© 1984 David Brazeau Kelly

DETECTION OF FAKING GOOD ON THE MMPI IN A
PSYCHIATRIC INPATIENT POPULATION

by

DAVID BRAZEAU KELLY, B.A., M.S.

A DISSERTATION

IN

PSYCHOLOGY

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

August, 1984

ML
801
T3
1954
No. 42
cop. 2

ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to Dr. Roger Greene, chairman of my dissertation committee, for his support and guidance throughout the dissertation process. I would also like to thank Drs. Clay George, Jim Clopton, Jane Winer, and Jim Maddux for their helpful suggestions.

I want to express my appreciation to the staff members of Big Spring State Hospital for their assistance in the data collection. Special thanks are extended to Dr. Harry Davis, Dr. Rick Stafford, Ruth Payne, Connie Edgemon, Susan Welch, Kathy Downey, and Marilyn Clark. Without their help this study could not have been completed.

Finally, I would like to thank the members of my family for their continuing moral support.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES	vi
I. REVIEW OF THE LITERATURE	1
The Traditional Validity Scales	2
Cannot Say scale	3
Lie scale	4
F scale	6
K scale	9
Consistency of Item Endorsement	10
Additional Validity Indicators	12
F-K scale	13
Mp scale	14
Dissimulation scale	14
Subtle and obvious scales	15
Social desirability indexes	18
Faking Bad Studies	21
Faking Good Studies	22
Role suggestion studies with students	23
Role simulation studies with students	28
Social desirability instructions with students	32
Studies using non-student populations	36
Studies using naturalistic situations	41
Summary of Fake Good Studies	45
Methodological Considerations	47
Summary of Methodological Consider- ations	52
Statement of Research Question	52
Hypotheses	53
II. METHOD	54
Subjects	54
Materials	58
Procedure	58

III. RESULTS	60
Comparison of MMPI Profiles	60
Major Experimental Results	62
Supplemental Analyses	69
<u>L</u> , <u>F</u> , and <u>K</u> scales	69
Hit rates	69
IV. DISCUSSION	74
Discussion of Hypotheses	74
Supplemental Analyses	79
Comparison of the <u>L</u> , <u>F</u> , and <u>K</u> scales	79
Comparison of hit rates	79
Limitations and Suggestions for Future Research	81
Conclusions and Implications	83
REFERENCES	86
APPENDICES	95

LIST OF TABLES

1.	DESCRIPTIVE DATA	55
2.	DIAGNOSES OF PATIENTS	56
3.	CALIFORNIA ACHIEVEMENT TEST VOCABULARY SECTION SCORES	57
4.	SUMMARY OF DESCRIPTIVE STATISTICS FOR THE <u>Mp</u> AND <u>X-0</u> SCALES	64
5.	SUMMARY OF ANOVAs FOR THE <u>Mp</u> and <u>X-0</u> SCALES . .	65
6.	POST HOC COMPARISON BY ORDER X INSTRUCTIONS FOR THE <u>Mp</u> SCALE	68
7.	POST HOC COMPARISONS BY INSTRUCTIONS FOR <u>X-0</u> . .	68
8.	SUMMARY OF DESCRIPTIVE STATISTICS FOR THE <u>L</u> , <u>F</u> , AND <u>K</u> SCALES	70
9.	SUMMARY OF ANOVAs FOR THE <u>L</u> , <u>F</u> , AND <u>K</u> SCALES . .	71
10.	POST HOC COMPARISON BY ORDER FOR THE <u>L</u> SCALE . .	72
11.	POST HOC COMPARISONS BY INSTRUCTIONS FOR THE <u>L</u> , <u>F</u> , AND <u>K</u> SCALES	72
12.	SUMMARY OF MANOVA FOR STANDARD PROFILES AND FAKE GOOD PROFILES	101
13.	SUMMARY OF ANOVAs FOR MMPI SCALES	102
14.	FREQUENCY OF HIGH POINT PAIRS	103
15.	<u>F</u> -max FOR THE <u>Mp</u> , <u>X-0</u> , <u>L</u> , <u>F</u> , AND <u>K</u> SCALES . . .	104
16.	SUMMARY OF MANOVA FOR THE <u>Mp</u> and <u>X-0</u> SCALES . .	105
17.	SUMMARY OF MANOVA FOR THE <u>L</u> , <u>F</u> , AND <u>K</u> SCALES . .	106
18.	HIT RATES FOR THE <u>Mp</u> and <u>X-0</u> SCALES	107
19.	HIT RATES FOR THE <u>L</u> , <u>F</u> , AND <u>K</u> SCALES	108
20.	RAW DATA FOR THE CLINICAL SCALES	109
21.	RAW DATA FOR THE VALIDITY SCALES	110

LIST OF FIGURES

1.	MMPI median profiles by condition	61
2.	Median standard profiles for inpatient subjects and a non-subject inpatient sample	63
3.	<u>Mp</u> scale scores by condition	67

CHAPTER I

REVIEW OF THE LITERATURE

The Minnesota Multiphasic Personality Inventory (MMPI) was developed by Hathaway and McKinley (1940) to aid in the assessment of abnormal behavior. The MMPI consists of 566 descriptive statements to which an individual responds "true" or "false." A series of 10 clinical scales were empirically derived from the items and the relative scores achieved on these scales provide the basis for generating a personality description of the individual. In addition to the clinical scales, several validity scales were developed to assess the person's test-taking attitudes. The concept of validity has traditionally referred to the degree to which a test actually measures what it purports to measure. When used in the context of the validity scales on the MMPI, however, validity describes the test-taking attitude of the examinee and whether the person has given an accurate and consistent self-appraisal in endorsing the items.

Research concerning the detection of distortion in MMPI profiles has continued over the past 35 years. A number of validity indicators, which function with varying degrees of success, have been developed. Studies continue to be carried out in this area since the ability to detect distorted MMPI profiles is still less than satisfactory, especially

with defensive or fake good profiles. Since administration of the MMPI is a standard component of many psychological evaluations, it is crucial to develop sensitive validity indicators and be aware of their accuracy with various populations.

The Traditional Validity Scales

Patients can distort responses in such a way as to either appear better adjusted than they really are or less well adjusted than in actuality. A patient might feign psychopathology ("fake bad") for a number of reasons such as to receive disability payments, to obtain treatment more quickly, or to escape responsibility for past actions. Conversely, some patients may want to feign good adjustment ("fake good") for reasons such as gaining release from hospitalization, retaining custody of children, or to pass employment screenings. In addition to consciously faking good, some patients may characteristically deny problems without consciously intending to deceive the examiner. Validity scales are also intended to detect this type of responding. The term "fake good" will be used to describe a test-taking set in which the person is consciously trying to appear better adjusted than is actually the case. The four MMPI validity indexes which have been most commonly used are the Cannot Say (?) scale, the Lie (L) scale, the F scale, and the K scale.

Cannot Say scale. The Cannot Say (?) scale score is the total number of unanswered or double-marked test items. As Greene (1980) noted, it is not a "scale" in the sense of L, F, or K since no specific items comprise it. In many research studies, subjects are told not to leave items blank since this decreases the validity of the results. Lachar (1974) presented a number of possible reasons why a patient might obtain a high Cannot Say score: reading difficulties, poor comprehension, marked confusion, an obsessive or ruminative approach, a highly defensive response set, an intentional resistance to cooperation, or psychomotor slowing in cases of depression. More research needs to be implemented to determine the most common reasons why clients in particular populations omit items. Investigations concerning whether omitted items are usually a result of defensiveness (Dahlstrom, Welsh, & Dahlstrom, 1975; Tamkin & Scherer, 1957) have not been empirically substantiated.

The T-scores for Cannot Say listed on the standard profile sheet were not derived from any normative group, but rather were estimated using clinical judgment (Hathaway & McKinley, 1967). Clopton and Neuringer (1977) investigated the frequency with which items were omitted and also the degree of profile distortion which occurred as a result of omitting various numbers of items in a random fashion. The MMPI records of Veterans Administration (VA) psychiatric

patients, outpatients from a regional mental health center, and applicants for jobs as firefighters and police officers were examined. It was found that the great majority of people in the three groups completed all or nearly all of the items. The job applicants as a whole yielded the greatest proportion of answer sheets with all items completed. The VA group produced the highest mean number of omissions with only 4.68, far below the 30 omissions which Hathaway and McKinley placed at the 50th percentile. These findings were in agreement with other studies (Ball & Carroll, 1960; Greene, 1980; Rosen, 1958; Tamkin & Scherer, 1957) in which only a small percentage of those taking the test omitted 30 or more items. Clopton and Neuringer also observed that omitting 30 items can cause a change in the high point pair for as many as 25% of the profiles. They suggested that 5 omitted items should be set at a T-score of 50 and that 30 omitted items be set at a T-score of 60 to indicate that some caution needs to be exercised in interpreting the profile, even though it may not be considered invalid.

Lie scale. The L scale consists of 15 items which were adapted from Hartshorne and May's (1928) research. These items, which were selected on a rational basis, consist of highly socially desirable behaviors to which few individuals can truthfully admit. Persons who answer a large number of L scale items in the deviant direction might be attempting

to portray themselves in a desirable manner. It is generally thought that the type of faking which occurs on the L scale is not very sophisticated, since most test-takers would be aware that few people could truthfully endorse many such items in the deviant direction. The L scale, nevertheless, has demonstrated moderate effectiveness in discriminating fake good from honest profiles in certain populations (Exner, McDowell, Pabst, Stackman, & Kirk, 1963; Match & Wiggins, 1974; Wiggins, 1959).

Marks, Seeman, and Haller (1974) reported that the mean number of L items answered in the deviant direction by normals is four, and it is rare to answer seven or more deviantly. Gravitz (1970) examined the records of over 11,000 adults to assess whether the L scale actually measured socially desirable but unlikely behavior. He found that large numbers of adults (ranging from about 20% to 85% in various groups) endorsed more than half of the items in the deviant direction. He also found that all but two of the items were differentially answered by males and females, which suggests that separate scales are needed. It should be noted that the increased number of L scale items endorsed in the deviant direction in Gravitz's study could be explained by the fact that the subjects were tested as part of the process of applying for vocational positions. In such circumstances it is likely that they would want to portray themselves very favorably.

The L scale, like the other validity scales, is thought to have additional value in identifying persons with certain personality characteristics. Marks et al. (1974) stated that rigidity, naivete, and lack of insight are frequent descriptions of persons scoring high on the L scale. Coyle and Heap (1965) maintained that an additional group in which a high L was often seen was in paranoid persons with pronounced grandiosity, but this hypothesis has not received empirical support. Burish and Houston (1976) found that persons who scored high on the L scale performed better under stress than those whose scores were lower. Matarazzo (1955) also observed that male medical students with low anxiety levels tended to have higher L scores than high anxiety students.

The L scale has thus shown some usefulness, but is not considered to be one of the most sensitive validity indicators. It may not detect more sophisticated forms of faking since the items are obvious in nature. The L scale does, however, provide some important descriptive information about the patient such as degree of psychological mindedness and use of denial.

F scale. The Frequency or F scale consists of 64 items which tap varying degrees of maladjustment. The F scale is designed to detect an atypical manner of responding to the

MMPI items. Only 10% of the normal sample in the standardization group endorsed a given F scale item in the deviant direction. Since the scale was constructed empirically, the items are not particularly homogeneous in content. Typical items are concerned with peculiar beliefs, unusual experiences, social isolation, and apathy. Less than 50% of the items, however, appear on the clinical scales. Elevation of the F scale could indicate scoring errors, intellectual or reading difficulties, atypical response sets (all true, all false, or random answering), an attempt at malingering, or an exaggeration of actual symptoms as in a "cry for help" (Lachar, 1974).

Profiles with a score greater than 16 on the F scale were originally considered invalid (Hathaway & McKinley, 1951). Numerous studies since that time, however, have found that F scores above 16 often yield accurate personality descriptions (Blumberg, 1967; Gynther, 1961; Gynther & Shimkunas, 1965b; Kazan & Sheinberg, 1945; Schneck, 1948). High F scores also have been shown to have some predictive validity for inpatient behavior. Post and Gasparikova-Krasnec (1979) found that hospitalized psychiatric patients who produced a "cry for help" profile (raw F score at least 11 points higher than raw K score) were perceived by the hospital staff as "acting out" more frequently than patients with normal or defensive validity configurations. Those

patients also accounted for a large proportion (77%) of the incidents of destructive behavior and were much more likely to be put in seclusion.

In a 7-year study of white hospitalized psychiatric patients, Gynther and Shimkunas (1965a) observed that F scores became lower with increasing age for those patients with high or low intelligence quotients but remained relatively constant for those with average intelligence. Sex of the patient, diagnosis, and educational level did not influence the F scores. Gynther, Altman, and Warbin (1973) demonstrated that racial factors should be considered when using the F scale. While descriptions of Caucasians with T-scores greater than 98 (raw scores greater than 25) were relatively accurate and consistent, no such characteristics could be found for black patients who scored in that range. A separate F scale for blacks has been developed (Gynther, Lachar, & Dahlstrom, 1978) which consists of items rarely endorsed by normal blacks.

In summary, the F scale items represent a broad range of abnormal behavior. Elevated F scores have demonstrated some predictive validity and accuracy in personality description. Further research needs to be performed regarding behavioral correlates to high scores with minority populations.

K scale. The K scale is comprised of 30 items which were empirically selected to aid in identifying defensive psychiatric inpatients who obtained normal profiles. It is now also used with non-inpatient populations as a measure of defensiveness. Fractions of the K scale are added to the raw score of five of the clinical scales in an effort to improve their validity. Most of the items on the K scale also contribute to the clinical scales. The item content assesses feelings about oneself, one's family, and relationships with others. The behavioral correlates often seen with a marked elevation of the K scale in a clinical population include general defensiveness, denial of problems, rigidity, and lack of personal insight (Lachar, 1974). Sweetland and Quay (1953) and Smith (1959) suggested that a moderate elevation of K in a normal population may be an indication of healthy emotional adjustment.

Studies with college populations (Heilbrun, 1963; Tyler & Michaelis, 1953; Yonge, 1966) have challenged the usefulness of the current K-corrections in that population. McKinley, Hathaway, and Meehl (1948) noted that college students generally obtain high K scores and suggested that this was a function of socioeconomic status. As Greene (1980) pointed out, the interpretation of a particular elevation of the K scale depends upon factors such as educational level and socioeconomic class of the client as well

as situational variables (e.g., is the client being tested for psychological treatment, screening for employment, research, etc.).

Several investigators have examined the utility of K-corrections in hospital settings. In a study by Hunt, Carp, Cass, Winder, and Kantor (1948) judges sorted inpatient profiles into psychotic and non-psychotic groups using both K-corrected and non K-corrected scores. They found no significant difference in predictive accuracy between the two types of profiles. Using an inpatient population, Sines, Baucom, and Gruba (1979) reported that judges' knowledge of the validity scale scores in K-corrected profiles did not significantly increase correct classification over profiles where the validity scales and amount of K-correction were withheld.

In summary, the K scale is used as a measure of defensiveness. The interpretation of a moderate elevation of K may be modified by factors such as socioeconomic status and reasons for testing. Although K-corrections have been used for a number of years, there is a growing body of literature which suggests that the corrections may not increase accuracy in all populations.

Consistency of Item Endorsement

There are several indexes which examine the consistency of item endorsement, which must be established before

assessing the validity. These indexes include the test-retest (TR) index and the Carelessness scale. The TR index is the number of 16 duplicate items on the MMPI that have been endorsed inconsistently by the subject. Buechley and Ball (1952) suggested that the TR index could be used to help rule out the possibility of random responding when a high F score is seen. A score of three or more on the index was suggested to be indicative of an invalid record.

Jones, Neuringer, and Patterson (1976) examined the TR index scores for brain-damaged and non-brain-damaged patients across various diagnostic categories (e.g., schizophrenic, alcoholic, neurotic). The brain-damaged patients showed the most inconsistency, but their TR scores were still low (overall mean of 2.81). Brain-damaged schizophrenics had the highest mean (3.59), while normals had the lowest (1.99). The only statistically significant comparison was between brain-damaged and non-brain-damaged patients. The authors concluded that because of the rarity of a significantly high TR score, the index has limited value in assessing validity. Greene (1979) gathered normative data on four populations: VA psychiatric patients, clients at a psychology clinic, adolescents on juvenile probation, and college students. The TR index for the juvenile probation group was higher than the other groups with a mean score of 4.14. The other three samples all had mean scores less

than 2. Overall, the TR index appears to be helpful in detecting random responding, at potentially clarifying the significance of elevated F scores, and as a signal that results may be distorted due to inconsistent responses.

The Carelessness scale (Greene, 1978) contains 12 pairs of items which were empirically selected based on the judgment that they were psychologically opposite in content. The optimal cutting score for detecting invalid profiles was determined to be four or more deviant responses. Greene (1980) suggested that the subtle nature of the index may enable it to detect more sophisticated patterns of response distortion. It also is helpful in identifying all true or all false response sets. A third condition which could elevate the scale is mental confusion. Further research is needed to determine the amount of error which occurs in profile interpretations of clients with high Carelessness scores.

Additional Validity Indicators

A number of alternative validity scales and indexes have been formulated in an attempt to increase the efficiency with which faking bad and faking good can be detected. Many of the indexes were developed using college students and have not been examined with psychiatric populations. Norman (1972) pointed out that the extent of generalizability of each validity index is not yet known. He offered as an

example a test which he designed that was very effective for detecting faking in Air Force training applicants but was much less successful for screening Peace Corp volunteers. He suggested that it may be necessary to derive a set of scoring keys, each of which is useful for a certain population and context. At the current time the research needed to construct such keys is lacking. In the present section, descriptions of a number of alternate validity indexes will be given together with data concerning their performances in the original studies. The further use of these scales in later investigations will be discussed in another section.

F-K scale. The F-K scale is commonly used by clinicians to determine the validity of obtained profiles. Gough (1947, 1950) suggested that subtracting the raw score on the K scale from the raw score on the F scale would successfully discriminate a high percentage of faked records from authentic ones. Gough (1947) had 11 mental health professionals take the MMPI under three conditions: honestly, attempting to simulate severe neurotic pathology, and attempting to simulate paranoid schizophrenia. Using the F-K index, 82% of the feigned profiles were discriminated from actual neurotic profiles which were similar in clinical configuration. The success rate for detecting the simulated psychotic profiles using F-K was also 82%. Gough (1950) reviewed a number of studies concerning F-K and concluded that the index was more

successful at detecting fake bad profiles than ones that are fake good. He recommended a cutoff score of +9 for maximizing fake bad discrimination. Other researchers (Anthony, 1971; Drasgow & Barnette, 1957; Exner et al., 1963) have found different F-K cutoff scores to be optimal with various populations.

Mp scale. The Mp scale (Cofer, Chance, & Judson, 1949) contains 34 items and was developed to detect fake good profiles. It was formulated based upon an item analysis which indicated that these items are sensitive to positive distortions but not to negative ones. Cofer et al. speculated that the items represent "things which a majority of people have to admit about themselves, although they are aware that these matters have a negative social value" (p. 497). Since they are endorsed in the socially undesirable direction by most people, faking bad would not change the responses to these items, but faking good would produce a change.

Cofer et al. found that the Mp scale had a high detection rate with college students who were attempting to fake good. Using the optimal cutting score, the Mp scale identified 96% of the honest student records and 86% of those which were fake good.

Dissimulation scale. The Ds scale consists of 40 items which are sensitive to fake bad response sets. Gough (1954)

constructed the Ds scale by contrasting the item endorsement of neurotic patients with those of normals who were simulating neurosis. Gough concluded that there was a great degree of misinformation about neurosis which could be used as an indicator of faking bad. Upon cross-validation, Gough found that 93% of the dissemblers scored at or above the optimal cutoff score as compared to only 6% of the clinical cases and 2% of a normal student population taking the test honestly.

Subtle and obvious scales. Weiner (1948) identified a set of subtle and obvious items for the MMPI. Items can be classified into those which are easily detectable as indicating psychological disturbance (obvious items) and those which do not seem especially pathological in content (subtle items). The subtle and obvious scales are a result of the empirical method of test construction where items may be seen to relate to a particular behavior empirically but not content-wise. Weiner determined empirically that the deviant response for a high percentage of the subtle items was in the direction opposite of that which would be rationally expected from the content. Subtle and obvious subscales were derived for five of the clinical scales. In looking at the subtle and obvious items in honest and fake good profiles of students, Weiner found that on the fake good tests there

was a decrease in obvious scores and an increase in subtle scores.

In addition to Weiner's scales, two other sets of subtle and obvious scales have been derived which are similar but not identical. Wales and Seeman (1968), in attempting to develop a reliable index for detecting fake good profiles, suggested that the use of a combination of zero items and X items might be effective. Zero items are a subset of subtle items which are scored in the deviant direction even though the subject answers in the manner characteristic of the majority of normals. The reason for this paradoxical scoring is that a particular item may be endorsed by 60% of the normal population and 90% of a particular pathological population. It is therefore more likely that a person in the pathological group will endorse it than a normal person. X items are answered by a minority of subjects and are endorsed in the direction keyed for psychopathology. In a study using undergraduates in a fake good test-retest design, Wales and Seeman (1968) were able to identify 100% of the falsified profiles and 71% of the honest profiles using an index of the number of X items endorsed minus the number of zero items endorsed. There were methodological problems in this study, however, because there was no counterbalancing of administration to rule out order effects. As will be discussed later, X and zero items have received support as

relatively effective indicators of faking good in several studies.

Christian, Burkhart, and Gynther (1978) formulated another set of subtle and obvious items by having university students rate each MMPI item on a 5-point scale of subtle-obvious. This method was employed in order to determine whether items were subtle or obvious from a psychologically naive point of view. Another benefit was that very subtle and very obvious items could be differentiated from those that were less extreme. Burkhart, Christian, and Gynther (1978) examined the pattern of subtle and obvious endorsements under fake good and fake bad response sets. Their findings were consistent with research (Fortova, 1979; Vesprani & Seeman, 1974; Wales & Seeman, 1968) using the other sets of subtle and obvious items in which the endorsing of obvious items was directly related to the instructional set, while endorsement of subtle items was an inverse function. In other words, under fake good instructions the endorsement of obvious items decreases and endorsement of subtle items increases. Under fake bad instructions, relatively more obvious and less subtle items are endorsed. While continued research has been performed concerning the predictive validity of the subtle and obvious items upon individual clinical scales (Burkhart, Gynther, & Fromuth, 1980; Gynther, Burkhart, & Hovanitz, 1979; Hovanitz & Gynther,

1980), no studies have examined the ability of Christian et al.'s (1978) subtle-obvious items to discriminate faked profiles from honest ones.

The controversy over the relative value of subtle and obvious items has continued for many years. Meehl (1945) and Seeman (1952) described the lack of transparency of many MMPI items as one of the strengths of the MMPI. Others such as Jackson (1971) criticize empirically based tests and reject the subtle items as lacking any predictive validity. Several studies (Duff, 1965; Wrobel & Lachar, 1982) have demonstrated experimental evidence that obvious items have much more predictive and interpretive validity than subtle items. In addition to predictive capabilities, subtle and obvious items seem to hold promise for detecting faking on the MMPI, but much more research needs to be performed.

Social desirability indexes. A number of measures of social desirability have been developed and subsequently employed in attempts to detect falsification. The underlying assumption in these studies is that the test-taking attitude and behavior, whether honest or faking, can be characterized by a "readiness or lack of readiness to respond to socially desirable items" (Fordyce, 1956, p. 174). Hanley (1957) derived a scale to measure test-taking defensiveness by selecting 53 items which had been endorsed by 36% to 64% of Hathaway's normative group of college students.

The items were then judged by psychology students on a 9-point scale of desirability. The items which were judged to be either desirable or undesirable (rather than neutral) formed the 26-item Ex scale and were keyed in the desirable direction. Hanley hoped to create a scale of items which had a low correlation between their social desirability and their probability of endorsement. He did not test the scale's potential to detect faking, but several later researchers (e.g., Baker, 1967; Wiggins, 1959) did employ it in that regard.

The Social Desirability (SD) scale (Edwards, 1957; Fordyce, 1956) is another index that was not intended to be a validity scale, but has been used in that manner. Fordyce presented 149 items consisting of the F, K, and Taylor Manifest Anxiety scales to 10 judges. The judges were asked to select those items which would be endorsed by all test subjects because of their socially desirable content. The SD scale was made up of the 79 items upon which all judges agreed as to selection and direction of scoring. It was subsequently reduced to 39 items (Edwards, 1957) which are keyed in the desirable direction.

Wiggins and Rumrill (1959) developed the Sd-A and Sd-R indexes based upon the Welsh (1956) factor scales of anxiety (A) and repression (R). Scale A has a heavy loading on the general maladjustment factor of the MMPI, while scale R

parallels the repression factor. Wiggins and Rumrill had 181 undergraduates rate the social desirability of the items of these two scales. The items from A tended to have low social desirability while those from R were more neutral. The Sd-A consists of 39 items keyed "true" and the Sd-R contains 40 items keyed "false." As will be discussed later, the discriminative ability of these indexes with falsified records was tested in several studies (Baker, 1967; Wiggins, 1959).

Another index of social desirability, the Sd scale (Wiggins, 1959), was developed upon a sample of 440 undergraduate men. Some of the subjects were instructed to answer honestly, while others were to answer in the socially desirable direction. The scale consists of 40 items which discriminated the experimental from the control group. As would be expected, Sd outperformed all other validity scales in the sample upon which it was developed. It also has been relatively successful in some of the other studies in which it has been used (e.g., Boe & Kogan, 1964; Skrzypek & Wiggins, 1966).

The Desirability (Dy) scales were derived by Jackson and Messick (1961) in order to assess both the tendencies to acquiesce and to respond desirably. There are five Dy scales which vary systematically in their level of social desirability, but vary randomly in content. Using previously

computed desirability values (Heineman, 1953), each MMPI item was placed into one of five categories ranging from extremely desirable to extremely undesirable based upon the item's average desirability rating. Overlap between each category and the clinical scales was limited to six items. The category of highest desirability was labeled Dy 1, the lowest was Dy 5, and Dy 3 was a scale of items judged neutral in desirability. Jackson and Messick did not utilize the Dy scales to detect faking, but Dies (1968) did apply them for that purpose, as will be discussed later.

Faking Bad Studies

Considerable research has been devoted to the detection of fake bad profiles. While the majority of these studies have used college students (Cofer et al., 1949; Exner et al., 1963; Harvey & Sipprelle, 1976; Lanyon, 1967), some research with psychiatric patients and prison inmates also has been performed (Anthony, 1971; Hunt, 1948; Johnson, 1982). Most of the research (Anthony, 1971; Exner et al., 1963; Grow et al., 1980; Hunt, 1948) has indicated that the F scale, the Ds scale, and F-K are all effective discriminators of fake bad from standard profiles. The optimal cutting scores for F and Ds have been relatively consistent across studies. It appears that for the F scale a cutting score of +14 or +15 will produce good results and for the Ds scale a +20 or +21 raw score is the optimal cutting point. The F-K index

has shown less stability of cutting scores and for this reason may not be as effective as the other two scales for general clinical use.

At the present time, the ability to detect faking bad on the MMPI in most populations appears to be adequate. In contrast, the detection of faking good is still unsatisfactory. There are several promising indexes but there is a lack of research which establishes their efficacy with psychiatric groups. It seems evident that future research in the area of faking needs to be focused more upon fake good indexes.

Faking Good Studies

A number of investigations have been performed concerning the ability of validity indicators to detect fake good response sets. Because of the widespread use of the MMPI for clinical purposes in both inpatient and outpatient settings, it is important that a maladjusted person cannot feign good emotional adjustment without being detected. Gough (1950) concluded that the traditional validity indexes were more successful at detecting fake bad profiles than fake good. Since that time, various other indicators have been developed and examined with different populations. The most commonly used population has been students.

Researchers have employed several different types of fake good instructions. A frequent practice is to instruct

the subjects to pretend that they are applying for a very desirable job and that it is necessary to portray themselves in the most favorable light to maximize the probability of being hired. This approach could be called "role suggestion" since a situation is suggested to the subject, but there is no detailed description of the specific types of behavior to endorse. Instructions which are very similar to this would be to portray "your ideal self" or to "try to make a very good impression." A second type of instruction is role simulation. In this case the subject is instructed to portray a role such as "Air Force officer" (Kroger & Turnbull, 1975) and answer as that type of person would. A description of a typical representative of that role is often read one or more times before and during the test administration to remind the subjects of the characteristics that they are supposed to present. A third type of instruction is for the subject to answer in the "socially desirable" way or in the manner that "most people would consider desirable." It has not been established which of these instructions produces responses most similar to those of persons feigning good adjustment in clinical situations. In examining the studies using these different approaches, however, it appears that the instructions do produce differences in the results.

Role suggestion studies with students. Vincent, Linsz, and Greene (1966) instructed 100 college students to attempt

to "make a good impression" by hiding any unfavorable qualities while taking the MMPI. The students also were told to circle those items which they falsified. The experiment was designed specifically to look at the utility of the L scale and F-K. While the L scale detected the majority of falsifiers, 10% of the students were able to distort their MMPI results considerably without being identified by either the L or F-K scales. Another 5% were undetected in a lesser degree of distortion. Vincent et al. concluded that the L scale items should be more subtle since in follow-up interviews, the successful students stated that they tended to avoid endorsing the most obvious items.

Hiner, Ogren, and Baxter (1969) had students take the MMPI under standard instructions, and then as their ideal selves. They compared the ideal profiles produced by the students with those obtained in several other fake good studies and concluded that there was a typical "ideal self" profile which might alert clinicians to faking. They also looked at the subtle and obvious items (Weiner, 1948) and found that under the ideal self condition, fewer obvious items were endorsed, but an increased number of subtle ones were endorsed. While Hiner et al. did not specifically examine the ability of the traditional validity scales to discriminate between the two types of responding, both the L and K scales were highly elevated in the mean ideal self profile.

The bulk of studies with students use instructions in which the subject is asked to imagine that he or she is applying for a job or in some other circumstance where it would be desirable to appear well adjusted. Cofer et al. (1949) tested college sophomores under an honest and a fake good condition. Those who were faking were asked to imagine themselves as interested in entering midshipman training in the Navy and attempting to produce a profile which would make a good impression. Cofer et al. inspected the results and concluded that L+K (sum of the L and K raw scores) might be an effective index for detecting fake good profiles. Using the L+K index they had moderate success in identifying false profiles, but the K score alone was slightly more successful. Cofer et al. also performed an item analysis which revealed a group of 34 items (the Mp scale) which were found to differentiate fake good records from honest ones with a high degree of accuracy. As will be discussed, the Mp scale has continued to be relatively successful across a number of studies.

Exner et al. (1963) examined the effectiveness of several indicators, including L+K, to discriminate fake good MMPIs from those taken honestly. Male and female college students were asked to respond to the items as they might if they were a job or school applicant. They were asked later to take the test again anonymously and be honest in their

answers. Exner et al. found considerable overlap between the honest and fake good scores for the F-K and L+K indexes. A number of other linear combinations of L, F, and K plus some clinical scales also were tried, but none was successful at greater than a 65% accuracy. Exner et al. concluded that faking good is relatively undetectable and that the L+K index used by Cofer et al. was not supported. Cofer's Mp scale was not tested in this study.

Wales and Seeman (1968) did score the Mp scale as well as looking at the usefulness of zero and X items to detect faking in male and female undergraduates. One group first took the MMPI under instructions to answer as if they were applying for a desirable job and wanted to make a good impression. They were administered the MMPI a second time under standard instructions. The other group was tested under standard instructions both times. Using X minus zero with a cutoff score of -4, 100% of the fake good records and 71% of the honest records were detected. The Mp scale, using Cofer's cutoff score of +20, detected 68% of the fake good records and 94% of the honest records. Wales and Seeman (1968) also examined multiple cutoff scores for X and zero items, and the use of zero scores alone, but these were not as successful as the X minus zero index. They concluded that both the X minus zero and the Mp scales can be used successfully to discriminate fake good from honest profiles

with students. The lack of a counterbalanced administration for the experimental group may be a methodological flaw since Wales and Seeman (1972) found that order effects influenced scores of zero and X items.

Several other studies examined the behavior of various types of subtle and obvious items under conditions of falsifying. Harvey and Sipprelle (1976) instructed undergraduate students to pretend that they were applying for a very desirable job. This group was compared with subjects who were faking bad. As expected, the Weiner subtle scores were higher than the obvious ones for the fake good group and the reverse was true for those who were faking bad. Burkhart, Christian, and Gynther (1978) utilized a 5-point scale of subtle-obvious which they had previously developed (Christian et al., 1978) in a test-retest situation where students took the MMPI both honestly and under fake good instructions. They found that for the fake good records, the very subtle and somewhat subtle items were endorsed more frequently than under standard instructions. The neutral, somewhat obvious, and very obvious items were endorsed with significantly more frequency by those taking the test honestly.

Grow et al. (1980) examined a number of validity indicators with a student population. In the fake good condition, the students were instructed to imagine a situation in which it was to their advantage to appear completely

normal such as applying for a good job, release from a mental hospital, and early release from prison. The control group students took the test anonymously and were instructed to answer truthfully. The index which was most successful was $F-K \leq -11$, which was able to identify 96% of the faked records and also detect 62% of the honest ones. Other validity indexes which were examined included the Mp scale, Weiner's subtle and obvious scales, and L+K. The Mp scale identified 46% of the fake good records and 96% of the honest ones. The remaining scales were not successful in discriminating between the two groups. A shortcoming of Grow et al.'s study is that cutoff scores recommended by earlier researchers were used rather than calculating an optimal cutoff. A low detection rate may be due to an inaccurate cutting score rather than to insensitivity of the scale.

Role simulation studies with students. Role simulation usually involves giving the subject a detailed description of the type of role to be played unless there already is some familiarity with the role. The description of the role to be simulated is much more developed than in the "role suggestion" studies. Some researchers (e.g., Kroger & Turnbull, 1975) believe that role simulation more accurately characterizes the actual processes of a person who is faking on the MMPI in a clinical situation. They maintain that the

patient takes on a role and answers in a manner consistent with it rather than just denying pathology.

Gough (1947) and Lanyon (1967) used forms of role simulation in fake bad studies, but relatively few studies of faking good have been performed which utilize this type of instruction. Lawton (1963) tested high school and college students with the items of the Pd scale, K scale, and the Taylor Manifest Anxiety Scale. Each student first took the test honestly, then simulated a delinquent who was trying to cover up his delinquency. Finally, the student was asked to take the test again and respond in the manner of a delinquent who was being honest. Lawton gave the students a clinical description of a sociopathic personality in order to clarify the role that the student was to play. In the fake good instructions, the student was to attempt to cover up any tendency to look like a "criminal type." It was found that both groups of students, regardless of sex, were able to successfully manipulate the Pd scores. Intelligence and school level influenced success to some extent. Lawton concluded that the Pd scale could be faked. He admitted, however, that with a full MMPI it might be more difficult to separate out the items which are critical to the Pd scale.

Kroger and Turnbull (1975) instructed male undergraduates to attempt to simulate either an Air Force officer or creative artist without being detected. No well-developed

role description was provided to the students. The group that was to simulate the Air Force officer was able to achieve a mean profile similar to that of actual officers without being identified by the validity scales (using F, K, and F-K). Those students who were to simulate creative artists were not able to produce a profile resembling that of actual artists. It was hypothesized that perhaps they failed because of erroneous views or lack of knowledge about the personalities and lifestyles of creative artists. Kroger and Turnbull (1975) performed a second experiment in which a group of 19 male art students was tested both with the MMPI and a questionnaire dealing with the values, personality characteristics, and lifestyles of creative artists. From their responses on the questionnaire an accurate role description was generated, and the MMPI was used to obtain a correct profile for the artists. A sample of undergraduates was then divided into two groups, one of which was given the detailed role description of creative artists and the other of which was given only the same instructions used in the first experiment. The group which was given an accurate role representation was able to produce valid (using F-K) MMPI profiles similar to those of the art students. The group using the same instructions as in the first experiment yielded profiles similar to those obtained in that experiment.

Kroger and Turnbull suggested that traditional validity scales are unable to detect faking when role simulation is used. A limitation of this study is that all the subjects were male. The results also did not provide any information relevant to the question of whether patients who fake on the MMPI use a role simulation strategy. Previous studies (Kroger, 1967; Kroger & Turnbull, 1970), which did not use the MMPI, gave some evidence that situational variables generate a set of role demands for the subject which influence test responses. Kroger and Turnbull (1970) found that male students who were given the Strong Vocational Interest Blank and Taylor Manifest Anxiety Scale and were told that they were taking military officer tests answered in a significantly different manner than students who were told that they were taking tests of creativity. The first group tended to answer like military officers, while the second group's profiles resembled those of artists, even though no suggestion was made directly that the students should answer in a particular way. There is no published research which examines whether a similar influence occurs with the MMPI. At the present time, the significance of the role simulation studies in relation to actual faking of a patient in a clinical situation is still unclear.

Social desirability instructions with students. A

number of studies were performed based upon Edwards' concept of social desirability. Edwards (1964) maintained that the MMPI measures only the subject's "tendency to give socially desirable responses in self-description" (p. 306). The debate concerning this issue has not continued as strongly since 1965, when critics such as Heilbrun (1964) and especially Block (1965) successfully countered the social desirability argument.

The MMPI studies which instruct the subject to answer in a socially desirable manner assume that this process occurs when a person attempts to fake good on the MMPI. If faking bad, the person would answer most questions in the socially undesirable manner. Based upon this assumption, it would be logical to look at the effectiveness of social desirability scales to detect faking. Wiggins (1959) examined a number of validity scales, as well as Edwards' Social Desirability (SD) scale, with regard to their efficacy in detecting faking. Wiggins tested 440 undergraduates, some of whom took the MMPI under standard instructions and some of whom were told to answer in a socially desirable manner as "people in general" would consider to be desirable. The indexes which were most effective at differentiating the two groups were the Mp scale, the L scale, and Wiggins' Sd scale, which was constructed on this sample and therefore would give

inflated results. Edwards' SD scale was not successful and Wiggins concluded that it must not be an accurate measure of social desirability. The SD scale was found to correlate substantially with Fricke's (1957) response bias scale which Wiggins considered to be a measure of acquiescence. He suggested that the SD and some of the other rationally constructed scales actually were measuring acquiescence.

Hanley (1961) challenged Wiggins' use of the response bias scale and provided substantial evidence that the scale was influenced by social desirability and therefore was not an appropriate measure of acquiescence. He also maintained that Wiggins' instructions to the students might account for the accuracy of the Mp and Sd scales. Hanley argued that when instructed to answer all items in a socially desirable manner, the students will not distinguish between subtle items and obvious items. An actual client who is attempting to fake good in a reasonably sophisticated manner, however, may often refrain from endorsing many of the obvious L scale-type items in a socially desirable way. Hanley (1961) had students act as judges to select the most obvious items from the L, Sd, Mp, Ex, and K scales. Hanley concluded that the scales found to be most effective in Wiggins' study were those which were most transparent, and that their seeming accuracy was an artifact of the type of instructions used.

Walker (1962) criticized Wiggins' (1959) study on three points: (1) the SD scale was constructed to detect social

desirability on the MMPI under standard instructions, not necessarily under instructions to answer in a socially desirable response set; (2) the expectation that social desirability responses will increase substantially when the student is instructed to answer in that manner assumes that very little social desirability responding occurs under standard instructions; and (3) differences in the SD scale between the "standard" and "socially desirable" instructional sets may have been minimized due to a failure of the students to fully understand the instructions. Walker (1962) used instructions which were modeled upon those employed by Edwards in the development of the SD scale. This set emphasized to the students that they were supposed to create the most favorable impression that they could and gave examples for further clarification. Walker found that with these instructions there was significantly more of a shift in the SD score than in Wiggins' study, although still not as much as the shift seen for the Sd scale.

Wiggins (1963) maintained that Walker's results were due to a methodological flaw because a test-retest design was used without counterbalancing the order of administration. Wiggins had avoided this pitfall by not using a test-retest design. Skrzypek and Wiggins (1966) performed an experiment which examined the effects of test-retest (non-counterbalanced) upon social desirability scales using

Walker's instructions. The scores were in fact higher (except for the F scale) for the retest group than for the group which had not been tested previously. The differences reached significance for the L, K, Ex, and Sd-R scales but not for F, SD, Sd-A, Mp, and Sd. So while the test-retest design did influence the scores, it did not act to significantly inflate the SD scale. It also was observed that Walker's instructions did lead to higher social desirability scores than Wiggins' (1959), independently of any test-retest factors. In Skrzypek and Wiggins' (1966) study, a comparison of the ability of various validity indicators to discriminate a group which was initially tested with standard instructions from a group initially tested with Walker's instructions yielded a ranking of scale efficacy similar to the original Wiggins (1959) results. The three most efficient scales were Sd, Mp, and L, while the least efficient were K and SD.

Dies (1968) used the Dy scales (Jackson & Messick, 1961), five scales of varying desirability, to determine if they could be helpful in detecting faking. The category of highest desirability was labeled Dy 1, the lowest was Dy 5, and Dy 3 was a scale of items which were neutral in desirability. Dies used a counterbalanced test-retest design in which the experimental group took the MMPI once under standard instructions and once by assuming a socially desirable response set. He found that the success of the Dy scales in

identifying authentic tests ranged in order from 92.5% for Dy 1 to 75% for Dy 5, except for Dy 3 which showed no discriminatory ability (50%), as would be expected of neutral items. No percentages of correctly identified fake good records were given, so the actual discriminatory capacities of the scales cannot be judged. No other studies using the Dy scales since Dies were found in the literature.

Overall, the most powerful criticism of the social desirability studies in general is that of Hanley (1961), which seemed to be largely ignored by the other researchers. The use of instructions which tell the subject to answer in the manner which is most socially desirable is likely to encourage the subject to endorse obvious items which a reasonably sophisticated patient attempting to fake good would not endorse since they are too obvious. It is not surprising that the L scale performs better in the social desirability studies than in ones using other types of instructions. Because of this shortcoming, the relevance of this set of studies to clinical faking is very questionable.

Studies using non-student populations. Very little research has been performed investigating the efficacy of validity scales in outpatient psychiatric groups. Vesprani and Seeman (1974) examined the behavior of zero and X items under different response sets with 14 male and 14 female outpatients at a psychiatric clinic. A counterbalanced

test-retest design was used. During one session, the patients took the test under standard instructions and at a later session as their "ideal selves," i.e., to respond as they would like to feel. The hypothesis, based upon previous studies with other populations (Anthony, 1971; Wales & Seeman, 1968, 1969), was that under the ideal response set there would be an increase in the zero item scores and a decrease in the X scores on various clinical scales. They found that in nearly all of the clinical scales that contain X and zero items, there were significant changes in the predicted direction between the honest and ideal self response sets. The only exceptions were the X items on the K scale and the zero items on the Mf and Si scales. They suggested that studies which look at changes in the MMPI clinical scales under different instructional sets should examine the X and zero items for each scale since their shifts in opposite directions could offset each other and not be detected if only the T-scores of the scales were used.

Hunt (1948) compared Navy court-martial prisoners and male Army Specialized Training Program (ASTP) students in ability to fake successfully on the MMPI. The ASTP students took the test once honestly and once under instructions to conceal any abnormality, as if they were attempting not to be excluded from the military for psychiatric reasons. The

prisoners took the test anonymously under three conditions: standard instructions, fake good, and fake bad. The order of test-taking was counterbalanced in both the student and prisoner groups. In the fake good situation, the instructions for the prisoners were the same as for the students except that being restored to active duty was an additional goal to be envisioned while faking. Based upon clinical judgment, Hunt (1948) set the cutting score for F-K at -11 or below for fake good. He divided the prison group into two samples in order to cross-validate within that group. The F-K index was not successful at discriminating fake good from honest profiles. While 100% of the ASTP students' faked profiles were identified correctly, 93% of their honest profiles also were identified as faked. Since Hunt did not specify whether the ASTP students took the test anonymously (as the prisoners did), it is possible that at least some of the "honest" profiles were highly defensive in nature out of concern that an abnormal profile could lead to exclusion from the Army training program. For the first prisoner sample, the cutting score of -11 correctly identified 62% of the fake good profiles while misclassifying 18% of the honest profiles. With the second sample of prisoners, 55% of the faked profiles were identified and 12% of the honest ones erroneously classified. F-K was very accurate in identifying the successfully faked prisoner profiles

profile was assessed by a reduction in total T-score on the clinical scales. Seventy-three percent of the patients were able to improve the profile when simulating normalcy. Although a considerable number of patients "softened" their profile, 62% did not produce a change in diagnostic category. Only 11% were able to produce normal profiles. Grayson and Olinger found that the amount of improvement in the profile was a positive prognostic indicator for early hospital discharge regardless of the severity of the original profile. They caution, however, that it cannot be ruled out that early discharge may be a function of a patient's ability to feign improvement rather than actual improvement. Validity indicators were not examined in this study. In addition, order effects probably inflated the ability to simulate normalcy. Support for the relationship between the capacity to role play and improved prognosis also came from Bloom and Arkoff (1961) and Rapaport (1958), although the trends noted in these studies did not reach significance.

Another study of psychiatric inpatients was undertaken by Boe and Kogan (1964) in which patients were examined using "social desirability" type instructions. The purpose of the study was to re-examine the findings of Wiggins (1959) concerning the lack of effectiveness of the SD scale, as compared with the Sd and Mp scales, in discriminating standard from socially desirable profiles. A group of male inpatients

was tested with items from several of the validity indexes that had been previously used by Wiggins. The patients were mostly diagnosed as neurotic and non-brain-damaged psychotic. Only items from the scales to be scored were used. For each statement, the patients were to decide if it was socially desirable or undesirable and mark either "D" or "U." Patients were randomly assigned to either the experimental group or a control group. For the controls, standard MMPI instructions were given and the patients marked the answers true or false as the items applied to them. When optimal cutting scores were used, it was found that the Sd and Mp scales were still the most accurate discriminators between the groups but the difference between these indexes and the SD scale were not statistically significant. The optimal cutting scores also were not the same as in Wiggins' study, which supports the position that different cutting scores may be needed for different populations.

Studies using naturalistic situations. A number of studies have attempted to use subjects who were in a situation where the desire to fake good came from more internal motivations rather than external causes such as instructional sets. Maclean, Tait, and Catterall (1953) examined the MMPI records of 100 candidates for nursing who were randomly drawn from a larger population. The candidates had been tested as part of the procedure for applying to a nursing

program. It was assumed that they were motivated to present themselves in a positive light. Support for this assumption came from the fact that few T-scores on the clinical scales exceeded 60 and none exceeded 70. Maclean et al. specifically were looking at which cutting scores for F-K would identify the distorted profiles. The scores from the normal population of adults obtained by Gough (1950) were used for honest profiles. Maclean et al. suggested that a cutting score of -17 be employed since it would classify 31% of the nurse candidates as faking good and 12% of the normal sample. They speculated that much of that 12% may have been those with a tendency to dissemble rather than individuals who were especially well adjusted. The fact that no external criterion was used to assess faking and no control group of nurses was tested leaves the accuracy of the results of this study uncertain.

Johnson et al. (1977) tested veterans requesting treatment at a VA hospital using an MMPI and a structured mental status exam to determine whether the validity indicators would be helpful in predicting faking during the structured interview. No significant relationships were found between the validity indicators and a judgment of "trying to appear as favorably as possible" on the mental status exam. There was a relationship between both F and F-K and ratings of "faking bad" in the interview. The authors, however,

thought that the correlation was not great enough to warrant the time and effort to use the MMPI to predict interview faking. Johnson et al. assumed that the observations during the structured interview were accurate and therefore used those as the criteria with which to compare the MMPI. It may be, however, that the source of error was in the clinical observations rather than the validity indicators.

Baker (1967) made an attempt to test a number of validity scales under conditions where the subjects would be assumed to have internal motivations for attempting to look as good as possible. She tested education students who were required to take the MMPI as a part of their program and could be dropped from the program if the results were overly pathological. There were two groups, one that took the MMPI under the "ego involved" condition (i.e., as a part of their program), then repeated the MMPI anonymously as part of a psychology class, and a group that took the tests in the reverse order. There was a 3-month interval between the test and retest. The critical methodological problem in the study was that in the "anonymous" testing the students were required to furnish information such as birthdate, home town, year in the program, major, and other identifying characteristics which would make it easy to trace their identities (which in fact Baker did with the information). It is likely that such an easily detected pretense of

anonymity only served to heighten the defensiveness and suspicion of the students. Defensiveness probably influenced the results most strongly in the group that took the ego involved test first. The students who initially took the "anonymous" test may not have been aware that the MMPI was the same test used for psychological screening later in their program. They would therefore be more likely than the first group to answer honestly. The test results support this view. For the group which took the ego involved MMPI first, there were only three clinical scales significantly lower than for the anonymous test, probably because of the defensiveness displayed in both testings. For the group which took the anonymous test first, seven clinical scales decreased significantly during the ego involved test.

Baker (1967) concluded that the second group distorted the ego involved responses more than the first due to order effects. She failed to consider, however, that distortion may have occurred in the first group because of defensiveness. The influence of order effects would probably not be too great since 3 months elapsed between test and retest. The methodological flaws are serious enough to warrant disregarding the data for the second testing of each group. Assuming that the groups were comparable, however, a comparison can be made between the first testing of each group. Baker compared the effectiveness of 11 validity scales to

discriminate between the two first testings, one of which was ego involved and the other anonymous. The most effective scale was Edwards' SD scale, followed by Wiggins' SD-A, and the K scale. The least effective were L, Sd-R, Mp, and Sd. Because the students were not randomly assigned to groups and there was no check made as to whether the groups were comparable, the results of this study need to be replicated with the methodological errors corrected before conclusions can be drawn.

Summary of Fake Good Studies

In general, no one indicator has been effective in the identification of fake good profiles across all studies. Several scales, however, performed well in the majority of studies in which they were examined. The most consistently successful validity indicator was Cofer's Mp scale. It was shown to be effective with college students using either role suggestion type instructions (Cofer et al., 1949; Grow et al., 1980; Wales & Seeman, 1968) or social desirability instructions (Skrzypek & Wiggins, 1966; Wiggins, 1959). Boe and Kogan (1964) also found it to be successful with psychiatric inpatients. Unfortunately, they used social desirability instructions which tend to produce more obvious faking than other types of instructions such as role suggestion. Although the Mp scale was not supported by all studies (e.g., Baker, 1967), it appears to be very promising

as an indicator of faking good. Further research needs to be performed, however, using psychiatric populations rather than college students.

The X minus zero index (Wales & Seeman, 1968) also has shown promise in detecting fake good profiles, although it has not been examined in many studies. The one study (Wales & Seeman, 1968) in which the percentage of correct classifications was calculated indicated that X minus zero is effective in discriminating between honest and fake good profiles. While the percentage of correctly identified honest profiles was less than that achieved by the Mp scale, the percentage of faked profiles detected was appreciably greater than for Mp. Wales and Seeman (1968) also found that the X minus zero index was more effective than either X or zero items used alone.

Several other studies (Vesprani & Seeman, 1974; Wales & Seeman, 1969) demonstrated that the totals of X and zero items endorsed in the deviant direction change significantly depending upon whether the subject is instructed to answer under standard conditions or as his/her "ideal self." This effect was seen in psychiatric outpatients and college students. The outcomes of these studies suggest that the use of X and zero items for detecting fake good profiles should be explored further.

Other indicators such as the Sd, L, and SD scales have demonstrated mixed results in the ability to detect faking

promise, there is consideration of the L scale, and Sd has not performed well. The L scale has achieved success where the faking has been

The SD scale has been very more reactive to the type of error scales.

Issues have arisen in the area addressed in any research. The role of order effects has been discussed in the previous review. Windle

(1955) found that subjects who took the MMPI twice within a 2-month period tended to show a shift in the "normal" direction regardless of whether any intervention such as therapy had occurred. Windle looked at the specific scales which declined and also interviewed subjects in order to determine the probable reason for the change. He concluded that the "improvement" was due to a decrease in test-taking anxiety which, in turn, caused some of the scales which are sensitive to anxiety to be lower. Most studies have attempted to deal with this effect through either counterbalancing the order of administration or not using a test-retest design.

Wales and Seeman (1972) also discovered that X and zero items were influenced by order of administration of the

MMPI. Three groups of nurses each were tested three times at 2-week intervals with the MMPI. There was a different instructional set at each administration: fake good, ideal self, or "yourself one year from now." The order of instructional sets was varied among the groups and the obtained $\bar{X}$ and zero totals were examined. The order of instructional sets had a strong effect on the zero scores and a lesser (but still significant) effect on the $\bar{X}$ scores. The authors suggested that these influences had not been seen previously because the $\bar{X}$ and zero score changes tend to offset each other by shifting in opposite directions. The $\bar{X}$ and zero items for the "one year from now" instructional set showed especially disparate scores depending upon whether it was the first set used or followed one of the other two. It was concluded that more attention needed to be accorded to this phenomenon when $\bar{X}$ and zero items were being examined in designs using different instructional sets.

A factor which has not been assessed to any great degree is the type of instructions used. The literature suggests that social desirability instructions tend to produce the most easily detected fake good records and that role simulation instructions produce records that are much more difficult to identify as fake. Role suggestion instructions seem to fall in between these two extremes. A more important question, however, is which of these instructions

induce subjects to answer in a manner similar to that of an actual patient in a clinical situation, where the motivation for faking is more internal. Kroger and Turnbull (1975) maintain that role simulation is the process which is used by patients attempting to fake, but little evidence is available to support this view. Role simulation is a complex process which would require that the patient be able to consciously adopt an unfamiliar role and play it consistently. There is some question as to whether most patients could successfully accomplish such a task. The shortcomings of social desirability instructions, on the other hand, have been mentioned already. They tend to produce extreme faking which is detected by even the less subtle indexes such as the L scale. The role suggestion instructions, such as "pretend that you are applying for a desirable job and want to make a good impression," seem to be more "true to life" than the other types of instructions. The subject is able to employ a form of self-reference, rather than playing a totally different role, but does distort aspects of him/herself in such a manner as to appear more favorably. At the same time the subject is not prohibited by the instructions from using a more sophisticated, less obvious form of faking. Overall, it appears that the role suggestion type of instructions are the most likely to produce responses similar to those seen in actual faking situations.

Another important issue is whether the results from investigations using students can generalize to other populations. At the present time it appears questionable that validity cutoff scores can generalize from one population to another very well. Some scores (e.g., +20 for the Mp scale) have had better generalizability than others. However, further research needs to be performed in this area. A related question concerns studies which instruct normal subjects to fake good. Since the group already obtains scores predominantly within the normal range under standard instructions, it is doubtful whether the results of fake good studies which use this population can generalize to pathological groups.

Variables such as race, age, intelligence, sex, and education also need to be considered as possible influential factors. Gynther and Shimkunas (1965a) examined the F scores of 400 hospitalized psychiatric patients over a 7-year period. No significant sex differences were found. They did observe that F scores tended to decrease with age for those with low or high IQ's, but remained consistent for those with average IQ's. A number of studies (Ball, 1960; McDonald & Gynther, 1962) have found that blacks generally score higher than whites on the F scale. Such a difference could obviously have ramifications for the behavior of validity scales in detecting faking with blacks. Johnson

(1982) tested the efficacy of various validity indexes in detecting faking in black and white prisoner groups. All of the validity scales (L, SD, Sd, and K) were about 70% successful at detecting fake good and honest records in the white group. None of the indicators, however, was effective with the black group at a significantly greater than chance level. The validity indicators (F, Ds, and Weiner obvious scale) were about 90% successful in identifying fake and honest profiles of the white prisoners who were faking bad. In the black group the hit rate ranged from 80% to 90%, with the F scale being the most successful at 90%. It appears then, that racial differences are an especially salient issue in fake good studies and less of an issue in fake bad investigations.

Burkhart, Gynther, and Christian (1978) examined the relationship of psychological mindedness (as measured by the California Personality Inventory) and intelligence with the amount of subtle item endorsement under standard and faking (good or bad) conditions. They found that subjects who scored high in psychological mindedness tended to endorse more subtle and very subtle items than those who scored low in psychological mindedness. Intelligence was related to the endorsement of subtle and obvious items only under fake bad instructions, where individuals with higher intelligence tended to endorse fewer somewhat subtle items and more very obvious items than those of lesser intelligence.

Summary of methodological considerations. Some of the factors which need to be taken into account when performing research with the validity scales include distortion of results produced by repeated testings, the influence of instructional sets, and subject characteristics such as race, intelligence, and psychological sophistication. While the task of attempting to determine which variables are influential in the validity studies is formidable, continued research needs to be conducted to more clearly delineate the influences of different factors. Ultimately the clinician may have a body of research that gives guidance as to which validity index is most accurate for a given client and situation.

Statement of Research Question

Since the MMPI is widely used by clinicians in psychological test batteries, an effective indicator of the accuracy of the test results is essential. It has been especially difficult to identify records in which the patient attempts to appear more psychologically healthy than is actually the case. Two indexes which have had success in detecting fake good MMPIs produced by students are the Mp scale and the X minus zero scale. Little research has been performed, however, which evaluates their ability to detect faking in actual psychiatric populations.

The present study sought to determine the degree to which these scales could accurately differentiate MMPIs produced by a psychiatric inpatient population under standard and fake good instructions. These results were compared with supplemental analyses for the L, F, and K scales.

Hypotheses

1. The Mp scale score will be significantly higher for MMPI records produced under fake good instructions than for those produced under standard instructions.
2. The X minus zero scale will be significantly lower for MMPI records produced under fake good instructions than for those produced under standard instructions.

CHAPTER II

METHOD

Subjects

The subjects used in this study were 30 psychiatric inpatients at the Big Spring State Hospital, Big Spring, Texas. Most of the patients were residing in the admissions unit at the time of testing. The sample consisted of Caucasians who were at least 18 years of age and who had a vocabulary level of seventh grade or higher as measured by the California Achievement Test (CAT). Descriptive data and diagnoses of the patients are presented in Tables 1 and 2. The means and standard deviations of the patients for the Vocabulary section of the CAT are provided in Table 3. Patients with a primary diagnosis of substance abuse, those who were too confused to be tested, or those who had been hospitalized more than twice during the past 2 years were excluded from the sample.

Each patient was administered the MMPI once under fake good instructions and once under standard instructions. Recruitment of patients continued until 30 usable test-retest pairs of MMPIs were obtained. For the purposes of this study, completed MMPIs were considered usable unless the patient scored below a seventh-grade vocabulary level on the CAT, failed to answer the manipulation check correctly,

TABLE 1
DESCRIPTIVE DATA

	Group 1 ^a	Group 2 ^b	Total
Sex			
Male	7	8	15
Female	8	7	15
Cultural Background			
Anglo-American	9	13	22
Hispanic-American	6	2	8
Age			
Mean	31.33	35.00	33.17
S.D.	9.34	11.31	10.36
Education			
Mean	12.70	12.87	12.79
S.D.	1.98	1.55	1.74
Type of Admission			
Involuntary	14	12	26
Voluntary	1	3	4

^aGroup 1 = Standard/Fake Good Order

^bGroup 2 = Fake Good/Standard Order

TABLE 2
DIAGNOSES OF PATIENTS

Diagnosis	Group 1 ^a	Group 2 ^b	Total
Schizophrenia, Paranoid	7	3	10
Major Depression	2	3	5
Bipolar Disorder, Manic	3	1	4
Organic Brain Syndrome	1	3	4
Schizoaffective Disorder	1	1	2
Schizophreniform Disorder	1	1	2
Atypical Psychosis	0	1	1
Brief Reactive Psychosis	0	1	1
Dysthymic Disorder	0	1	1

^aGroup 1 = Standard/Fake Good Order

^bGroup 2 = Fake Good/Standard Order

TABLE 3

CALIFORNIA ACHIEVEMENT TEST VOCABULARY SECTION SCORES

	Group 1 ^a	Group 2 ^b	Total
Grade Level			
Mean	11.11	11.42	11.26
S.D.	1.96	2.20	2.06

NOTE: Maximum grade level attainable is 13.6.

^aGroup 1 = Standard/Fake Good Order

^bGroup 2 = Fake Good/Standard Order

or appeared to have deviated obviously from the instructions (e.g., answering all items true). Patients who elevated more than one clinical scale above a T-score of 70 under fake good instructions also were eliminated on the premise that they were not able to follow the directions to appear well adjusted. Between 100 and 150 patients on the admissions unit were screened informally over a 6-week period for education level, fluency in English, number of previous admissions, and absence of substance abuse. Fifty-five of the patients who passed the initial screening agreed to be tested. Nine patients failed to complete testing due to subsequent refusal, medication problems, or inability to concentrate. Sixteen additional patients were excluded: four patients deviated obviously from the directions; five

patients failed the instruction check; and seven patients failed to score above a seventh-grade vocabulary level on the CAT.

Materials

The Vocabulary section of the CAT was administered to obtain an estimate of grade level for the patient's verbal skills. The results of the CAT were used to determine whether the patient had sufficient verbal capabilities to comprehend and respond accurately to the MMPI.

Patients were administered the first 399 items of Form R of the MMPI plus 22 additional items used to complete the validity scales. Seventeen patients were administered the group booklet form of the MMPI for the standard instructions testing because the test results were utilized by the hospital psychologists in addition to being used in the study.

A manipulation check (Appendix D) was completed upon finishing the MMPI. The purpose of the manipulation check was to ascertain whether the patient understood and followed the instructions.

Procedure

Patients who fit the criteria for inclusion in this study were approached and asked if they would be willing to participate. If the patient agreed to take part, a consent

form (Appendix A) was explained to the patient and signed. Groups of patients were assigned randomly to one of two orders of test administration. Half of the 30 patients were tested under standard instructions (Appendix B) and retested under fake good instructions (Appendix C), while the other half were tested in the reverse order. The test-retest interval was always less than 1 week and was usually 1 day. In the first session the patients were tested with the CAT prior to taking the MMPI. After finishing the MMPI on each occasion, the manipulation check was completed.

CHAPTER III

RESULTS

The results of the current investigation are presented in two sections. The first section consists of a comparison of MMPI profiles obtained under each experimental condition and the results which pertain to the major experimental hypotheses. The second section presents supplemental analyses and hit rates.

Comparison of MMPI Profiles

A multivariate analysis of variance (MANOVA) was performed to ascertain whether there was a significant overall difference between the MMPI profiles produced under the two instructional sets (see Table 12 in Appendix E). As recommended by Olson (1976), Pillai's Trace was used to determine the approximate F -statistic. There was a very significant difference between the two groups for the instruction effect and no significant differences for order or the interaction between order and instructions. Graphs of the median profiles for each combination of instructions and order are presented in Figure 1. Median profiles were plotted since they are less distorted by extreme scores than mean profiles. Since the MANOVA was significant, analyses of variance (ANOVAs) were performed for each MMPI scale (see Table 13 in Appendix E). The only scales which did not show a

Figure 1. MMPI median profiles by condition

significant difference between the scores produced under the two experimental conditions were the Mf and the Ma scales. The high point pairs obtained under each type of instructions are presented in Table 14, Appendix F.

The MMPIs of 25 psychiatric inpatients who were not part of the present study were obtained from the files at Big Spring State Hospital. The MMPIs were selected from patients who had been tested by the hospital staff during the 4 months prior to the beginning of the experimental testing. The median MMPI profile from this sample was compared with the median standard profile of the patients who were tested as a part of the present study (see Figure 2).

Major Experimental Results

The two major hypotheses concerned the ability of the Mp and X minus zero scales to discriminate MMPI profiles produced under standard instructions from those produced under instructions to fake good. A summary of descriptive statistics for the Mp and X minus zero scales is presented in Table 4. The F-max values for each of the validity scales indicated that the variances were homogeneous (Table 15, Appendix G). A MANOVA was performed for the Mp and the X minus zero scales in order to determine whether there were significant overall effects for order, type of instructions, and the interaction between order and instructions (see Table 16 in Appendix H). Pillai's Trace was used to

Figure 2. Median standard profiles for inpatient subjects and a non-subject inpatient sample

determine the approximate F-statistic. There was a highly significant effect for type of instructions and a significant effect for the interaction between order and instructions.

TABLE 4
SUMMARY OF DESCRIPTIVE STATISTICS FOR
THE Mp AND X-0 SCALES

Variable	Standard Instructions	Fake Good Instructions
<u>Mp</u> Scale		
Mean	13.33	22.20
S.D.	4.89	4.99
<u>X-0</u> Scale		
Mean	73.60	1.97
S.D.	35.68	20.00

NOTE: Means and standard deviations are expressed as raw scores.

Hypothesis 1 stated that the Mp scale score would be significantly higher for MMPI records produced under fake good instructions than for those produced under standard instructions. Since the MANOVA was significant, an ANOVA was performed for each scale and the results are presented in Table 5. The interaction between order and instructions was significant for the Mp scale. The interaction is

TABLE 5
SUMMARY OF ANOVAs FOR THE M_p AND X-0 SCALES

Source	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
<u>M_p</u> Scale				
Between Subjects				
Order	1	91.27	3.11	.0887
Error	28	29.34		
Within Subjects				
Instructions	1	1179.27	78.79	.0001
Order X Instructions	1	81.67	5.46	.0269
Error	28	14.97		
<u>X-0</u> Scale				
Between Subjects				
Order	1	673.35	.98	.3302
Error	28	685.58		
Within Subjects				
Instructions	1	76970.02	75.38	.0001
Order X Instructions	1	58.02	.06	.8133
Error	28	1021.16		

portrayed graphically in Figure 3. A post hoc analysis of the interaction (Table 6) indicated that there was a significant order effect for the Mp scores produced under fake good instructions. The group of patients who took the test under standard instructions first and then under fake good instructions scored significantly higher on the Mp scale under fake good instructions than the group which took the MMPIs in the reverse order. There was no difference between the two groups in Mp scores produced under standard instructions. Consequently, hypothesis 1 was only partially confirmed in that the Mp scores were significantly higher under fake good instructions than under standard instructions.

Hypothesis 2 stated that the X minus zero index would be significantly lower for MMPI records produced under fake good instructions than for those produced under standard instructions. The ANOVA for the X minus zero scale is presented in Table 5. The main effect for instructions was significant at the .0001 level. Hypothesis 2, therefore, was confirmed. Neither the order effect nor the interaction between order and instructions was significant. A post hoc analysis (Table 7) confirmed the significant effect for instructions. There were also no order or interaction effects when the X scores and zero scores were analyzed separately.

Figure 3. $\underline{Mp}$ scale scores by condition

TABLE 6
POST HOC COMPARISON BY ORDER X INSTRUCTIONS
FOR THE Mp SCALE

Variable	Group 1 ^a Mean	Group 2 ^b Mean	Difference Between Means
<u>Mp</u> Scale			
Standard	13.40	13.27	.13
Fake Good	24.60	19.80	4.80*

NOTE: Means are expressed as raw scores.

^aGroup 1 = Standard/Fake Good Order

^bGroup 2 = Fake Good/Standard Order

*p < .05 using Tukey's HSD test.

TABLE 7
POST HOC COMPARISONS BY INSTRUCTIONS FOR X-0

Variable	Standard Mean	Fake Good Mean	Difference Between Means
<u>X-0</u> Scale	73.60	1.97	71.63*

NOTE: Means are expressed as raw scores.

*p < .05 using Tukey's HSD test.

Supplemental Analyses

L, F, and K scales. The performance of the L, F, and K scales as validity indicators will be examined in this section. The descriptive statistics for these scales are presented in Table 8. A MANOVA (see Table 17 in Appendix H) was performed in order to test the overall effectiveness of the three scales in discriminating between the standard and fake good profiles. Pillai's Trace was used to determine the approximate F-statistic. The effect for type of instructions was significant at the .0001 level. There were no significant effects for order or the interaction between order and instructions.

Since the MANOVA was significant for instructions, ANOVAs were performed for the L, F, and K scales (Table 9). Only the L scale showed a significant order effect. The effect for type of instructions was significant at the .0001 level for all three scales. None of the scales showed a significant interaction between order and instructions. The difference between the means for order on the L scale failed to reach significance at the .05 level using Tukey's HSD test (Table 10). The differences between means obtained under the two instructional sets were significant for the L, F, and K scales (Table 11).

Hit rates. It is important for a validity indicator to be able to discriminate reliably between standard and faked

TABLE 8
SUMMARY OF DESCRIPTIVE STATISTICS FOR THE
L, F, AND K SCALES

Variable	Standard Instructions	Fake Good Instructions
<u>L</u> Scale		
Mean	4.07	7.90
S.D.	2.41	3.20
<u>F</u> Scale		
Mean	12.10	4.13
S.D.	6.25	2.94
<u>K</u> Scale		
Mean	10.90	18.33
S.D.	4.60	4.06

NOTE: Means and standard deviations are expressed as raw scores.

TABLE 9
SUMMARY OF ANOVAs FOR THE L, F, AND K SCALES

Source	<u>df</u>	<u>MS</u>	<u>F</u>	<u>p</u>
<u>L</u> Scale				
Between Subjects				
Order	1	46.82	4.17	.0508
Error	28	11.23		
Within Subjects				
Instructions	1	220.42	61.55	.0001
Order X Instructions	1	2.82	.79	.3827
Error	28	3.58		
<u>F</u> Scale				
Between Subjects				
Order	1	1.35	.04	.8387
Error	28	31.98		
Within Subjects				
Instructions	1	952.02	54.70	.0001
Order X Instructions	1	.15	.01	.9267
Error	28	17.40		
<u>K</u> Scale				
Between Subjects				
Order	1	20.42	.76	.3911
Error	28	26.90		
Within Subjects				
Instructions	1	828.82	73.08	.0001
Order X Instructions	1	.15	.01	.9093
Error	28	11.34		

TABLE 10

POST HOC COMPARISON BY ORDER FOR THE L SCALE

Variable	Group 1 ^a Mean	Group 2 ^b Mean	Difference Between Means
<u>L</u> Scale	6.87	5.10	1.77

NOTE: The difference between means was not significant using Tukey's HSD test at the .05 level. Means are expressed as raw scores.

^aGroup 1 = Standard/Fake Good Order

^bGroup 2 = Fake Good/Standard Order

TABLE 11

POST HOC COMPARISON BY INSTRUCTIONS FOR THE
L, F, AND K SCALES

Variable	Standard Mean	Fake Good Mean	Difference Between Means
<u>L</u> Scale	4.07	7.90	-3.83*
<u>F</u> Scale	12.10	4.13	7.97*
<u>K</u> Scale	10.90	18.33	-7.43*

NOTE: Means are expressed as raw scores.

*p < .05 using Tukey's HSD test

profiles. One process involves setting a cutting score for which scores falling above the cut are classified differently than those falling below. Various cutting scores were evaluated for each validity scale and the percentage of correctly classified MMPIs in the standard and fake good categories were calculated. Although these percentages should be viewed as tentative due to the relatively small number of patients and lack of cross-validation, the hit rates give some general idea about the efficacy of each scale.

The highest overall hit rate was achieved by the X minus zero scale using +35 as a cutting score (Table 18, Appendix I). This scale was able to correctly classify 83.3% of the standard profiles and 96.7% of the fake good profiles. Because of the interaction effects between order and instructions, the hit rate for the Mp scale varied according to the order in which the tests were administered. For the standard/fake good order, the optimal cutting score of 17 identified 86.7% of the standard profiles and 100% of the fake good ones (Table 18, Appendix I). For the fake good/standard order, the optimal cutting score of 17 identified 86.7% of the standard profiles and 66.7% of the fake good ones. Of the supplementary scales, the F scale had the highest hit rate followed by the K scale then the L scale (Table 19, Appendix I).

CHAPTER IV

DISCUSSION

In the first section of this chapter, the results which pertain to the central hypotheses will be discussed. The second section will examine the supplemental analyses. Finally, the conclusions and suggestions for future research will be discussed.

Discussion of Hypotheses

In order to judge the efficacy of MMPI validity scales in discriminating standard and fake good records, it must first be established that there were differences in the profiles produced under those instructions. The results of the MANOVA indicated that there was a very significant overall difference between the profiles depending upon type of instructions used. The psychiatric inpatients were able to judge whether items were pathological and then answer according to instructions. In the one other study (Boe & Kogan, 1964) which instructed psychiatric inpatients to fake on the MMPI, there was also a large difference between the standard and fake good profiles. The extreme differences are due to the highly elevated standard profiles produced by inpatients rather than to the fake good profiles. The median fake good profile produced by the inpatients in the current study is similar in elevation and configuration to fake good profiles

produced by college students (e.g., Cofer et al., 1949; Exner et al., 1963; Hiner et al., 1969).

In the present study, the only individual scales which did not change significantly due to instructions were the Mf and Ma scales. The Ma scale was the most elevated clinical scale in a large number of the fake good profiles. Patients apparently did not perceive the items on the Ma scale to be as pathological as those for the other clinical scales. This impression is not unusual since the Ma scale is also the most often elevated scale in normal populations taking the test under either standard or fake good instructions.

The median MMPI profile of the inpatient sample which was used for comparative purposes was very similar in configuration to the median standard profile of the inpatients used in the study. This similarity suggests that the patients in the study were representative of the larger population of inpatients. Because of the screening procedures, it is to be expected that the median profile for the patients in the study is less elevated than that of the sample group.

Having established that there was a significant difference between the profiles depending upon type of instructions, the efficacy of certain validity indicators in discriminating the two types of profiles can be examined. The significant MANOVA for the main effects of instructions indicated that the Mp and X minus zero scales were sensitive to the experimental manipulation.

Hypothesis 1 predicted that the Mp scale score would be significantly higher for the MMPI records produced under fake good instructions than for those produced under standard instructions. The Mp scale was the only one of the primary or supplementary scales to show a significant interaction between order and instructions. The post hoc analysis indicated that the interaction effect occurred only with the fake good instructions. The group who took the MMPI first under standard instructions produced a higher Mp score on the fake good instructions than those who were tested in the reverse order.

The reasons for the interaction effect on only the Mp scale are unclear. Characteristics of the Mp scale such as having a high number of subtle items or socially desirable items might be linked to the interaction, but other scales also share these characteristics. One unique aspect of the Mp scale is that it is the only one of the validity indicators whose items were empirically selected based upon a sample of subjects who were intentionally faking good. Cofer et al. (1949) used two groups, each of which took the MMPI under both standard and fake instructions. As in the present study, the order of administration was counterbalanced. Cofer compared the overall profiles of the two groups and saw no differences, so the results of the groups were combined. It was from this combined group that the Mp scale

was derived by item analysis. Cofer apparently did not examine whether there was an order or interaction effect specifically for the Mp scale. It is possible that such effects did exist for the Mp scale even though there was no order or interaction effect for the overall profile.

In the present study, the fake good records produced under the fake good/standard order of administration would most closely resemble those produced in a clinical setting. Since a patient in a clinical setting who is attempting to fake good is usually administered the MMPI only once, there would be no recent prior exposure to the test (unlike the standard/fake good order). The results of the fake good/standard order of administration will be used for interpretation and comparison purposes since it is least likely to have been distorted by the interaction effects. There was less change in Mp scores due to instructions under the fake good/standard condition than in the reverse order. Calculation of separate hit rates for each order demonstrated that the Mp scale hit rate for detecting fake good records was much lower (66.7%) for the fake good/standard order than for the standard/fake good order (100%).

Most of the studies which have examined the Mp scale have either not used repeated testing designs (Boe & Kogan, 1964; Grow et al., 1972; Wiggins, 1959) or have made comparisons of the initial testings for each group in repeated

testing designs (Skrzypek & Wiggins, 1966; Wales & Seeman, 1968). Across these studies the Mp scale identified about 70% of the fake good profiles and 90% of the standard ones. These hit rates are similar to the ones in the present study under the fake good/standard order where 66.7% of the faked tests and 86.7% of the standard ones were identified. While the occurrence of interaction effects does not negate the usefulness of the Mp scale (since repeated testing within a short time span usually does not occur in clinical practice), it does suggest that repeated measure designs should not be used with the Mp scale. It should be noted, however, that Skrzypek & Wiggins (1966) did not find a significant interaction effect for order and type of instructions on the Mp scale. Since the evidence concerning the influence of repeated testings is inconsistent, the most prudent approach would be to avoid that type of design with the Mp scale.

Hypothesis 2 predicted that the X minus zero index would be significantly lower for MMPI records produced under fake good instructions than for those produced under standard instructions. The highly significant main effect for type of instructions confirmed this hypothesis. There were no significant effects for order or the interaction between order and instructions. The sensitivity of the X minus zero scale to faking good is consistent with studies using outpatient and student populations.

The additional finding that there were no order or interaction effects when the X scores and zero scores were analyzed separately did not confirm the results of Wales and Seeman (1972). They concluded that order of instructional sets had a strong influence on the number of zero items endorsed and a lesser (but still significant) effect on the X item endorsement.

Supplemental Analyses

Comparison of the L, F, and K scales. The MANOVA indicated that there was a significant overall effect for instructions for the L, F, and K scales. There were no interaction or order effects. The ANOVAs for each of the three scales showed a very significant effect for type of instructions. The only scale to show an order effect was the L scale. The non-significant post hoc analysis, however, suggested that this result may have been due to chance or sampling error rather than being an actual effect. There were no significant interactions for any of the three scales.

Comparison of hit rates. Except for the Mp scale, the hit rates in the present study were higher than in many other studies. This result was probably due to the relatively large differences between the standard and the fake good scores. The X minus zero scale achieved the best overall hit rate. The relatively good performance by the F

scale is surprising and is not supported by the literature in the case of fake good, although it has shown effectiveness in identifying fake bad profiles.

Another result of the extreme scores was that the optimal cutting scores for most of the scales were appreciably different in this study than in some of the others. This outcome was especially true for the X minus zero score. Wales and Seeman (1968) determined that an optimal cutting score was -4 for their group of undergraduates. In the present study the optimal cutting score was +35. This difference indicates that the psychiatric inpatient group endorsed more X items overall than did Wales and Seeman's undergraduates. This fact is not surprising (since X items tend to be more overtly pathological than other items) and reaffirms the idea that different cutting scores for different populations may need to be determined.

The Mp scale, on the other hand, continued to be the one scale which has a relatively consistent cutting score across populations and types of instructions. The cutting score recommended by Cofer et al. (1949) was +20. Wales and Seeman (1968) applied this cutoff to undergraduates and found that the Mp scale identified 68% of the fake good profiles and 94% of the standard ones. Grow et al. (1980) also used +20 with a student sample and found that the Mp scale identified 54% of the fake good profiles and 90% of the standard

ones. In the present study, the optimal cutting score was +17. Using the cutoff of +20 (for the fake good/standard administration) yielded results similar to those of Grow et al. in that 46.7% of the fake profiles and 86.7% of the standard ones were correctly classified. Boe and Kogan (1964) found that an optimal cutting score of +18 identified 81% of the fake good profiles and 71% of the standard profiles in an inpatient psychiatric population.

Overall, it appears that the Mp scale is less subject to wide fluctuations in cutoffs than are some of the other scales. A +20 cutting score generally produces good results, but with psychiatric inpatients the optimal cutting score seems to decrease slightly. While the X minus zero cutting score appears to be less stable, in the present study it achieved a considerably higher hit rate than the Mp scale. It is possible that future research will establish cutting scores for specific populations and thus increase the efficacy of validity scales.

Limitations and Suggestions for Future Research

An issue which pertains to most investigations of faking on the MMPI is whether instructing patients to fake good produces the same results as faking due to the intrinsic motivation of the patient. The literature concerning this question has not provided any clear answers. It seems that

the most judicious approach would be to perform studies in naturalistic settings and situations whenever possible. An example would be a situation where individuals believe they will realize some benefit by portraying themselves in an idealized manner. When MMPIs are used to test job applicants, there is a tendency for the applicants to try to give very good impressions of themselves. The MMPIs of this group could be compared with those of a comparable group of people taking the test anonymously. Situations also occur with psychiatric populations where patients might see a benefit in trying to look good on the MMPI (e.g., child abuse cases where the parent is attempting to regain custody). By examining such groups, it will be possible to determine which scales can discriminate those who have a high probability of faking good from those who have no motive to fake. The widespread use of college students who are instructed to fake on tests seems to have little more to contribute to the faking literature.

Future research also needs to concentrate upon establishing norms for various groups. The present study and others suggest that optimal cutting scores change in varying degrees depending upon the population being tested and the scale being used. In most studies the optimal cutting score of the Mp scale has varied only from +17 to +20 while in the few studies using the X minus zero scale, the cutting score

has varied to a much greater degree. Until norms for individual populations are established, some balance between high hit rates and relatively stable cutting scores will need to be achieved in order for the scales to be practical for use by clinicians. A great deal more research needs to take place in order to arrive at adequate normative data. Such research is important, however, as long as the MMPI continues to be widely used in psychological test batteries.

Conclusions and Implications

The X minus zero scale was able to differentiate between the standard and fake good records at a highly significant level. The results of this study provide support for the view that the X minus zero scale is a promising indicator for the detection of faking good on the MMPI. The results concerning the Mp scale are not as clear due to the significant interaction effect. The Mp scale generally did not perform as well as the X minus zero scale. The hit rate for the fake good/standard order of the Mp scale was considerably below that for the X minus zero scale. Supplemental analyses indicated that the L, F, and K scales also were not as effective as the X minus zero scale, but still reached significance in differentiating between the standard and fake good MMPIs. It is recommended that continued research be performed with the X minus zero scale to determine its effectiveness in more naturalistic settings.

It is possible that with the growing use of computers to score MMPIs, more sophisticated means of detecting faking will be developed. Systems which use combinations of scales weighted according to characteristics of the patient (e.g., population, reason for testing) may achieve a higher hit rate than any single scale. Such a system would, of course, depend upon research which establishes norms and hit rates for combinations of scales with different populations.

An additional use for the Mp and X minus zero scales will be the establishment of behavioral correlates for high and low scores on each scale. As mentioned in Chapter I, descriptive characteristics of high and low scores on the L, F, and K scales have been discussed in the literature for a number of years. Such behavioral correlates for the Mp and X minus zero scales may have implications for diagnosis and treatment.

In addition to examining validity indexes, the current study also supported the view that higher functioning inpatients are able to accurately identify pathological behaviors. Having patients take the MMPI under fake good instructions may be a potentially helpful tool in diagnosing the severity of illness in a patient. The ability to fake good already has been shown to be correlated to early release from inpatient settings (Grayson & Olinger, 1957).

Analysis of the results may be useful in treatment planning and repeated testing could indicate progress in recovery. Because of the potential usefulness of the fake good MMPI with inpatients, more research needs to be performed in order to establish an empirical base upon which to make diagnostic and treatment decisions.

REFERENCES

- Anthony, N. Comparison of clients' standard, exaggerated, and matching MMPI profiles. Journal of Consulting and Clinical Psychology, 1971, 36, 100-103.
- Baker, J. N. Effectiveness of certain MMPI dissimulation scales under "real life" conditions. Journal of Counseling Psychology, 1967, 14, 286-292.
- Ball, J. C. Comparison of MMPI profile differences among Negro-white adolescents. Journal of Clinical Psychology, 1960, 16, 304-307.
- Ball, J. C., & Carroll, D. Analysis of MMPI Cannot Say scores in an adolescent population. Journal of Clinical Psychology, 1960, 16, 30-31.
- Block, J. The challenge of response sets. New York: Appleton-Century-Crofts, 1965.
- Bloom, B. L., & Arkoff, S. Role playing in acute and chronic schizophrenia. Journal of Consulting Psychology, 1961, 25, 24-28.
- Blumberg, S. MMPI F scale as an indicator of severity of psychopathology. Journal of Clinical Psychology, 1967, 23, 96-99.
- Boe, E. E., & Kogan, W. S. Effect of social desirability instructions on several MMPI measures of social desirability. Journal of Consulting Psychology, 1964, 28, 248-251.
- Buechley, R., & Ball, H. A new test of "validity" for the group MMPI. Journal of Consulting Psychology, 1952, 16, 299-301.
- Burish, T. G., & Houston, B. K. Construct validity of the Lie scale as a measure of defensiveness. Journal of Clinical Psychology, 1976, 32, 310-314.
- Burkhardt, B. R., Christian, W. L., & Gynther, M. D. Item subtlety and faking on the MMPI: A paradoxical relationship. Journal of Personality Assessment, 1978, 42, 76-80.

- Burkhart, B. R., Gynther, M. D., & Christian, W. L. Psychological mindedness, intelligence, and item subtlety endorsement patterns on the MMPI. Journal of Clinical Psychology, 1978, 34, 76-79.
- Burkhart, B. R., Gynther, M. D., & Fromuth, M. E. The relative predictive validity of subtle vs. obvious items on the MMPI depression scale. Journal of Clinical Psychology, 1980, 36, 748-751.
- Christian, W. L., Burkhart, B. R., & Gynther, M. D. Subtle-obvious ratings of MMPI items: New interest in an old concept. Journal of Consulting and Clinical Psychology, 1978, 46, 1178-1186.
- Clopton, J. R., & Neuringer, C. MMPI Cannot Say scores: Normative data and degree of profile distortion. Journal of Personality Assessment, 1977, 41, 511-513.
- Cofer, C. N., Chance, J., & Judson, A. J. A study of malin-gering on the MMPI. Journal of Psychology, 1949, 27, 491-499.
- Coyle, F. A., & Heap, R. F. Interpreting the MMPI L scale. Psychological Reports, 1965, 17, 722.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. An MMPI handbook. Vol. 1. Clinical interpretation. (Rev. ed.). Minneapolis: University of Minnesota Press, 1972.
- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. An MMPI handbook. Vol. II. Research applications. (Rev. ed.). Minneapolis: University of Minnesota Press, 1975.
- Dies, R. R. Detection of simulated MMPI records using the desirability (Dy) scales. Journal of Clinical Psychology, 1968, 24, 335-337.
- Drasgow, J., & Barnette, W. L. F-K in a motivated group. Journal of Consulting Psychology, 1957, 21, 399-401.
- Duff, F. L. Item subtlety in personality inventory scales. Journal of Consulting Psychology, 1965, 29, 565-570.
- Edwards, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. Journal of Abnormal and Social Psychology, 1957, 55, 394-395.
- Edwards, A. L. Social desirability and performance on the MMPI. Psychometrika, 1964, 29, 295-308.

- Exner, J. E., McDowell, E., Pabst, J., Stackman, W., & Kirk, L. On the detection of willful falsification in the MMPI. Journal of Consulting Psychology, 1963, 27, 91-94.
- Fordyce, W. E. Social desirability in the MMPI. Journal of Consulting Psychology, 1956, 20, 171-175.
- Fortova, V. [Validity of indicators of simulation and dissimulation in MMPI] Ceskoslovenska Psychologie, 1979, 23, 50-60. (Psychological Abstracts, 1980, 65, no. 2461.)
- Fricke, B. G. Subtle and obvious test items and response set. Journal of Consulting Psychology, 1957, 21, 250-252.
- Gough, H. G. Simulated patterns on the MMPI. Journal of Abnormal and Social Psychology, 1947, 42, 215-225.
- Gough, H. G. The F minus K dissimulation index for the MMPI. Journal of Consulting Psychology, 1950, 14, 408-413.
- Gough, H. G. Some common misconceptions about neuroticism. Journal of Consulting Psychology, 1954, 18, 287-292.
- Gravitz, M. A. Validity implications of normal adult MMPI "L" scale endorsement. Journal of Clinical Psychology, 1970, 26, 497-499.
- Grayson, H. M., & Olinger, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI, Journal of Consulting Psychology, 1957, 21, 73-77.
- Greene, R. L. An empirically derived MMPI carelessness scale. Journal of Clinical Psychology, 1978, 34, 407-410.
- Greene, R. L. Response consistency on the MMPI: The TR index. Journal of Personality Assessment, 1979, 43, 69-71.
- Greene, R. L. The MMPI: An interpretive manual. New York: Grune & Stratton, 1980.
- Grow, R., McVaugh, W., & Eno, T. D. Faking and the MMPI. Journal of Clinical Psychology, 1980, 36, 910-917.

- Gynther, M. D. The clinical utility of "invalid" MMPI F scores. Journal of Consulting Psychology, 1961, 25, 540-542.
- Gynther, M. D., Altman, H., & Warbin, R. Interpretation of uninterpretable MMPI profiles. Journal of Consulting and Clinical Psychology, 1973, 40, 78-83.
- Gynther, M. D., Burkhart, B. R., & Hovanitz, C. Do face valid items have more predictive validity than subtle items? The case of the MMPI Pd scale. Journal of Consulting and Clinical Psychology, 1979, 47, 295-300.
- Gynther, M. D., Lachar, D., & Dahlstrom, W. G. Are special norms for minorities needed? Development of an MMPI F scale for blacks. Journal of Consulting and Clinical Psychology, 1978, 46, 1403-1408.
- Gynther, M. D., & Shimkunas, A. M. Age, intelligence, and MMPI F scores. Journal of Consulting Psychology, 1965, 29, 383-388. (a)
- Gynther, M. D., & Shimkunas, A. M. More data on the MMPI F>16 scores. Journal of Clinical Psychology, 1965, 21, 275-277. (b)
- Hanley, C. Deriving a measure of test-taking defensiveness. Journal of Consulting Psychology, 1957, 21, 391-397.
- Hanley, C. Social desirability and response bias in the MMPI. Journal of Consulting and Clinical Psychology, 1961, 25, 13-20.
- Hartshorne, H., & May, M. A. Studies in the nature of character. I. Studies in deceit. New York: The Macmillan Co., 1928.
- Harvey, M. A., & Sipprelle, C. N. Demand characteristic effects on the subtle and obvious subscales of the MMPI. Journal of Personality Assessment, 1976, 40, 539-544.
- Hathaway, S. R., & McKinley, J. C. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. Journal of Psychology, 1940, 10, 249-254.
- Hathaway, S. R., & McKinley, J. C. MMPI Manual. (Rev. ed.). New York: Psychological Corporation, 1951.

- Hathaway, S. R., & McKinley, J. C. Minnesota Multiphasic Personality Inventory manual (Rev. ed.). New York: Psychological Corporation, 1967.
- Heilbrun, A. B. Revision of the MMPI K correction procedure for improved detection of maladjustment in a normal college population. Journal of Consulting Psychology, 1963, 27, 161-165.
- Heilbrun, A. B. Social-learning theory, social desirability, and the MMPI. Psychological Bulletin, 1964, 61, 377-387.
- Heineman, C. E. A forced-choice form of the Taylor Manifest Anxiety Scale. Journal of Consulting Psychology, 1953, 17, 447-454.
- Hiner, D. L., Ogren, D. J., & Baxter, J. C. Ideal-self responding on the MMPI. Journal of Projective Techniques and Personality Assessment, 1969, 33, 389-396.
- Hovanitz, C. A., & Gynther, M. D. The prediction of impulsive behavior: Comparative validities of obvious vs. subtle MMPI hypomania (Ma) items. Journal of Clinical Psychology, 1980, 36, 422-427.
- Hunt, H. F. The effect of deliberate deception on MMPI performance. Journal of Consulting Psychology, 1948, 12, 396-402.
- Hunt, H. F., Carp, A., Cass, W. A., Winder, C. L., & Kantor, R. E. A study of the differential diagnostic efficiency of the MMPI. Journal of Consulting Psychology, 1948, 12, 331-336.
- Jackson, D. N. The dynamics of structured personality tests: 1971. Psychological Review, 1971, 78, 229-248.
- Jackson, D. N., & Messick, S. Acquiescence and desirability as response determinants on the MMPI. Educational and Psychological Measurement, 1961, 21, 771-789.
- Johnson, J. H., Klingler, D. E., & Williams, T. A. An external criterion study of the MMPI validity indices. Journal of Clinical Psychology, 1977, 33, 154-156.
- Johnson, P. Black-white differences on the faking indexes of the MMPI in a prison sample. Unpublished doctoral dissertation, Texas Tech University, 1982.

- Jones, F. W., Neuringer, C., & Patterson, T. W. An evaluation of an MMPI response consistency measure. Journal of Personality Assessment, 1976, 40, 419-421.
- Kazan, A. T., & Sheinberg, I. M. Clinical note on the significance of the validity score (F) in the MMPI. American Journal of Psychiatry, 1945, 102, 181-183.
- Kroger, R. O. Effects of role demands and test-cue properties upon personality test performance. Journal of Consulting Psychology, 1967, 31, 304-312.
- Kroger, R. O., & Turnbull, W. Effects of role demands and test-cue properties on personality test performance: Replication and extension. Journal of Consulting and Clinical Psychology, 1970, 35, 381-387.
- Kroger, R. O., & Turnbull, W. Invalidity of validity scales: The case of the MMPI. Journal of Consulting and Clinical Psychology, 1975, 43, 48-55.
- Lachar, D. The MMPI: Clinical assessment and automated interpretation. Los Angeles: Western Psychological Services, 1974.
- Lanyon, R. I. Simulation of normal and psychopathic MMPI personality patterns. Journal of Consulting Psychology, 1967, 31, 94-97.
- Lawton, M. P. Deliberate faking on the psychopathic deviate scale of the MMPI. Journal of Clinical Psychology, 1963, 19, 327-330.
- Lawton, M. P., & Kleban, M. H. Prisoners' faking on the MMPI. Journal of Clinical Psychology, 1965, 21, 269-271.
- MacLean, A. G., Tait, A. T., & Catterall, C. D. The F minus K index on the MMPI. Journal of Applied Psychology, 1953, 37, 315-316.
- Marks, P. A., Seeman, W., & Haller, D. L. The actuarial use of the MMPI with adolescents and adults. Baltimore: Williams & Wilkins, 1974.
- Matarazzo, J. D. MMPI validity scores as a function of increasing levels of anxiety. Journal of Consulting Psychology, 1955, 19, 213-217.

- Match, J., & Wiggins, N. Individual viewpoints of social desirability related to faking good and desirability estimation. Educational and Psychological Measurement, 1974, 34, 591-604.
- McDonald, R. L., & Gynther, M. D. MMPI norms for southern adolescent Negroes. Journal of Social Psychology, 1962, 58, 277-282.
- McKinley, J. C., Hathaway, S. R., & Meehl, P. E. The MMPI: The K scale. Journal of Consulting Psychology, 1948, 12, 20-31.
- Meehl, P. E. The dynamics of "structured" personality tests. Journal of Clinical Psychology, 1945, 1, 296-303.
- Norman, W. T. Psychometric considerations for a revision of the MMPI. In J. Butcher (Ed.), Objective personality assessment: Changing perspectives. New York: Academic Press, 1972.
- Olson, C. L. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 1976, 83, 579-586.
- Post, R. D., & Gasparikova-Krasnec, M. MMPI validity scales and behavioral disturbance in psychiatric inpatients. Journal of Personality Assessment, 1979, 43, 155-159.
- Rapaport, G. M. "Ideal self" instructions, MMPI profile changes, and the prediction of clinical improvement. Journal of Consulting Psychology, 1958, 22, 459-463.
- Rosen, A. Differentiation of diagnostic groups by individual MMPI scales. Journal of Consulting Psychology, 1958, 22, 453-457.
- Schneck, J. M. Clinical evaluation of the F scale on the MMPI. American Journal of Psychiatry, 1948, 104, 440-442.
- Seeman, W. "Subtlety" in structured personality tests. Journal of Consulting Psychology, 1952, 16, 278-283.
- Sines, L. K., Baucom, D. H., & Gruba, G. H. A validity scale sign calling for caution in the interpretation of MMPIs among psychiatric inpatients. Journal of Personality Assessment, 1979, 43, 604-607.

- Skrzypek, G. J., & Wiggins, J. S. Contrasted groups vs. repeated measurement designs in the evaluation of social desirability scales. Educational and Psychological Measurement, 1966, 26, 131-138.
- Smith, E. E. Defensiveness, insight, and the K scale. Journal of Consulting Psychology, 1959, 23, 275-277.
- Sweetland, A., & Quay, H. A note on the K scale of the MMPI. Journal of Consulting Psychology, 1953, 17, 314-316.
- Tamkin, A. S., & Scherer, I. W. What is measured by the "Cannot Say" scale of the group MMPI? Journal of Consulting Psychology, 1957, 21, 370.
- Tyler, F. T., & Michaelis, J. U. K-scores applied to MMPI scales for college women. Educational and Psychological Measurements, 1953, 13, 459-466.
- Vesprani, G. J., & Seeman, W. MMPI X and zero items in a psychiatric outpatient group. Journal of Personality Assessment, 1974, 38, 61-64.
- Vincent, N. M. P., Linsz, N. L., & Greene, M. I. The L scale of the MMPI as an index of falsification. Journal of Clinical Psychology, 1966, 22, 214-215.
- Wales, B., & Seeman, W. A new method for detecting the fake good response set on the MMPI. Journal of Clinical Psychology, 1968, 24, 211-216.
- Wales, B., & Seeman, W. What do MMPI zero items really measure--an experimental investigation. Journal of Clinical Psychology, 1969, 25, 420-424.
- Wales, B., & Seeman, W. Instructional sets and MMPI items. Journal of Personality Assessment, 1972, 36, 282-286.
- Walker, J. N. An examination of the role of the experimentally determined response set in evaluating Edwards' Social Desirability Scale. Journal of Consulting Psychology, 1962, 26, 162-166.
- Weiner, D. N. Subtle and obvious keys for the MMPI. Journal of Consulting Psychology, 1948, 12, 164-170.
- Welsh, C. E. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), Basic readings on the MMPI in psychology and medicine. Minneapolis: University of Minnesota Press, 1956.

- Wiggins, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. Journal of Consulting Psychology, 1959, 23, 419-427.
- Wiggins, J. S. Social desirability under role playing instructions: A reply to Walker. Journal of Consulting Psychology, 1963, 27, 107-111.
- Wiggins, J. S., & Rumrill, C. Social desirability in the MMPI and Welsh's factor scales A and R. Journal of Consulting Psychology, 1959, 23, 100-106.
- Windle, C. Further studies of test-retest effect on personality questionnaires. Educational and Psychological Measurement, 1955, 15, 246-253.
- Wrobel, T. A., & Lachar, D. Validity of the Weiner subtle and obvious scales for the MMPI: Another example of the importance of inventory-item content. Journal of Consulting and Clinical Psychology, 1982, 50, 469-470.
- Yonge, G. D. Certain consequences of applying the K factor to MMPI scores. Educational and Psychological Measurement, 1966, 26, 887-893.

APPENDICES

- A. CONSENT FORM
- B. STANDARD INSTRUCTIONS
- C. FAKE GOOD INSTRUCTIONS
- D. INSTRUCTION CHECK
- E. ANALYSES FOR PROFILES
- F. HIGH POINT PAIRS
- G. F-MAX
- H. MANOVAs FOR VALIDITY SCALES
- I. HIT RATES
- J. RAW DATA

APPENDIX A: CONSENT FORM

NAME OF PROJECT: Detection of Faking Good on the MMPI in a Psychiatric Inpatient Population

I hereby give my consent and agree to participate in the research project entitled, "Detection of Faking Good on the MMPI in a Psychiatric Inpatient Population." The duration of the entire research project will be approximately one year. However, my participation will take less than four (4) hours. This time will be divided into two (2) sessions which will each last less than two (2) hours. The sessions will be on different days. I also give my consent for David Kelly, Principal Investigator, to have access to my hospital records. I understand that my participation in this study is voluntary and my choice to participate or not will in no way affect my treatment at this hospital. If I decide to participate, my performance and records will be kept confidential.

The objective of this study has been explained: To examine the ability to detect faking on a psychological test.

I understand that the procedure requires my taking a short vocabulary test (requiring approximately 10 minutes) and answering a personality inventory (requiring approximately 90 minutes). At the second session the personality inventory will again be taken but with different instructions. This procedure has been judged to provide no discomfort or risks to the subjects who participate.

I understand that this experiment will not harm me in any way and should not be upsetting in any manner. However, I am free to discontinue participation in this project at any time without affecting my treatment.

I understand that I may not receive any direct benefits as a result of my participation in this study. The investigator, David Kelly, will be available to answer any additional questions I may have concerning this study.

I understand that I may consult with a member of the Institutional Review Board at any time concerning my treatment and welfare by writing P. O. Box 231, Big Spring, Texas 79720 or by calling the I.R.B. chairman at (915) 267-8216, Ext. 314, at Big Spring State Hospital.

I understand that I may consult with a member of the Public Responsibility Committee at any time concerning my treatment and welfare by writing P. O. Box 1892, Big Spring, Texas 79720 or by calling (915) 267-8216. The Public Responsibility Committee is a group of volunteers who work to protect the rights and interests of patients.

I understand there is no special provision for compensation and medical treatment for research subjects who are physically injured as a result of participation in a research project. Compensation and medical treatment are available to research subjects on the same basis that compensation and medical treatment are available to other patients of Big Spring State Hospital. Further information concerning the availability of compensation and medical treatment, should I be physically injured as a result of participating in a research project, may be obtained from the Chairman of the Big Spring State Hospital Institutional Review Board by writing P. O. Box 231, Big Spring, Texas 79720 or by calling (915) 267-8216, Ext. 314.

Page 2

RE: Consent Form

A. Certificate of person giving consent:

I understand each of the above items relating to the participation of

_____ in the research of "Detection of Faking Good on
(Name of Project)
the MMPI in a Psychiatric Inpatient Population", under the care of David Kelly,

Principal Investigator, and I hereby consent to _____ participation
(my/his/her)
in the research project.

Signature of Person Giving Consent_____
Date_____
Relation to Patient (Patient/Parent/Guardian)

B. Certification of person explaining proposal

I have explained the above items to _____
(Name of person giving consent)
and believe that _____ understands each of the items.
(he/she)

Investigator's Signature_____
Date

We were present at the explanation of the above items to _____
(Name of client)
and we believe that _____ understands each of the above items.
(he/she)

Witness_____
Date_____
Witness_____
Date

APPENDIX B: STANDARD INSTRUCTIONS

This test consists of numbered statements. Read each statement and decide whether it is true as applied to you or false as applied to you.

You are to mark your answer on the answer sheet that you have. If a statement is True or Mostly True, as applied to you, blacken the circle with the T. If a statement is False or Mostly False, as applied to you, blacken the circle with the F.

Remember to give your own opinion of yourself. Do not leave any blank spaces. In marking your answers on the answer sheet, be sure that the number of the statement agrees with the number on the answer sheet. Make your marks heavy and black. Erase completely any answer that you wish to change.

Remember, try to make some answer to every statement. If you have any questions about the instructions, raise your hand and I will try to answer your questions.

(These instructions were taken in part from the standard instructions on the group form booklet by Hathaway and McKinley, 1970; and from Johnson, 1982)

APPENDIX C: FAKE GOOD INSTRUCTIONS

This test consists of numbered statements to which you will answer true or false. In taking the test, however, do not answer the way that you really are. Instead, pretend that you are trying to make a very good impression and try to answer as a very well-adjusted person would. Answer the test so that a psychologist looking at the results would be convinced that you are very well adjusted.

You are to mark your answers on the answer sheet that you have. If you think that answering True or Mostly True to a question would convince the psychologist that you are very well adjusted, blacken the circle with T. If you think that answering False or Mostly False would convince the psychologist that you are very well adjusted, blacken the circle with F. Do not leave any blank spaces.

In marking your answer on the answer sheet, be sure that the number of the statement agrees with the number on the answer sheet. Make your marks heavy and black. Erase completely any answer you wish to change.

Remember, try to make some answer to every statement and answer in such a way as to appear very well adjusted.

If you have any questions about the instructions, raise your hand and I will try to answer your question.

(These instructions were adapted in part from Hathaway & McKinley, 1970; and Johnson, 1982)

APPENDIX D: INSTRUCTION CHECK

You have just completed the psychology test. Please answer the following questions:

1. The instructions that you received for the test were to:

(check one)

_____ a. Pretend that you are trying to make a good impression and appear very well adjusted.

_____ b. Answer randomly.

_____ c. Answer each statement honestly as it applies to you.

2. How well were you able to follow these instructions?

(check one)

_____ a. I generally was able to follow them.

_____ b. I generally was not able to follow them.

(This instruction check was adapted from Johnson, 1982)

APPENDIX E: ANALYSES FOR PROFILES

TABLE 12
SUMMARY OF MANOVA FOR STANDARD PROFILES
AND FAKE GOOD PROFILES

Source	<u>df</u>	<u>F</u>	<u>p</u>
Between Subjects			
Order	1	1.66	.1673
Error	28		
Within Subjects			
Instructions	1	10.02	.0001
Order X Instructions	1	1.67	.1648
Error	28		

NOTE: Pillai's Trace was used to determine the approximate F-statistic.

TABLE 13
SUMMARY OF ANOVAS FOR MMPI SCALES

Scale	Order		Instructions		O x I	
	<u>MS</u>	<u>F</u>	<u>MS</u>	<u>F</u>	<u>MS</u>	<u>F</u>
<u>Hs</u>	16.02	.18	1316.02	15.06**	33.75	.39
<u>D</u>	29.40	.11	6242.40	33.26***	173.40	.92
<u>Hy</u>	198.02	1.50	608.02	8.70*	1.35	.02
<u>Pd</u>	355.27	2.28	5078.40	45.97***	9.60	.09
<u>Mf</u>	459.27	2.47	.27	.01	86.40	1.74
<u>Pa</u>	2.82	.02	2926.02	31.47***	14.02	.15
<u>Pt</u>	86.40	.51	4788.27	28.20***	81.67	.48
<u>Sc</u>	48.60	.30	5841.07	28.24***	2.40	.01
<u>Ma</u>	2.40	.02	6.67	.07	38.40	.39
<u>Si</u>	12.15	.15	3985.35	43.72***	.82	.01

NOTE: For each scale $df = 1, 28$.

* $p < .01$.

** $p < .001$.

*** $p < .0001$.

APPENDIX F: HIGH POINT PAIRS

TABLE 14

FREQUENCY OF HIGH POINT PAIRS
(Excluding the Mf Scale)

Standard Instructions		Fake Good Instructions	
2 or more scales above 70		2 or more scales above 70	
2-7 & 7-2	4	-----	
4-8 & 8-4	2		
4-9 & 9-4	2		
4-2	2	1 scale above 70	
4-6	2		
6-9	2	9-8 & 8-9	4
8-2	2	9-6	2
4-3	1		
4-8/9	1		
7-2/8	1	No scale above 70	
8-7	1		
9-8	1	9-4 & 4-9	6
		3-4	3
		9-6	3
		9-8	3
		4-2	2
		4-7/8	1
		4-8	1
		6-7	1
		8-6	1
		9-3	1
		9-4/6	1
		0-2	1
1 scale above 79			
9-4/8	2		
4-0	1		
No scale above 70			
2-4	1		
6-9	1		
8-2	1		
8-7/0	1		
8-9	1		
0-9	1		

NOTE: A / indicates that the two scales had equal T-scores.

APPENDIX G: F-MAX

TABLE 15
F-max FOR THE Mp, X-O, L, F, AND K SCALES

Scale	<u>F</u> -max Score
<u>Mp</u>	.39
<u>X-O</u>	1.01
<u>L</u>	1.36
<u>F</u>	1.46
<u>K</u>	1.75

NOTE: Since the F-max values do not exceed the value of 2.86 for F-max .05; 2,14, the hypotheses of homogeneity of variance are not rejected.

APPENDIX H: MANOVAS FOR VALIDITY SCALES

TABLE 16
SUMMARY OF MANOVA FOR THE M_p AND X-0 SCALES

Source	<u>df</u>	<u>F</u>	<u>p</u>
Between Subjects			
Order	1	1.88	.1723
Error	28		
Within Subjects			
Instructions	1	44.94	.0001
Order X Instructions	1	5.27	.0117
Error	28		

NOTE: Pillai's Trace was used to determine the approximate F-statistic.

TABLE 17
SUMMARY OF MANOVA FOR THE L, F, AND K SCALES

Source	<u>df</u>	<u>F</u>	<u>p</u>
Between Subjects			
Order	1	1.40	.2657
Error	28		
Within Subjects			
Instructions	1	36.09	.0001
Order X Instructions	1	.35	.7868
Error	28		

NOTE: Pillai's Trace was used to determine the approximate F-statistic.

APPENDIX I: HIT RATES

TABLE 18

HIT RATES FOR THE Mp AND X-0 SCALES

Cutting Score	Standard % Correctly Classified	Fake % Correctly Classified
<u>Mp</u> Scale (Standard/Fake Order)		
14	60.0	100.0
15	80.0	100.0
16 ^a	86.7	100.0
17 ^a	86.7	100.0
18	86.7	93.3
19	86.7	86.7
20	86.7	86.7
<u>Mp</u> Scale (Fake/Standard Order)		
14	73.3	80.0
15	73.3	80.0
16	73.3	73.3
17 ^a	86.7	66.7
18	86.7	53.3
19	86.7	46.7
20	86.7	46.7
<u>X-0</u> Scale		
20	90.0	76.7
25	90.0	83.3
30	83.3	93.3
35 ^b	83.3	96.7
40	80.0	96.7
45	80.0	100.0
50	76.7	100.0

^aMost efficient cutting score for the Mp scale.

^bFor the X-0 scale, cutting scores of 31 through 39 all yielded the same hit rate since no score fell within this range.

TABLE 19
HIT RATES FOR THE L, F, AND K SCALES

Cutting Score	Standard % Correctly Classified	Fake % Correctly Classified
<u>L</u> Scale		
2	26.7	93.3
3	40.0	93.3
4	60.0	86.7
5 ^a	76.7	73.3
6	90.0	63.3
7	90.0	53.3
8	93.3	43.3
<u>F</u> Scale		
5	96.7	60.0
6	96.7	70.0
7	90.0	80.0
8 ^a	80.0	90.0
9	70.0	93.3
10	63.3	96.7
11	56.7	96.7
<u>K</u> Scale		
11	60.0	93.3
12	63.3	90.0
13	73.3	83.3
14 ^a	80.0	80.0
15	83.3	76.7
16	86.7	73.3
17	90.0	70.0

^aMost efficient cutting score.

APPENDIX J: RAW DATA

TABLE 20

RAW DATA FOR THE CLINICAL SCALES

SUBJ	a		b									
	ORDER	INSTR	HS	D	HY	PD	HF	PA	PT	SC	MA	SI
1	1	1	47	68	55	64	51	47	46	53	50	58
2	1	1	59	48	58	69	53	62	66	69	88	40
3	1	1	59	80	55	74	61	62	81	80	48	65
4	1	1	62	106	62	95	61	76	103	109	65	78
5	1	1	52	47	47	79	51	62	56	66	81	45
6	1	1	48	63	59	74	43	70	68	63	45	69
7	1	1	68	69	80	81	26	70	79	74	58	54
8	1	1	78	84	77	90	43	76	68	89	81	68
9	1	1	39	51	42	53	61	53	50	60	55	58
10	1	1	48	47	42	71	45	70	55	74	60	64
11	1	1	75	89	73	71	49	73	99	86	68	67
12	1	1	62	80	64	95	69	62	75	73	65	58
13	1	1	47	53	45	76	57	85	64	80	81	50
14	1	1	44	53	45	81	47	73	61	69	58	51
15	1	1	62	61	66	76	49	65	60	67	70	66
1	1	2	52	60	56	67	55	53	52	63	60	50
2	1	2	52	41	49	60	53	53	36	44	68	41
3	1	2	49	53	56	55	57	53	52	50	55	44
4	1	2	52	58	60	48	55	44	54	55	65	32
5	1	2	52	53	59	69	41	56	61	61	58	45
6	1	2	44	42	56	55	61	44	46	52	53	46
7	1	2	52	46	57	60	48	56	48	58	65	43
8	1	2	54	51	54	53	43	53	50	58	65	40
9	1	2	48	34	47	48	55	50	43	52	81	44
10	1	2	33	44	38	50	49	50	43	46	65	43
11	1	2	49	58	53	62	61	50	52	53	55	40
12	1	2	52	46	55	57	59	44	48	59	68	39
13	1	2	49	65	45	60	59	67	46	69	63	51
14	1	2	50	44	47	60	49	50	51	55	60	36
15	1	2	44	49	38	57	59	59	46	47	58	49
16	2	1	70	101	65	97	74	76	97	111	70	72
17	2	1	62	89	75	76	80	65	85	78	50	71
18	2	1	47	106	64	81	84	65	91	88	50	78
19	2	1	62	67	70	79	47	59	65	75	75	54
20	2	1	54	58	53	53	61	44	54	61	48	46
21	2	1	50	63	50	69	47	64	66	69	81	54
22	2	1	67	99	67	93	69	85	81	111	60	70
23	2	1	80	92	78	81	84	79	101	10	58	72
24	2	1	52	59	42	39	49	56	61	66	58	61
25	2	1	78	82	70	83	41	73	69	74	50	65
26	2	1	52	65	50	76	51	53	60	63	55	66
27	2	1	39	58	47	53	59	47	42	38	60	61
28	2	1	60	49	72	74	45	105	65	83	96	43
29	2	1	76	42	73	64	43	65	53	81	86	47
30	2	1	39	41	40	46	51	62	42	44	58	41
16	2	2	49	44	60	64	59	44	56	65	70	37
17	2	2	54	63	60	64	71	56	54	53	60	52
18	2	2	49	51	64	60	73	44	54	51	50	36
19	2	2	54	38	61	67	49	59	50	64	81	44
20	2	2	57	63	62	57	69	67	64	59	63	42
21	2	2	46	40	56	60	55	47	51	51	58	37
22	2	2	44	48	53	39	63	56	54	55	65	46
23	2	2	47	41	56	57	57	62	50	53	73	35
24	2	2	44	40	47	46	53	47	48	49	68	44
25	2	2	44	47	52	46	49	56	48	52	63	41
26	2	2	42	57	45	55	51	53	50	54	50	64
27	2	2	52	48	58	48	59	50	62	65	70	58
28	2	2	52	47	54	43	57	59	55	55	70	32
29	2	2	42	46	52	55	39	53	43	46	63	43
30	2	2	49	41	49	39	47	50	60	71	65	49

NOTE: Scores are expressed as T-scores.

^a₁ = Standard/Fake Good Order; 2 = Fake Good/Standard Order.

^b₁ = Standard instructions; 2 = Fake Good instructions.

TABLE 21
RAW DATA FOR THE VALIDITY SCALES

SUBJ	ORDER ^a	INSTRUCTIONS ^b	MP	X-0	L	P	R
1	1	1	15	17	6	4	13
2	1	1	21	63	2	11	9
3	1	1	14	73	3	12	11
4	1	1	7	116	1	12	13
5	1	1	15	16	5	9	18
6	1	1	8	46	5	8	14
7	1	1	14	50	10	15	22
8	1	1	8	110	3	11	8
9	1	1	16	62	6	8	6
10	1	1	15	83	5	10	6
11	1	1	12	115	2	13	6
12	1	1	14	62	6	12	17
13	1	1	21	78	4	26	12
14	1	1	10	70	4	13	76
15	1	1	11	78	9	16	9
1	1	2	23	10	8	5	20
2	1	2	22	29	2	4	9
3	1	2	30	-17	11	0	20
4	1	2	19	-27	5	1	23
5	1	2	26	-15	10	9	27
6	1	2	26	-14	11	2	20
7	1	2	22	-24	13	1	23
8	1	2	22	1	5	1	21
9	1	2	28	3	10	6	18
10	1	2	18	18	5	3	10
11	1	2	29	-14	11	1	20
12	1	2	28	4	10	4	19
13	1	2	25	43	10	13	14
14	1	2	27	-13	12	2	22
15	1	2	24	10	12	7	17
16	2	1	7	127	3	23	10
17	2	1	8	96	1	13	11
18	2	1	9	77	4	7	13
19	2	1	17	59	6	8	16
20	2	1	26	12	8	7	20
21	2	1	14	84	0	10	5
22	2	1	13	156	4	27	5
23	2	1	4	111	0	9	10
24	2	1	14	59	2	13	6
25	2	1	14	106	5	12	9
26	2	1	9	39	3	6	15
27	2	1	12	26	2	2	8
28	2	1	21	84	4	26	14
29	2	1	14	106	4	14	6
30	2	1	17	27	5	6	6
16	2	2	13	-14	1	6	21
17	2	2	18	15	6	3	18
18	2	2	19	-30	6	0	22
19	2	2	16	22	6	8	16
20	2	2	22	22	10	6	21
21	2	2	17	-6	7	2	20
22	2	2	26	29	11	7	13
23	2	2	18	-24	4	2	20
24	2	2	22	-17	7	3	19
25	2	2	30	-8	12	5	20
26	2	2	13	30	5	4	12
27	2	2	13	27	4	5	13
28	2	2	26	2	7	7	19
29	2	2	22	-2	8	3	16
30	2	2	22	19	8	4	15

NOTE: Scores are expressed as raw scores.

^a1 = Standard/Fake Good Order; 2 = Fake Good/Standard Order.

^b1 = Standard instructions; 2 = Fake Good instructions.