CA: Pacific Books.
- Rice, M. E., Arnold, L. S., & Tate, D. L. (1983). Faking good and bad adjustment on the MMPI and overcontrolled-hostility in maximum security psychiatric patients. *Canadian Journal of Behavioural Science*, 15, 43-51.
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, 35, 651-665.
- Snyder, C. M., & Graham, J. R. (1984). The utility of subtle and obvious MMPI subscales based on scale-specific ratings. *Journal of Clinical Psychology*, 40, 981-985.
- Taylor, D. M. (1990). *The role of obvious and subtle items in the detection of exaggeration on the MMPI*. Unpublished doctoral dissertation, Oklahoma State University, Stillwater.
- Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity*. New York: McGraw-Hill.
- Timbrook, R. E., Graham, J. R., Keiller, S. W., & Watts, D. (1993). Comparison of the Wiener-Harmon subtle-obvious scales and the standard validity scales in detecting valid and invalid MMPI-2 profiles. *Psychological Assessment*, 5, 53-61.
- Turner, J. A., & Romano, J. M. (1984). Self-report screening measures for depression in chronic pain patients. *Journal of Clinical Psychology*, 40, 909-913.
- Vesprani, G. J., & Seeman, W. (1974). A new method for detecting the fake-good response set on the MMPI. *Journal of Clinical Psychology*, 24, 211-216.
- Wales, B., & Seeman, W. (1968). A new method for detecting the fake good response set on the MMPI. *Journal of Clinical Psychology*, 24, 211-216.

- Wales, B., & Seeman, W. (1969). What do MMPI zero items really measure: An experimental investigation. *Journal of Clinical Psychology*, 25, 420-424.
- Wales, B., & Seeman, W. (1972). Instructional sets and MMPI items. *Journal of Personality Assessment*, 36, 282-286.
- Ward, L. C. (1986). MMPI item subtlety research: Current issues and directions. *Journal of Personality Assessment*, 50, 73-79.
- Weed, N. C., Ben-Porath, Y. S., & Butcher, J. N. (1990). Failure of Wiener and Harmon Minnesota Multiphasic Personality Inventory (MMPI) subtle scales as personality descriptors and as validity indicators. *Psychological Assessment*, 2, 281-285.
- Wiener, D. N. (1948). Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory. *Journal of Consulting Psychology*, 12, 164-170.
- Wiener, D. N., & Harmon, L. R. (1946). *Subtle and obvious keys for the MMPI: Their development*. (Advisement Bulletin No. 16, Regional Veterans Administration Office, Minneapolis).
- Woodworth, R. S. (1920). *Personal Data Sheet*. Chicago: Stoelting.
- Worthington, D. L., & Schlottmann, R. S. (1986). The predictive validity of subtle and obvious empirically derived psychological test items under faking conditions. *Journal of Personality Assessment*, 50, 171-181.
- Woychyshyn, C. A., McElheran, W. G., & Romney, D. M. (1992). MMPI validity measures: A comparative study of original with alternative indices. *Journal of Personality Assessment*, 58, 138-148.
- Wrobel, T. A., & Lachar, D. (1982). Validity of the Wiener subtle and obvious scales for the MMPI: Another example of the importance of inventory-item content. *Journal of Consulting and Clinical Psychology*, 50, 469-470.
- Zuckerman, M. (1977). *Preliminary manual with scoring keys and norms for Form V of the Sensation Seeking Scale*. Unpublished manuscript, University of Delaware, Newark.
- Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12, 63-70.

APPENDIX A

VOLUNTEER FORM

VOLUNTEER INFORMATION

THE BASIC EXPERIMENT

The experiment described in this form is part of an independent study which has gained the consent of the Psychological Services Center (PSC) to present this opportunity to you. We are looking for volunteers to complete a questionnaire two times. If you decide to participate in this study, completion of the questionnaires will take about 1½ to 2½ hours; compensation for your time and effort is discussed below.

THE QUESTIONNAIRE

The questionnaire is a personality inventory called the Minnesota Multiphasic Personality Inventory (MMPI). This questionnaire is a widely used personality test that is mainly used in vocational, medical, and psychological settings (like the PSC). It includes statements about personal attitudes, beliefs, and behaviors which require a True-False answer.

CONFIDENTIALITY

Your decision to participate in this study will not be communicated to your therapist unless you so request. This study is not designed to be part of your therapy; however, you can request for your MMPI results to be sent to your therapist for use in therapy. Your association with the PSC will also be kept confidential and all test data that is not used for your therapy will be coded to ensure your anonymity.

YOUR COMPENSATION

All volunteers will be paid \$10.00 for their time and effort in this research project. The compensation will be in the form of a direct cash payment (in order to protect confidentiality) immediately following your participation. If you are planning on integrating the MMPI into therapy, the usual cost of administration and scoring is covered for you.

TEST RESULT FEEDBACK

The interpretation of your MMPI can either be given by your therapist or an experimenter (a supervised Doctoral Student in Clinical Psychology). If you chose to use the MMPI in therapy, you may discuss the mode of interpretation with your therapist. If you are not using the MMPI in therapy, an experimenter will provide your feedback and arrange a convenient time and day with you.

YOUR DECISION

[] **Yes** I wish to participate in this study and request that my therapist be informed of my decision.

[] **Yes** I wish to participate in this study and I do not want this information communicated to my therapist.

(appendix continues)

APPENDIX A (cont.)

- [] **Maybe** I wish to obtain more information.
[] Please contact me (include telephone # below)
[] I will contact the experimenter (request the number from the
PSC receptionist
- [] **No** I do not wish to participate in this study. (if you checked this
option, do not include your name or number below)

If you marked yes or the first maybe choice, please continue. . .

By signing below, I understand that an experimenter will contact me about
my decision to seek more information or to participate. By signing below,
I understand that an experimenter will be aware of my connection to the
clinic associated with this study.

I consent to allow an experimenter to contact me about this study.

(your printed name)

(telephone number)

(signature)

APPENDIX B
INFORMED CONSENT FORMS

INFORMED CONSENT STATEMENT

Project Title: Analysis of an Empirical Method of Test Construction

Experimenters: Robert S. Schlottmann, Ph.D. and Joni L. Mihura, M.S.

I, (print name) _____, hereby authorize Robert S. Schlottmann, Ph.D. and Joni L. Mihura, M.S., or assistants of their choosing, to perform the procedures listed here.

A. Purpose: The present study is designed to investigate the method of test construction used by the MMPI.

B. Procedures: In participating in this experiment, you will be involved in the following activities:

1. You will be asked to complete the MMPI twice. The test results will not be made available to your therapist prior to or during treatment, nor will your therapist be made aware of whether or not you have decided to participate in the experiment.

2. The results of the MMPI may be interpreted by the experimenter on the request of the client whose name appears above. The scoring and interpretation by the experimenter will take at least two weeks to prepare.

3. A debriefing will be provided at the end of the test administration. The purpose of the experiment will be discussed and any relevant questions will be answered.

C. Duration of participation: Your participation will require 2-3 hours of your time.

D. Confidentiality: All files of the experiment's data will be numerically coded and kept in a secure place. The results of this study may be presented at professional meetings or in publications in the form of statistical information. Your anonymity will be preserved.

E. Risks: The risks in this study are minimal and do not exceed those ordinarily encountered in daily life. However, the MMPI includes items about personal attitudes and behaviors and some may be considered embarrassing.

F. Benefits: As a research participant, you will be exposed to the conduct of scientific psychological research and may gain insight into your own psychological profile as assessed by your scores on the MMPI. Through research like this, psychological tests are made more reliable and valid which help in the assessment, diagnostics, and treatment of people seeking psychological help.

G. Compensation for participation: The costs will be covered for the MMPI, which is a somewhat expensive and time consuming personality inventory to administer, score, and interpret. Also, you will be paid a total of \$10.00 in cash for the completion of the second MMPI.

(appendix continues)

APPENDIX B (cont.)

Consent Form for "Analysis of an Empirical Method of Test Construction"

I have been fully informed about the procedures listed here. I am aware of what I will be asked to do and of the risks and benefits in this study.

I also understand the following statements:

I certify that I am 18 years of age or older.

My participation today is part of an investigation entitled "Analysis of an Empirical Method of Test Construction"

The purpose of this study is to investigate the empirical method of test construction used by the MMPI.

I understand that my participation is voluntary, that there is not penalty for refusal to participate, and that I am free to withdraw my consent and participation in this study at any time without penalty.

I understand that I may contact any of the experimenters at the following address and telephone number should I desire to discuss my participation in this study and/or to request information pertaining to the study's outcome: 215 North Murray, department of Psychology, Oklahoma State University, Stillwater, OK 73078-0250, (405)744-6027. Additionally, I understand that I may contact LeAnn Prater, University Research Services, 001 Life Sciences East, Oklahoma State University, Stillwater, OK 74078, (405)744-5700.

I have read and fully understand this consent form. I sign it in free will and voluntarily. A copy of this form has been given to me. I hereby give permission for my participation.

Signature of Participant

Date

Time: _____ am pm

Signature of Witness

Date

I certify that I have personally completed all the blanks in this form and have explained them to the subject before requesting the subject sign this consent form.

Signature of Project Director or Authorized Representative

APPENDIX B (cont.)

INFORMED CONSENT STATEMENT

Project Title: Analysis of an Empirical Method of Test Construction

Experimenters: Robert S. Schlottmann, Ph.D., Joni L. Mihura, M.S., Anne B. Scott, M.S. and David G. Brunetti, M.S.

I, (print name) _____, hereby authorize Robert S. Schlottmann, Ph.D., Joni L. Mihura, M.S., Anne B. Scott, M.S., and/or David G. Brunetti, M.S. to perform the procedures listed here.

A. Purpose: The present study is designed to investigate the method of test construction used by the MMPI.

B. Procedures: In participating in this experiment, you will be involved in the following activities:

1. You will be asked to complete the MMPI twice. The test results from the first MMPI will subsequently be made available to your therapist for use in your treatment.

2. The results of the MMPI may be interpreted by the experimenter or therapist on the request of the client whose name appears above. The scoring and interpretation by the experimenter will take at least two weeks to prepare.

3. A debriefing will be provided at the end of the test administration. The purpose of the experiment will be discussed and any relevant questions will be answered.

C. Duration of participation: Your participation will require 2-3 hours of your time.

D. Confidentiality: All files of the experiment's data will be numerically coded and kept in a secure place. Upon your agreement, your MMPI results will also be released to your therapist. The results of this study may be presented at professional meetings or in publications in the form of statistical information. Your anonymity will be preserved.

E. Risks: The risks in this study are minimal and do not exceed those ordinarily encountered in daily life. However, the MMPI includes items about personal attitudes and behaviors and some may be considered embarrassing.

F. Benefits: As a research participant, you will be exposed to the conduct of scientific psychological research and may gain insight into your own psychological profile as assessed by your scores on the MMPI. You and your therapist will be provided administration and scoring free of time and cost. Through research like this, psychological tests are made more reliable and valid which help in the assessment, diagnostics, and treatment of people seeking psychological help.

G. Compensation for participation: The costs will be covered for the MMPI, which is a somewhat expensive and time consuming personality inventory to administer, score, and interpret. Also, you will be paid a total of \$10.00 in cash for the completion of the second MMPI.

(appendix continues)

APPENDIX B (cont.)

Consent Form for "Analysis of an Empirical Method of Test Construction"

I have been fully informed about the procedures listed here. I am aware of what I will be asked to do and of the risks and benefits in this study.

I also understand the following statements:

I certify that I am 18 years of age or older.

My participation today is part of an investigation entitled "Analysis of an Empirical Method of Test Construction"

The purpose of this study is to investigate the empirical method of test construction used by the MMPI.

I understand that my participation is voluntary, that there is not penalty for refusal to participate, and that I am free to withdraw my consent and participation in this study at any time without penalty.

I understand that I may contact any of the experimenters at the following address and telephone number should I desire to discuss my participation in this study and/or to request information pertaining to the study's outcome: 215 North Murray, department of Psychology, Oklahoma State University, Stillwater, OK 73078-0250, (405)744-6027. Additionally, I understand that I may contact the University Research Services, 001 Life Sciences East, Oklahoma State University, Stillwater, OK 74078, (405)744-5700.

I have read and fully understand this consent form. I sign it in free will and voluntarily. A copy of this form has been given to me. I hereby give permission for my participation.

Signature of Participant

Date

Time: _____ am pm

Signature of Witness

Date

I certify that I have personally completed all the blanks in this form and have explained them to the subject before requesting the subject sign this consent form.

Signature of Project Director or Authorized Representative

2

VITA

Joni Lynn Hollrah

Candidate for the Degree of
Doctor of Philosophy

Dissertation: THE MMPI SUBTLE SUBSCALES: A TEST OF VALIDITY

Major Field: Clinical Psychology

Biographical:

Personal Data: Born in Stillwater, Oklahoma, On September 25,
1960, the daughter of Jim and Fran Mihura.

Education: Graduated from Byng High School, Byng, Oklahoma in
May 1978; received Bachelor of Science degree in Psychology and a
Master of Science degree in General Psychology from Oklahoma State
University, Stillwater, Oklahoma in December 1987 and December
1990, respectively. Completed the requirements for the Doctor of
Philosophy degree with a major in Clinical Psychology at Oklahoma
State University in December 1995.

Experience: Psychotherapy and testing experience through the
Psychological Services Center and the Marriage and Family Clinic
at Oklahoma State University, Meadowlake Hospital at Enid,
Oklahoma, and Harvard Medical School/Massachusetts Mental Health
Center, Boston, Massachusetts. Research experience at Oklahoma
State University with Drs. Robert Weber (1987-88), Donald K.
Fromme (1989-90), Larry Hochhaus (1990 to present), and Robert S.
Schlottmann (1991 to present). Teaching assistant, lab
instructor, and course instructor at Oklahoma State University
from 1989 to 1993.

Professional Memberships and Honors: American Psychological
Association, Society for Personality Assessment, Ford Foundation
Predoctoral Fellow (1990-93), OSU Graduate College Leadership
Award

OKLAHOMA STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
FOR HUMAN SUBJECTS RESEARCH

Proposal Title: ANALYSIS OF AN EMPIRICAL METHOD OF TEST CONSTRUCTION

Principal Investigator: ROBERT S. SCHLOTTMANN/ JONI L. MIHURA

Date: 9-10-92 IRB # AS-93-010

This application has been reviewed by the IRB and

Processed as: Exempt [] Expedite [X] Full Board Review []

Renewal or Continuation []

Approval Status Recommended by Reviewer(s):

Approved [X]

Deferred for Revision []

Approved with Provision []

Disapproved []

Approval status subject to review by full Institutional Review Board at
next meeting, 2nd and 4th Thursday of each month.

Comments, Modifications/Conditions for Approval or Reason for Deferral or
Disapproval:

COMMENT:

Please remove LeAnn Prater's name as IRB contact person in informed
consent form. No name need be inserted, just list the IRB office
and phone #.

On the volunteer information statement, participation time is stated as
1/2-2 1/2 hours. The informed consent says 2-3 hours - needs to be
consistent.

Signature: Maria L. Tilley

Chair of Institutional Review Board

Date: 9-17-92