1983

Minnesota Multiphasic Personality Inventory (Mmpi) Response Patterns Among Prisoners Under Varied Instructional Sets.

Rafael Francisco Salcedo

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/3936

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Salcedo, Rafael Francisco

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI) RESPONSE PATTERNS AMONG PRISONERS UNDER VARIED INSTRUCTIONAL SETS

The Louisiana State University and Agricultural and Mechanical Col. Ph.D. 1983

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106
**PLEASE NOTE:**

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark \( \checkmark \).

1. Glossy photographs or pages ______
2. Colored illustrations, paper or print ______
3. Photographs with dark background ______
4. Illustrations are poor copy ______
5. Pages with black marks, not original copy ______
6. Print shows through as there is text on both sides of page ______
7. Indistinct, broken or small print on several pages ______
8. Print exceeds margin requirements ______
9. Tightly bound copy with print lost in spine ______
10. Computer printout pages with indistinct print ______
11. Page(s) _______ lacking when material received, and not available from school or author.
12. Page(s) _______ seem to be missing in numbering only as text follows.
13. Two pages numbered _______. Text follows.
14. Curling and wrinkled pages ______
15. Other ____________________________________________

**University Microfilms International**
MINNESOTA MULTIPHASIC PERSONALITY INVENTORY (MMPI)
RESPONSE PATTERNS AMONG PRISONERS UNDER VARIED
INSTRUCTIONAL SETS

A Dissertation

Submitted to the Graduate Faculty
of the Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

in

The Department of Psychology

by

Rafael F. Salcedo
B.S., Florida State University, 1978
M.S., Louisiana State University, 1980
December, 1983
ACKNOWLEDGEMENTS

I would like to express my appreciation to several persons whose assistance contributed greatly to the completion of this dissertation. I especially want to thank my major professor, Dr. Ralph M. Dreger, for his support and guidance throughout every phase of this project. Dr. Thomas C. Fain, of the Feliciana Forensic Facility in Jackson, Louisiana, provided the initial impetus and intellectual challenge necessary to formulate and clarify the ideas tested in this study. Drs. Scott Moss, Alpha Curry and Michael Marvin of the U.S. Camp and Penitentiary in Lompoc, California provided invaluable assistance during the data collection phase of this project. Dr. George Stanton, of the California Polytechnic State University, assisted greatly with the analysis of the data. Finally, I want to thank Dee A. Hoffman, R.M.T., and Linda Brown, for their assistance in the finalizing and typing of this dissertation.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Description and brief historical overview</td>
<td>4</td>
</tr>
<tr>
<td>Cannot say (?) scale</td>
<td>10</td>
</tr>
<tr>
<td>Lie (L) scale</td>
<td>19</td>
</tr>
<tr>
<td>F scale</td>
<td>24</td>
</tr>
<tr>
<td>K scale</td>
<td>31</td>
</tr>
<tr>
<td>Validity scale configuration</td>
<td>41</td>
</tr>
<tr>
<td>Special response measures</td>
<td>53</td>
</tr>
<tr>
<td>Studies using prisoner samples</td>
<td>77</td>
</tr>
<tr>
<td>Method</td>
<td>86</td>
</tr>
<tr>
<td>Subjects</td>
<td>86</td>
</tr>
<tr>
<td>Procedure</td>
<td>92</td>
</tr>
<tr>
<td>Results</td>
<td>96</td>
</tr>
<tr>
<td>Discussion</td>
<td>128</td>
</tr>
<tr>
<td>References</td>
<td>136</td>
</tr>
<tr>
<td>Appendix A</td>
<td>144</td>
</tr>
<tr>
<td>Appendix B</td>
<td>146</td>
</tr>
<tr>
<td>Appendix C</td>
<td>147</td>
</tr>
<tr>
<td>Vita</td>
<td>148</td>
</tr>
</tbody>
</table>
ABSTRACT

The purpose of this study was to examine the effect of various instructional sets on the MMPI performance of prisoners. Specifically, an attempt was made to examine whether any significant biases existed in the scales and ratios used to measure the prisoner's attitudinal set concerning the MMPI. Ninety-six subjects were selected from the U. S. Camp and Penitentiary in Lompoc, California. The subjects were randomly assigned to three groups of 32 subjects each. One group was requested to take the MMPI under instructions to malinger psychopathology. Another group was instructed to take the MMPI under instructions to deny or conceal any psychological problems. The third group was administered the MMPI under standard instructions so as to serve as a control group. The results indicated that subjects were able to feign maladjustment on the MMPI when instructed to do so. At the same time, the validity indicators were effective in detecting the malingering subjects. Under instructions to feign hyper-adjustment, the subjects were considerably less successful at manipulating the clinical scales. Likewise, the validity measures were not uniformly effective in detecting attempts to fake good. Appropriate cut-off points were devised for each of the validity indicators which successfully produced significantly different distribution between the experimental and control group subjects. The results were discussed in terms of various forensic and legal issues concerning the use of the MMPI with prisoners.
INTRODUCTION

During recent years, psychologists have become increasingly involved in working within various facets of the criminal justice system (Ziskin, 1981). A perusal of recent literature reveals a large increase in the number of publications in this area, suggesting an increased interest among researchers in the interface of psychology and law. Although the functions of psychologists within the criminal justice system are many and varied (Cooke, 1980), clinicians are frequently requested by the courts to assess the credibility of a litigant. To this end, a variety of techniques have been developed which make use of standardized tests such as the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1967). As will be described later in detail, the MMPI includes several indicators of test-taking attitudes which have direct relevance to the evaluation of a respondent's credibility. Although by no means the only method available to achieve this purpose, the MMPI seems particularly appropriate given its frequent utilization within the criminal justice system, and the large body of empirical data available on the various scales (Dahlstrom, Welsh & Dahlstrom, 1975).

The issue of credibility is an especially important consideration in the assessment of an individual's competency to stand trial. A person charged with crimes may view it to his advantage to feign
or malinger psychopathology in order to avoid or postpone criminal prosecution on the basis of mental incompetency. Similarly, the problem of credibility becomes relevant in making decisions concerning the release from institutional confinement of persons considered mentally ill and/or dangerous. In this case, it seems reasonable to assume that some individuals might attempt to conceal or deny any psychological abnormalities which may in fact be present in order to increase the likelihood of being released.

There is a large body of data indicating that subjects can falsify their responses on the MMPI so as to appear either more disturbed or more adjusted psychologically than they are upon being instructed to do so (Dahlstrom et al., 1975). However, there are several scales and indices which are sensitive to these types of test distortion and which can accurately detect feigned records. One major problem with these techniques is that, for the most part, they have been developed and standardized on non-prisoner samples (e.g., college students, psychiatric patients). Only three published studies were found in the literature in which the effectiveness of the various deception-detecting techniques was evaluated with prisoner samples, and these are inadequate in several respects (cf., p. 71 ff.). The appropriateness of using these techniques to evaluate prisoners is thus still in question. Given that the question of competency to stand trial in particular, as well as the issue of release from a confining institution, frequently arises in the context of the penal system, it seems necessary to evaluate with more
suitable methods the effectiveness of the available MMPI techniques for assessing credibility with a prisoner sample. Such is the intent of this study.

The following sections of this review will be concerned with a description and evaluation of research studies on the various MMPI faking detection techniques. Following a brief description and a review of the history of the development of the MMPI, the focus will shift toward an examination of research on the standard validity scales (? , L, F & K), the validity scale configurations, and the special response measures currently available.
DESCRIPTION AND BRIEF HISTORICAL OVERVIEW

The MMPI consists of 566 items which are sorted into one of three categories, namely, "true," "false," or "cannot say." In the original version of the test, it was administered on an individual basis. The statements were printed on separate cards, and the respondent placed each card in one of three groups, one for each category. Subsequently, a group form was devised, in which the statements are printed in a test booklet and the examiner records the responses on an answer sheet. The items included in the test vary widely in content, and cover areas such as the following: general health, neurological disorders, psychosomatic symptoms, motor disturbances; sexual, religious, political, and social attitudes; educational, occupational, family and marital questions; and several neurotic or psychotic behavior manifestations, such as obsessive and compulsive states, ideas of reference, hallucinations and delusions (Anastasi, 1976).

The MMPI was developed by Stark Hathaway and J. C. McKinley during the late 1930's and early 1940's. It was partly a reaction to the lack of demonstrated success of earlier instruments derived on a rational basis, such as the Woodworth Personal Data Sheet (Woodworth, 1920), and the Bernreuter Personality Inventory (Bernreuter, 1933). In the construction of such rationally derived
tests, items were included on a given scale if, on the basis of clinical experience, they were believed to measure a particular trait or construct. Greene (1980) has reviewed a number of early studies which seriously question the reliability and validity of these and similarly constructed inventories.

Hathaway and McKinley intended to develop an inventory which could overcome the shortcomings of the earlier instruments. To this end, they utilized an empirical approach to scale construction. They began by assembling an item pool of over 1,000 statements (Hathaway & McKinley, 1940). These items were selected from psychiatric textbooks, other personality inventories and the clinical experience of the developers, to generate a large group of personality descriptors. After elimination of duplicate items and items with other difficulties, 504 items were retained; these provided the basis for much of the derivational work that followed.

Using this 504-item pool, Hathaway and McKinley constructed a series of scales that would be useful in diagnosing behavioral disorders. They collected samples of normal men and women and selected patient groups in the clinics and wards of the University of Minnesota Hospitals. In order to select items for a particular scale, the items had to be answered differently by the criterion group as compared with normal groups. The specific procedure in the derivation of the MMPI scales involved a number of steps. Initially, an appropriate criterion group was established for each of the diagnostic categories of interest (e.g., Hypochondriasis). Then appropriate normative groups
were assembled to control for diverse "nuisance" variables, such as age, socioeconomic status, and education, and to provide a data base for comparing response frequencies. Once the criterion and normative groups were established, item selection for particular scales was determined by differential response frequencies between the various criterion groups and the reference groups.

An attempt to cross-validate the scales followed in which new groups of normal subjects and clinical subjects with particular diagnoses were selected and the scales were administered to them. If significant differences were obtained between scores for the normal and clinical groups being considered, the scale was assumed to have been adequately validated and ready for use.

This process resulted in the initial development of the following seven clinical scales: Scale 1 (Hypochondriasis\(^1\)); 2 (Depression); 3 (Hysteria); 4 (Psychopathic Deviate); 6 (Paranoia); 7 (Psychasthenia); and Scale 8 (Schizophrenia). Subsequently, two additional clinical scales were derived. Scale 5 (Masculinity-Femininity) was originally intended to distinguish between homosexual males and heterosexual males. Because a small number of items were obtained that discriminated between homosexual and heterosexual males, other items were added that were differentially endorsed by normal male and female subjects. Thus, the criterion group of male homosexuals could not be compared to the original normative group. Instead, 54 male

\(^1\)Presently, common terminology utilizes the scale number rather than the name.
soldiers were used as one of the normative groups in the construction of this scale.

Scale 0, the Social Introversion Scale, was developed by Drake (1946) and has come to be included among the MMPI clinical scales. This scale was constructed by selecting items from the original item pool that differentiated between college students who tended to participate in many extracurricular activities from those who were not very socially participative. Although the initial derivation was conducted separately for males and females, the norms were highly similar and thus the normative data for the two groups was combined.

Table 1 presents a listing of the ten clinical scales currently used in a routine fashion in MMPI scoring and interpretation.

One special feature of the MMPI is its utilization of four scales designed to assess test-taking attitudes, known commonly as the validity scales. Although previous test developers had addressed the importance of assessing test-taking attitudes, the MMPI represents the first attempt to measure such attitudes directly through the use of scales developed in an empirical manner. The four scales routinely employed to assess deviant response tendencies are the following: the "Cannot Say" (?) Scale, which represents the total number of items the client omits; the Lie Scale (L), which is

As used here, the term "validity" refers to the acceptability or appropriateness of any given administration of the test (Dahlstrom, Welsh & Dahlstrom, 1972), as opposed to the broader concepts of validity used in psychometric theory.
<table>
<thead>
<tr>
<th>Scale Number</th>
<th>Scale Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypochondriasis</td>
</tr>
<tr>
<td>2</td>
<td>Depression</td>
</tr>
<tr>
<td>3</td>
<td>Hysteria</td>
</tr>
<tr>
<td>4</td>
<td>Psychopathic Deviate</td>
</tr>
<tr>
<td>5</td>
<td>Masculinity-Femininity</td>
</tr>
<tr>
<td>6</td>
<td>Paranoia</td>
</tr>
<tr>
<td>7</td>
<td>Psychasthenia</td>
</tr>
<tr>
<td>8</td>
<td>Schizophrenia</td>
</tr>
<tr>
<td>9</td>
<td>Hypomania</td>
</tr>
<tr>
<td>0</td>
<td>Social Introversion</td>
</tr>
</tbody>
</table>
based on a group of items that make the respondent appear in a favorable light but are unlikely to be truthfully answered in the favorable direction; the F Scale, which is based on responses to a set of items very infrequently answered in the scored direction by the original standardization group; and the K Scale, which like the L Scale, provides a measure of defensiveness, but which is believed to be more subtle.

The remaining portions of this review will focus specifically on research dealing with the four validity scales described above as well as other scales designed to assess test-taking attitudes. An attempt will be made to cite pertinent findings which will allow for an assessment of the effectiveness of the various scales. Further, studies reporting data concerning non-test correlates of these scales will be reviewed.
CANNOT SAY (?) SCALE

As mentioned earlier, the Cannot Say score is simply the number of items omitted by a respondent, including items answered both true and false. The effect of item omission on a given test protocol is to lower the general elevation of the profile, since omitted items are considered to be answered in the non-deviant direction. Specific patterns of item omission may result in a lowering of the score for a particular scale, particularly if the examinee is having difficulty with certain types of items. Clopton and Neuringer (1977) examined the effect of omitting a certain number of items on the MMPI profile obtained. They selected 180 fully completed (no omissions) protocols and randomly omitted groups of 5, 30, 55, 80, 105 and 130 items. They found that as more items were omitted, there was a progressive reduction in the elevation of the profile. For each of the levels of omitted items described above, there was a corresponding drop in the clinical scales of an average of .45, 2.74, 5.61, 7.70, 9.09 and 11.54. Further, there were changes in the profile high-point pairs, and the number of changed profiles increased as the number of omissions increased. The effect of item omission is thus a negative one and should be taken into account in the interpretation of individual profiles as well as in conducting MMPI research.
An important consideration in dealing with the problem of item omission concerns the respondents' motivation for failing to answer given items. Dahlstrom, Welsh and Dahlstrom (1972) provide a number of possible reasons. One of these is that some items may not apply to certain types of examinees. For example, an individual orphaned at an early age may find items dealing with family relations inapplicable to her/him. A subject might also find the content of certain items intrusive upon her/his privacy, such as items bearing on sexual or religious practices, or those dealing with intimate bodily functions.

In an attempt to examine whether item content is related to item omission, Gravitz (1967) examined item omission patterns in 4,816 females and 7,149 males who took the MMPI as part of a pre-employment screening procedure. He found that the most frequently omitted items fell into one of the following six content areas: personal attitudes and interest, sex, family, religion, politics and law and order, and fears. Similarly, Butcher and Tellegen (1966) identified four item categories which a group of college students found objectionable. These included "sex," "religion and religious beliefs," "family relationship," and "bladder and bowel movements." These studies suggest that subjects may be prone to omit certain items because of the content being tapped, that is, they may find certain items too personal or intrusive and hence refuse to answer them.
Another factor which could account for excessive item omission is the tendency of some examinees to adopt a defensive approach to taking the MMPI. Dahlstrom, Welsh and Dahlstrom (1972) suggest that in an effort to deliberately tailor responses to project a certain image, subjects may opt to omit a number of items rather than answer them one way or the other. Tamkin and Scherer (1957) hypothesized that a high number of item omissions could be related to a defensive attitude on the part of subjects and to symptoms of depression and psychasthenia. In order to test their hypotheses, they performed chi-square analyses between "Cannot say" scores and each of the following scales: L, F, 2 and 7. They found no significant differences between the scales in terms of item omissions, and concluded that a high number of item omissions does not necessarily represent a defensive attitude. It is not clear why the authors in this study used chi-square analyses in order to assess the relationship between "Cannot say" scores and scales L, F, 2 and 7. If this was their main purpose, some type of correlational analysis may have been more appropriate. The findings are thus difficult to interpret. Another problem with this study is that item omissions result in an attenuation of scores on all of the test scales, including L and K, and thus the results from each scale taken separately are difficult to compare.

In an effort to examine whether the occurrence of MMPI item omissions varied according to the specific respondent group involved, Clopton and Neuringer (1977) examined the frequency of omission among
439 Veterans Administration Hospital psychiatric patients, 162 mental health center outpatients, and 112 job applicants. They found that a significantly higher percentage of complete MMPI protocols (no omissions) was produced by the job applicants in comparison to the two other groups. The mean number of omitted items was 4.68 for V.A. inpatients, 2.24 for mental health center clients, and 2.45 for job applicants. In addition, they found that the vast majority of subjects answered all 566 items; the percentage of people omitting five or fewer items was 85.6 for V.A. inpatients, 88.9 for mental health center clients, and 93.7 for job applicants. Ball and Carrol (1960) collected data on 224 ninth grade public school students and 38 incarcerated juvenile delinquents. Although they did not find significant differences between the public school students and the delinquents, they found a significant effect for sex, with "Cannot say" scores being significantly lower for girls than boys. The mean "Cannot say" score for the various groups ranged from 1.07 to 10.14.

Greene (1980) presents normative data on frequency of item omission for four subject groups. The first sample consisted of 415 clients treated at a psychology clinic in a major southwestern university. The second sample was composed of 241 patients treated by a specialist in internal medicine in private practice in a large metropolitan area in the southwest. These subjects were administered the MMPI as a routine screening procedure. The third sample consisted of 200 prisoners sentenced to a state penitentiary in the western United States who took the MMPI as part of a routine battery.
The fourth sample was comprised of 209 freshmen and sophomores at a major southwestern university who took the MMPI in order to provide normative data on the performance of normal college students. His findings are reported in Table 2.

As can be seen from examining Table 2, the vast majority of subjects omitted fewer than five items. Roughly five percent of subjects in each of the four samples omitted more than 30 items. These data suggest that a large number of omissions occurs rather infrequently. Nonetheless, the fact that they do occur requires that some systematic method of dealing with them be available. Brown (1950) proposed a method of including the "Cannot say" items in the scoring of the clinical and validity scale. He reasoned that subjects respond as "Cannot say" for one of the following reasons: the items do not apply to her/him or she/he does not know about them; an inability to be decisive, or "pathological irresoluteness;" or a deliberate unwillingness to respond "true" or "false." He suggested that the only ones that should remain unscored are those in the first category. Those in the other two groups would be given a weight of one-half point.

There are some difficulties inherent in Brown's method of dealing with item omissions. It presupposes that the clinician has the time and/or expertise to make dependable judgments about how a given subject interprets an item or how it relates to particular circumstances in her/his life history. Further, his reasoning is based on the notion that the added weights given to unanswered items provide
<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Clinic Patients (n=415)</th>
<th>Medical Patients (n=241)</th>
<th>Prison Inmates (n=200)</th>
<th>University Students (n=209)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>54.9%</td>
<td>30.5%</td>
<td>60.5%</td>
<td>74.6%</td>
</tr>
<tr>
<td>1 - 5</td>
<td>27.3%</td>
<td>37.5%</td>
<td>25.0%</td>
<td>21.1%</td>
</tr>
<tr>
<td>6 - 30</td>
<td>11.9%</td>
<td>24.7%</td>
<td>10.0%</td>
<td>3.3%</td>
</tr>
<tr>
<td>&gt; 30</td>
<td>2.5%</td>
<td>7.4%</td>
<td>3.5%</td>
<td>.5%</td>
</tr>
</tbody>
</table>

for more accurate clinical profiles. No data are provided by the author which address these difficulties. Hovey (1958) has investigated the advantages of a similar correction procedure by contrasting profiles created by either leaving the unanswered items out or by scoring them with full credit on the scales in which they fall. He also compared profiles obtained using standard scoring procedures with those generated by having subjects guess on items they were unsure of. He found that profiles obtained by using either of these correction procedures were more in agreement with non-test evaluations of the respondents than were those obtained under the standard system. Further, the main effect of either correction procedure was on the elevation of the profile rather than the profile configuration, suggesting that the corrected profiles corresponded more closely with the judged severity of the disorder without influencing the form of the disorder as portrayed by the configural pattern.

Any procedure designed to correct for excessive item omission is likely to possess an element of arbitrariness, and the possible gains to be obtained from it will have to be weighed against the loss of accuracy which may result. It is preferable to attempt to eliminate the presence of a large number of omissions by enlisting the full cooperation of the subject. Dahlstrom, et al. (1972) also suggest that should a large "Cannot say" score be obtained, the examiner might interview the subject about her/his possible reasons for item omissions, or perhaps request that she/he go over unanswered items. Clarification of doubts or explanation of the meaning
of words might be helpful in this regard.

The raw score on the "Cannot say" scale can be converted to T scores by using the tables provided by Hathaway and McKinley (1967). A raw score of 30 would thus correspond to a T score of 50, one of approximately 65 would equal a T score of 60, and one above 100 would be considered equal to or greater than a T score of 70. Greene (1980) considers these T score conversions to be out of line with more recent data. He notes that in his normative data, a raw score of 30 on the "Cannot say" scale occurs in about five percent of cases. This would mean that a raw score of 30 is close to a T score of 70 rather than 50, as proposed by Hathaway and McKinley (1967). Likewise, he proposes that a raw score of 100 is approximately a T score of 80 rather than 70. Until more normative data can be obtained, it seems prudent to use the T score equivalents available for the "Cannot say" raw scores with caution, although it seems clear that a large number of item omissions makes the accuracy of a given protocol suspect.

The following three sections of this review will be concerned with research studies investigating the L, F and K scales. Unlike the "Cannot say" scale, which does not include a specific set of items, each of the three major validity scales contains a specified group of items which contribute to the total raw score. One major problem inherent in the validity scales (as well as the clinical scales) is the presence of items which are scored on two or more of the scales, i.e., some items contribute to the raw score on more
than one scale. The result is that the scales are not completely independent. This creates major difficulties in interpreting research findings on MMPI scales in general, but is particularly problematic in studies examining inter-scale relationships without any external validational criteria. This issue should be kept in mind in interpreting the research findings to be discussed below.
LIE (L) SCALE

The MMPI Lie Scale was designed originally to detect deliberate efforts to avoid answering the test in an honest manner (Dahlstrom et al., 1972). It consists of 15 items selected on a rational basis dealing with content areas such as denial of aggression, bad thoughts, minor personal dishonesties, and weakness of character. Some examples of items on this scale and the deviant responses are listed below:

"I do not like everyone I know." (false)

"I do not read every editorial in the newspaper every day." (false)

"I get angry sometimes." (false)

The items on the L scale are all scored in the deviant direction when the examiner responds "false." This feature makes it susceptible to certain unsophisticated deviant response sets, such as "all false" responding, or to a general acquiescence set in responding to items, i.e., the tendency of some respondents to give one or the other response alternatives available without regard for content. On the other hand, more sophisticated deviant response sets may not be detected by the L scale, perhaps because certain subjects realize it would be unconvincing to respond to L items in the scored direction (Greene, 1980).

Although the 15 items comprising the L scale were selected on a judgmental basis, early findings supported the notion that a high
score indicated an unusual response pattern. In the original Minnesota normative sample, most of the items were endorsed in a deviant direction by less than twenty percent of subjects (Dahlstrom, Welsh & Dahlstrom, 1975). More recent data generally support the findings obtained from the Minnesota group. Gravitz (1970) administered the MMPI to 6,686 males and 4,717 females who were presumably normal voluntary applicants for various occupational positions. His findings indicate that the majority of male subjects responded in the non-defensive direction on all but four items, whereas females did so on all but five items. Further, Gravitz found significant differences between males and females in response frequencies to twelve of the 15 items, suggesting the need for separate norms for each sex. One possible limitation with this study is that it is reasonable to suspect at least moderate levels of defensiveness among a group of job applicants. The author addresses this issue by proposing that such a defensive set "would probably apply to other non-clinical situations where the MMPI is administered, not excluding research settings" (Gravitz, 1970, p. 497).

It is of interest to examine the specific L scale items frequently endorsed in a deviant direction by a majority of normal subjects. Gravitz (1970) found that items No. 15, 135, 165 and 255 were answered in the affirmative by most subjects in his sample in both sex groups. Similarly, these items were not acknowledged by a majority of the Minnesota college normals (Dahlstrom et al., 1972). The content of these items is listed below:
No. 15. Once in a while I think of things too bad to talk about.

No. 135. If I could get into a movie without paying for it and be sure I was not seen, I would probably do it.

No. 165. I like to know some important people because it makes me feel important.

No. 255. I gossip a little at times.

The reasons for subjects unwillingness to endorse these items are not clear, and to date, there are no data specifically addressing this issue. Nonetheless, the appropriateness of including these items on the L scale is questionable.

As mentioned earlier, although the L scale may detect relatively unsophisticated deviant response patterns, it is not sensitive to more sophisticated attempts to manipulate the test. Several studies have found that the L scale fails to detect sophisticated subjects instructed to falsify their responses (Greene, 1980; see also discussion below on special response measures). Thus, the L scale appears to be sensitive to factors influencing psychological sophistication, such as socioeconomic status and education. Persons from high socioeconomic classes and college educated subjects rarely obtain a high score on the L scale (Graham, 1977). Hence, such factors need to be taken into consideration when interpreting L scale elevations.

High L scale elevations generally have a suppressive effect on the clinical scales. Dahlstrom et al. (1972) provide data indicating that there are very few records with any clinical scales over 40 T which also have L scale T scores of 66 or more. Further, the
relative frequency of records with high points on the lower numbered scales (scales 1, 2, 3 and 4) tends to increase with L score elevations, whereas the relative frequencies of the higher numbered scales generally decrease.

Relatively few studies have investigated empirical correlates of the L scale. Matarazzo (1955) examined the relationship between scores on the Taylor Scale of Manifest Anxiety, a scale derived from the MMPI, and scores on scales L, F and K. He found that in his sample of 119 medical students, L scale scores were inversely related to scores on the Taylor scale \( r = -0.32 \). These results, although statistically significant, are of limited practical usefulness since the two measures only shared approximately ten percent common variance. Burish and Houston (1976) have provided somewhat stronger validational data for the L scale. They found a significant positive correlation \( r = +0.55 \) between the L scale and another MMPI measure of defensiveness, the denial (Dn) scale (Little & Fisher, 1958) in a group of 66 male introductory psychology students. (It is noted that the 27 item Dn scale and the 15 item L scale share one item, No. 30, which is scored in the same direction on both scales.) They also found the L scale to be unrelated to scales 1 (Hypochondriasis) and 8 (Schizophrenia). (No items are shared by the L scale and either scale 1 or 8.) These data were seen as providing evidence of both convergent and discriminant validity. These authors also found that high L scale scorers exhibited less stress in a threatening situation (avoidable versus unavoidable shock) than low scorers; they interpret
this finding as suggesting that high L scores are associated with a tendency to employ defensive maneuvers against threat.

Reliability data for the L scale indicate that test-retest coefficients among various subject groups range from .73 to .85 for a one-day interval between testings, from .66 to .79 for a one- to two-week interval, from .35 to .61 for a one-year interval, and from .39 to .50 for a three-year interval (Dahlstrom et al., 1975).

The standard MMPI profile form developed by Hathaway and McKinley (1967) indicates that a raw score of four on the L scale is equivalent to a T score of 50, and a raw score of ten equals a T score of 70. However, in the 1967 MMPI manual, they note that clinical experience has shown that a raw L score of approximately 7 should equal a T score of 70. Further, Rosen (1958) has suggested that a raw score of ten should equal a T score of 80.
F SCALE

The F scale was originally developed to detect unusual or atypical ways of responding to the test items. It has been referred to as the frequency (or infrequency) scale, the confusion scale, or the validity scale. It consists of 64 items selected on the basis of the frequency with which they were endorsed by the original Minnesota normative samples. Specifically, an item was included if no more than ten percent of the normative sample endorsed it in the deviant direction. The item frequency counts used to select potential F-scale items came from an early subsample of the Minnesota normal subjects (Gynther, Lachar & Dahlstrom, 1978). Subsequent analysis based on the complete data of the Minnesota samples revealed that five of the 64 items (20, 54, 112, 115 and 185) do not meet the ten percent or below criterion for inclusion, and an additional three items for females and eleven items for males do not meet the criterion (Greene, 1980). Further, there are 38 items that could have been included in the F scale in that they met the ten percent or less criterion but were excluded by the test authors for unknown reasons.

The content of the F-scale items is fairly unambiguous and quite varied. Content areas covered by these items include bizarre sensations, strange thoughts, peculiar experiences, feelings of alienation and isolation, atypical attitudes toward laws, religion or authority,
and a number of unlikely or contradictory beliefs, expectations and self-descriptions (Dahlstrom et al., 1972). Example of F-scale items are listed below:

"When I am with people, I am bothered by hearing very queer things." (true)

"My soul sometimes leaves my body." (true)

"At times, I am all full of energy." (false)

Initially, high scores on the F scale (typically a raw score greater than or equal to 16) raised the questions of the possible invalidity of the resulting profile. Gynther (1961) examined a group of 246 protocols and found that 39 of these had F raw scores greater than 16. Of these, the majority (about two-thirds) were obtained from individuals labeled as behavior disordered by independent criteria. Since these subjects were court referrals, they may have had reason to dissemble or "fake bad," and thus the resulting profiles were not accurate representations of the subjects, i.e., they were technically invalid. Nonetheless, Gynther suggests that the High F score can be given a characterological interpretation, and may be considered a valid measure of antisocial tendencies. Similar results were obtained by Gynther and Shimkunas (1965a).

Other studies have indicated that F scores may be sensitive indicators of severity of psychopathology. Blumberg (1967) found that in a sample of 347 psychiatric admissions to a temporary, acute treatment center, 70 (twenty percent) had F scores greater than 16. The group with F scale elevations tended to remain hospitalized in
the acute unit over longer periods of time, had a higher rate of subsequent discharge to state mental institutions, and had a higher proportion of diagnoses of psychosis. Gauron, Severson and Engelhart (1962) also found that a majority of a psychiatric patient sample with F raw scores over 16 were diagnosed as psychotic. Gynther and Shimkunas (1965b) obtained parallel results in their psychiatric patient group.

Elevations on the F scale may also be related to the age of the subject population, although findings in this area are not consistent. Gauron et al. (1962) indicated that when they divided their subject sample on the basis of age, high F scale scorers under the age of 40 were more likely to be diagnosed as behavior disordered, whereas those over 40 were almost always classified as psychotic. Gynther and Shimkunas (1965a) found that F scores decreased with increasing age for low- and high-IQ subjects, but they remained relatively constant for subjects within the average range. On the other hand, Blumberg (1969) obtained no significant differences between subjects with high F scores divided into a younger (age 22 and under) and an older (age 23 and over) group.

Addressing a related issue, McKegney (1965) proposed that the high F scale scores frequently observed among juvenile delinquents may be a realistic reflection of certain deviant attitudes, feelings and behaviors actually present in this population as a group. To test this hypothesis, he had three professional staff members (two psychologists and one social worker) and three direct care staff
members (correctional officers) respond to the 64 F scale items according to their expectations of how the average delinquent would honestly answer. In addition, he administered the MMPI to 29 delinquent boys. The results indicate that the delinquent group obtained elevations on the F scale ($\bar{X} = 15.4$), consistent with the notion that as a group they tend to get higher scores. Further, the six judges responded in the deviant direction to an average of 21 F scale items, suggesting that persons experienced in working with delinquents predict that the average delinquent can be expected to obtain F scale elevations above the usual validity levels. Of particular interest is that only certain meaningful F items contributed to the delinquent's high F score. Item frequencies were significantly higher for certain content categories, such as "Attitude Toward Law and Religion" and "Impulse Control" than for other categories such as "Somatic Concerns" and "Peculiar Thoughts or Beliefs." McKegney suggests that these content categories accurately and realistically reflect the attitudes, feelings and behavior of delinquents as a group.

Another important variable to consider in understanding the meaning of F scale elevations is that of race. Gynther (1972), in a review of the literature on MMPI black-white differences, concludes that both presumably normal and institutionalized blacks generally obtain higher scores on scale F than white subjects. Gynther, Altman and Warbin (1973) examined the correlates of MMPI profiles with F raw scores greater than or equal to 26, obtained from 55
white and 15 black psychiatric hospital patients. They identified and cross-validated seven descriptors among the white subjects, including: inability to understand proverbs, delusions of reference, auditory hallucinations, disorientation for place, short attention span, and poor recent memory. The two remaining descriptors portrayed these subjects as monosyllabic and as not knowing why they were hospitalized. Descriptors that distinguished the black subjects with an F score greater than 26 from lower scoring blacks failed to hold up on replication, indicating there were no replicable differences between high and low scoring blacks on any of the descriptors. These results suggest that whereas high F scale elevations appear to reflect severity of psychopathology for white subjects, for blacks the meaning of such elevations is not clear and seems to be substantially different from that associated with whites.

In response to the need for special MMPI norms for black subjects, Gynther, Lachar and Dahlstrom (1978) have developed and F scale for use with this population. Using a sample of 882 normal black adults (321 males and 561 females), they identified 33 items that met the 10 percent or less endorsement criterion used to develop the MMPI F scale. Their data also indicate that although white and black subjects responded similarly to the 33 items of the F scale for blacks (i.e., in terms of frequency of endorsement), comparison of blacks and whites on the standard F scale revealed considerable disagreement. Specifically, black endorsement patterns agreed with only 37 percent of the standard F scale items. These results
suggest that the new F scale may yield a more accurate measure of correlates associated with endorsement of deviant items than the standard F scale among black subjects.

The results from the studies cited above and other similar investigations indicate that F scale elevations may have different interpretive significance depending on the clinical status (e.g., court referred versus psychiatric patient), age, race, and other demographic factors of the various subject groups. Dahlstrom et al. (1972) have suggested other possible variables which may influence F scale scores as well as scores on other scales. Factors affecting a subject's attention and concentration, such as poor testing facilities, may hamper his/her ability to provide accurate responses. Inability to read or comprehend test items could also affect the results obtained. Similarly, patients experiencing a toxic reaction, or those in heavily drugged states may be unable to give interpretable test results. A test subject may also impose special meaning on the testing process, and utilize it as a "cry for help," i.e., an opportunity to call attention in a dramatic fashion to a need for assistance. Other reasons for F scale elevations include the operation of specific response sets or patterns, such as "all true" or "all false" responding, attempts to manipulate the test so as to appear in a highly unfavorable light, or acquiescing to any symptom or difficulty suggested. These response sets will be discussed in detail in later sections of this review.
Elevations on the F scale are associated with an overall elevation of the clinical profile. With increasing F scale scores, elevations on scales 1, 2, 3, 4, 7 and 9 become relatively less frequent, whereas increased scores on scales 6 and 8 are much more frequent (Dahlstrom et al., 1972).

The test-retest reliabilities for the F scale range from .80 to .97 for one- to two-day intervals, from .62 to .87 for one- to two-week intervals, from .51 to .61 for an eight-month interval, from .63 to .76 for a one-year interval, and from .45 to .49 for a three-year interval.

The T scores for F raw score values were arbitrarily assigned by Hathaway and McKinley (1967). They originally proposed that a raw score of three should equal a T-score of 50, a raw score of twelve should equal a T-score of 70, and a raw score of 16 should be equal to a T-score of 80. However, based on their clinical experience, they later suggested that a raw score of 16 should equal a T-score of 70 (Hathaway & McKinley, 1967).
K SCALE

The K scale consists of 30 empirically derived items developed with the intent to include on the MMPI an additional validity indicator sensitive to more subtle kinds of test distortion. It was derived largely by finding items which distinguished between presumably normal subjects and hospitalized psychiatric patients who showed normal test profiles and elevated L scale scores, the assumption being that the occurrence of a normal profile was suggestive of a defensive attitude on the part of the patients (Meehl & Hathaway, 1946; McKinley, Hathaway & Meehl, 1948). The L scale elevation supported the notion that a defensive attitude was operating. Twenty-two items were selected which differentiated between these two groups, and an additional eight items were added which improved detection of depressive and schizophrenic symptoms, resulting in a total of 30 items. In addition, various fractions of this scale were empirically derived which when added to the clinical scales would maximize the discrimination between the criterion and the normative groups. These optimal weights are currently routinely employed with five clinical scales, namely, scales 1, 4, 7, 8 and 9. It was found that the addition of K to the other clinical scales 2, 3, 5 and 6 resulted in a loss of discriminative power, and thus they were not K-corrected.

The content of the K scale items is quite heterogeneous,
covering descriptions of the subject's mental health, stability and control, feelings and expectations about others, and family relationships. Examples of K scale items are listed below:

"I find it hard to set aside a task that I have undertaken, even for a short time." (false)

"I have very few quarrels with members of my family." (false)

"At times I feel like smashing things." (false)

As noted earlier, the nature of the criterion group used to derive the K scale (i.e., psychiatric patients with normal test profiles and elevated L scores) suggested that it might be tapping a defensive test-taking attitude whereby the subjects were attempting to present themselves in an unduly favorable light. In order to test this notion empirically, Nakamura (1960) compared the MMPI scores of a group of university students under two levels of motivation, one being a relatively non-stressful situation and a second where subjects could be expected to be highly motivated to fake a good test result. The non-stressful situation consisted of a routine administration of the MMPI upon entrance to a university. The experimental subjects (high motivation to present a good impression) were students referred to a disciplinary bureau for violation of university regulations. Nakamura's results indicate that the K scale was significantly higher for the experimental subjects compared to the controls, which supports the notion of a relationship between K scale elevations and test-taking defensiveness.

Other research indicates that while K scale elevations may represent some measure of defensiveness among clinical groups, in
normal subjects such elevations may be associated with psychological health or adequate adjustment. Heilbrun (1961) compared the K scores of a group of 289 university counseling service clients with those of 350 presumably normal college students. Although no significant differences were obtained between the two groups, the females in the normal group ($N = 153$) obtained significantly higher K scores than a subset of most seriously maladjusted females ($N = 43$). In addition, this study found a significant positive correlation (.64) between K scores and the defensiveness scale of the Adjective Checklist for the counseling service subjects, supporting the hypothesis that K scale elevations reflect a pattern of defensiveness for maladjusted subjects. The correlational findings for normal subjects were inconsistent, with the male group ($N = 92$) obtaining a smaller but significant positive correlation (.35) and the females ($N = 141$) obtaining a significant negative correlation (-.36) between K scores and scores on the defensiveness scale. These results, although far from being conclusive, lend some support to the notion that K scale elevations are associated with defensiveness in clinical groups, whereas among normal subjects such elevations may not have the same correlates. Further, these data suggest that K scores may have different meanings for normal male and female subjects.

Similarly, Matarazzo (1955) obtained a significant negative correlation between K scores and scores on the Taylor Manifest Anxiety scale in a group of medical students, supporting the notion that in
a normal sample the K scale may be a measure of psychological health. Further, Smith (1959), using a group of college students and corporate supervisors, found that elevated K scores were positively correlated to measures of insight, and Sweetland and Quay (1953) observed that K scores were associated with measures of social security and emotional adjustment among college students. Addressing the issue of the relationship between K scores and psychological adjustment from a different angle, Ries (1966) hypothesized the K scores might be related to clinical ratings of improvement in a group of 60 state hospital patients. He found a significant correlation (.66) between K scores falling within the range of nine to 15 and ratings of improvement. K scores outside of this range, either higher or lower, were related to ratings of not improved. These findings suggest that in clinical groups, K scores within a certain range may represent greater psychological health, whereas more extreme scores suggest the opposite. The data also support the idea that K scores do not have the same correlates for clinical and normal subjects, in that for clinical subjects, the relationship between K scores and psychological adjustment appears to be curvilinear, whereas the same does not hold for normals.

Other researchers have investigated the effectiveness of using the various K corrections for scales 1, 4, 7, 8 and 9 as a means for improving the validity of these scales. In general, the findings have not been encouraging, and some writers have questioned the appropriateness of using the K corrections indiscriminately in the absence
of empirical evidence to support this practice (e.g., Greene, 1980; Yonge, 1966). In an early study addressing this issue, Hunt, Carp, Cass, Winder and Kantor (1948) compared the effectiveness of using K-corrected versus uncorrected profiles in differentiating between psychotic and non-psychotic male psychiatric patients. They found that K correction failed to improve the accuracy of such classifications significantly.

In an attempt to measure the usefulness of both the K scale and the K corrections on the accuracy of diagnostic classifications, Silver and Sines (1962) had two clinical psychologists sort the profiles of a sample of 100 male and 100 female state hospital patients into four diagnostic categories, namely, affective psychotic, schizophrenic, neurotic and personality disorder. They prepared four profiles for each subject, either excluding or including the K score, and either having the scales K-corrected or uncorrected. They found that neither knowledge of the K score, nor the use of K-corrected profiles, increased the accuracy of diagnostic classification. Results consistent with these were obtained by Yonge (1966) in his comparisons of K-corrected and uncorrected clinical scales vis-a-vis measures of social-emotional adjustment.

Other research in this area has addressed the problem of devising different K-correction weights for use with varied subject groups. The reasoning here follows from findings that, as mentioned earlier, the correlates of K scale values differ across diverse populations. It might not be surprising that the optimal K corrections
for MMPI scales might also vary across different subject groups. Heilbrun (1963) addressed this issue by attempting to devise a system of K weighting for the clinical scales which would improve their usefulness as measures of adjustment within normal college populations. Using 2-group discriminant analysis, he determined the K values which maximized discrimination between maladjusted and adjusted college students. The resulting weighting system differed from the standard system, the main differences being a negative weighting for scale 3 (-0.7K for males and -0.5K for females), and deletion of weights from scales 1, 4 and 9. Smaller changes in weighting were obtained for scales 7 and 8, and scales 2, 5, 6 and 0 continued to be unweighted. Similarly, Fricke (1956) has suggested that the validity of scale 3 is increased by subtracting a fraction of K (.6). Using a group of 63 clinically diagnosed conversion hysterics, he found that the discriminant validity of scale 3 was increased through the use of this K correction procedure.

Another investigation has produced findings which argue against the indiscriminate use of the standard K-correction system across different subject groups. Ruch and Ruch (1967) gave the MMPI to 182 sales representatives who had been categorized into an upper and a lower criterion group in terms of job effectiveness. They found that of the five clinical scales which normally undergo K-correction three (1, 4, and 8) were more valid predictors, at statistically significant levels, of criterion group membership when they were uncorrected. The net effect of applying the K-correction was a
decrease in accuracy of classification for this particular subject group.

The findings from these studies raise doubt as to the appropriateness of using the standard K-correction weights with all populations. Although more research is clearly needed, MMPI users need to be aware of possible differences existing across subject groups both in terms of the meaning of K scores and the appropriateness of existing K-correction procedures.

Another interesting line of research with the K scale concerns the Normal K+ profile (Marks & Seeman, 1963; Marks, Seeman & Haller, 1974). In these records, the only significant clinical or validity scale elevation is on the K scale. However, in the Marks et al. (1963) sample, these patients had diagnoses of psychosis (48 percent), brain syndrome (24 percent), psychoneurosis (14 percent) and personality disorder (14 percent). These patients are described as shy, anxious, inhibited, and defensive about admitting that their problems might be psychological in nature. They tend to fear emotional involvement and thus avoid close interpersonal relationships. At the same time, they are easily suggestible and submissive, and are readily dominated by others. There seems to be a schizoid element in these patients; they are seen as spending a good deal of time in fantasy and daydreaming. Often their stream of thought is incoherent and they frequently appear perplexed (Marks et al., 1974).

Research on the K+ profile has been scanty and inconclusive. Newmark, Gentry, Simpson and Jones (1978) found that out of 350
patients diagnosed as schizophrenics through the use of a standardized interview and other criteria, only 17 obtained the normal K+ profile. Gynther and Brilliant (1968), on the other hand, failed to replicate Marks et al. (1963) original findings. They found that out of 1,155 profiles obtained at a mental health center, 42 (3.6 percent) obtained a K+ profile. However, they were unable to obtain significant differences between the K+ and the non-K+ profiles on any behavioral or psychological criteria. These authors interpret their failure to replicate as reflecting possible differences between their sample and that of Marks et al. (1963). Specifically, they found that their K+ sample differed from Marks et al. in terms of frequency of various diagnostic categories, marital status, education, age, race and intelligence. The Newmark et al. (1978) sample, in turn, seems more similar to the Marks, et al. sample on many of these variables; this similarity might explain their positive findings. However, the findings at this point do not yield a consistent pattern, and further attempts at replication with different populations seem necessary before definite conclusions can be made.

Other issues relevant to a discussion of the MMPI K scale merit brief mention. One of these is that, as Greene (1980) points out, subjects may achieve a high score on a K-corrected clinical scale in two different ways. They may either endorse a large number of items in the deviant direction, or they may have a large K-correction added to the scale. Depending on which is the case, elevations on these scales are likely to have different behavioral and psychological
correlates. Another important consideration is that the socioeconomic and educational background of the subject affects the interpretation of K scores (McKinley, Hathaway & Meehl, 1948; Dahlstrom et al. 1972). In addition, as indicated by studies cited above, the client population and the setting in which testing takes place are factors to consider in understanding the implications of K scale scores.

As Dahlstrom et al. (1972) point out, K scale elevations are generally associated with lower profile elevations, while the opposite is true for lower K values. However, it is important to note that the K score itself enters into the determination of the T score value on five of the clinical scales (scales 1, 4, 7, 8 and 9), and thus, it influences the results obtained when examining these relationships.

The test-retest reliabilities for the K scale range from .46 to .89 for a one- to two-day interval, from .71 to .96 for a one- to two-week interval, from .64 to .67 for an eight-month interval, from .42 to .72 for a one-year interval, and from .52 to .56 for a three-year interval (Dahlstrom et al. 1975).

In contrast to the other validity scales, for the K scale, there is no specific score that indicates that a given profile is invalid, i.e., there are no specific cutting scores for K. Dahlstrom et al. (1972, p. 163) have provided the following general interpretive levels to be used as guidelines for making inferences concerning K score values:
<table>
<thead>
<tr>
<th>Raw-Score</th>
<th>T-Score</th>
<th>Interpretive Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>Range</td>
<td></td>
</tr>
<tr>
<td>0-4</td>
<td>27-35</td>
<td>Markedly Low</td>
</tr>
<tr>
<td>5-9</td>
<td>36-44</td>
<td>Low Range</td>
</tr>
<tr>
<td>10-15</td>
<td>45-55</td>
<td>Middle Range</td>
</tr>
<tr>
<td>16-20</td>
<td>57-64</td>
<td>High Average</td>
</tr>
<tr>
<td>21-25</td>
<td>65-74</td>
<td>Moderately Elevated</td>
</tr>
<tr>
<td>26-30</td>
<td>75-83</td>
<td>Markedly Elevated</td>
</tr>
</tbody>
</table>

The interpretation given to any K scale value should consider not only the level of a given score, but as stated earlier, various subject and setting characteristics.
VALIDITY SCALE CONFIGURATIONS

In evaluating the validity of a given MMPI protocol, it is often more clinically productive to examine the relationship among the major validity scales (?, L, F and K) rather than simply noting single scale scores. For the majority of subjects, there are four validity scale configurations which occur frequently enough to warrant discussion (Greene, 1980). The most common of the four is called the caret-shape configuration, and is diagrammed in Figure 1.

As shown on the figure, this validity scale configuration is characterized by L and K scale T-scores below 50 and an F scale T-score above 60. Greene (1980) describes clients with this configuration as willing to admit to personal and emotional difficulties. They seem to be requesting assistance with these problems, and may be unsure of their capabilities for dealing with these problems. Greene also indicates that with increasing F scale elevations, the subject may either be experiencing greater difficulties, may be exaggerating symptoms in order to get help, or may be simulating psychopathology. Lachar (1974) indicates that this pattern is frequently obtained by individuals labeled character disorder and psychotic, but is rare among neurotics. He notes that it suggests open admission of problems, emotional instability, a poor self-concept and dysphoria.
Figure 1. Validity scale configuration: caret
In an effort to investigate some of the behavioral correlates of the validity scale configuration depicted in Figure 1, Post and Gasparikova-Krasnec (1979) examined the incidence of "acting out" behavior in patients with this configuration vis-a-vis patients with other validity scale patterns (see below). They found that those patients were perceived as acting out more frequently, in that they accounted for 77 percent of the incidents of inappropriate, destructive behavior, and 83 percent of confinement to a seclusion area in their sample. Similarly, Gross (1959) found that severely behaviorally disturbed subjects obtained this validity scale pattern. These findings suggest that the configuration of the validity scales may be a useful indicator of general behavioral disorganization.

Greene (1980) has presented normative data on the frequency of validity scale configuration in the four samples described earlier (see page 13). His findings indicate that the caret-shaped pattern was obtained by 67 percent of the clinic client sample (N = 415), 31 percent of the medical patient sample (N = 241), 48 percent of the prison inmates sample (N = 200) and 51 percent of the university student sample (N = 209).

The next of the four most frequently obtained validity scale configurations is called the "inverted caret" pattern and is shown in Figure 2.

As can be observed, this pattern is characterized by L and K scale T-scores of 60 or above, and F scale T-scores near or below 50. Greene (1980) describes subjects who obtain the "inverted caret"
Figure 2. Validity scale configuration: inverted caret
pattern as attempting to avoid or deny unacceptable feelings, impulses and problems. They are presenting themselves in the best possible light, and tend to view the world in simplistic terms. Lachar (1974) notes that this pattern is frequently obtained among normal defensive subjects or among those labelled hysteric or hypochondriac. He adds that deliberate defensiveness and falsification may be indicated. Similarly, Graham (1977) suggests this configuration may reflect a tendency toward "faking good."

Research investigating the behavioral correlates of this validity scale pattern generally support the descriptors presented above. For example, Hiner, Ogren and Baxter (1969) found that college students instructed to take the MMPI under "ideal-self" instructions produced the inverted-caret pattern. Similar results were obtained by Gloye and Zimmerman (1967), Grayson and Olinger (1957), and other studies examining the effect of varied instructional sets on MMPI performance, to be discussed in a later section of this review. Gross (1959) obtained data indicating that subjects who produced this configuration were rated less behaviorally disturbed in comparison with those producing a caret-shaped pattern.

Greene (1980) has provided normative data on the frequency of this validity scale pattern among his four samples. The data indicate that 10 percent of the clinic clients (N = 415), 35 percent of the medical patients (N = 241), 21 percent of the prison inmates (N = 200) and seven percent of the university students produced this pattern.
The third commonly encountered validity scale configuration is shown on Figure 3.

In this configuration, the validity scales have a positive slope in which the score on the L scale is less than that on the F scale, which in turn is less than the score on the K scale. Greene (1980) describes subjects obtaining this pattern as normal individuals who have the appropriate resources for dealing with difficulties and who are not under any major stress. He notes that a job applicant or a prison inmate trying to appear in a favorable light might obtain this configuration. Lachar (1974) views this pattern as reflecting sophisticated defensiveness or "conforming responses" among subjects from higher socioeconomic or educational level. This configuration is often accompanied by lowered clinical scale scores and a possible scale 5 elevation for males.

The least frequently obtained validity scale configuration is shown on Figure 4.

In this pattern, the three validity scales have a negative slope, with the L scale score being larger than F score, and the F score being larger than the K score. Greene (1980) described subjects obtaining this pattern as naive and unsophisticated but as attempting to appear in a favorable light. They typically have little education and come from lower socioeconomic classes. Greene also notes that these subjects are unlikely to admit their problems, and when they do, they lack the interest or motivation to change. This pattern is frequently observed in conjunction with elevations on
Figure 3. Validity scale configuration: ascending slope
Figure 4. Validity scale configuration: descending slope
scales 1, 2 and 3 (the neurotic triad) and a low scale 5 for males (Lachar, 1974).

Figures 5, 6 and 7 provide additional examples of the validity scale configurations obtained by individuals approaching the testing situation in a deviant manner. Figure 5 shows the theoretically expected configuration of a random response set. Figure 6 shows the pattern obtained by an "all-true" response set, and Figure 7 illustrates the configuration obtained when answering all items "false."
Figure 5. Random response set
(Source: From The MMPI: Clinical Assessments Interpretation, by David Lachar, Los Angeles, California: Western Psychological Services, 1974.)
Figure 6. "All true" response set
(Source: From The MMPI: Clinical Assessment and Interpretation, by David Lachar, Los Angeles, California: Western Psychological Services, 1974.)
Figure 7. "All false" response set
(Source: From The MMPI: Clinical Assessment and Interpretation, by David Lachar, Los Angeles, California: Western Psychological Services, 1974.)
SPECIAL RESPONSE MEASURES

In addition to the four standard MMPI validity scales, researchers have developed a number of techniques and measures designed to detect certain deviant response sets. One line of work has dealt with what is known as an acquiescent response set, that is, the tendency of some subjects to give one or the other of the responses available to her/him with little regard for the content of the items (Dahlstrom et al., 1972). Investigators have also dealt with the problem of social desirability responding, which consists of an attempt on the part of the subjects to respond to items so as to create a socially favorable impression rather than in a frank and honest manner (Edwards, 1957). Still another line of research has been concerned with the effect of explicit instructional sets on MMPI test performance (e.g., Grow, McVaugh & Eno, 1980). The discussion to follow will be concerned with the development of procedures designed to detect and evaluate the influence of these deviant response styles.

Acquiescence Response Set Measures

Originally, the term acquiescence referred to a tendency to give responses such as "agree," "yes," "like" and "true," that is a tendency to agree more than disagree (Jackson, 1967). However, the term is now viewed as referring to a tendency for some subjects to
consistently give one or the other of response alternatives available with little regard for the content of the item (Dahlstrom et al., 1972). This response set may be of particular significance for questionnaires such as the MMPI, which have answers to items in a true-false format (c.f. Jackson & Messick, 1958).

In an effort to investigate the problem of response acquiescence, Weiss and Moos (1965) assessed the serial dependencies in MMPI responses and in responses to an instructional set to be random in a non-content guessing task. They found that for MMPI responses, the effect of item content clearly outweighed any tendency to develop sequential dependencies, whereas for the non-content guessing tasks, it had a significant impact. These results suggest that the effect of response acquiescence may not be as significant as would be expected given the format of the MMPI, and that any tendency to develop a response bias is countered by the influence of item content.

Other investigators (e.g., Wiggins, 1962; Block, 1965) have generally been unable to demonstrate that acquiescence is a significant factor in test distortion. Nonetheless, some subjects may attempt to sabotage the test results deliberately by giving an "all true" or "all false" response pattern. As noted above, these response sets yield clearly recognizable clinical and validity scale patterns (see Figures 5 and 6). In addition, as Jackson (1967) points out, items are not uniform in the extent to which they elicit acquiescence. It is possible that some ambiguous, unclear or inapplicable items may have a greater likelihood of eliciting a biased response set.
Thus, although response acquiescence may not be a major source of error in general, it may be that certain types of items have a higher potential for eliciting acquiescence. Researchers who have developed measures of response acquiescence have generally used items rated high on "controversiality," that is, items which have about a 50 percent endorsement frequency by normal subjects in a given direction (e.g., "true"). This would, in principle, maximize the acquiescence eliciting potential of a scale consisting of such items.

Several scales exist which represent attempts to measure response acquiescence, including the B scale (Fricke, 1957), the Bn scale (Hanley, 1957), the Rb scale (Wiggins, 1962), the ATS scale (Shaffer, 1963), the AT scale (Hanley, 1961) and the Acq scale (Fulkerson, 1958). A detailed discussion of these scales is beyond the scope of this paper; for further treatment the reader is referred to other sources such as Dahlstrom et al. (1972).

Before proceeding to a discussion of other topics, it seems important to mention another area of investigation in the response bias literature. This concerns the issue of the tendency of some subjects to consistently deviate from established norms, which in a sense represents the opposite of response acquiescence. This phenomenon has been discussed at length by Berg (1955, 1957), who observed that when subjects respond to stimuli, the distribution of their responses often does not follow a normal distribution pattern. For instance, he notes that on the first toss of a coin, 80 percent of subjects pick "heads" as opposed to "tails." Likewise, when
asked to choose a number from the list 1, 2, 3 and 4, about 60 percent will choose "3," and from a list of the letters A, B, C and D, about 60 percent will choose "B." There are a number of subjects, however, who consistently deviate from established norms. Berg (1957) claims that these deviant response patterns tend to be general, that is, some subjects tend to deviate regardless of the stimulus pattern provided. Thus, whether a subject deviates from a norm on coin toss selection or on a personality questionnaire item is unimportant. According to Berg, what is significant is the fact that the subject deviates, and it is likely that deviation from the norm in one stimulus pattern is likely to be accompanied with deviations in other stimulus patterns. Accordingly, the value of a personality assessment procedure lays in its ability to show the extent to which a subject deviates from an established pattern regardless of content or other stimulus properties.

A number of problems with Berg's (1957) "Deviation Hypothesis" in relation to the MMPI have become apparent. One of these is that a number of items endorsed by a majority of the original Minnesota normal subjects were included in the clinical scales. That is, items were identified for membership on scales not only on the basis of a high frequency of endorsement by clinical groups of the minority response but also on the basis of excessive conformity to the majority response (Dahlstrom, 1969). The occurrence of these items on MMPI scales goes against the expectations of the deviation hypothesis. Another problem concerns the issue of the generalizability of
of deviations. Some subjects, according to the hypothesis, would be expected to generate a very large number of deviant responses throughout the MMPI. However, as Dahlstrom et al. (1972) indicate, such a response pattern yields a dramatically unusual clinical and validity scale pattern. The fact that subjects may show deviation from the norm in some scales and not others does not seem to fit with the expectations of the deviation hypothesis.

Other problems with the original formulation of the deviation hypothesis have been discussed elsewhere in the literature (c.f. Dahlstrom, 1969). In addition, it is noted that this hypothesis has undergone major revisions; these are described in detail in an article by Berg (1967). The set of assumptions and postulates of the hypothesis described in this article seem less extreme and more in accord with research findings.

**Measures of Social Desirability**

As mentioned earlier, several researchers have investigated the issue of social desirability responding, i.e., the tendency of subjects to respond to items so as to create a socially favorable impression. The result of these investigations has been the development of a number of scales designed to detect such a tendency. One of the better known of these scales was developed by Edwards (1953, 1957). He derived his scale by instructing ten judges to respond to 149 items from the L, F, K and Taylor Manifest Anxiety scales in such a way as to give the most socially desirable picture of themselves. Unanimous agreement among the ten judges was obtained for
79 items which were reduced to 39 items by item analysis. This 39 item scale formed the basis for an extensive series of studies attempting to evaluate both the usefulness of the scale and the existence of social desirability responding. In their review of the literature in this area, Dahlstrom et al. (1972) conclude that there are insufficient data which justify the use of this scale as a measure of response slanting. They add: "The available data indicate that this scale has more relevance for personological interpretation than for estimation of profile validity" (p. 149).

Other attempts at developing social desirability scales for the MMPI have been made. Jackson and Messick (1961) developed five scales by placing each MMPI item into one of five categories based on its average desirability rating. The resulting scales, Dy 1 through Dy 5, were constructed so that Dy 1 contains 50 items rated extremely desirable and Dy 2, Dy 3, Dy 4 and Dy 5 each contain 60 items rated moderately desirable, neutral, moderately undesirable, and extremely undesirable, respectively. Dies (1968) examined the effectiveness of these scales in detecting responses of college students given instructions to assume a social desirability set. The results showed significant increases in the endorsement of socially desirable items and the rejection of socially undesirable items under social favorability instructions. Further, the relative efficacy of the scales in detecting subjects in the social desirability instructions group ranged from 92 percent for Dy 1 to 75 percent for Dy 5. The Dy 3 scale, which was neutral in rated social desirability,
failed to detect the dissimulated records. The author interpreted this last finding as consistent with the notion that items ranking neutral in social desirability are more subtle and thus less susceptible to manipulation.

In contrast to the two scales described above, which were constructed on a rational basis, Wiggins (1959) developed a social desirability measure in an empirical manner. This scale consists of 40 items which discriminated between 178 students instructed to answer in a socially desirable direction and a comparable control group of 140 students. He compared this scale to ten other dissimulation measures in terms of their effectiveness in detecting college students instructed to answer the MMPI in a socially favorable direction from students taking the test under standard instructions (N = 250). He found that his scale performed better than any of the other measures, correctly identifying 75 percent of the simulated records and 98 percent of the authentic ones. He also found that some of the rationally derived scales, including Edward's (1957) social desirability scales, performed poorly in terms of correctly classifying subjects. Specifically, the Edwards scale identified 46 percent of the simulated and 85 percent of the authentic records accurately. This provides additional data indicating that the Edwards scale may not be an effective measure for estimating profile validity, and that other procedures might perform this task with a greater degree of accuracy. In the development of future social desirability scales for the MMPI, researchers could follow the
technique and methodology employed by Marlowe and Crowne (1961) in their development of a non-MMPI social desirability measure.

The F - K Index

Another technique designed to assess profile invalidity is the F - K index, which employs the raw score difference between the standard F and K validity scales. Gough (1947), in a study on MMPI simulation, instructed eleven clinical workers to feign two psychiatric syndromes, an acute, severe anxiety neurosis and a paranoid schizophrenic psychosis. He found that a combination of the F raw score minus the K raw score was effective in detecting ten of the eleven simulated records in both conditions. He determined that an F - K cutting score of plus four and over was adequate for identifying the dissembled neurotic profiles, whereas a score of plus 16 and over was appropriate for detecting the psychotic records. In a subsequent article, Gough (1950) suggested that in general, an F - K cutting score of plus nine or above is optimal for identifying malingered or "faking bad" profiles. For "faking good" profiles, i.e., records in which the subject was trying to deny psychopathology, an F - K of zero or less was considered appropriate. Gough (1950) recognized that the F - K index, although quite effective in detecting "fake bad" profiles, was less sensitive in detecting positive dissimulation.

A number of studies have produced findings suggesting that although the F - K index is effective in identifying "fake bad" profiles, Gough's cut-off of nine or above is too low, resulting in the failure
to detect a large number of malingered profiles (Anthony, 1971). Likewise, the relative ineffectiveness of the F-K index in identifying "fake good" profiles has been corroborated by more recent studies (Johnson, Klinger & Williams, 1977). A more extensive discussion of several research studies concerning the F-K index will be postponed until a later section of this review dealing with comparisons of different techniques for detecting profile invalidity.

Gough Dissimulation Scale

Another MMPI scale designed to measure test-taking attitudes is the Gough Dissimulation (Ds) scale (Gough, 1954, 1957). In deriving it, Gough compared the performance of patients diagnosed as psychoneurotic with that of 50 college students and 11 professional clinical workers requested to assume the role of someone experiencing a psychoneurotic reaction. He obtained 74 items which effectively discriminated between the actual and the feigned records. These 74 items were subsequently reduced to 40, resulting in the current version of the scale (Gough, 1957). An interesting finding in this study was that the items on the Ds scale do not necessarily pertain to neuroticism and maladjustment, but rather to prevailing stereotypes about neuroticism. That is, the professional clinical workers and students scored significantly higher on the Ds scale than neurotic patients, suggesting that they were endorsing items which appear to relate to neuroticism when in fact they may not. In addition, the finding that the professional workers were only slightly better than students at feigning neuroticism suggests that similar
(and apparently erroneous) stereotypes are held even by trained individuals.

In a related study, Mehlman and Rand (1960) found that subjects were unable to identify the scales to which a group of MMPI items belonged regardless of education and training in clinical work. They asked 20 clinical psychologists, 20 graduate students in clinical psychology, and 14 undergraduate students enrolled in an introductory psychology course to identify the clinical scale from which 45 MMPI items were obtained. They found that none of the groups had better than chance success on their identifications; there were no differences among groups in their ability to identify the scale on which the items appeared. These results suggest that both clinically trained and untrained persons are likely to have difficulty producing a profile pattern which resembles closely that of a bona fide patient.

A discussion of other research findings concerning the Ds scale will be deferred until a later section of this review dealing with comparisons of the effectiveness of various techniques for assessing profile validity.

The Test-Retest Index

The test-retest (TR) index provides another measure to detect deviant test taking attitudes. Originally introduced by Buechley and Ball (1952), this index makes use of the 16 items in the current versions of the MMPI which are repeated in identical form. The number of conflicting responses to these repeated items constitutes the
total score for the index. The TR scale is useful in that it gives an estimate of the subject's ability to answer items consistently and that it is sensitive to random responding.

Buechley and Ball (1952) suggested that a TR score of 3 or more was indicative of profile invalidity. They found that the TR score provided an additional means for discriminating between adolescents who obtained high F scores due to random responses from subjects whose responses may be validly and consistently deviant. Dahlstrom, et al. (1972), in their review of the literature, suggested that a TR score of four or greater was sufficient to seriously question the appropriateness of a profile.

In a more recent investigation of the TR index, Coche and Steer (1974) compared the response consistencies of presumably normal, neurotic and psychotic women. Their normal sample consisted of 110 applicants to a school of nursing. Their patient sample consisted of 42 women from the psychiatric ward of a Veterans Administration hospital and 100 women from a small, private psychiatric hospital. Among the private hospital patients, there were 55 subjects with a diagnosis of neurosis and 45 with a diagnosis of psychosis. Their results indicate that both patients samples obtained significantly higher TR scores (the VA patient sample mean was 2.5, and the private patient sample mean was 2.7) than the nursing school applicants ($\overline{X} = 0.7$). The authors suggested that a score of five or more on the TR index could indicate a careless approach in taking the MMPI.
Similarly, Gravitz and Gerton (1976) provided normative data on response consistency to the TR items of 2,000 presumably normal subjects, one-half from each sex, who took the MMPI during pre-employment screening for a variety of vocational positions. They found that among the male group, three of the item pairs showed significant differences in endorsement frequencies. Among females, four item pairs showed significant differences. Jones, Neuringer and Patterson (1976) have provided response consistency data on the TR index for brain damaged and nonbrain-damaged schizophrenic, alcoholic, neurotic and presumably normal subjects (N = 22 for each of the eight groups). They found that the TR score for the group ranged from 3.59 for the brain-damaged schizophrenic sample to 1.99 for the normal sample. In addition, they found that regardless of psychiatric diagnosis, brain-damaged subjects were more response inconsistent than nonbrain-damaged subjects. For reasons that are not clear, these investigators used only 14 of the 16 TR items, and thus their results have limited usefulness.

Greene (1979) has also provided normative data on the TR index based on MMPI data from 200 subjects. Those subjects were equally divided among four different populations: Veterans Administration psychiatric inpatients, university psychology clinic clients, adolescents seen at a juvenile probation office, and students enrolled in an introductory psychology class. He found that the highest number of inconsistent responses was given by the juvenile probation sample, with a mean score of 4.14 on the TR index. The total TR
scores for the other three samples were fairly consistent, and clustered around a mean of 2. The author views the findings for the adolescent sample as probably due to general uncooperativeness and lack of motivation. He also found positive and significant correlations between TR scores and F scale elevations. He suggested that the TR scale may help in discriminating between profiles in which F scale elevations may genuinely reflect the subjects distress from those in which the high F score is a result of random or careless responding.

The Carelessness Scale

The Carelessness scale consists of 12 pairs of MMPI items that were judged to be psychologically opposite in content. Developed by Greene (1978), it was conceptually based on earlier work by Haertzen and Hill (1963), in which they constructed a carelessness scale for the Addiction Research Center Inventory. Haertzen and Hill found that items which were psychological opposites were more sensitive than items that were simply repeated in detecting the inability or unwillingness of subjects to complete the test appropriately.

Greene (1978) developed the Carelessness scale by using MMPI data from three samples of 50 subjects each, including a group of patients from a Veterans Administration (VA) hospital, clients at a university psychology clinic, and college students enrolled in introductory psychology courses. Using a computer program, the author selected all possible non-redundant pairs of items that were answered in a consistent direction more than 90 percent of the time for each
group. This process yielded 271 pairs of items in the VA sample, 643 in the psychology clinic sample, and 140 pairs in the college student sample. Following this, 12 pairs of items were selected which represented psychological opposites by using three judges, resulting in the 12 item Carelessness scale. The author notes that for some items, the deviant response involves the same responses to both items in a pair, whereas for other items, the deviant response is counted if the answers to the item pair is different. Greene suggests that an optimal cutting score for this scale is four or more deviant responses.

Other than Greene's (1978) original derivational work, to date there has been no further empirical work conducted on the Carelessness scale. One problem with Greene's investigation is that he did not use an external (non-MMPI) criterion of profile invalidity. Future research could examine the sensitivity of the Carelessness scale in detecting subjects instructed to respond to the MMPI in a deviant (e.g., random) manner. In addition, more work needs to be done in establishing norms and appropriate cut-off points for diverse subject groups.

Subtle-Obvious Scales

Another approach taken to identify deviant response sets has been the examination of responses to subtle and obvious items on the MMPI. The earliest published description of an attempt to develop subtle and obvious scales was undertaken by Wiener and Harmon (Wiener, 1948). These investigators, using a rational approach, divided all of the items on the MMPI into two groups, those to which significant
responses were relatively easy to detect as indicating emotional disturbance, and those to which they were relatively difficult to detect. This judging process yielded 146 obvious items and 110 subtle items. The authors then examined the frequencies of responses to subtle and obvious items among 139 normal males. They found that subtle items were answered in a significant direction approximately twice as frequently as obvious items. Further, their results indicated that only eight of the items identified as obvious were scored in a reverse direction from the judge's expectations, whereas 65 of the 110 subtle items were found to be scored in the direction opposite to the judge's expectations. These findings provided evidence concerning the subtle and obvious nature of these items.

Initially, Wiener and Harmon attempted to develop subtle and obvious scales for all scales of the MMPI, but they were successful in doing so only for scales 2, 3, 4, 6, and 9. This was apparently a function of the obvious nature of the other scales.

In an investigation designed to examine the sensitivity of the subtle and obvious items in detecting deviant test taking attitudes, Harvey and Sipprelle (1976) administered the MMPI to 40 college students under two different instructional sets. Twenty subjects were asked to take the MMPI while imagining they were applying for a job, and the other 20 were asked to take it while imagining they were applying for psychotherapy. Their results indicate that for the job groups, the subtle score was significantly higher than the obvious score, whereas for the therapy group significant differences were
obtained in the opposite direction. In addition, the subtle scores were significantly higher for the job group in comparison to the therapy group, and the obvious score was significantly higher for the therapy group when compared to the job group. These and other findings to be discussed later (e.g., Anthony, 1971; Grow et al., 1980) provide evidence for the usefulness of the Wiener-Harmon subtle-obvious scales in detecting deviant response sets.

Other investigators have used a somewhat different approach in their work with subtle and obvious items. This line of research has focused on the use of MMPI X and Zero items in the detection of deviant response sets (Vesprani & Seeman, 1974; Wales & Seeman, 1972, 1969, 1968). The Zero items are those which are scored in the direction of pathology even though a majority of the normative group endorsed the item in the deviant direction, because they were endorsed in the deviant direction by even a larger proportion of the criterion group. For example, 53 percent of a normal sample might endorse an item like "I sometimes tease animals" as true, and thus the "normal" response to this item would be true. However, if this item were endorsed as "true" by 93 percent of a depressed sample, then the item yields a greater probability of depression than of normality (Wales & Seeman, 1972). In contrast, X items are endorsed by a minority of subjects in the normative samples. There are 84 Zero items on the MMPI and the majority of them are subtle in content. In turn, there are 315 X items, and most of them are obvious in content. Thus Zero and X items are roughly analogous to Wiener's subtle and
obvious items (Wales & Seeman, 1972).

Vespram and Seeman (1974) hypothesized that Zero and X items would be influenced in opposite directions under varied instructional sets. They had 28 psychiatric outpatients take the MMPI under standard and "ideal self" instructions. The subjects in the "ideal self" group were instructed to respond "not as you actually feel now but as you would like to feel" (p. 62). Their results indicate that under "ideal self" instructions, subjects were able to manipulate X items to produce a less pathological profile, i.e., the X scores declined. In contrast, the Zero items tended to increase under ideal self instructions, and thus they moved in the pathological direction. Results consistent with these have been obtained in other studies (e.g., Wales & Seeman, 1969, 1968), which suggests that X and Zero items are sensitive to attempts on the part of subjects to portray themselves in an unduly favorable light. In addition, Wales and Seeman (1968) presented data indicating that under instructions to "fake bad," subjects increase their X item score, whereas their Zero item score decreases, suggesting these items may be helpful in identifying malingering records.

In an empirical approach to defining subtle and obvious items, Christian, Burkhart and Gynther (1978) used college students to generate obvious and subtle ratings for the MMPI. They utilized a five-point scale ranging from very obvious (5) to very subtle (1), with a score of 3 indicating a neutral rating. Using these item subtlety ratings, Burkhart, Gynther and Christian (1978) examined the
endorsement patterns of subjects asked to respond to the MMPI under standard, "fake good" and "fake bad" instructional sets. They found that, in general, under instructions to "fake bad," subjects endorsed more very obvious, obvious and neutral items and fewer very subtle items. Under instructions to fake good, subjects endorsed more very subtle and subtle items and fewer neutral, obvious and very obvious items. These findings, as well as those from the other studies reviewed above, suggest that comparison of the subtle and obvious subscales can provide an effective method of identifying deviant test-taking attitudes. However, more research is needed for the development of adequate norms for use with different subject populations. In addition, studies comparing the efficacy of the various subtle-obvious subscales available would be useful and informative.

The M Scale

In an early investigation of deception on the MMPI, Cofer, Chance and Judson (1949) examined response patterns of 81 college students under instructions to feign abnormal and "normal" MMPI results. The subjects in the fake good group were asked to "imagine themselves as being desirous of entering midshipman training in the navy and, therefore, as wishing to make the best possible impression through their test scores" (p. 494). The subjects in the malingering group were told to "answer the questions as they thought an emotionally disturbed person would answer them" (p. 494), and specifically, "to answer the questions in such a way as to avoid being drafted into the army" (p. 484). A comparison control group was
was asked to take the MMPI under standard instructions.

The findings revealed that subjects in the fake good group (positive malingerers) were readily detectable using a 39-item scale derived by item analysis. Using a raw score cut-off of 20, this scale correctly identified 86 percent of the faked records and 96 percent of the honest records. However, these items were insusceptible to negative malingering, i.e., they failed to differentiate between the fake bad and the honest records. This scale has come to be known as the $M^p$ scale (apparently representing "Malingering-positive"). Research on the $M^p$ scale has been scanty. In a study to be discussed later in detail, Grown, et al. (1980) found that the $M^p$ scale was successful in differentiating between honest and fake good records. However, the proportion of variance that could be accounted for by this detection strategy was relatively small.

Comparative Studies of Faking Detection Techniques

Several investigators have attempted to evaluate the relative efficacy of the various faking detection techniques by comparing their sensitivity in identifying dissimulated records. In general, the findings obtained from these studies, although not grossly inconsistent, indicate that the effectiveness of the different approaches varies according to the direction of the simulation (i.e., faking "good" or "bad"), the specific method used by the investigators in eliciting a deviant response set, and the nature of the subject population sampled. In addition, the cut-off points for the various techniques were also dependent on these factors. The details of these
comparative studies will be reviewed below.

Using 50 college students, Exner, McDowell, Pabst, Stockman and Kirk (1963) examined the usefulness of the three standard validity scales, Gough's Ds scale, and the F minus K index in distinguishing between malingered, fake good and honestly reported profiles. Twenty-five subjects were instructed to answer the MMPI "in such a manner as to appear normal or socially desirable as would an attractive job or school applicant" (p. 92). The other 25 subjects were asked to respond "in a manner as to appear sufficiently deviant to be exempt from some social responsibility such as military service but not so deviant that institutionalization would be required" (p. 92). All subjects were requested to take the MMPI on a second occasion, and to respond in an honest manner. Their results indicate that the group instructed to mangle was considerably more successful in manipulating the clinical scales than the group instructed to fake good. For the malingering group, there were significant differences on all of the clinical scales between the pre-test and the post-test scores, whereas for the fake good group, significant differences existed for only one of the clinical scales. With one exception, the scores on the three validity scales were significantly different from pre-test to post-test in the expected direction. For the Malingering group, there was a slight but non-significant increase in the L scale (from a T score of 44 to one of 47) which is difficult to account for. For the Malingering group, significant differences existed between the pre-test and post-test scores on the Ds scale, with 24 out of the 25
honest records scoring below a raw score of 20 and all of the malingered records scoring above 20. Significant differences existed for the malingering group pre- and post-test scores on the F-K index; a cut-off score of +12 detected 24 of the 25 malingered records. The F scale alone proved to be the most sensitive indicator of malingering; using an F raw score of 12, all of the malingered and honest records were correctly classified. The identification of fake good records was considerably less successful. There was a great deal of overlap between the pre-test and post-test scores on the F - K index, as well as on the three standard validity scales taken separately. Other procedures for detecting the fake good records were also unsuccessful. These findings suggest that whereas malingering may be more readily detectably by use of existing procedures, the identification of fake good remains a difficult and uncertain process.

In a subsequent investigation, Anthony (1971) examined the applicability of several MMPI faking detection techniques for clients with nonpsychotic disturbances. He administered the MMPI to 40 U.S. Air Force male clients with nonpsychotic diagnoses twice, once under standard instructions, and a second time under instructions to "exaggerate on the test whatever difficulties had brought them to the clinic and to appear in worse condition than they actually were" (p. 101). The exaggerated profiles were then matched with 32 similar profiles from other psychiatric clients. The results indicate that the experimental subjects were successful in exaggerating
their pathology, as reflected by significant differences between the pre- and post-test scores on all the standard clinical scales. Using an F - K score of 0, 80 percent of the 40 standard and 40 exaggerated records were correctly identified. An F raw score of 10 correctly classified 81 percent of the profiles. With a cut-off score of 21, the Ds scale correctly classified 86 percent of the records. The author also used a 146-item Zero (subtle) scale and a 110-item X (obvious) scale for assessing profile validity. Using 100 as a cut-off score, the X scale correctly classified 86 percent of the profiles, and was generally lower for this sample than that obtained with other samples in previous investigations, and this in turn affected the F - K index. In addition, the clinical profile pattern obtained by the malingerer group was not as erratic and was of less overall elevation than that observed by other investigators, suggesting they accomplished the exaggeration with a greater degree of sophistication.

Another major comparative investigation has been carried out by Grow (1980), in which he used MMPI records from 150 undergraduate psychology students to evaluate the effectiveness of 13 MMPI faking detection techniques. The subjects were randomly assigned to one of three experimental conditions. The fake bad group was told to respond to the MMPI under instructions to "imagine a situation in which it would be to your advantage to appear as if you had psychopathology of one form or another. Examples of such a situation could be applying for Rehabilitation Services, trying to qualify for disability
benefit, or trying to beat a legal charge on the grounds of insanity" (p. 911). The fake good group was given instructions to "imagine a situation in which it would be to your advantage to appear as if you were completely normal and sane. Examples of such a situation could be: trying to secure an early release from prison, trying to secure a release from a mental hospital, or applying for a good job" (p. 912). The third group was instructed to take the MMPI anonymously and in an honest fashion.

Consistent with other studies, the authors found that the various strategies were generally successful in producing significant differences across groups. Their results indicate that malingering could best be identified by use of either $F \geq 15$, or $F - K \geq 7$. The next most effective procedures for detection of faking bad were the Ds scale, Wales and Seeman's (1968) X scale, Wales and Seeman's (1968) X minus Zero difference score, and Wales and Seeman's (1968) Zero scale, in that order. Also consistent with findings from other studies, detection of faking good was less successful than identification of malingerers. The most effective technique for detection of malingering, i.e., the F raw score, accounted for 80 percent of the variance associated with students faking bad. In contrast, the most effective technique for identifying faking good was the $F - K$ index (using $< - 11$ was a cut-off), which accounted for 36 percent of the variance associated with faking good. The next most effective techniques for identifying fake good records were in order, the addition of the L and K scale scores, the M scale, X minus Zero difference
score, the X scale, and the Zero scale.

The authors cross validated these results with MMPI records from files of a state mental hospital and an outpatient psychological clinic. They obtained 16 records in which the MMPI and other file data suggested an attempt on the part of the subject to mangle. Further, they found 24 records in which the MMPI suggested less psychopathology than other file data. The results from the cross validation indicate that although almost all of the faking detection techniques were effective with the clinical sample, the mean scores for the different techniques, as well as the proportion of variance that could be attributed to faking, were somewhat different from the student sample. Nonetheless, the best techniques to detect faking good or bad were the same for both samples, namely the use of $F - K > 7$ or $F > 15$ for identifying malingered profiles, and the use of $F - K < -11$ for faking good.

As stated earlier in the introduction of this review, the MMPI is used extensively in correctional settings. Further, the various techniques for assessing profile validity are routinely employed in those settings during evaluation procedures. Given these patterns of clinical practice, it is surprising that relatively little research has been conducted on the adequacy of existing norms and cutoff points for the various validity indicators for use with prisoner populations. Following a thorough search in the literature, only three publications of research on validity indicators with prisoner samples were found. These will be reviewed below.
In an early study concerned with deception on the MMPI, Hunt (1948) compared the performance of a sample of psychology students and a group of U.S. Navy court-martialed prisoners under instructions to falsify their responses. Fifty-three students responded to the MMPI under instructions to "conceal their personality abnormalities as much as possible so that they would be certain not to be excluded from induction to the military services on psychiatric grounds" (p. 396). These subjects made up a fake good group. The malingering group consisted of 56 subjects instructed to feign "sufficiently severe personality abnormality to insure a neuropsychiatric discharge or psychiatric disqualification for military service" (p. 396). There were also "honest" profiles on all of the students obtained for comparison purposes. The prisoner group consisted of 74 subjects who took the MMPI on three occasions, once under each condition. The results indicate that the vast majority of subjects in both the student and the prisoner samples were successful in malingering psychopathology on the clinical scales. In the fake good condition, the majority (84 percent) of students were successful in manipulating the clinical scales whereas only a minority of the prisoners (36 percent) were able to do so. In addition, the prisoners showed greater variability than the students in the distortions produced under both
faking conditions. Using an F-K score of +11 or more, 88 percent of the student and 85 percent of the prisoner malingering groups were correctly indentified. As other investigators since then have observed, the identification of fake good profiles was much less successful, with considerable overlap of F-K scores between honest and faked records. Since many of the techniques for detection of faking have been developed subsequent to this study, no data are provided on the other scales and indices discussed in this review.

Another study examining deceptive response patterns on the MMPI with prisoners was conducted by Lawton and Kleban (1965). They hypothesized that prisoners, who presumably possess sociopathic traits to a significant degree, because of their lack of empathy, would be unable to recognize their sociopathy and would therefore be unable to decrease their scale 4 scores even when instructed to do so. The authors used 32 prisoners who took the MMPI under standard instructions and again under instructions to respond "in terms of the way a person who had had no trouble with the law" (p. 270). Their results indicate that although the prisoners were able to lower their mean T scores on several of the clinical scales, they were unable to single out the specific scale 4 items to manipulate. Thus, the effect of the deceptive instructional set was a general lowering of the profile as a whole, and not an isolated decrease in scale 4 values. Further, under both conditions, the majority of subjects showed either scale 4 or 9 as their high or second highest point scale. These results are consistent with other findings showing that
it is generally difficult to attempt to manipulate MMPI test responses in the direction of faking good. However, since no evaluation was made of the effectiveness of any of the faking detection techniques, the findings are only partly relevant to the present review.

The only other published study evaluating the effect of varied instructional sets on MMPI performance in prisoners was done by Gendreau, Irvine and Knight (1973). They had 23 prisoners take the MMPI under three instructional sets in counterbalanced order. One group was told to feign maladjustment using the following instructions: "For example, you know that in this prison, if you are diagnosed as having a lot of problems on psychological tests you may be able to get more psychological counseling or obtain a transfer to other institutions which have more or different treatment facilities... I want you to try to give a bad impression of yourself. I want you to try to make yourself look worse than you really are. Try to show that you have a lot of 'problems.' In other words, fake this test so that the results will show there are a lot of things wrong with you" (p. 185). The instructional set for the fake good adjustment was phrased as follows: "For example, you know that in this prison if you give a very favorable impression of yourself on the psychological tests this fact may be able eventually to help you obtain a transfer to another institution that has less security and more recreational facilities. Also, within this institution you realize you may be able to get a better job, etc., if you appear as being a very 'normal' person on the tests... I want you not to tell
the truth, the way you really are. I want you to try to give an extra good impression of yourself. . . In other words, fake this test so that the results will show that there is nothing wrong with you" (p. 185). A third record was also obtained for each subject using standard instructions.

The results indicate that the prisoners were successful in feigning maladjustment on the MMPI, as reflected by significant differences between the fake bad and honest records on all of the clinical scales except scale 5. Of the faking indices, the F, F-K, Ds, and Wiener Obvious items all significantly discriminated between the honest and malingered records. Scales L, M_p, K and Wiener Subtle items were little influenced by instructions to fake bad. The successful classification rate for the F scale and the F-K index was 100 percent. For the Ds scale, the hit rate was 96 percent, and for the Wiener Obvious items it was 88 percent. In contrast to findings from other studies, the subjects in this study were also able to manipulate the clinical scales in the direction of better adjustment. Scales 4, 7, 8 and 9 were significantly lower for the fake adjustment records in comparison with the honest ones, whereas scales 1, 2, 3, 5 and 6 remained unaffected. All of the faking indices significantly differentiated adjustment from maladjustment. The overall percentage of correct classification for each technique was as follows:

\[4\] It is noted that the use of percentages in this study may be inappropriate given that the total sample size was 23.
M = 92 percent, F - K = 85 percent, L = 83 percent, K = 81 percent, Wiener Obvious items = 81 percent, Wiener Subtle items = 78 percent, F = 75 percent, and Ds = 74 percent. The findings obtained in this study on the faking good adjustment and the detection thereof is in contrast to those obtained in other investigations (e.g., Lawton & Kleban, 1965), and thus should be viewed as suggestive.

The results of the Gendreau et al. (1973) study are encouraging, but the methodology employed in the investigation contains serious flaws. One major deficiency is the small sample size used. Data were obtained for 16 dependent measures (the MMPI clinical scales and the validity indicators) on 23 subjects across three test administrations. It would seem that the number of dependent variables examined is excessively large in relation to the sample size. In addition, the authors do not provide any data on the racial composition of the sample, nor do they indicate whether or not an effort was made to control for this variable. Given the importance of race in the interpretation of MMPI scales (c.f. Gynther, Lachar & Dahlstrom, 1978), the omission of this information limits the generalizability of the results obtained. Another problem with this study concerns the MMPI profile obtained from the subjects under standard instructions. The highest scale points for the mean profile under the honest condition were, in order, 9, 8 and 4. The occurrence of the 4 - 9 and 9 - 4 code as the highest points in the profile of prisoners is not surprising; documentation exists in the literature to indicate that such a profile pattern is not uncommon in this population (c.f. Greene,
1980, and Dahlstrom et al., 1972). However, the elevation of scale 8 on the mean subject profile raises some questions, particularly in light of the purpose of this study. The possible presence of a significant degree of psychological maladjustment in at least some of the subjects used in a study on how prisoners feign maladjustment makes interpretation of the results difficult. The subject selection problems in this study are aggravated by the fact that each subject served as his own control, i.e., no external criterion groups were used. This problem may have further confounded the results by augmenting any contaminating effects due to the idiosyncratic nature of the sample.

Statement of the Problem

The findings obtained in several studies indicate the MMPI faking-detection techniques available are useful in discriminating between honestly obtained records and those generated under deviant instructional sets for a variety of "normal" and psychiatric samples. However, there is a marked paucity of data on the MMPI performance of prisoners instructed to respond to the test under varied instructional sets. The data that are available are either obsolete (Hunt, 1948), tangentially relevant (Lawton & Kleban, 1965); or are the result of experimental procedures with serious limitations (Gendreau et al., 1973).

Given the extensive utilization of the MMPI in correctional settings (Dahlstrom et al., 1975), it seems that more research concerning the manner in which prisoners respond to this instrument
under various conditions is necessary. In particular, since the MMPI is frequently used in the assessment of the credibility of subjects in criminal-judicial proceedings (Ziskin, 1981), it seems appropriate to investigate further the issue of deceptive responding using controlled experimental procedures. The purpose of this investigation is to generate new data concerning the manner in which prisoners respond to the MMPI when instructed to feign psychological maladjustment and "hyper-adjustment." It is hoped that such data will aid in establishing the effectiveness of various faking detection techniques in a correctional setting and that it will provide some basis for determining appropriate cut-off points for use with prisoner samples.

Based on the findings obtained in the studies reviewed above, the following hypotheses concerning the manner of responding of prisoners instructed to feign "hyper-adjustment" (fake good) and maladjustment (malingering) can be made:

1. The experimental subjects will be able to fake maladjustment, as evidenced by higher scores on the clinical scales than those of a control group.

2. The experimental subjects will be able to fake good, as evidenced by lower scores on the clinical scales than those of a control group.

3. The experimental subjects in both faking conditions will be identified by the L scale. The experimental subjects in the Malingering group will score lower, and those in the Fake Good group will score higher, than a control group.
4. The experimental subjects in both faking conditions will be identified by the F scale. The experimental subjects in the Malingerer group will score higher, and those in the Fake Good group will score lower than a control group.

5. The experimental subjects in both faking conditions will be identified by the K scale. The experimental subjects in the Malingerer group will score lower, and those in the Fake Good group will score higher, than a control group.

6. The experimental subjects in both faking conditions will be identified by the F - K index. The experimental subjects in the Malingerer group will score higher, and those in the Fake Good group will score lower, than a control group.

7. The experimental subjects in both faking conditions will be identified by Wiener's Obvious scale. The experimental subjects in the Malingerer group will score significantly higher, and those in the Fake Good group will score lower, than a control group.

8. The experimental subjects in both faking conditions will be identified by Wiener's Subtle scale. The experimental subjects in the Malingerer group will score lower, and those in the Fake Good group will score higher, than a control group.

9. The experimental subjects in both faking conditions will be identified by Wiener's Obvious minus Subtle score. The experimental subjects in the Malingerer group will score
higher, and those in the Fake Good group will score lower, than a control group.

10. The experimental subjects in the Malingerer group will be identified by a higher score on the Ds scale than those obtained by subjects in a control group.

11. The experimental subjects in the Fake Good group will be identified by a higher score on the M_p scale than those obtained by subjects in a control group.

The eleven hypotheses stated above are presented in symbolic form in Table 5 (p. 99).
METHOD

Subjects

Ninety-six subjects were obtained from the United States Camp and Penitentiary, a federal facility in Lompoc, California. Initially, an attempt was made to recruit all of the subjects from the Camp area, which is a minimum security section of the institution. Due to difficulties encountered in subject cooperation, only 40 Camp subjects were obtained; the remaining fifty-six were recruited from the Penitentiary, a maximum security area. Prisoners in the Camp area are generally considered less of a security risk; most of them either have relatively short sentences or are approaching the end of longer terms, in which case they might have been transferred to the Camp from a more secure setting. In contrast, the subjects obtained from the Penitentiary were quite variable in terms of the security risk they represent. These subjects were recruited from a special section of the Penitentiary, called the "Hold-over" unit. This unit houses inmates who are in a transitional stage, usually awaiting transfer to other federal institutions. As a group, they are quite heterogeneous, varying in terms of type of offense, length of sentence, prior offenses, etc.

Given that the subjects were recruited from two different locations (i.e., Camp and Penitentiary) which house different types of
inmates, the decision was made to examine whether the subjects from each location were comparable in terms of demographic and other characteristics. The reason for this was to evaluate the possibility of a bias in the results of the study due to subject selection procedures, that is, an untoward effect on the dependent variables due to pre-existing differences between subjects from the two settings.

Data were obtained for the Camp and Penitentiary subjects for each of the following demographic variables: type of offenses committed; length of sentence; security level; education; and ethnicity. For the offense categories, distributions were obtained by assigning subjects into those categories which had relatively large frequencies. Drug Related offenses included possession, possession with intent to distribute, transportation and manufacture of a variety of substances. Crimes included under the Fraud category included mail fraud, wire fraud, income tax fraud, false application for passport, obstruction of mail and similar offenses. The vast majority of crimes included under the Robbery category were bank robbery. Weapons-related offenses included crimes such as possession and distribution of firearms, possession of a destructive device, and manufacture of firearms. Under the category of Violent Crimes Against Persons were included conspiracy to commit murder and manslaughter.

In order to assess the significance of possible differences between the subjects in the two settings, the variables of age, sentence
and education were analyzed by way of analysis of variance. Specifically, for the purpose of analysis, subject setting was treated as an independent variable with two levels (i.e., Camp and Penitentiary), and the variables of age, sentence and education were treated as dependent measures of a continuous nature. The remaining variables (namely, security level, offense category, previous residence, and ethnicity), being categorical in nature, were subjected to chi square analysis with setting as the independent variable.

Significant differences were found between the Camp and Penitentiary subjects on several of the demographic variables (see Table 4, p. 97). For this reason, the demographic characteristics of the subjects from each of the settings will be discussed separately. The analyses indicated that significant differences existed between the Camp and Penitentiary subjects in terms of sentence, offense category, and security level. The subjects in the Penitentiary had longer sentences, a higher proportion of more serious offenses, and larger frequencies among the higher security levels. Since security levels are determined in part by the type of offense committed and the length of sentence, it seems reasonable that these three factors would vary together.

Significant differences also existed on the education variable, the Camp subjects having slightly more education. There was a greater

5 Security levels are numerical values representing security risk, with Level 1 representing the lowest, and Level 5 the highest threat to security.
proportion of Hispanic subjects in the Penitentiary group relative to those of the Camp. However, all subjects reported good command of the English language, with the vast majority indicating that English was their primary tongue. In addition, all subjects were caucasian. No significant differences were found for the age and geographical location of the subjects (geographical location being the subject's last residence). The average age of the entire sample was 35.90 (S.D. = 9.88). The majority of subjects had last resided in the West and Northwest regions of the United States.

All of the subjects used in this study were volunteers; they were informed that neither participation nor refusal to participate would affect their status in any way. Initially, 104 volunteers from the Camp signed up to participate in the study. Of these, only 40 arrived for the scheduled testing sessions, whereas the others failed to show up in spite of re-scheduling of additional testing sessions. Due to the substantial subject attrition, it was necessary to recruit 56 subjects from the Penitentiary. The subjects from the Camp were offered snacks (e.g., doughnuts, potato chips) and soft drinks in exchange for their participation, whereas the subjects in the Penitentiary were given extra time in a recreation room. Six subjects from the Camp and two from the Penitentiary refused to complete the experiment. Reasons given included boredom, lack of time, the length of the test, and fear of jeopardizing themselves because of the test results. All subjects were required to sign an Informed Consent Statement prior to participation in the study. (A copy of
the Statement is included in Appendix A).

Procedure

The Camp subjects were administered the standard booklet form of the MMPI in a classroom on the Camp grounds. The testing took place over a four-week period with two testing sessions per week, both in the evenings. Testing session attendance ranged from two to thirteen subjects, with all sessions being supervised by this writer. Forty completed protocols were obtained; the investigator was not able to elicit further cooperation from the remaining Camp volunteers.

The Penitentiary subjects were administered the standard booklet form of the MMPI in a recreation room in the Hold-over unit. The testing took place over a three-week period with one testing session per week conducted during the afternoon. Testing session attendance ranged from 23 to 31; again all testing sessions were supervised by this writer. Fifty-six completed protocols were obtained from 58 volunteers, the remaining two records being from two subjects who refused to complete the experiment (see above).

As stated earlier, in the beginning phases of this project, 104 Camp volunteers were obtained; of these 96 were randomly assigned to three groups of 32 subjects each. Following the non-participation of 56 of the Camp subjects, the 56 subjects from the Penitentiary were selected to replace them. (Although these subjects were also randomly assigned to the groups, the possibility of a non-random distribution of subjects across experimental groups cannot be ruled out, since they were selected from two naturally assembled groups.)
This procedure yielded three groups of 32 subjects each. One of the groups had 10 subjects from the Camp and 22 from the Penitentiary. This group was told to respond to the MMPI in such a way as to feign maladjustment, i.e., to mangle. The specific instructions delivered to the group were as follows:

In this study, I am interested in finding out how people fake mental illness on psychological tests. After you read the instructions on the booklet in front of you, I want you to answer the test not as you really feel, but how you imagine someone who was trying to fake mental illness would answer it. Specifically, I want you to imagine that you are awaiting trial for charges of committing a crime, and that the answers you give on this test will determine whether you are sent to a mental hospital or to prison. You are to also imagine that you very much prefer to be sent to a mental hospital, and so you want to appear as mentally ill as possible on this test. Does anyone have any questions?

The second group consisted of 15 subjects from the Camp and 17 from the Penitentiary. This group was told to respond to the MMPI as a person who has absolutely no problem whatsoever. The specific instructions given to the group were as follows:

In this study, I am interested in how people fake psychological tests when they want to appear like someone who has no psychological problems whatsoever. After you read the instructions on the booklet in front of you, I want you to answer not as you really feel, but how you imagine someone who was trying to appear as having absolutely no problems would answer. Specifically, I want you to imagine that you are confined to an institution such as a mental hospital, and that the answers you give on this test will determine whether you are released or not. The idea is that you want to appear as "normal" or "sane" as possible on this test so that you increase your chances of being released. Does anybody have any questions?

The third group also consisted of 15 subjects from the Camp and 17 from the Penitentiary. They were given the MMPI under
standard conditions. The specific instructions given to the group were as follows:

In this study, I am interested in how prisoners respond to psychological tests. The test results will not affect you in any way, since this is an experimental research project. Therefore, you can feel free to answer in a completely honest manner without fearing for the test results. Go ahead and read the instructions, and ask me if you have any questions.

One problem encountered during the data collection process was that some subjects reported doubts concerning the confidentiality of the results obtained from the study. An attempt to deal with this problem consisted of including reassuring statements in the Informed Consent form (see Appendix A) which the subjects were requested to read and sign prior to participating in the study. Also, during the subject recruitment phase at the Camp, the investigator was accompanied by an inmate who would attempt to reassure potential subjects concerning the confidentiality of the results. (In the Penitentiary, this was not possible, since due to security considerations, all subjects were recruited by a staff psychologist). Nonetheless, some subjects still expressed reservations about the issue of confidentiality. Similarly, some subjects questioned the validity of the study, pointing out that the results could be easily sabotaged by refusing to follow the instructions. Although an attempt was made to deal with this issue by including a five-point scale intended to assess the subjects' attitude toward the study (see Appendix B), this investigator found no adequate way of dealing with this problem.
All of the MMPI records obtained were scored for the ten clinical scales, the three validity scales, the Ds scale, the M^p scale and Wiener's Obvious and Subtle scales. In addition, the F - K index and the Wiener Obvious minus Subtle difference score was computed for each record.

In order to examine how the results obtained in this study compared with the MMPI scores the subjects might obtain under more natural circumstances, an attempt was made to secure MMPI records taken by the inmates as part of their routine evaluation procedure. For the Penitentiary subjects, this investigator could not obtain previous MMPI records. As mentioned earlier, these subjects were recruited from a Hold-over unit, meaning that all of them were in transit either to other institutions or other units. Thus, there was little information available on these subjects, and no MMPI data. For the Camp subjects, 23 prior MMPI records were obtained. The remaining records were not available either because the subject refused to take the MMPI previously, because none had been administered, or because it was not in the subject's file.

All of the subjects were requested to complete a demographic data form for the purpose of obtaining information concerning age, education, offense type, etc. (see Appendix C). Also, as mentioned earlier, they were requested to answer a five-point scale concerning their attitudes about their participation in this study (see Appendix B).
In this investigation, the instructional set given to each of the three groups was the independent variable; it had 3 levels, namely, instructions to malinger, instructions to fake hyper-adjustment, and standard instructions. The dependent variables in the design were the ten clinical scales and the nine validity indicators. The effect of the instructional set was examined for each of the dependent measures. The clinical scales were evaluated in order to address the issue of whether the subjects could either malinger or feign hyper-adjustment (see statement of hypotheses above, p. 83). The validity indicators were examined to see how they were influenced by the instructional sets, i.e., to see how well they detected attempts to malinger or to feign hyper-adjustment.

In order to assess the significance of the differences between the experimental and control groups on the dependent measures, the data were analyzed using a combination of multivariate analysis of variance (MANOVA) and univariate analysis of variance (ANOVA). This procedure follows from the recommendations of Hummel and Sligo (1971), in which they encourage this approach to analyzing multivariate data. The use of MANOVA is a way of dealing with the problem of analyzing a large number of dependent variables which may not be independent of one another, a situation which occurs in the case of MMPI scales. In this study, this procedure allowed for an examination of the effect of the instructional sets on the dependent measures, while taking into
account the lack of independence between them.

The analysis was conducted through the use of computer facilities at California Polytechnic State University in San Luis Obispo, California. The SPSS standard MANOVA program described in Hull and Nie (1981) was used. This procedure computes four statistics used for significance tests: Roy's largest root, Wilks' lambda, Hotelling's trace, and Pillai's criterion. It also generates univariate ANOVAs for each of the dependent variables. Further analyses were conducted in order to make specific pairwise comparisons, i.e., between the experimental and control groups, using ANOVA. This procedure permitted an evaluation of the significance of the differences between the control group and each of the experimental groups on the dependent measures, a step that was necessary in order to address the research hypotheses stated above (p. 83). In addition, the possibility that significant differences might exist between the experimental groups on the demographic variables was also investigated, since the presence of such differences could have an effect on the results obtained. The age, education and sentence variables were analyzed using ANOVA and the remaining variables were subjected to chi square analysis.

Optimal cut-off scores were derived for several of the validity indicators. 6 This was done by finding the score for each scale which yielded the highest percentage of correct classification for the subjects in each of the experimental conditions vis-a-vis the control

---

6 This procedure was done only for those scales which yielded significantly different distributions for the experimental and control groups.
group subjects. Specifically, the score which yielded the highest percentage of correct classification of the Malingerer group subjects (true positive) and Control group subjects (true negative) was determined to be the optimal cut-off score for that scale for the detection of malingering. Similarly, for the detection of faking hyper-adjustment, the score which yielded the highest percentage of correct classification of subjects in the Fake Good group (true positive) and Control group subjects (true negative) was determined to be the optimal cut-off score for that scale.

In order to assess the effect of location of the subject (i.e., Camp versus Penitentiary) on the dependent measures, the MANOVA program included location as an independent variable. This allowed for an evaluation of the effect of location independent of the instructional set effect, and also yielded tests for significance of the location-by-instructional set interaction.

Results

Table 4 presents the demographic data for the two subject settings discussed in the Subjects section. The general trend of the results obtained in this study are summarized in Table 5. The table includes a restatement of the original research hypotheses presented in Table 3 (p. 86) and an indication as to whether the hypotheses were confirmed, partially confirmed, or disconfirmed by the data.

As indicated in Table 5, all of the hypotheses were either partially or fully confirmed. However, as will be presented later in detail, the data supported the hypotheses for the Malingerer group more consistently than for the Fake Good group.
Table 4
Demographic Data for Subjects in the Camp and in the Penitentiary

<table>
<thead>
<tr>
<th></th>
<th>Camp (N = 40)</th>
<th>Penitentiary (n = 56)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M.</strong></td>
<td><strong>S.D.</strong></td>
<td><strong>M.</strong></td>
</tr>
<tr>
<td>Age</td>
<td>38.12</td>
<td>11.65</td>
</tr>
<tr>
<td>Sentence (in months)</td>
<td>53.63**</td>
<td>56.06</td>
</tr>
<tr>
<td>Education (in years)</td>
<td>13.35*</td>
<td>2.68</td>
</tr>
<tr>
<td><strong>Frequency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Security Level&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Unknown</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Offense Category&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug Related</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Fraud</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Robbery</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Weapons Related</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Violent Crimes Against Persons</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other or Unknown</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>
### Table 4
(Continued)
Demographic Data for Subjects in the Camp and in the Penitentiary

<table>
<thead>
<tr>
<th>Previous Residence</th>
<th>Camp (N = 40)</th>
<th>Frequency</th>
<th>Penitentiary (n = 56)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>24</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northwest</td>
<td>9</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southwest</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anglo</td>
<td>37</td>
<td>39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>3</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( ^a \) chi square showed significant differences in frequency distribution \( (p < .05) \).

\( ^b \) chi square showed significant differences in frequency distribution \( (p < .01) \).

\* \( p < .05 \)

\** \( p < .01 \)
Table 5

Summary of Results: Confirmation and Disconfirmation of Research Hypotheses

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $M^{a} &gt; CN^{b}$ on clinical scales</td>
<td>$C^{d}$</td>
</tr>
<tr>
<td>2. $G^{c} &lt; CN$ on clinical scales</td>
<td>$PC^{c}$</td>
</tr>
<tr>
<td>3. $M &lt; CN &lt; G$ on L scale</td>
<td>$PC$</td>
</tr>
<tr>
<td>4. $G &lt; CN &lt; M$ on F scale</td>
<td>$PC$</td>
</tr>
<tr>
<td>5. $M &lt; CN &lt; G$ on K scale</td>
<td>$G$</td>
</tr>
<tr>
<td>6. $G &lt; CN &lt; M$ on F - K index</td>
<td>$PC$</td>
</tr>
<tr>
<td>7. $G &lt; CN &lt; M$ on Wiener's Obvious scale</td>
<td>$C$</td>
</tr>
<tr>
<td>8. $M &lt; CN &lt; G$ on Wiener's Subtle scale</td>
<td>$PC$</td>
</tr>
<tr>
<td>9. $G &lt; CN &lt; M$ on Wiener's Obvious minus Subtle scale</td>
<td>$C$</td>
</tr>
<tr>
<td>10. $M &gt; CN$ on Ds scale</td>
<td>$C$</td>
</tr>
<tr>
<td>11. $G &gt; CN$ on $M_p$ scale</td>
<td>$C$</td>
</tr>
</tbody>
</table>

$^{a}M =$ Malingerer group

$^{b}CN =$ Control group

$^{c}G =$ Fake Good group

$^{d}C =$ Confirmation of hypothesis

$PC =$ Partial confirmation of hypothesis
As mentioned previously, one of the first issues dealt with during the analysis of the data was whether the instructional sets had an effect on the MMPI scores of the different groups, given the large number of dependent variables and the likelihood of non-independence among them. Table 6 gives the results of the MANOVA tests computed to establish the effect of the instructional sets on the K-corrected MMPI scales.

The results indicate that there were highly significant differences among the experimental and control groups on the various dependent measures. The univariate ANOVAs revealed highly significant differences on all of the scales across the three groups. Table 7 gives the results of the MANOVA tests using the non-K-corrected clinical scales.

Given that the MANOVAs and the univariate ANOVAs showed significant differences, it was decided to test the specific research hypotheses.\(^7\) The first hypothesis stated that the prisoners would be able to fake maladjustment, as evidenced by higher scores on the standard clinical scales than those of the control groups. The results indicate that the subjects in the Malingerer group were quite successful at feigning maladjustment. As shown in Table 8, there were highly significant differences between the Malingerer and the Control

\(^7\)The univariate ANOVAs generated by the MANOVA program did not answer the research hypotheses, since they tested the effect of instructional set across the three groups; the hypotheses require pairwise comparisons (e.g., between the Malingerer and Control groups on scale F).
### Table 6

**Results of MANOVA Tests to Determine Effect of Instructional Set on K-corrected MMPI Scales**

<table>
<thead>
<tr>
<th>Test Name</th>
<th>Value</th>
<th>Approximate F</th>
<th>Significance of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>1.114</td>
<td>4.832</td>
<td>.000</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>6.616</td>
<td>12.362</td>
<td>.000</td>
</tr>
<tr>
<td>Wilk's</td>
<td>.102</td>
<td>8.039</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Note: Roy's statistic is not presented since the MANOVA output does not provide F approximations or probability levels for its value.*
Results of MANOVA Tests to Determine Effect of Instructional Set on Non K-corrected MMPI Scales

<table>
<thead>
<tr>
<th>Test Name</th>
<th>Value</th>
<th>Approximate F</th>
<th>Significance of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>1.097</td>
<td>4.67</td>
<td>.000</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>6.500</td>
<td>12.14</td>
<td>.000</td>
</tr>
<tr>
<td>Wilk's</td>
<td>.106</td>
<td>7.84</td>
<td>.000</td>
</tr>
</tbody>
</table>
Table 8
Effect of Instructional Set on Standard Validity
and K-corrected Clinical Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Malingerer (n = 32)</th>
<th>Control (n = 32)</th>
<th>Fake Good (n = 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M.</td>
<td>S.D.</td>
<td>M.</td>
</tr>
<tr>
<td>L</td>
<td>48.28&lt;sup&gt;a&lt;/sup&gt;</td>
<td>11.15</td>
<td>54.22</td>
</tr>
<tr>
<td>F</td>
<td>136.75&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20.86</td>
<td>63.62</td>
</tr>
<tr>
<td>K</td>
<td>42.41&lt;sup&gt;b&lt;/sup&gt;</td>
<td>7.75</td>
<td>54.81</td>
</tr>
<tr>
<td>1</td>
<td>92.56&lt;sup&gt;a&lt;/sup&gt;</td>
<td>14.05</td>
<td>59.47</td>
</tr>
<tr>
<td>2</td>
<td>90.31&lt;sup&gt;b&lt;/sup&gt;</td>
<td>15.48</td>
<td>61.44</td>
</tr>
<tr>
<td>3</td>
<td>79.25&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.21</td>
<td>59.47</td>
</tr>
<tr>
<td>4</td>
<td>86.28&lt;sup&gt;b&lt;/sup&gt;</td>
<td>12.37</td>
<td>70.49</td>
</tr>
<tr>
<td>5</td>
<td>69.44&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.17</td>
<td>59.28</td>
</tr>
<tr>
<td>6</td>
<td>105.06&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11.04</td>
<td>62.87</td>
</tr>
<tr>
<td>7</td>
<td>93.56&lt;sup&gt;b&lt;/sup&gt;</td>
<td>10.75</td>
<td>58.62</td>
</tr>
<tr>
<td>8</td>
<td>137.41&lt;sup&gt;b&lt;/sup&gt;</td>
<td>20.94</td>
<td>67.91</td>
</tr>
<tr>
<td>9</td>
<td>85.72&lt;sup&gt;b&lt;/sup&gt;</td>
<td>11.58</td>
<td>67.25</td>
</tr>
<tr>
<td>0</td>
<td>69.84&lt;sup&gt;b&lt;/sup&gt;</td>
<td>9.78</td>
<td>51.84</td>
</tr>
</tbody>
</table>

<sup>a</sup>Significantly different from control group (p < .05).

<sup>b</sup>Significantly different from control group (p < .01).

<sup>c</sup>Scale values are presented in T-scores.
groups on all of the clinical scales. This effect is illustrated in Figure 8.

The effect of the instructions to malingering on the non K-corrected clinical scales (scales 1, 4, 7, 8 and 9) was also examined; the findings are shown in Table 9 and Figure 9. These results parallel those of the K-corrected clinical scales in terms of showing a highly significant difference between the Malingerer and control groups. For the control group, the effect of the K-correction appears to have been one of increasing the T-score values on all of the scales. For the Malingerer group, the same was true for scales 7, 8 and 9, whereas for scales 1 and 4, the K-correction decreased the T-score values.

The second research hypothesis stated that subjects would be able to feign hyper-adjustment (fake good), as evidenced by lower scores on the clinical scales than those of the control group. The results indicate the subjects were, for the most part, unable to accomplish this objective. As shown in Table 8, no significant differences were found between the Fake Good and the Control groups on nine out of ten clinical scales. The only scale showing a significant difference was scale 9. This effect is illustrated in Figure 8.

The effect of K-correction was also examined for the Fake Good group; the results are presented in Table 9 and Figure 9. Significant differences between the Fake Good and the control groups were found for the non K-corrected scales 4, 8 and 9, in contrast to the K-corrected scales, which showed significant differences only for
Figure 8. Effect of instructional set on standard validity and K-corrected clinical scales.
<table>
<thead>
<tr>
<th>Scale</th>
<th>M.</th>
<th>S.D.</th>
<th>M.</th>
<th>S.D.</th>
<th>M.</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.66b</td>
<td>13.96</td>
<td>56.47</td>
<td>15.99</td>
<td>51.09</td>
<td>11.33</td>
</tr>
<tr>
<td>4</td>
<td>93.69b</td>
<td>15.11</td>
<td>69.34</td>
<td>16.91</td>
<td>60.91a</td>
<td>11.93</td>
</tr>
<tr>
<td>7</td>
<td>88.34b</td>
<td>17.05</td>
<td>53.81</td>
<td>11.57</td>
<td>49.44</td>
<td>8.43</td>
</tr>
<tr>
<td>8</td>
<td>118.81b</td>
<td>19.69</td>
<td>60.06</td>
<td>17.17</td>
<td>51.37a</td>
<td>11.80</td>
</tr>
<tr>
<td>9</td>
<td>83.69b</td>
<td>11.90</td>
<td>64.19</td>
<td>11.17</td>
<td>58.19a</td>
<td>8.00</td>
</tr>
</tbody>
</table>

\(^a\)Significantly different from control group (p < .05).

\(^b\)Significantly different from control group (p < .01).

\(^c\)Scale values are presented in T-scores.
Figure 9. Effect of instructional set on standard validity and non K-corrected clinical scales.
scale 9. For the Fake Good group, the effect of the K-correction seems to have been a general increase in the T-score values for five scales.

The remaining nine research hypotheses were all related to the various MMPI validity indicators. Hypotheses 3, 4 and 5 dealt with scales L, F and K, respectively. It was anticipated that the subjects in the Malingering group would be identified by higher scores on the F scale and lower scores on the L and K scales than subjects in the control group. Conversely, it was hypothesized that subjects in the Fake Good group would score lower on the F scale and higher on the L and K scales than the Control group.

As shown in Table 8, the hypotheses were supported by the data for the Malingering group. The subjects in this group scored significantly higher on the F scale and lower on scales L and K relative to the control subjects. In contrast, the subjects in the Fake Good group scored significantly higher than the control subjects on the K scale. For scales L and F, no significant differences were found. Thus, for the Fake Good group, only the hypothesis pertaining to the K scale was supported.

The sixth hypothesis stated that the experimental subjects would be identified by the $F - K$ index. It was anticipated that the subjects in the Malingering group would have a higher, and those in the Fake Good group a lower, $F - K$ score than the Control subjects. The results are presented in Table 10.
Table 10
Effect of Instructional Set on the Special Validity Scales

<table>
<thead>
<tr>
<th>Group</th>
<th>Malingerer (n = 32)</th>
<th>Control (n = 32)</th>
<th>Fake Good (n = 32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>M.</td>
<td>S.D.</td>
<td>M.</td>
</tr>
<tr>
<td>F - K</td>
<td>35.84^b</td>
<td>13.53</td>
<td>-5.84</td>
</tr>
<tr>
<td>OB</td>
<td>470.00^b</td>
<td>60.50</td>
<td>289.91</td>
</tr>
<tr>
<td>SU</td>
<td>261.41^b</td>
<td>28.38</td>
<td>282.34</td>
</tr>
<tr>
<td>O - S</td>
<td>208.59^b</td>
<td>76.64</td>
<td>7.66</td>
</tr>
<tr>
<td>Ds</td>
<td>106.66^b</td>
<td>14.08</td>
<td>57.87</td>
</tr>
<tr>
<td>M</td>
<td>51.59^c</td>
<td>12.10</td>
<td>56.03</td>
</tr>
</tbody>
</table>

^aSignificantly different from control (p < .05).

^bSignificantly different from control (p < .01).

^cNot compared with control.

^dFor this set of data, only the Ds and the M. scales are presented with T-score values. The F-K index is the difference between the F and K scale raw scores. The OB and SO scales are derived by adding the T-score values for the scales which have subtle and obvious subscales, namely scales 2, 3, 4, 6 and 9. The O-S scale is the difference between the score of the OB and SU scales.
The data indicate that a significant difference existed between the Malingerer and control groups on the F - K index in the predicted direction. The subjects in the Fake Good group did not obtain a significantly lower F - K score relative to the control subjects, and thus the hypothesis that this measure would identify the subjects in the Fake Good group was not supported.

The seventh and eighth hypotheses predicted that the subjects in both faking conditions would be identified by their scores on Wiener's Obvious and Subtle scales. It was anticipated that the subjects in the Malingerer group would score higher and those in the Fake Good group lower, than subjects in the control group. Conversely, it was predicted that subjects in the Malingerer group would score lower, and those in the Fake Good group would score higher, than the subjects in the control group.

As shown on Table 10, the Obvious (OB) scale performed as predicted. The Malingerer group had a significantly higher mean than the control, which in turn was higher than the mean for the Fake Good subjects. The Subtle (SU) scale, in contrast, produced significant differences only for the Malingerer group. Thus, the hypothesis for the OB scale was supported by the data, whereas the hypothesis for the SU scale received support only for the Malingerer group.

The ninth hypothesis stated that the subjects in both faking conditions would be identified by the difference of the OB and the SU scale, i.e., the Obvious minus Subtle score (O-S). It was predicted that the subjects in the Malingerer group would score higher,
and those in the Fake Good group would score lower, than the control group subjects.

The data in Table 10 indicate that the 0-S scale performed as predicted. The Malingering group had a significantly higher score than the Control, which in turn had a higher mean than the Fake Good group.

The tenth hypothesis predicted that the subjects in the Malingering group would be identified by a higher score on the Ds scale relative to the control group. As shown in Table 10, the mean Ds score for subjects in the Malingering group was significantly higher than for subjects in the control group, a finding which supports the hypothesis.

The eleventh hypothesis stated that the subjects in the Fake Good group would be identified by a higher score on the $M_p$ scale than those obtained by subjects in the control group. The results presented in Table 10 indicate that the subjects in the Fake Good group scored significantly higher than those in the Control group, providing support for the hypothesis.

Table 11 presents information on the magnitude and significance of the differences between the groups on all of the dependent variables. F values and probability levels are provided for each of the comparisons made.

As stated earlier, optimal cut-off scores were derived for each of the validity indicators that succeeded in producing significantly different distributions for the experimental and control groups. This
Table 11
Univariate F Values and Probability Levels for Malingerer-Control Group and Fake Good-Control Group Comparisons

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Malingerer-Control</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>68.25</td>
<td>.001</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>55.17</td>
<td>.001</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>55.69</td>
<td>.001</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>23.45</td>
<td>.001</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>19.38</td>
<td>.001</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>154.34</td>
<td>.001</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>126.81</td>
<td>.001</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>174.50</td>
<td>.001</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>36.85</td>
<td>.001</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>59.27</td>
<td>.001</td>
</tr>
<tr>
<td>U₄</td>
<td></td>
<td>96.20</td>
<td>.001</td>
</tr>
<tr>
<td>U₅</td>
<td></td>
<td>36.88</td>
<td>.001</td>
</tr>
<tr>
<td>U₇</td>
<td></td>
<td>89.90</td>
<td>.001</td>
</tr>
<tr>
<td>U₈</td>
<td></td>
<td>157.46</td>
<td>.001</td>
</tr>
<tr>
<td>U₉</td>
<td></td>
<td>45.68</td>
<td>.001</td>
</tr>
<tr>
<td>L</td>
<td></td>
<td>4.86</td>
<td>.001</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>244.79</td>
<td>.001</td>
</tr>
<tr>
<td>K</td>
<td></td>
<td>32.66</td>
<td>.001</td>
</tr>
<tr>
<td>F - K</td>
<td></td>
<td>191.28</td>
<td>.001</td>
</tr>
<tr>
<td>OB</td>
<td></td>
<td>133.76</td>
<td>.001</td>
</tr>
<tr>
<td>SU</td>
<td></td>
<td>7.76</td>
<td>.007</td>
</tr>
<tr>
<td>O - S</td>
<td></td>
<td>112.42</td>
<td>.001</td>
</tr>
<tr>
<td>Ds</td>
<td></td>
<td>173.49</td>
<td>.001</td>
</tr>
</tbody>
</table>
Table 11 (Continued)
Univariate Values and Probability Levels for Malingering-Control Group and Fake Good-Control Group
Comparisons

<table>
<thead>
<tr>
<th>Comparison</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.23</td>
<td>.635</td>
</tr>
<tr>
<td>2</td>
<td>1.61</td>
<td>.209</td>
</tr>
<tr>
<td>3</td>
<td>0.20</td>
<td>.655</td>
</tr>
<tr>
<td>4</td>
<td>2.81</td>
<td>.099</td>
</tr>
<tr>
<td>5</td>
<td>0.64</td>
<td>.427</td>
</tr>
<tr>
<td>6</td>
<td>0.47</td>
<td>.495</td>
</tr>
<tr>
<td>7</td>
<td>0.02</td>
<td>.888</td>
</tr>
<tr>
<td>8</td>
<td>3.14</td>
<td>.081</td>
</tr>
<tr>
<td>9</td>
<td>4.27</td>
<td>.043</td>
</tr>
<tr>
<td>0</td>
<td>1.43</td>
<td>.236</td>
</tr>
<tr>
<td>U1</td>
<td>2.41</td>
<td>.126</td>
</tr>
<tr>
<td>U4</td>
<td>5.32</td>
<td>.024</td>
</tr>
<tr>
<td>U7</td>
<td>2.99</td>
<td>.089</td>
</tr>
<tr>
<td>U8</td>
<td>5.33</td>
<td>.024</td>
</tr>
<tr>
<td>U9</td>
<td>6.10</td>
<td>.016</td>
</tr>
<tr>
<td>L</td>
<td>2.61</td>
<td>.111</td>
</tr>
<tr>
<td>F</td>
<td>0.46</td>
<td>.499</td>
</tr>
<tr>
<td>K</td>
<td>4.35</td>
<td>.041</td>
</tr>
<tr>
<td>F - K</td>
<td>3.78</td>
<td>.056</td>
</tr>
<tr>
<td>OB</td>
<td>4.79</td>
<td>.032</td>
</tr>
<tr>
<td>SU</td>
<td>1.50</td>
<td>.225</td>
</tr>
<tr>
<td>O - S</td>
<td>5.37</td>
<td>.024</td>
</tr>
<tr>
<td>M</td>
<td>5.70</td>
<td>.020</td>
</tr>
</tbody>
</table>

*The letter U prior to a scale number indicates it is not K-corrected.*
was achieved by establishing the score which yielded the largest percentage of correct classification for each of the groups. The optimal cut-off scores for each of the scales are presented in Table 12. The scales were then compared for their efficiency at correctly classifying subjects using their optimal cut-off points, and were ranked accordingly. These results are presented in Table 13.  

As mentioned earlier, difficulties encountered in subject recruitment resulted in having to use subjects from two different settings, namely, the Camp and the Penitentiary. This situation raised the concern that the setting from which the subjects were obtained might have an effect on the results obtained. In order to explore this possibility, the location variable was included in the MANOVA program and was treated as an independent variable for the purpose of analysis. The results of the MANOVA analyzing the effect of location independent of the effect of instructional set are presented for the K-corrected and non K-corrected sets of MMPI data in Table 14.

These results indicate there was a significant effect for the location of the subjects across the dependent measures. Tables 15 and 16 present the mean T-score values on all of the dependent measures for the Camp and Penitentiary subjects independent of the instructional set effect. These results are also illustrated in

---

8It is noted that these results are tentative; further research designed to cross validate the cut-off scores is necessary to establish their adequacy.
Table 12
Optimal Cut-off Scores for the Validity Indicators

<table>
<thead>
<tr>
<th>Scale</th>
<th>Malingering</th>
<th>Faking Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>46</td>
<td>--</td>
</tr>
<tr>
<td>F</td>
<td>100</td>
<td>--</td>
</tr>
<tr>
<td>K</td>
<td>46</td>
<td>57</td>
</tr>
<tr>
<td>F - K</td>
<td>15</td>
<td>--</td>
</tr>
<tr>
<td>Ds</td>
<td>75</td>
<td>--</td>
</tr>
<tr>
<td>M</td>
<td>--</td>
<td>57</td>
</tr>
<tr>
<td>P</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>OB</td>
<td>352</td>
<td>260</td>
</tr>
<tr>
<td>SU</td>
<td>274</td>
<td>--</td>
</tr>
<tr>
<td>O - S</td>
<td>62</td>
<td>25</td>
</tr>
</tbody>
</table>

Note: Only those scales which produced significantly different distributions for the experimental and control groups are included.
### Table 13

**Numbers of Subjects Correctly Classified**

**Using Optimal Cut-off Scores for Validity Indicators**

<table>
<thead>
<tr>
<th>Scale</th>
<th># of S's Faking Good (n = 32)</th>
<th># of S's Control (n = 32)</th>
<th>Rank&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>M&lt;sub&gt;p&lt;/sub&gt;</td>
<td>26</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>OB</td>
<td>21</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>K</td>
<td>22</td>
<td>19</td>
<td>3.5</td>
</tr>
<tr>
<td>O - S</td>
<td>22</td>
<td>19</td>
<td>3.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scale</th>
<th># of S's Malingering (n = 32)</th>
<th># of S's Control (n = 32)</th>
<th>Rank&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>O - S</td>
<td>32</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>31</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>OB</td>
<td>32</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>Ds</td>
<td>32</td>
<td>28</td>
<td>4.5</td>
</tr>
<tr>
<td>F - K</td>
<td>29</td>
<td>32</td>
<td>4.5</td>
</tr>
<tr>
<td>K</td>
<td>26</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td>SU</td>
<td>19</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>L</td>
<td>29</td>
<td>23</td>
<td>8</td>
</tr>
</tbody>
</table>

<sup>a</sup>The scales were rank-ordered according to total proportions of correct classification.
### Table 14

**MANOVA Test Results for the Effect of Location on K-corrected and Non K-corrected MMPIs**

<table>
<thead>
<tr>
<th>Test Name</th>
<th>K-Corrected Value</th>
<th>Approx. F</th>
<th>P</th>
<th>Non K-corrected Value</th>
<th>Approx. F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.342</td>
<td>1.97</td>
<td>.021</td>
<td>0.365</td>
<td>2.18</td>
<td>.009</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>0.520</td>
<td>1.97</td>
<td>.021</td>
<td>0.576</td>
<td>2.18</td>
<td>.009</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.658</td>
<td>2.97</td>
<td>.021</td>
<td>0.634</td>
<td>2.18</td>
<td>.009</td>
</tr>
</tbody>
</table>
Table 15
Effect of Inmate Location on the Standard Validity
and K-corrected Clinical Scales

<table>
<thead>
<tr>
<th>Location (Setting)</th>
<th>Camp</th>
<th></th>
<th>Penitentiary</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scale</td>
<td>M.</td>
<td>S.D.</td>
<td>M.</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>53.80</td>
<td>11.10</td>
<td>53.60</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>77.72</td>
<td>38.99</td>
<td>93.80</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>54.70</td>
<td>12.20</td>
<td>50.69</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>65.51*</td>
<td>22.14</td>
<td>75.14</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>63.05*</td>
<td>21.78</td>
<td>74.37</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>61.95*</td>
<td>14.34</td>
<td>69.64</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>67.55**</td>
<td>15.08</td>
<td>78.57</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>62.72</td>
<td>8.84</td>
<td>63.61</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>67.92**</td>
<td>23.02</td>
<td>82.05</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>63.87*</td>
<td>20.34</td>
<td>74.59</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>76.90</td>
<td>38.18</td>
<td>96.66</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>66.92*</td>
<td>14.73</td>
<td>74.87</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>54.40</td>
<td>13.28</td>
<td>58.89</td>
</tr>
</tbody>
</table>

*p < .05

**p < .01
Table 16

Effect of the Inmate Location on the Non K-corrected and Special Validity Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Location (Setting)</th>
<th>Camp M.</th>
<th>Camp S.D.</th>
<th>Penitentiary M.</th>
<th>Penitentiary S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td></td>
<td>59.20**</td>
<td>24.40</td>
<td>72.70</td>
<td>21.26</td>
</tr>
<tr>
<td>U4</td>
<td></td>
<td>66.22*</td>
<td>20.74</td>
<td>80.66</td>
<td>17.71</td>
</tr>
<tr>
<td>U7</td>
<td></td>
<td>58.70</td>
<td>25.18</td>
<td>67.55</td>
<td>18.03</td>
</tr>
<tr>
<td>U8</td>
<td></td>
<td>65.87**</td>
<td>32.85</td>
<td>84.52</td>
<td>35.58</td>
</tr>
<tr>
<td>U9</td>
<td></td>
<td>64.05*</td>
<td>15.21</td>
<td>72.00</td>
<td>14.29</td>
</tr>
<tr>
<td>F - K</td>
<td></td>
<td>0.50</td>
<td>24.32</td>
<td>10.68</td>
<td>22.65</td>
</tr>
<tr>
<td>Ds</td>
<td></td>
<td>64.42</td>
<td>29.27</td>
<td>76.48</td>
<td>26.66</td>
</tr>
<tr>
<td>M P</td>
<td></td>
<td>57.22</td>
<td>15.28</td>
<td>56.70</td>
<td>10.75</td>
</tr>
<tr>
<td>OB</td>
<td></td>
<td>300.17**</td>
<td>115.25</td>
<td>367.98</td>
<td>96.48</td>
</tr>
<tr>
<td>SU</td>
<td></td>
<td>277.40</td>
<td>29.17</td>
<td>279.21</td>
<td>34.09</td>
</tr>
<tr>
<td>O - S</td>
<td></td>
<td>22.85*</td>
<td>133.29</td>
<td>88.75</td>
<td>116.85</td>
</tr>
</tbody>
</table>

\(^a\)The letter U prior to a scale number means that it is not K-corrected.

\(^b\)These scales have both positive and negative values.

\(^*P < .05\)

\(^{**}P < .01\)
Figure 10. Effect of location on standard validity and K-corrected clinical scales.
Figure 11. Effect of location on standard validity and non K-corrected clinical scales.
Figures 10 and 11. In general, the Penitentiary subjects appeared
to have higher mean scores on most of the scales, and some of these
differences reached significant levels (i.e., for scales 1, 2, 3, 4,
6, 7, 8, 9, U1, U4, U8, U9, OB and O-S). The issue then became
whether the effect of location was interacting in some way with the
effect of instructional set to create a bias in one of the experi­
mental groups. Since the MANOVA processed both the instructional
set and the location as independent variables, it also provided
tests for the interaction effect. The results of the MANOVA tests
for the effect of interaction on the K-corrected and non K-corrected
sets of data are shown in Table 17.

The results indicate there was no significant effect on the de­
pendent measures due to an interaction of the location and instruc­
tional set factors. Nonetheless, there was the problem of a possible
overrepresentation of subjects from one location in one of the ex­
perimental groups. It may be recalled that in the Malingering group
there were 10 subjects from the Camp and 22 from the Penitentiary.
In the control and Fake Good groups, there were 15 and 17 subjects
from the Camp and the Penitentiary, respectively. A chi square
analysis was performed to see if these frequency distributions were
significantly different. The resulting chi square was 2.14 (P = .342),
indicating no significant differences existed. In addition, no sig­
nificant differences were found between the experimental and control
groups on any of the demographic variables (i.e., age, sentence, edu­
cation, offense category, previous residence, and ethnicity).
Table 17

MANOVA Test Results for the Effect of Instructional Set and Location Interaction on K-Corrected and Non K-corrected MMPIs

<table>
<thead>
<tr>
<th>Test Name</th>
<th>K-corrected Value</th>
<th>Approx. F</th>
<th>P</th>
<th>Non K-corrected Value</th>
<th>Approx. F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai's</td>
<td>0.434</td>
<td>1.06</td>
<td>.390</td>
<td>0.488</td>
<td>1.24</td>
<td>.183</td>
</tr>
<tr>
<td>Hotelling's</td>
<td>0.613</td>
<td>1.15</td>
<td>.280</td>
<td>0.753</td>
<td>1.41</td>
<td>.079</td>
</tr>
<tr>
<td>Wilk's</td>
<td>0.600</td>
<td>1.10</td>
<td>.332</td>
<td>0.549</td>
<td>1.32</td>
<td>.122</td>
</tr>
</tbody>
</table>
One other matter that was explored was the question of how the scores obtained in this study under the experimental and control conditions described above relate to the scores these same subjects would obtain under more natural conditions. As stated previously, only twenty-three previously administered MMPI's were obtained for the Camp subjects, and none were available for the Penitentiary subjects. Table 18 presents the data for these subjects on all of the dependent measures. The K-corrected and non K-corrected profiles for this group are illustrated in Figures 12 and 13.

Comparison of these profiles with those of the control group (c.f., Figures 1 and 2) shows a general resemblance between them in terms of overall pattern elevations.

A five-point scale was administered to each subject concerning his attitudes toward the task required of him (a copy is included in Appendix B). Of the 96 subjects, 95 reported they understood the instructional set given to them concerning how to take the test. The subject who reported not understanding the instructions was part of the Malingerer group. Similarly, 95 subjects reported they were able to follow the instructions. The subject who reported not being able to follow the instructions was also part of the Malingerer group. Six subjects reported they had answered in a random fashion. Of these, three were part of the Malingerer group, two were in the control group, and one was in the Fake Good group. In order to maintain the intactness of the groups and the randomization procedure, and given the relatively small number of these subjects, their records
### Table 18
Mean T-score Values for Camp Inmates Prior to the Experiment

<table>
<thead>
<tr>
<th>Scale</th>
<th>M</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>50.87</td>
<td>8.25</td>
</tr>
<tr>
<td>F</td>
<td>57.00</td>
<td>9.29</td>
</tr>
<tr>
<td>K</td>
<td>56.27</td>
<td>10.22</td>
</tr>
<tr>
<td>1</td>
<td>54.22</td>
<td>10.06</td>
</tr>
<tr>
<td>2</td>
<td>57.22</td>
<td>9.46</td>
</tr>
<tr>
<td>3</td>
<td>56.78</td>
<td>6.86</td>
</tr>
<tr>
<td>4</td>
<td>65.65</td>
<td>8.06</td>
</tr>
<tr>
<td>5</td>
<td>61.04</td>
<td>6.98</td>
</tr>
<tr>
<td>6</td>
<td>57.17</td>
<td>9.69</td>
</tr>
<tr>
<td>7</td>
<td>54.22</td>
<td>8.55</td>
</tr>
<tr>
<td>8</td>
<td>59.61</td>
<td>11.70</td>
</tr>
<tr>
<td>9</td>
<td>60.91</td>
<td>8.48</td>
</tr>
<tr>
<td>0</td>
<td>49.66</td>
<td>8.01</td>
</tr>
<tr>
<td>U1</td>
<td>50.78</td>
<td>9.02</td>
</tr>
<tr>
<td>U4</td>
<td>62.48</td>
<td>12.29</td>
</tr>
<tr>
<td>U7</td>
<td>49.48</td>
<td>10.24</td>
</tr>
<tr>
<td>U8</td>
<td>52.78</td>
<td>11.73</td>
</tr>
<tr>
<td>U9</td>
<td>56.35</td>
<td>9.80</td>
</tr>
<tr>
<td>OB</td>
<td>259.52</td>
<td>39.69</td>
</tr>
<tr>
<td>SU</td>
<td>284.56</td>
<td>25.17</td>
</tr>
<tr>
<td>OB-SU</td>
<td>-25.04</td>
<td>57.26</td>
</tr>
<tr>
<td>Ds</td>
<td>51.26</td>
<td>11.73</td>
</tr>
<tr>
<td>M</td>
<td>55.04</td>
<td>8.52</td>
</tr>
</tbody>
</table>

**Note:** n = 23

*a* The letter U prior to a scale indicates that it was not K-corrected.

*b* These scales are a summation of T-score values.
Figure 12. Mean K-corrected MMPI profile pattern for Camp inmates prior to the study.
Figure 13. Mean non K-corrected MMPI profile pattern for Camp inmates prior to the study.
were included in the sample.

Discussion

The findings obtained in this study will be discussed in terms of the subject's ability to manipulate the MMPI clinical scales in the desired direction and in terms of the effectiveness of the various validity indicators in detecting the fakers. Following this, the discussion will turn to the problem of the location of the subjects used (i.e., Camp, Penitentiary) and some of the implications raised by their differences on the dependent measures. The focus will then turn to the issue of how these data relate to the use of the MMPI for the assessment of an individual's credibility, a task which forensic psychologists are frequently required to perform.

The results indicate that the subjects were able to manipulate the clinical scales of the MMPI when instructed to mangle psychopathology. The consistency of the difference between the score of the Malingerer group and the control group on all the clinical scales, as well as the magnitude of some of these differences (e.g., scale 8) suggest they were quite successful at this task. In contrast, the subjects did not appear to be able to manipulate the scales in the direction of hyper-adjustment since only one scale (Scale 9) differentiated between the subjects in the Fake Good and control groups. At the same time, the Malingerer group subjects were readily identified by all of the faking measures used in this study. The detection of faking good, in turn, was much less successful, in that only four measures (M_p, OB, O-S and K) out of nine, were able to discriminate between the Fake
Good and the control subjects. However, a conceptual issue is raised by the performance of the Fake Good group, in that if they were not able to fake hyper-adjustment, then it would follow that there should not be anything for the faking indices to detect. Nonetheless, given that most of the validity indices are measures of defensiveness or denial theoretically independent of the clinical scales, their poor performance in this study raises questions as to the utility of these indices for the detection of faking good.

The findings discussed so far have some resemblance to those obtained by other investigators. In studies using prisoners, Gendreau et al. (1973) and Hunt (1941) both reported their subjects were quite successful at feigning psychopathology, and also that the validity indicators readily identified these subjects. Results consistent with these were reported by Grow (1980) and Exner et al. (1973) using college students and by Anthony (1971) using Air Force personnel as subjects.

With regard to faking good, the pattern of findings reported by different investigators is more complex. In the Gendreau et al. (1973) study, the authors reported that their subjects were able to lower their scores on scales 2, 7, 8 and 9 under instructions to feign hyper-adjustment. Similarly, Lawton and Kleban (1965) reported that under instructions to fake good, their groups of prisoners were able to significantly lower all their scale scores except for scale 4. For the sample used in the present study, significant differences existed between the Fake Good and control subjects only for scale 9. However,
using the non K-corrected scales, significant differences emerged between the groups on scales 4, 8 and 9. (The effect of applying the K-correction factor on the results obtained in this study will be discussed more fully below.) The inconsistency observed across these studies may be due to differences in populations sampled, procedures used, etc., or it may be that feigning hyper-adjustment is a complex phenomenon difficult to quantify and investigate.

The findings reported in the literature concerning the detection of positive faking are also somewhat inconsistent. Gendreau et al. reported that all of the measures they used (M, F - K, L, F, K, OB, SU, F and Ds) distinguished between the fake good and control group scores. Hunt (1948), in contrast, reported a failure to identify fake good subjects using the F - K index. Using a non-prisoner sample, Exner et al. (1963) reported that none of the validity indicators they used were successful in identifying fake good profiles. Similarly, Grow (1981) reported that among college students the identification of fake good records was much less successful than for the malingered records. In the present study, the M, OB, O-S and K scales were able to discriminate between the Fake Good and the control group subjects, although relative to the Malingering group, the magnitude of the differences was considerably less. Thus, it appears that the detection of faking good is also a more complex task than the detection of malingering. Other researchers (e.g., Greene, 1980; Dahlstrom et al., 1975) have suggested that the subtle and obvious scales might have some promise for use in the detection of faking
good; the results of the present study support this suggestion. However, more research is needed on these and other faking good indicators in order to establish their usefulness.

As mentioned earlier, the results obtained in this study were influenced by whether or not the K-correction factor was applied. The net effect of applying the K-correction to the scales for the control and Fake Good groups was a general increase in their T-score values. For the Malingering group, the effect was an increase in the value of scales 7, 8 and 9, and a decrease for scales 1 and 4. The interesting finding was that, for the Fake Good and control group comparisons, the non K-corrected records showed significant differences on scales 4, 8 and 9, whereas the K-corrected ones showed significant difference only for scale 9. This finding would argue for the continued routine use of K-corrected scores, given that they appear to be somewhat more resistant to attempts to fake good.

One of the possible threats to the internal validity of this study is the fact that the subjects were recruited from two different locations and the analyses revealed a significant difference between the two groups on several of the dependent measures, independent of the instructional set effect. In addition, the two groups showed significant differences on many of the demographic variables. The possibility also exists that complete randomization was not achieved, since the Penitentiary inmates, although randomly assigned, were

9The K-correction factor is applied to scales 1, 4, 7, 8 and 9.
replacement subjects. On the other hand, the MANOVA for the location-by-instructional set interaction was not significant, and the chi square to see if significant differences existed between the experimental groups in terms of number of subjects each had from the two locations was also non-significant. Thus, although the internal validity may have been compromised somewhat, the experimental groups appeared to be roughly comparable in terms of the more relevant characteristics.

Before concluding this section, there are a number of legal issues which should be mentioned pertaining not only to the use of the MMPI but also to psychological testing in general. These issues revolve around three common potential risks associated with psychological testing: invasion of privacy, self-incrimination, and unfair discrimination. According to Schwitzgebel and Schwitzgebel (1980), invasion of privacy may take the form of "public disclosure of private facts, publicly placing a person in a false light, commercial exploitation, and intrusion into the person's physical or mental solitude" (p. 126). The risk of such an invasion of privacy does exist for individuals tested with the MMPI. In particular, studies cited earlier (see p. 12) suggest that some subjects

10 It is acknowledged that, even though the chi square was non-significant, the fact remains that there were only 10 subjects out of 32 from the Camp in the Malingering group. This probably accounts to some extent for the differences in profile patterns between the subjects in the two locations shown in Figures 3 and 4.

11 For a more comprehensive discussion of these and other legal issues, the reader is referred to Schwitzgebel and Schwitzgebel (1980).
experience many of the items on the MMPI as too personal or intrusive into their privacy. It is the responsibility of the test user to insure the confidentiality of the test results and to utilize testing procedures that are the least intrusive necessary to achieve their assessment goals.

The use of psychological tests in criminal proceedings raises concern over the possible infringement of an individual's privilege against self-incrimination. Responses to the test might constitute an admission of guilt to a crime, and the examinee, presuming confidentiality of the test results, could unwittingly incriminate him/herself. However, an individual can waive his/her privilege against self-incrimination by giving voluntary and informed consent to any testing procedures. In addition, where a defendant first raises the issue of insanity in a criminal trial, then he/she implicitly waives any objection to psychological testing if such is required.

The problem of unfair discrimination resulting from the use of psychological tests such as the MMPI is also germane to the present discussion. Where a test has been developed and standardized on a certain population (e.g., white, middle class subjects), it is likely to be inapplicable to members of another population. When used with subjects from non-normative populations, the test may be used to unfairly discriminate against them by yielding unfavorable results reflective of the inadequacies of the instrument for the particular population rather than actual subject characteristics. Although this problem arises more frequently in the context of cases involving
unfair employment practices or educational matters, it seems the same concerns are raised with regard to the use of psychological tests in clinical/forensic settings. With regard to the MMPI, several researchers (e.g., Costello, Fein & Blair, 1973; Gynther, 1972; Gynther & Lachar, 1978; see also discussion above, p. 28) have pointed out the inadvisability of using this test, which was developed and standardized on an all-white population, with non-white subjects. When a clinician uses a test such as the MMPI for making dispositional or other recommendations concerning patients or clients, he/she is responsible for knowing the relevance of the test for various subject populations and for supporting claims of validity and reliability (APA, 1978).

In conclusion, the results provided by this study allowed for an examination of the effectiveness of the various faking detection techniques with a prisoner sample, something which was needed, since these techniques were developed and have been studied mainly with non-prisoner samples. Given that the MMPI is used routinely in many correctional institutions (Dahlstrom et al. 1972), this represented a serious deficiency. The findings provided here by no means provide a new normative data base for use with prisoner populations. They do, however, give a better idea about which scales are useful for the purpose of assessing a subject's credibility and what is to be considered a significant score on each of these. Thus, it seems that with the data provided here, it may be possible to use the various faking detection techniques with a greater degree of certainty and
accuracy when assessing subjects from correctional settings.

This issue of developing data relevant to the populations being evaluated is especially important for psychologists who work within the criminal justice system (Schwitzgebel & Schwitzgebel, 1980; Cooke, 1980). As mentioned previously, psychologists are frequently requested by the legal system to conduct evaluations on individuals who, for reasons described above (c.f., p.p. 1 ff.), may want to present either an overly favorable or unfavorable portrayal of themselves. In such cases, the issue of credibility becomes relevant. Psychologists who perform these types of evaluations frequently include the MMPI as part of their assessment procedures, since the various validity indicators provide data relevant to the issue of credibility (Ziskin, 1981). The results presented here have added to the data base available for the MMPI, and therefore, will hopefully improve the quality of the decisions made by clinicians who utilize the MMPI in forensic settings.
REFERENCES


Edwards, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. Journal of Applied Psychology, 1953, 37, 90-93.


Fricke, B. G. A response bias (B) scale for the MMPI. *Journal of Consulting Psychology*, 1957, 4, 149-153.


Gough, H. G. The F minus K dissimulation index for the MMPI. *Journal of Consulting Psychology*, 1950, 14, 408-413.


Grayson, H. M., & Olinger, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *Journal of Consulting Psychology*, 1957, 21, 73-77.


Harvey, M. A., & Sipprelle, C. N. Demand characteristic effects on the subtle and obvious sub-scales of the MMPI. *Journal of Personality Assessment*, 1976, 40, 539-544.


Hovey, H. B. Correction of MMPI profiles for excessive item omissions. Unpublished materials, Veterans Administration Hospital, Salt Lake City, 1958.


Marks, P. A., & Seeman, W. The actuarial description of personality; an atlas for use with the MMPI. Baltimore: Williams & Wilkins, 1963.


Silver, R. J., & Sines, L. K. Diagnostic efficiency of the MMPI with and without the K correction. *Journal of Clinical Psychology*, 1962, 18, 312-314.


Wiener, D. N. Selecting salesmen with subtle-obvious keys for the MMPI. *American Psychologist*, 1948, 3, 364.


INFORMED CONSENT STATEMENT

A. General Information:

My name is Rafael Salcedo, and I am affiliated with the Louisiana State University in Baton Rouge, Louisiana. This study is entitled MMPI Response Patterns Among Prisoners Under Varied Instructional Sets. The purpose of this study is to find out how prisoners respond to a psychological test when given specific instructions on how to take the test, that is, when told to assume certain roles while answering the various questions. You will be requested to take the test under instructions to assume certain roles. The results will be used to increase our understanding of what happens to test scores when people try to assume certain roles. The risks and discomforts are minimal. They may include your becoming bored or restless from taking the test. Possible benefits to others from your participation in this study include an increase in our knowledge about how prisoners take psychological tests.

Your participation is completely voluntary and you may refuse participation at any time without penalty or prejudice. All research information will be handled in the strictest confidence and your participation will not be individually identifiable in any reports. If you are an inmate, your participation or non-participation in this research project will not affect your release date or parole eligibility. I will be happy to answer questions you have about the above items.

B. Signed Consent Portion:

I, _________________________________, understand the study entitled MMPI Response Patterns Among Prisoners Under Varied Instructional Sets as explained on page 1 and I consent to participate in the study. My participation is completely voluntary.

I consent to the following procedures (initial what you agree to; cross out what you do not agree to):
1. I authorize __________________________________________
   (institution)
   staff to release the information specified below to the
   researchers only for the purpose of this study and only until
   the completion of this project. I understand that I may revoke
   this consent in writing before the information is disclosed.

   ___________________________  Central File  Initials ________
   ___________________________  Other (specify)

2. I consent to complete written tests/questionnaires
   and/or to participate in an interview, and/or to ________

   ________________________________
   Initials ________

I understand that all research information will be handled in
the strictest confidence and that my participation will not be indi-
vidually identifiable in any reports. I understand that participation
or non-participation in this research project will not affect my
release date or parole eligibility. I further understand that there is
no penalty or prejudice of any kind for withdrawing from or not partici-
pating in the study.

   __________________________         __________________________
   (Signature)                             (Date)

   __________________________         __________________________
   (Register Number)            (Unit)

   __________________________         __________________________
   (Witness Typed Name and Signature)                     (Date)

cc: Research Project File
   Privacy File (only where the researcher is authorized access to the
   inmate's Central or Medical File)
   Subject (upon request)
APPENDIX B

ATTITUDE TOWARDS STUDY SCALE

PLEASE ANSWER THE FOLLOWING QUESTIONS USING THE SCALE SHOWN BELOW:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disagree</td>
<td>Somewhat Disagree</td>
<td>Neither Agree nor Disagree</td>
<td>Somewhat Agree</td>
<td>Agree</td>
</tr>
</tbody>
</table>

1. I understood the instructions concerning what I was expected to do.

2. I was able to follow the instructions given.

3. I did not read most of the items, that is, I answered randomly.
APPENDIX C

DEMOGRAPHIC DATA FORM

NAME: ___________________________ # __________________________ AGE: __________

DATE OF BIRTH: ________________ PLACE OF BIRTH ____________________________

CONVICTION __________________________________________________________________

SENTENCE ____________________________________ # OF YEARS SERVED _____________

HOW MUCH TIME LEFT TILL RELEASE ___________________________________________

LAST RESIDENCE (CITY & STATE) ____________________

FOR HOW LONG ____________________________________________

HIGHEST SCHOOL GRADE COMPLETED ___________ DEGREES _________________________

MARITAL STATUS ______________________________________________________________


RAFAEL F. SALCEDO
P. O. Box 2321
Atascadero, CA 93423
Telephone: 805-461-2219

PERSONAL: Birthday: Height: Weight: Health: Single
12/11/56 5'10" 160 lbs. Excellent Bilingual

EDUCATION: Louisiana State University, Baton Rouge, LA
Degree: Ph.D. in Clinical Psychology (Anticipated date of
   dissertation completion: September, 1983).
Minor: Behavioral Neurology
M.A. in Clinical Psychology (1980)

Florida State University, Tallahassee, FL
Degree: B.S. in Psychology (1978)
Minor: Cultural Anthropology, Speech Communication

RELEVANT EXPERIENCE:

9/82-Present Atascadero State Hospital, Atascadero, CA
Position: Clinical Psychology Intern
Responsibilities: Member of an interdisciplinary team responsible for treatment and dispositional decisions concerning judicially committed patients. Population includes patients found Incompetent to Stand Trial, Not Guilty by Reason of Insanity, and Mentally Disordered Sex Offenders. Forensic evaluation responsibilities include assessment of competency to stand trial, potential for reoffending and/or violence, and amenability to treatment. Other evaluation responsibilities include neuropsychological assessments using the Halstead-Reitan Battery and other procedures, assessment of acculturation (for hispanic patients), and more traditional assessment procedures conducted to provide treatment recommendations. Treatment responsibilities include individual and group therapy with a broad range of patients and culturally sensitive counseling with hispanic patients. Completed a six month rotation in an innovative family therapy program in which spouses and children were actively involved in treatment. Consulting functions have included coordinating and implementing a standardized evaluation procedure to assess hispanic patients' suitability for a culturally sensitive treatment program. Also designed and presented a group therapy format for Anger Management Training which was later implemented by level of care staff.

Didactic experiences include seminars in Forensic Psychology, Psychotherapy and Hypnotherapy. The internship also includes a rotation at the Counseling Center at California Polytechnic State University which involves counseling college students concerning personal and academic problems. Duties include career counseling and evaluations.
6/81-9/82 Feliciana Forensic Facility, Jackson, LA  
**Position:** Psychological Assistant III  
**Responsibilities:** Conducted evaluations for patients committed under criminal/judicial statutes. Assessments provided recommendations relating to various forensic issues, including competency to stand trial, potential for violence, and possible mental state at the time of the offense. Also engaged in neuropsychological assessments using the Halstead-Reitan Battery and the Luria-Nebraska with a broad range of patients. Under supervision from Chief Psychologist, completed two comprehensive literature reviews (unpublished), one concerning psychopathy and the other dealing with malingering on psychological tests.

6/81-9/82 The Runnymede Clinic, 10319 Old Hammond Hwy., Baton Rouge, LA 70815  
**Position:** Clinical Associate  
**Responsibilities:** Member of an interdisciplinary health care team in a private psychological firm. Clients included children, adolescents, adults and families. Engaged in a variety of treatment modalities including biofeedback, behavioral medicine procedures, and individual, group, and family therapy. Duties also included neuropsychological assessments of adults, adolescents and children (using the Halstead-Reitan), educational evaluations contracted for by the school system, and behavioral assessments, as well as more traditional psychological evaluations. Also provided consultation, evaluation and therapy services to St. Joseph's Children's Home, a residential setting for girls. Subspecialty area included assessment and treatment of Hispanic patients.

9/80-7/81 Don Lichtenstein, Ph.D., Inc., 625 Colonial Dr., Baton Rouge, LA 70813  
**Position:** Psychological Associate  
**Responsibilities:** Conducted psychological evaluation for a private psychological firm in contractual agreement with the Chemical Dependency Unit of Baton Rouge General Hospital. Patients included adolescents and adults experiencing a variety of substance abuse disorders. Provided recommendations to hospital staff for the purpose of treatment planning.

6/80-9/80 Greenwell Springs Hospital, Baton Rouge, LA  
**Position:** Psychiatric Aide  
**Responsibilities:** Supervised a group of emotionally disturbed and juvenile delinquent adolescents residing in a structured, secure setting. Patients included both males and females and exhibited a wide range of disorders. Duties included crises intervention, management of assultive behavior, modeling of appropriate behaviors, and other forms of intervention. Also conducted psychological evaluation utilized in treatment planning.
9/79-8/81  
Student Health Service, Louisiana State University, Baton Rouge, LA  
**Position:** Psychological Assistant  
**Responsibilities:** Provided academic and personal counseling to college students. Counseling modalities included individual, couple, family and group therapy. Treatment was most often short-term, problem or crisis oriented. Other duties included conducting psychological, neuropsychological and academic evaluation, and collaborating with other disciplines concerning counseling issues and operating policies.

8/79-6/80  
Psychology Department, Louisiana State University, Baton Rouge, LA  
**Position:** Teaching Assistant  
**Responsibilities:** Assisted in the preparation of lectures and examinations for upper level undergraduate psychology courses. Substituted professor in delivering lecture when necessary. Coordinated ongoing research projects for students working under professor.

1/79-6/79  
Developmental Disabilities Center, 408 Government Street, Baton Rouge, LA 70811  
**Position:** Program Evaluation  
**Responsibilities:** Conducted evaluations of several programs funded by this agency. Evaluation consisted of reviewing the design and implementation of the programs based on written descriptions and on-site visits.

**VOLUNTEER WORK**

1/79-6/79  
St. Joseph's Children's Home, Baton Route, LA  
**Responsibilities:** Conducted a program evaluation by assessing existing need area and assisting staff in establishing and implementing future goals.

9/78-6/79  
"The Phone", Crisis intervention and information referral hotline, Baton Rouge, LA  
**Responsibilities:** Completed a six week training period on crisis intervention techniques and served as a phone counselor.

**AFFILIATIONS**

American Psychological Association, student affiliate.  
San Luis Obispo County Psychological Association, student member.  
Atascadero State Hospital Psychology Organization, student member.  
Louisiana Psychological Association, student member.
RESEARCH

MMPI Response Patterns Among Prisoners Under Varied Instructional Sets.

Practice and Training Characteristics of Division 37 Members.
Co-authored with June M. Tuma, Ph.D., submitted for publication in Professional Psychology.

The Rorschach Experience Balance and Problem Solving Behavior.
Unpublished Master's Thesis (1980), Louisiana State University, Baton Rouge, LA.

WORKSHOP
PRESENTATIONS

"Communication Skills Training" Workshop conducted with Budd Babb, Ph.D., for students at California Polytechnic State University, May, 1983.

"Procedures for Determining Degree of Acculturations in Hispanic Patients and Suitability for Culturally Sensitive Treatment Program". Presented to Atascadero State Hospital Spanish Project staff, April, 1983.

"A Group Format for Anger Management Skills Training".
Presented to Atascadero State Hospital Staff, March, 1983.
Candidate: Rafael F. Salcedo

Major Field: Psychology

Title of Thesis: Minnesota Multiphasic Personality Inventory (MMPI)
Response Patterns Among Prisoners Under Varied Instructional Sets

Approved:

[Signatures]

Major Professor and Chairman
Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Felicia A. Pynn
Junge Suen
Harold Shernan
Carrol Tingle

Date of Examination:

September 2, 1983