



1979

Effects of Administrative Situational Variables on Minnesota Multiphasic Personality Inventory Validity Scores of Drug Abusers

Barry Joseph Weber
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

 Part of the [Psychology Commons](#)

Recommended Citation

Weber, Barry Joseph, "Effects of Administrative Situational Variables on Minnesota Multiphasic Personality Inventory Validity Scores of Drug Abusers" (1979). *Dissertations*. 1870.
https://ecommons.luc.edu/luc_diss/1870

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](#).
Copyright © 1979 Barry Joseph Weber

EFFECTS OF ADMINISTRATIVE SITUATIONAL VARIABLES ON
MINNESOTA MULTIPHASIC PERSONALITY INVENTORY
VALIDITY SCORES OF DRUG ABUSERS

by
Barry J. Weber

A Dissertation Submitted to the Faculty of the Graduate School
of Loyola University of Chicago in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

August

1979

ACKNOWLEDGMENTS

I am indebted to Drs. Alan DeWolfe, James Johnson and Daniel Barnes for their patient assistance to my efforts to complete the dissertation. The balance of criticism with helpful discussion and advice made my work easier.

My wife, Kathleen S. Weber, also receives my gratitude for carrying extra family responsibilities so that I could complete my work.

Dr. Carol Jean Rogalski assisted me in the early stages of data collection and made suggestions about better ways to organize my experimental role with my intern role so that the data collection was more efficient. Needless to say, I appreciated both forms of help very much.

Without the administrative approval of Dr. George Meschel, Chief of Psychology, and Dr. Robert Craig, Chief of Drug Dependency Treatment Program, both of West Side Veterans Hospital, collection of data would have been impossible. They both also offered emotional support and intellectual interest which was not only beyond the call of their official duties but was also facilitative of an excellent working research environment.

I also thank the patients and other staff of West Side Veterans Hospital in Chicago, Illinois, without whose cooperation this research would have been impossible.

VITA

The author, Barry Joseph Weber, is the son of Joseph Ralph Weber and Linda Ficklin Weber. He was born June 25, 1947, in St. Louis, Missouri.

His elementary and secondary education was obtained in the public schools of Villa Park, Illinois. He graduated from high school in 1965.

In September, 1965, he entered Hamilton College in Clinton, New York. In June of 1969, he received the Bachelor of Arts degree with a major and honors in psychology. While attending Hamilton, he co-authored two publications on sensor psychology and was elected to student membership in Sigma Xi, a national honorary society for research in science.

He attended the University of Rochester for one year, where he studied perceptual psychology at the Center for Visual Science. During that time, he was an assistant to Dr. John Lott Brown and a teaching assistant at the National Training Institute for the Deaf.

In September of 1970, he entered Bethany Theological Seminary. He received a Master of Divinity degree, passing two of his three qualifying examinations with distinction, from Bethany in June of 1973. Following seminary, he was pastor of a church in Springfield, Illinois for two years.

In September of 1975, he entered Loyola University of Chicago. During 1976, he completed clerkship training at Loyola University of

Chicago Counseling Center. The following year, he was an intern at West Side Veterans Hospital in Chicago, Illinois. In June of 1978, he received a Master of Arts degree from the graduate school for his successful completion of the Department of Psychology's requirements for that degree.

In September of the same year, he took a position at the Day Hospital of Northwestern University's Institute of Psychiatry, where he is currently employed, for additional experience with the treatment of chronic psychiatric patients.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES	vi
INTRODUCTION AND REVIEW OF THE LITERATURE	1
The Minnesota Multiphasic Personality Inventory	3
The Validity Scales of the MMPI	3
The L Scale	5
The F Scale	6
The K Scale	6
F-K	8
Non-distortion Interpretations of the Validity Scales	9
Are the Validity Scales Clinical Scales As Well?	9
Refusals to Complete the MMPI	11
Environmental Impact on Validity Scores	12
Instructions	12
Group or Individual Testing	15
Administrator Gender	17
Validity Scale Scores and Drug Abusers	21
The MMPI and Race	22
Pilot Study	23
Hypotheses of the Present Study	24
METHOD	25
Subjects	25
Materials	25
Instructions	25
Procedure	28
RESULTS	31
DISCUSSION	46
REFERENCES	56
REFERENCE NOTES	64

LIST OF TABLES

Table	Page
1. Some Demographic Variable Means for Each Treatment Group	26
2. Total Group Frequency Distributions for Nine Categorical Demographic Variables	27
3. Portion of Multivariate Analysis of Variance with L Score as Dependent Measure	32
4. Portion of Multivariate Analysis of Variance with F Score as Dependent Measure	33
5. Portion of Multivariate Analysis of Variance with K Score as Dependent Measure	34
6. Portion of Multivariate Analysis of Variance with F-K Score as Dependent Measure	35
7. Refusals to Take MMPI as a Function of Independent Variables	36
8. Mean Validity Scale Scores for the Total Sample	39
9. <u>Chi-square</u> Tests of Differences in Categorical Demographic Variables Across Treatment Groups	40
10. The <u>F</u> -Tests of Differences in Continuous Demographic Variables Across Treatment Groups	41
11. Frequencies of Occurrence of Certain K Scores as a Function of Alcohol Use	43
12. Analysis of Covariance with K Scale Score as a Dependent Measure and Alcohol Use as a Covariate	44

INTRODUCTION AND REVIEW OF THE LITERATURE

Psychological tests come in many varieties. Some, like the Rorschach and the Thematic Apperception Test are projective in nature, requiring a more impressionistic approach both by the person taking the test and by the examiner who interprets it. Others, like the Wechsler Adult Intelligence Scale (WAIS), an objective test, and the Minnesota Multiphasic Personality Inventory (MMPI), a subjective test, have standardized norms with which the results obtained from any individual can be compared for interpretation.

Most psychological tests, however, share a basic premise: that the test measures something intrinsic to the person. Variability in response due to situations may be acknowledged by test authors, but is considered "noise" that an examiner should attempt to minimize in the interests of obtaining a clear "signal" about the person's internal psychological state. Sometimes even certain intrapsychic factors are also considered "noise" in the interest of getting a more clear signal about other aspects of the person. For example, when a person's score on the Digit Span and Arithmetic subscales of the WAIS are much lower than scores on other subscales, the usual interpretation is not that a person's ability to manipulate numbers or his memory is impaired, but that anxiety interferes with his intellectual functioning. The most central function of the WAIS is to assess intellectual ability, which then would be estimated as to what it would be without impairment by emotional factors. The anxiety in certain types of functioning might be noted, but for the evaluation of general intellectual

ability, it would be considered a type of noise to be filtered out from the signal about the level of intellectual functioning.

The basic premise of psychological tests, then, is that they measure certain specified things that are true about an individual's internal functioning and do not measure unspecified or environmental influences. The present study examines some aspects of this basic premise for one important psychological test, the MMPI.

THE MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

The MMPI has been found to be the fifth most widely used psychological test in the United States (Lubin, Wallis & Paine, 1971). It has been the most widely used test in work with heroin addicts (Craig, 1976b). It has also been widely used with many other kinds of persons for a variety of purposes (Dahlstrom, Welsh & Dahlstrom, 1975, pp. 3-117).

The reliability and validity of the MMPI have been repeatedly measured and supported (Dahlstrom et al., 1972, pp. 99-154). Its accuracy when used to diagnose or describe a person's personality functioning has been shown to be as good or as superior to other clinical measures (Cowan & Walkins, 1975; Newark, Ziff, Finch & Kendall, 1978; Newark, Gentry, Simpson & Jones, 1978; Wildman & Wildman, 1975). Compared with other psychological tests, various addiction inventories and physiological tests, the MMPI was most successful in discriminating between addicts and non-addicts (Siegal, 1976). Therefore, the MMPI is a very important clinical tool. Because of its popularity and usefulness in work with addicts, it makes special sense to do MMPI research with this population.

The Validity Scales of the MMPI

The MMPI has an unique feature in that "validity" scales are incorporated into it. Researchers have generally shown that the accuracy of the self-report (Dahlstrom et al., 1972, pages 94-174). These scales are used by examiners for the important task of deter-

mining the validity of a particular report. If the scales indicate that a test is valid, the result will be considered as an accurate measure of the person's intrapsychic process. If the scales indicate that the test is invalid, the information about the person from the clinical scales will be considered untrustworthy. The validity scales, then, are used as the fundamental indicators of whether or not the basic premise of psychological testing has been met for a particular MMPI. In examining the possible impact of situational variables on the MMPI; that is, in examining the credibility of the basic premise that the MMPI measures something having to do with a person's intrapsychic process, it is sensible to consider the validity scales as dependent variables.

For the purposes of the present discussion, a more detailed description of the validity scales would be helpful. During the course of the test's development, three scales (L, F and K) were constructed and tested, which are useful in separating the valid from the invalid tests. Since the test is a self-report measure, each scale was developed to reveal some type of dissimulation in the self-report. Thinking about the meaning of validity scale variation may have begun to recently undergo change. As will be discussed later, some psychometricians are moving away from a simple "dissimulation" explanation of the scales. Even if this newer hypothesis finds support, the validity of the validity scales rests not only on the known elevations that occur for malingers and persons instructed to fake, but also on the usefulness of the scales in eliminating protocols that would incorrectly predict knowledge of the examinees obtained independently of the test. Determinations

of test validity are typically made by looking at the three scales together, but clearer understanding of the original and possible future understandings of the validity scales can be obtained through an independent discussion of each scale and the research connected with it.

The L Scale: The L or "lie" scale is composed of fifteen items "designed to identify deliberate or intentional efforts to evade answering the test frankly and honestly" (Dahlstrom et al., 1972, pp. 108-109). Endorsement of the items in scored direction indicates "denial of aggression, bad thoughts, weakness of character or resolve, poor self control, prejudices and minor dishonesties" (Dalstrom et al., 1972, p. 110). People who get elevations on L are disowning the presence of certain human aspects which almost everyone regards as bad but which are present in us all. Since the L items are embedded in the matrix of the rest of the MMPI items and are given no special emphasis or standing, it is reasonable to assume that a person who is denying these qualities within himself on the fifteen items is doing similarly on the rest of the items, thereby producing a distorted test result. Since the purpose of the test is to obtain an accurate picture of the person, not a picture of the person's ability to portray himself in good light, a protocol with an elevated L score is considered invalid. On the other hand, a reasonably honest person will generate a protocol with a low L score and more trust can be invested in that research as an accurate depiction of the person's internal world. Research has tested the L scale to see if it measures what it claims to measure and has been supportive (Dahlstrom et al., 1972, p. 110).

The F Scale: The F or "frequency" scale, sometimes called the confusion scale, was "designed to detect unusual responding or atypical ways of answering test items" (Dahlstrom et al., 1972, pp. 112-113).

Dahlstrom et al. (1972) continue to describe this scale, saying:

Low scores on this set of items . . . signify a general conformity to the response patterns of the standardization population over a wide range of experiential content. Conversely, answering a large number of these items in the scored direction reflects a deviation from the normative group in a number of different ways.

Unusual response patterns which result in an elevated F can stem from carelessness, superficial compliance, resentment, wish to appear in need of help, and psychosis (Dahlstrom et al., 1972, pp. 117-119; Newmark, Gentry, Simpson & Jones, 1978). The clinician who is interpreting the profile needs to examine the whole pattern of clinical and validity scales to determine what might be the cause of a particular elevated F. For example, if a moderately elevated F score occurs with an elevated schizophrenia scale score, the most reasonable interpretation would be that the unusual responding was due to a psychotic process. On the other hand, a highly elevated F occurring with an elevated L score would indicate faking or carelessness.

The K Scale: The K scale, sometimes called the "suppressor" factor, was generated out of research directly addressed to a more subtle kind of distortion. It indicates a characterological defensiveness, the sort where a person attempts to put himself in a

more socially acceptable light. A person who answers questions about which he feels somewhat ambivalent in a consistently socially acceptable direction would generate an elevated K. Alternatively, a person could consistently answer in a socially undesirable direction, producing an unusually low K and a profile which would exhibitionistically reveal every foible that has entered his won consciousness. Indeed, moderate elevation on K has been correlated with social adjustment (Campbell, Clarkson & Sensabaugh, 1977). Therefore, a middle range of scores on the K scale, resulting from no consistent attempt to slant one's self presentation in either direction, would be most desirable.

Among a sample of individuals who are dealing with the test in an open and forthright manner, then, there would be little reason to expect that these items would be answered with any uniformity. (Dahlstrom et al., 1972, p. 26)

Research on the validity of the K scale has been of three types. A few initial studies found the predicted relationship between K and defensive subjects (Dahlstrom et al., 1972, p. 125). Later studies examined the impact on the MMPI scale scores when subjects were instructed to fake their answers in one direction or another and found the K scale to be particularly sensitive in discriminating the faked from the genuine profiles (Boe & Kogan, 1964; Cofer, Chance & Judson, 1949; Gough, 1950; Harvey & Siprelle, 1976; Hunt, 1948; Marks & Seeman, 1963; Rappaport, 1958). A third type of research on the K scale has compared the predictive validity of regular profiles with those where the K score has been used as a corrective weight added to clinical scales. This research has been inconclusive (Dahlstrom et al., 1972, p. 127).

Since the K scale is being examined in this study as a measure

in itself, research testing the hypothesis that it is a useful weight for correcting clinical scales is not of particular interest. It is important to distinguish the types of studies, however, since the hypotheses concerning K as an indicator of faking or defensiveness have been supported and since these are the hypotheses concerning K that are of interest to the present study.

F Minus K: Validity scales are often examined not only in themselves, but in the configurations they stand with other validity scales. One configuration, F-K, has received special interest as an indicator of faking. The original F-K research found that profiles with F-K scores greater than nine were often generated by malingerers, whereas those with F-K scores less than nine were not (Gough, 1950). Later research has found that very low F-K scores are produced by persons who wish to appear particularly healthy (Dahlström et al., 1972, p. 170).

One study from research on faking is especially relevant to illustrate the present discussion. Special instructions to "fake" their report in a specified direction were given to undergraduate subjects, completing the test: a) imagining that they were applying for a job (fake-good), b) as if it were an accurate self-perception (control), or c) as if they were applying for psychotherapy (fake-bad). The L scale was unaffected in each group. The K scale was elevated for the job group and the F scale was elevated for the psychotherapy group and reduced for the job group (Harvey & Siprelle, 1976). Not only does this study show the functions of the F and K scores, an F-K score would have been very effective in discriminating the two types of "fakes"

from the control group.

Even though the F-K score is often used to detect faked profiles, no standard cut-off scores are established. It appears that both fake-good and fake-bad cut-off scores vary with the particular sociological and clinical groups so that unique cut-off scores need to be set for each population being tested (Dahlstrom et al., 1972, p. 172).

Non-distortion Interpretations of the Validity Scales

As stated before, the basic premise of the MMPI is that it measures something intrinsic to the person, that an undistorted profile portrays important facts about a person's intrapsychic world. At the same time, the basic premise about the validity scales is that they represent the distortion of an authentic report about a person's intrapsychic condition, distortion stemming from the person that clouds the picture of his psychological world.

The discussion of the meaning of F scale elevations and to some extent the K scale elevations already hints at another possible way of thinking of the validity scale scores: that validity scores not only measure distortion, but that they are also clinical indicators in themselves. A third possible explanation departs even further from the traditional definition of validity scales: that the distortion they measure is produced by environmental, as well as intrapsychic variables. The following discussion will examine both of these hypotheses in greater detail.

Are the validity scales clinical scales as well? One of the previously mentioned explanations for an elevated F worthy of special note

for a drug abuser population is that elevations result from resentment, hostility, or a behavior disorder. One study found that eliminating tests with F scores greater than 16 didn't significantly increase the predictability of alcoholism scores but rather decreased predictive significance to non-significance in one group (Apfeldorf & Hunley, 1976). In this study, the F scale proved to be the best discriminator of alcoholics from disciplinary problems and controls. A second study found that the F scale was best at discriminating disciplinary problems from other types of problems (Zuckerman, Soln, Masterson & Angelone, 1975). The authors pointed out that the data supported Rice's contention that the F score is a measure of overt hostility (Rice, 1968). Another author, examining completed tests of adolescents found support for the same hypothesis (Gynther, 1961).

Other scales seem to have clinical meaning as well. The implications of K for characterological defensiveness and for positive social adjustment have already been mentioned. Complex functions of validity and clinical scales have been found to predict success or failure in rehabilitation of disabled persons (Campbell, Clarkson & Sensabaugh, 1977). F-K has been found to be a useful indicator of psychopathic and character disorders (Gough, 1969).

Even if the validity scales are somewhat useful as clinical measures, it may be that the two types of interpretations can be integrated. For example, the F scale may be sensitive to the acting out of aggressive impulses, especially as they are expressed in a distorted self-report. The K scale may represent a type of defensiveness in social situations. Too little of this defensiveness may indicate a dimin-

ished ability to cope, whereas too much may indicate a rigid guardedness, as well as a suspect profile. The L scale may have its own clinical meaning, as well as implications for the test's validity, which can be elucidated in further research.

Precedent exists in the California Personality Inventory (CPI) for this kind of double meaning of validity scales. The CPI contains three measures, which can be interpreted clinically, as well as in terms of the particular test's validity (Gough, 1975, p. 16).

The present study may shed further light on the meaning of validity scale score elevations. For example, if F elevations are reduced by the special "Don't Fake" instructions which will be discussed later, support will accrue for the idea that the F score is a type of conscious distortion with the drug abuser population. Following a complete presentation of the background and the procedures employed in the present study, the import of the present study for the clinical interpretation of the validity scales can be discussed more clearly.

Refusals to Complete the MMPI

Some subjects might refuse to take the test, rather than attempt to distort the outcome. "Refusals," therefore, could be a measure of covert dissimulation: that is, another form of hiding one's true nature from others.

No research on the MMPI has used refusals as a dependent measure (Dahlstrom et al., 1972; 1975). It would be sensible to use "refusals" as a dependent measure in future MMPI studies.

Environmental Impact on Validity Scores

As stated before, abnormal validity scores are typically regarded as resulting from something within the person. This same hypothesis is typical of much of psychological research. Even though it seems to be a fundamental assumption, one might wonder if it is tenable.

A review of the actual findings of psychological experimentation concluded that variance attributable to the situation and to the situation by person interaction rather than variance attribute to the person account for more of the total experimental variance (Sarason, Smith & Diener, 1975). It seems reasonable, then, to wonder if situational variables might also influence the outcome of MMPI testing. A basic handbook on the MMPI suggests that such might be the case, at least in terms of the interpersonal situation between examiner and examinee:

. . . a subject easily senses a test administrator's attitude, especially as it may be reflected in superficiality or flippancy of manner, and he may respond with similar lightness or with an unwillingness to reveal personal feelings or socially unacceptable reactions to an apparently unsympathetic audience. (Dahlstrom & Welsh, 1972, p. 42)

If examiner variables can influence test outcome, it is not difficult to imagine other situational variables having similar effects. The rest of the introductory discussion will examine the possibility that three situational variables (instructions, group or individual testing and administrator gender) also affect validity scores for the MMPI.

Instructions: Validity scale scores have been repeatedly shown to be related to accuracy and dissimulation of the self-report, as pre-

viously discussed. Yet, much of this research has been devoted to measuring the impact of dissimulating instructions on the validity scores. As Dalstrom et al. say: "Relatively little research has been directed to the possibility for improving test validity through variations in test-taking instructions" (pp. 132-133, 1975). One study gave direct explanations of the derivation and method of scoring the MMPI to college students and found fewer items left omitted and an increase in the range of scores of several of the clinical scales, as compared with controls (Fink & Butcher, 1963). The results were interpreted to mean that the experimental subjects were less defensive. Any other research on the use of special instructions is unknown to this writer and to an expert on the MMPI who keeps a computerized bibliography of MMPI research (Dahlstrom, Note 1).

Variations in instructions have been shown to affect other tests. Instructions to fake-bad on a depression scale get significantly more depressed scores than standard or fake-good instructions (Mikesell & Calhoun, 1969). Fake-good instructions produced higher self-esteem scores on the Coopersmith Self-Esteem Inventory than standard or fake-bad instructions, though the lie scale score could discriminate the faked protocols (Mikesell, Calhoun & Lottman, 1970). One researcher told some of his subjects to complete the Bem Sex-Role Inventory as an accurate self-perception, as opposed to two other groups who were told to complete it as the most masculine or feminine person imaginable (Henrischen & Stone, 1978). Results were that the "masculine" group high masculine scores and the control group intermediate scores. For the purposes of this review, however, the study

was unique in that one set of instructions were out of the "Don't Fake" variety, as opposed to standard, fake-good and fake-bad and that these instructions produced scores significantly different from the other instructional groups.

Special instructions have been shown to strongly affect the Rorschach (Klopfer, Ainsworth, Klopfer & Holt, 1954, pp. 452-458). Telling the patient the test is one of creativity gets different results from describing it to him as a test of intelligence or of psychopathology. In a review of about one hundred studies, the authors claim that Rorschach results depend upon the total situation, but that even indirect suggestive instructions (such as telling examinees that high-status persons see certain things) and stating the purpose differently can produce distinctly different results. Another reviewer cited studies in which subliminal exposure to certain words and cognitive set produced changes in the Rorschach (Bespalec, 1977).

An even more objective measure like digit span can also be affected by instructions. More digits were produced for the same set of numbers under the Wechsler instructions than with the Stanford-Benet instructions (Hagen, Durham & Shannon, 1977).

It is clear that instructions can influence the outcome of psychological tests, even so-called objective tests. It also seems that even though many researchers have studied the effects of instructions to distort, very few have studied instructions to be especially truthful, and apparently none have done so for the MMPI.

Dahlstrom et al. (1975, p. 132) have said ". . . the behavior of a subject taking a personality inventory is controlled to a large but not perfect extent by the test instructions," and have suggested

researchers might profitably look for special instructions that maximize the validity of the MMPI protocols. The present study will begin the search for such instructions by comparing "Don't Fake" instructions with standard instructions. It is expected that the don't fake instructions will generate lower validity scores than the standard instructions. At the same time, it is expected that more subjects will refuse to take the test when they know that dissimulation is likely to be detected, as they will with the don't fake instructions.

Group or individual testing: The MMPI was first developed as a card sort test to be given individually, but later another form was developed which was initially for group testing. Later the later was called the "booklet form" and used for individual testing as well.

One author stated at the time the second form was developed:

It would be helpful to see some good studies of the validity of the two forms; the opinion of the test's authors is that when it is administered as an individual test, the subject considers each item (printed on a separate card) more carefully and responds more truthfully than when it is administered in the group (printed in booklets) and one item closely follows another." (Super, 1949, pp. 70-71)

Such research has been done, finding the two forms to provide comparable results, but the research allowed the test form variable to remain confounded with the situational variable of group vs. individual testing (Dahlstrom et al., 1960, pp. 25-26). No other research has been published that compares the effects of giving the test individually with giving it in groups (Dahlstrom, Note 1).

Other psychological tests have sometimes shown a difference between group and individual testing. If there is a difference, gener-

ally individual testing facilitates performance. The Machover hypothesis concerning self-esteem has been found to be upheld when the human figure drawing task is given individually, but not when given in groups (Wilee & Davis, 1976). Other studies showing the same findings were reviewed by Wilee and Davis, who summarized by saying that group testing inhibited the relationship between the size of the human figure drawn and self-esteem but that in individual testing self-esteem and figure size were directly related. With a measure of feminist and child-rearing attitudes, it has been found that group composition of the test-takers had a significant effect on the outcome (Shomer & Centers, 1970).

Tests of creativity have shown the same result. "Group administration of creativity tests is more stressful than individual and thereby inimical to the production of creativity," say two authors of a prominent creativity measure who have reviewed the experimentation in the area (Wallach & Kogan, 1965). Further support for Wallach and Kogan's contention was found with fifth grade southern Negro school children (Chambers, 1970). In another study, individual testing produced higher creativity scores than did group testing for non-gifted children, but had no effect with gifted children (Milgram & Milgram, 1976).

The well-known failure of a considerable body of research to support interpretive hypotheses of the Rorschach has been attributed by some reviewers to the group testing used in most of these experiments (Klopfer, Ainsworth, Klopfer & Holt, 1954, pp. 148-154). Klopfer et al. say that group, rather than individual testing is a different kind of test and cite evidence in support of the "color shock" Rorschach hypothesis when individual (but not when group) testing is used. However,

other reviewers, while agreeing that group and individual Rorschach testing are different examinations, summarize the research by saying that hypotheses are supported and not supported about equally (though differently) with both methods (Zubin, Eron & Schumer, 1965).

Though there does not seem to be any research comparing group with individual testing with the same procedure and form of the MMPI, Ligon was aware of some possible problems with group testing:

Once group morale is lost, it is very hard to regain. Let there be a few sighs, whistles, groans, shufflings of feet, low-intensity grumblings or catcalls and the situation for group testing is almost hopelessly lost. (1942, p. 398)

Group testing with the MMPI could, like the Rorschach and other tests, be a different test. The MMPI is commonly given both ways, to individuals and to groups. Based on the creativity and human figure drawing research above, it seems reasonable to predict that individual testing produces lower validity scores for the MMPI, since the subject is less distracted and is more assured of his importance in the testing. Fewer refusals could also be expected with the individual testing because subjects would probably perceive more caring and esteem communicated and so would feel more comfortable disclosing about themselves. Research comparing individual with group testing needs to be done to evaluate these hypotheses.

Administrator gender: The sex of the presenter is another situational variable that could well influence the outcome of MMPI testing and warrants investigation. There is no published study on this topic to date (Dahlstrom, Note 1).

As basic as this variable might be, the lack of research is not surprising. One reviewer of 226 empirical personality studies found

that less than half of these studies reported any tests of the gender variable (Carlson, 1971). Carlson cogently argued that by ignoring important personological variables, we are leaving the "person" out of personality research, and she called for much more investigation of the effects of such variables.

Returning to consideration of the MMPI, one notes that the importance of the MMPI examinees' gender has long been recognized. There are separate norms for male and female subjects. Nonetheless, the effect of examiner gender has neither been studied nor considered a variable worth taking into account when the test scores are interpreted (Dahlstrom et al., 1972; 1975).

A general review of experimenter variables in psychological research concluded that the gender effect is not simple (Rosenthal, 1966, p. 42-56). In some studies, it seems that the variable producing the most effect is aggressiveness, rather than sex. In motor tasks, it seems that males obtain a higher level of performance. In marble-dropping tasks, there seems to be an interaction, with females getting better performance with younger subjects, males with older subjects. In sensory deprivation experiments, sexual feelings are disclosed more readily to experimenters of the same sex.

Other experiments with the marble-dropping task have found that opposite sex experimenters obtained a higher level of performance (Stevenson, 1961; Stevenson & Allen, 1964). In a word association task, more pathology and more variability were found when an opposite sex experimenter was employed (McDonald & DeWolfe, 1976). Another experiment on self-disclosure found no main or interaction effect of ex-

perimenter gender but found a curvilinear relationship between self-concept and self-disclosure (Shapiro & Swenson, 1977). A study of feminist and child-rearing attitudes found that sex of administrator had no effect, while group composition and sex of subjects did (Shomer & Centers, 1970).

A few studies have been done on the examiner gender effect with projectives. One study on the Rorschach found no significant effects while another found that male subjects gave more sex and guilt responses to a male examiner as compared to a female examiner when they had spent time in a waiting room decorated with pictures of female nudes (Masling, 1960). Another study found no differences between the Thematic Apperception Test stories of subjects, regardless of gender, on level of plot, mood, outcome of story, or activity (Garfield, Blek & Melker, 1952).

In a classic study that focused on examiner effects in Rorschach testing, nine graduate students in psychology each gave 30 different college students the Rorschach (Saunders & Cleveland, 1965). Significant differences on 20 of the 38 Rorschach measures taken were found across the nine examiner groups. In particular, examiners rated as hostile or anxious by their supervisors and examinees obtained more hostile or anxious Rorschach responses from their examinees. This study, which was carefully designed and controlled, and aimed directly at discovering what effects examiner variables can have on Rorschach output, clearly shows that the examiner can have an effect on his subject's performance.

On the other hand, some of the cited research shows that some-

times a test administrator's gender does not have a measurable effect on his subject's responses. It may be that traits such as aggressiveness, which are often confounded with sexual role, produce confounded results.

One study on an objective test found that female examiners elicited higher Full-scale IQ's, Verbal IQ's, Comprehension, Similarity, and Vocabulary scores from children than did male examiners (Back & Dana, 1977). This result seems reminiscent of the research reviewed by Rosenthal with the marble dropping task, where female experimenters obtained better performance with young children and male experimenters better performance with older children.

In summary, there is limited evidence for four hypotheses concerning experimenter gender effects. Females seem to generate higher performance with young children, whereas male experimenters obtain better performance from older subjects. Opposite sex examiners seem to produce some kind of anxiety which can lead to better performance on a simple task or poorer performance on a complex task with psychiatric patients. Same sex experimenters seem to usually provoke more self-disclosure. Sometimes experimenter gender effects are obtained that upon finer analysis seem more attributable to certain personality characteristics such as aggressiveness rather than to sexual characteristics.

None of the hypotheses mentioned above is solidly documented and furthermore some of the research cited found no measurable effects. On the other hand, there is ample evidence that experimenter gender frequently produces some effect making an investigation of this variable as it concerns the MMPI seem reasonable. In particular, it is ex-

pected that male examiners permit more self-disclosure and elicit less anxiety with male subjects in the complex task of completing the MMPI, resulting in fewer refusals to disclose about self by taking the test and resulting in lower validity scores (more accurate self-disclosure) when the test is taken, than result with female examiners.

Validity Scale Scores and Drug Abusers

As mentioned before, the MMPI is an important clinical tool with drug abusers. In addition, abusers are often described as "con-men" or "manipulative" and so might be expected to attempt to fake their MMPI protocols, especially if financial or other benefits would result from a pathological report (Black & Heald, 1975; Craig, 1979a; Gough, 1969). However, the characteristic MMPI profile of heroin addicts is well known. In a review of 77 studies, the mean T scores of the validity scales were: L = 49; F = 63, and K = 52 (Craig, 1979b). With the possible exception of the F score, these results are not different from the normal population. One study found an average higher F score among addicts than reported for a comparable group of psychiatric patients in another study, but this comparison is far from rigorous, since the data were from two different studies using different patients at different locations (Collins, Burger & Taylor, 1977). Other indirect support of the idea that drug abusers have higher F scores comes from the finding that F elevations are correlated with elevations in the Psychopathic Deviate and Mania MMPI scales, which are commonly higher for drug abusers (Collins et al., 1977; Dahlstrom et al., 1972; Hill, Haertzen & Glaser, 1960). Even if such F ele-

vations are found in the addict groups, it is not clear what they mean, since, as was discussed before, such elevations seem to sometimes indicate resentment, psychosis, or character disorder, as much as a deliberate attempt to fake.

In either case, research on the MMPI validity scales with drug abusers would be especially interesting. On the one hand, their scores are typically much like those of the general population. On the other hand, F score differences among drug abusers subjected to different experimental manipulations may further illuminate the meaning of F scale elevations for this clinical group.

The MMPI and Race

The present study uses subjects which are predominantly black. If blacks were to complete MMPIs differently than whites or other races, one would have to understand the differences in order to properly interpret the results.

Early research found that there were indications of differences: MMPIs tended to portray blacks as more pathological than they were (Dreger & Miller, 1969; Gynther, 1972). A later study showed that this was true for only blacks having less than 12 years of education (Cowan, 1975).

However, other recent research has not supported the hypothesis that MMPI protocols are less accurate for blacks than for whites (Klinge & Strauss, 1976; Shore, 1976). Another study found no differences between the races but between diagnostic categories and argued that the blacks in the earlier research were simply more disturbed than the whites (Davis, 1975). It is well known that psychological disturbance

is negatively correlated with social class (Coleman, 1964, p. 270). Possibly blacks have acculturated or become more sophisticated in recent years because of civil rights progress, accounting for the finding of differences between earlier and later studies.

In summary, there is no clear or convincing evidence that MMPI protocols are different for blacks. Therefore, no special screening procedures seem to need consideration with a predominantly black population.

Pilot Study

Previous to the present study, it was found that 50 percent of the MMPIs administered by a female psychiatrist to groups of drug abusers were returned invalid. About two years later, 26 drug abusers were tested individually with the MMPI, using the "Don't Fake" instructions which are described later in the present study. Only eight percent of these later tests were returned invalid, using the same cut-off criterion.

The pilot work showed that the large number of MMPIs returned invalid could be reduced. However, the reason for the reduction was not clear. Sex of administrator, other administrator personality variables, instructions, number of persons being tested at once, time of testing, and subjects differed between the two groups tested.

The pilot results provided impetus for further investigation into the variables responsible for reducing MMPI invalidity. However, an experimental design which carefully controlled some of these variables in an independent manner was needed before any interpretation could be offered. The following study addresses that need.

Hypotheses of the Present Study

The present study investigates the effects of administrator gender, group vs. individual testing, and special instructions on the MMPI validity scales scores and number of refusals to complete the test for inpatient drug abusers who have volunteered for treatment. Although the subjects and the experimenters will be different, the materials, type of subject, and milieu will be the same as that of the pilot study.

It is hypothesized that lower validity scale scores and fewer refusals obtain with individual testing than with group testing.

Special "Don't Fake" instructions are expected to produce lower validity scale scores and more refusals than the standard instructions.

A male examiner is hypothesized to elicit lower validity scores and fewer refusals than a female examiner for the male subjects of this study.

METHOD

Subjects

A total of 148 patients from a volunteer inpatient drug dependency treatment program at a Veterans Administration Hospital in the Chicago area participated in the study. Each subject was assigned to one of eight treatment groups. Table 1 shows the number of subjects and some demographic variable medians for each group. Table 2 shows the frequencies of the entire sample for another set of categorical demographic variables.

An examination of the second set of discontinuous demographic variables by group showed the distribution across categories to be very similar to that of the total sample. Consequently, only the total sample data are reported in Table 2. As can be seen from the table, subjects were typically about 30 years old, had about 12 years of education, were black, single, and unemployed. Over 92 percent were long-term heroin users, most of whom had been treated before.

Materials

Form R, the "group" form of the MMPI, was given to each subject. An answer sheet was fixed to the back of the booklet and a pencil was supplied.

Instructions

In the present research, subjects were told:

Your doctor has requested that you take a psychological test.

TABLE 1

Some Demographic Variable Medians for Each Treatment Group

Group	n	Age	Edu- cation	Arrests during last year	Number of previous admissions	Months since last discharge	Years of use	Days of treat- ment
1	20	33.5	12.1	0.9	1.9	2.5	10	14.4
2	19	33.3	12.0	1.1	1.9	5.0	11	14.1
3	24	30.5	11.9	0.7	2.0	6.0	10	14.1
4	14	28.5	12.2	1.3	1.3	7.0	8	17.0
5	14	28.3	12.5	1.2	1.2	4.0	10	12.3
6	22	31.5	12.8	1.5	2.0	0.0	9	14.1
7	18	31.0	12.1	0.7	1.7	12.0	10	20.8
8	17	29.9	11.7	0.8	1.1	8.0	10	15.0
Total	148	30.9	12.1	1.0	1.7	5.3	10	14.4

Notes: The treatment conditions associated with each group are listed in the procedure section.

A statistical comparison of variables across groups is shown in Table 10.

TABLE 2

Total Group Frequency Distribution for Nine Categorical Demographic Variables

Race			
Black	White	Ori- ental	His- panic
128	15	1	4

Never Married	Married	Widowed	Divorced	Separated
53	36	2	23	34

Living Arrangements			
Live Alone	With Parents	With Spouse	With Others
43	49	23	33

Employment		
Unem- ployed	Part- time	Full- time
117	8	23

Also Use Alcohol?	
Yes	No
36	73

Abuse Other Drugs?	
Yes	No
73	36

Drugs Used			
Heroin	Other Opiates	Sedatives/ Hypnotics	Other
132	6	4	6

No Use	Once	Once/ Week	2-6 Times Weekly	Once Daily	Multiple Daily
12	1	8	15	19	93

Type of Discharge				
Completed Treatment Regular	Completed Treatment Out-patient Referral	Transfer	Disciplinary	Against Medical Advice Incomplete Treatment
32	69	2	40	40

Note: A statistical comparison of these variables by group is shown in Table 10.

It will provide the staff with information on what sort of person you are, so that your treatment can be addressed more to your own personal needs, rather than just to "patients in general." Upon completion of the testing, you will be offered feedback on what the test says about you, so you may also learn something about yourself, if you desire.

Standard instructions were also employed, plus the instructions to complete all items, answering the ones about which a patient felt ambivalent by whichever alternative (true or false) was mostly characteristic of him. Under the "don't fake" instructions, patients were also told the following:

This test has three different scales which detect different ways the test may be faked, answered at random or otherwise distorted. If you plan to do any of these, please refuse to take the test, since we will detect it and you will only waste your time and ours.

Procedure

Subjects were given the test under one of the following eight different conditions:

Group 1: Male administrator, group testing and standard instructions.

Group 2: Female administrator, group testing and standard instructions.

Group 3: Female administrator, individual testing, don't fake instructions.

Group 4: Male administrator, individual testing and standard instructions.

Group 5: Female administrator, group testing and don't fake instructions.

Group 6: Male administrator, individual testing and don't fake instructions.

Group 7: Female administrator, individual testing and standard instructions.

Group 8: Male administrator, group testing and don't fake instructions.

Presentation of all eight conditions resulted in a completely crossed two by two factorial design. To control for order effects, all eight conditions were presented twice, with the order of presentation in the second block exactly the opposite as that of the first.

Two administrators participated in the study. One was male and the other female. They differed in the following ways: the male administrator was younger (age = 29) and a graduate student doing his internship; the female administrator was older (age = 41), a Ph.D. with several years of experience, and director of the ward. They were both careful, conscientious persons who administered the test according to professional standards described in the handbook (Dahlstrom et al., 1972).

All subjects had volunteered for treatment, and at the time of admission were told that testing might be a part of their treatment. Nonetheless, in the context of the present study, they were informed that they had the right to refuse the testing, without penalty.

Testing was done on a locked hospital ward where only patients

desiring treatment for drug dependency were housed. The ward was generally busy and noisy, but examiners were asked to take the test in special rooms. The group testing was done in a small room which was typically used for listening to music. The examinees sat on couches and completed their test on tables in front of the couch. The door was closed and it was quiet inside during the examination. Individual tests were administered in the patient's own room. The patient typically sat on a chair near his bed and completed the test. When the room was shared, the other patients were asked to remain out of the room until the test was completed. Except for the furniture, the individual and group testing rooms were very similar.

RESULTS

The inclusion of a repetition of each treatment condition in a second block allowed for a fourth independent variable to be included in the analysis. First presentations of a condition were from second presentations so that time had two dimensions, early and late. This made the design a two by two by two by two factorial.

Likewise, a fourth non-independent variable, F-K, was included in the analysis. As was discussed in the introduction, F-K has established some reputation as a measure of "faking good" or "faking bad."

The results were initially analyzed by a four way unequal n's multivariate analysis of variance. In summary, there were four independent variables in the present analysis (time, instructions, group or individual presentation, and administrator gender) and four dependent variables (F, K, L, and F-K). The results of this analysis are presented in Tables 3 through 6. Because the multivariate analysis is an approximation of the F ratio, calculated differently, and not a true F ratio, it has no error term.

Differences in the one remaining dependent variable, the number of refusals across treatment groups are tested using the distribution free Chi-square statistic. Table 7 presents the number of refusals for each of the independent variables and the results of the Chi-square test. Each of the Chi-square tests has non-significant results.

TABLE 3

Portion of Multivariate analysis of Variance with L Score as
Dependent Measure

	<u>MS</u>	<u>F</u> (1,93)	<u>p</u>
Time (T)	.10	.02	.88
Administrator Gender (A)	8.47	1.92	.17
Group vs. Individual (G)	.02	.01	.95
Instructions (I)	2.07	.47	.50
TA	7.30	1.66	.20
TG	.28	.06	.80
TI	3.66	.83	.37
AG	2.91	.66	.42
AI	1.18	.27	.61
GI	.75	.17	.68
TAG	2.18	.50	.48
TAI	.91	.21	.65
TGI	.03	.01	.93
AGI	4.34	.94	.32
TAGI	30.23	6.86	.01

Note: Multivariate analysis of variance is an approximation of a true
F ratio and has no error term.

TABLE 4

Portion of Multivariate Analysis of Variance with F Score as
Dependent Measure

	<u>MS</u>	<u>F</u> (1,93)	<u>p</u>
Time (T)	11.03	.11	.74
Administrator Gender (A)	1.05	.01	.92
Group vs. Individual	7.08	.07	.79
Instructions (I)	235.64	2.33	.13
TA	59.94	.59	.44
TG	274.57	2.72	.10
TI	155.07	1.54	.22
AG	146.69	1.45	.23
AI	210.06	2.08	.15
GI	.35	.00	.95
TAG	185.11	1.83	.18
TAI	28.72	.28	.60
TGI	292.90	2.90	.09
AGI	46.11	.46	.50
TAGI	215.42	2.13	.15

Note: Multivariate analysis of variance is an approximation of a true
F ratio and has no error term.

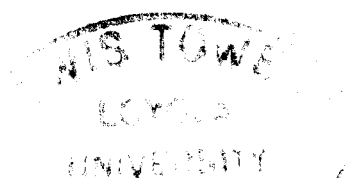


TABLE 5

Portion of Multivariate Analysis of Variance with K Score as
Dependent Measure

	<u>MS</u>	<u>F</u> (1,93)	<u>p</u>
Time (T)	.29	.02	.89
Administrator Gender (A)	6.42	.43	.51
Group vs. Individual (G)	8.95	.60	.44
Instructions (I)	3.91	.26	.61
TA	25.61	1.72	.19
TG	4.54	.30	.58
TI	4.13	.28	.60
AG	13.37	.90	.35
AI	16.67	1.12	.29
GI	3.02	.20	.65
TAG	.39	.03	.87
TAI	2.68	.18	.67
TGI	9.35	.63	.43
AGI	9.03	.61	.44
TAGI	2.12	.14	.71

Note: Multivariate analysis of variance is an approximation of a true
F ratio and has no error term.

TABLE 6

Portion of Multivariate Analysis of Variance with F-K Score as
Dependent Measure

	<u>MS</u>	<u>F</u> (1,93)	<u>p</u>
Time (T)	4.71	.03	.85
Administrator Gender (A)	3.81	.03	.81
Group vs. Individual (G)	29.87	.21	.64
Instructions (I)	281.55	2.05	.16
TA	9.86	.07	.79
TG	323.81	2.36	.13
TI	203.46	1.48	.23
AG	250.37	1.83	.18
AI	330.59	2.41	.12
GI	.33	.00	.96
TAG	154.59	1.13	.29
TAI	50.96	.37	.54
TGI	406.66	2.97	.09
AGI	105.33	.77	.38
TAGI	174.95	1.28	.26

Note: Multivariate analysis of variance is an approximation of a true
F ratio and has no error factor.

TABLE 7

Refusal to Take MMPI as a Function of Independent Variables

	Refused	Took	Total	χ^2	p<
<u>Time of Testing (T)</u>					
Early	17	56	73		
Late	22	53	75	.420	.50
<u>Administrator Gender (A)</u>					
Male	17	57	74		
Female	22	52	74	.557	.40
<u>Group vs. Individual (G)</u>					
Group	16	55	71		
Individual	23	54	77	.681	.40
<u>Instructions (I)</u>					
Standard	15	55	70		
Don't Fake	24	54	78	1.212	.20
Totals for Each Variable	39	109	148		

Notes: The totals for each independent treatment variable are the same because the sample was the same in every case.

The degrees of freedom (df = 1) are also the same for each Chi-square statistic.

None of the experimental hypotheses was supported. Lower validity scores were not obtained with individual testing ($F_L(1,93) = .01$, $p = .95$; $F_F(1,93) = .07$, $p = .79$; $F_K(1,93) = .60$, $p = .44$; $F_{F-K}(1,93) = .21$, $p = .64$). Fewer refusals were not obtained with individual testing ($\chi^2_{(1)} = .681$, $p \approx .43$). Lower validity scores were not obtained with special "don't fake" instructions ($F_L(1,93) = .47$, $p = .50$; $F_F(1,93) = 2.05$, $p = .16$). Fewer refusals were not obtained with special instructions ($\chi^2_{(1)} = 1.212$, $p = .29$). A male administrator did not cause lower validity scores ($F_L(1,93) = 1.92$, $p = .17$; $F_F(1,93) = .01$, $p = .92$; $F_K(1,93) = .43$, $p = .51$; $F_{F-K}(1,93) = .03$; $p = .81$). Nor were fewer refusals obtained with a male administrator ($\chi^2_{(1)} = .557$, $p \approx .47$).

Order of presentation, represented by the "time" variable also produced no significant differences in validity scores ($F_L(1,93) = .02$, $p = .88$; $F_F(1,93) = .11$, $p = .74$; $F_K(1,93) = .02$, $p = .89$; $F_{F-K}(1,93) = .03$; $p = .85$). There was also no significant difference in the number of refusals obtained in the first vs. the second block of presentations ($\chi^2_{(1)} = .420$, $p \approx .51$).

The fourth order interaction was significant with L as a dependent variable at the p less than .01 level. Table 3 shows the mean square, F test and p value for this interaction. There was no hypothesis for this result.

Three other results approached significance. A trend toward significance occurs for the time by group (TG) interaction with F as a dependent measure ($F(1,93) = 2.72$, $p = .10$) and for the time by group by instruction (TGI) interaction with F ($F(1,93) = 2.90$, $p = .09$) and with F-K ($F(1,93) = 2.97$, $p = .09$) as dependent measures.

No other result was significant. Neither did any other result approach significance.

Table 8 presents the mean validity scale scores for the total sample. Average raw scores and the standard T-score conversion are included.

Although subjects were assigned randomly to treatment groups and the assumption was, therefore, that the subjects were equivalent across treatment groups, enough demographic data was available to test this assumption in some detail.

With treatment groups as the independent variable, Chi-square tests compared the various categorical demographic variables across the groups. The results are summarized in Table 9. Table 10 summarizes analyses of variance with treatment groups as the independent variable conducted upon the continuous demographic variables.

Differences among the treatment groups were significant for one of the categorical subject variables: alcohol use ($\chi^2(7) = 19.431$, $p = .01$). Differences among the treatment groups were significant for one of the continuous subject variables since last discharge from a drug treatment program ($F(7) = 2.43$, $p = .02$). There were no other significant differences.

The Pearson correlation coefficient was used to examine the relationship between the months since last discharge from a drug treatment program and the dependent variables of the study; L, F, K, and F-K. The number of cases, 109, was significant ($r_L = .06$, $p = .27$; $r_F = .00$, $p = .49$; $r_K = .13$, $p = .09$; $r_{F-K} = .04$, $p = .36$).

The relationship between the categorical demographic variable

TABLE 8

Mean Validity Scale Scores for the Total Sample

<u>Validity Scale</u>	<u>Raw Score</u>	<u>T-Score</u>
L	3.7	49
F	14.1	76
K	11.3	48
F-K	2.8	

Note: A T-Score of 50 means that 49% of the normal population received a lower score and 49% received a higher score. There is no established T-Score conversion for F-K.

TABLE 9

Chi-Square Tests of Differences in Categorical Demographic Variables
Across Treatment Groups

<u>Variable</u>	<u>Chi-Square</u>	<u>df</u>	<u>p</u>
Race	37.160	28	.12
Marital Status	27.239	21	.17
Living Arrangement	31.371	21	.44
Employment	30.547	21	.08
Drug Abused	41.895	35	.20
Frequency of Use	41.924	56	.92
Reason for Discharge	27.226	28	.51
Alcohol Use	19.431	7	.01
Other Drug Abuse	10.907	7	.14

TABLE 10

F-Tests of Differences in Continuous Demographic Variables Across
Treatment Groups

<u>Variable</u>	<u>F-Ratio</u>	<u>df</u>	<u>p</u>
Age	1.355	7	.23
Education	.717	7	.66
Arrests During Last Year	1.088	7	.38
Previous Admissions	.494	7	.84
Months Since Last Discharge From Drug Treatment Program	2.429	7	.02
Year of First Drug Use	1.355	7	.23
Days in Hospital During This Treatment	1.070	7	.39

that differed for subjects across treatment groups was examined by the Chi-square statistic. Frequency distributions of the L, F, K, and F-K scores were used to find cut-off scores for 10 categories for each variable that would allow for an equal number of scores within each category. Since many subjects received the same score, equivalency could only be approximated. Only seven such categories could be formed for L. The number in each category ranges from nine to 27 for L, nine to 14 for F, six to 16 for K, and seven to 15 for F-K. Since too many of the cells had n's less than 4.5, adjacent categories were combined in the Chi-Square test of K as a function of alcohol use.

A statistically reliable trend approaching significance relating K to alcohol use is obtained ($\chi^2 = 9.45$, $p \approx .052$). Alcohol use was not significantly related to any of the other validity scales.

Table 11 presents the raw data, means and medians for this relationship. As can be seen, drug abusers who also use alcohol had lower scores.

An analysis of covariance was then performed with K as a dependent measure and alcohol use as a covariate. Since no demographic subject variable was significantly related to L, F, and F-K, no analysis of covariance were conducted for those dependent variables. The results are presented in Table 12.

As can be seen by examining Table 12 and comparing it to Table 5. the results of the analysis of covariance were not much different from those of the multivariate analysis of variance. None of the main effects was significant, so that none of the experimental hy-

TABLE 11

Frequencies of Occurrence of Certain K Scores as a Function of
Alcohol Use

<u>K-Scores</u>	<u>Use of Alcohol</u>	
	<u>Yes</u>	<u>No</u>
1 to 7	13	10
8 to 10	4	16
11 to 12	11	13
13 to 14	6	14
15 to 21	5	17
Totals	39	70
Mean	10.41	11.83
Median	11	12

TABLE 12

Analysis of Covariance with K Scale Score as a Dependent Measure and
Alcohol Use as a Covariate

<u>Source of Variation</u>	<u>Mean Square</u>	<u>F-Ratio</u>	<u>p</u>
Alcohol Use	36.04	2.43	.12
Time (T)	.07	.01	.95
Administrator Gender (A)	5.45	.37	.55
Group vs. Individual (G)	2.39	.16	.69
Instructions (I)	14.09	.94	.33
TA	15.02	1.01	.32
TG	3.56	.24	.63
TI	.67	.05	.83
AG	10.39	.70	.41
AI	17.77	1.20	.28
GI	7.41	.50	.48
TAG	.02	.00	.97
TAI	.01	.00	.98
TGI	13.37	.90	.35
AGI	4.38	.30	.59
TAGI	1.33	.09	.77
Residual	14.84		

pothesis was supported. Furthermore, in this analysis, when alcohol abuse was considered as an independent variable, it could be seen that significant differences in K scores do not occur.

DISCUSSION

None of the experimental hypotheses were supported. This means that an important part of the basic premise of the MMPI, that it measures something intrinsic to the person and not environmental influences, is left unchallenged. No significant differences occurred for validity scale scores or in number of refusals to take the test, regardless of manipulations in the type of instructions, differences in administrator gender, number of persons taking the test at one time or the times in the experimental sequence when the test was given.

It is true that the Null hypothesis (in this case that the validity scores reflect internal process and therefore are not open to external manipulations) cannot be proven. It is also possible that other external manipulations not included in the present study might affect validity scores. Nonetheless, it is important to note that the hypotheses of this study represent a strong challenge to traditional thinking about the MMPI and that this challenge was not substantiated.

As the literature review indicated, there was reason to believe that the independent variables of the present study would affect the validity scores. In other studies already cited, instructions, group testing and administrator gender had been strong enough variables to show an effect on certain psychological tests. The present study did not show such an effect of extrinsic variables on the MMPI validity scales, however. It is possible that future research will demonstrate

variables that are not intrapsychic that do significantly influence the accuracy of the MMPI self-report, but clear suggestions of what these variables might be do not emerge, either in the review of the relevant literature or in the present study.

The only significant result, a fourth order interaction, is difficult to interpret in a meaningful way. Not only is there no hypothesis concerning it, the complexity of a four-way interaction is difficult to explain. Furthermore, with the sheer number of statistical tests, one expects at least one significant result out of twenty tests, simply on the basis of chance.

The three results that approached significance are best explained in terms of the pattern of non-significant findings. The one interaction (TGI) with F-K as a dependent variable is not independent of the test on the same interaction with F as the dependent variable, a test that also approached significance, so that the near significance of the former could be primarily attributed to the latter. The other nearly significant results suggest that some complex interaction such as ward attitudes changing towards special instructions and group testing may effect the F score, but a priori research would need to find significant differences in the same direction before any explanation could be attempted. Even if the interaction were significant, the effect is complex, open to many explanations, including intrapsychic interpretations, so that there is still no clear suggestion of external variables that affect validity scale scores.

One might wonder if design flaws caused the lack of findings, but this seems unlikely with the possible exception of the administrator gender variable. The group vs. individual testing variable and

the instruction variable were carefully controlled and presented in a completely crossed factorial design. Not only was the order of presentation of each condition balanced but also in the analysis of variance, time (as measure of the order of presentation) yielded no significant differences. So it seems unlikely that weaknesses in the design or confounds concerning these variables are reasonable explanations of the lack of findings.

The administrator gender variable was one exception. Since, as it turned out, only two presenters were available, administrator gender was confounded with administrator age, professional experience, possible personality differences and other administrator variables. It could be that some administrator variable or combination of them acted in a fashion contrary to the predicted gender effect, so that the effects cancelled.

For example, it has been found that subjects are more receptive to instructions about therapy from a "high status" therapist than from a "low status" therapist (Childress & Gillis, 1977). It may be that the predicted effect of the male experimenter was offset by the fact that he was an intern and that the female experimenter was an older doctoral level psychologist who was the director of the ward.

Regardless, even if other important administrator variables were confounded with gender in a way that their effects cancel, the effect of this confound was limited to the gender variable. The measurement of the effects of the other independent variables was unaffected. Future research on the same topic should use more administrators of both sexes so that gender and other administrator variables can be systematically studied.

Subject variables that cancel, or mask the effects of the independent variables or otherwise spuriously influence the dependent variables are also not a reasonable explanation for the lack of findings in this study. Subjects in different treatment groups differed significantly on only two of the 16 demographic variables. Only one of these, alcohol use, approached a statistically significant relationship to differences in one of the validity scales, K. The analysis of covariance showed that even when variance associated with alcohol use was parceled out, the findings remained essentially the same: none of the hypotheses was supported. The use of random assignment and the careful study of the relationship of subject differences to group assignment and to the dependent variables all strongly point to the same interpretation: it is unlikely that the findings can be attributed to subject variables.

An unexpected finding is that alcohol use in drug abusers is significantly related to a lower K score (see page 49 and Table 11). However, when an F test was performed using raw K scores rather than categories of K scores, and alcohol use was considered a covariate, the result no longer approached significance. Since the analysis of covariance is typically considered a more powerful test than the Chi-square, these results are somewhat troubling. It may be, though, that the relationship between alcohol use in drug abusers and a low K score only holds in a gross sense or that when the precision of the K score is increased beyond that of the global "yes or no" alcohol measure, the relationship seems to weaken. A priori research that uses a more precise measure of alcohol use and that compares tests

of precise measurements with those of global measurements could systematically study the relationship between alcohol use in drug abusers and low K scores. Then it would be more sensible to make interpretive comments.

Although the results clearly do not challenge the part of the basic premise of the MMPI that validity scores reflect internal functioning and not environmental influence, one might wonder at the meaning of the pilot study findings. In the context of the larger body of the present study, the pilot study may represent the limits to the hypothesis that the environment does not influence MMPI output. The earlier testing was conducted by a female psychiatrist untrained in MMPI or any psychological test administration. It may be that she did not provide the minimal conditions for standard testing. The quotations on pages 12 and 15 state how important the authors of the primary source book on the MMPI feel it is for the administrator to create a favorable testing atmosphere. Aside from possible differences in atmosphere, the only other known differences between the groups are represented by the variables of this study and possible subject differences. In the absence of a replication of the pilot result, the only thing that can be said is that future research may reveal the conditions necessary for such gross differences in validity scores. Until that research is conducted, one can only guess that the pilot study signals some limit to the basic premise that the environment does not influence the validity of MMPI testing or that certain subject differences produce great validity score differences.

The second portion of the basic premise, that the internal dynam-

ics that the validity scales measure have been correctly named is more suspect. It is surprising that the don't fake instructions do not reduce validity scores among a population known for its characteristic "conning" (Craig, 1979a). This makes it more difficult to believe that conscious distortion is what is represented in the validity scores.

It may be that conscious distortion is an explanation applicable to more extreme ranges of validity scores. With the exception of the F score, the average validity scores obtained were very much like that of the normal population. As the T scores on Table 9 indicate, the mean L and K scores of the present study fell around the median T scores for the L and K scales for the normal population. On the average, then, the L and K scores obtained in the present study were not extreme.

Of course there were extreme scores that were averaged to reach the "normal" averages, and one might expect fewer high extreme scores with the special instructions. If there had been fewer, the average of the L and K scores for the special instructions would have been lower than the averages for the standard instructions. The results showed that this did not occur, indicating that whatever the validity scores represent, they were not influenced by the special instructions of the present study. One possible interpretation is that the validity scales represent some internal process other than conscious distortion (Apfeldorf & Hunley, 1976; Campbell, Clarkson & Sensabaugh, 1977; Gynther, 1961; Rice, 1968).

It was also expected that the special attention of individual testing would reduce a more unconscious type of distortion: that due

to the feeling that the testing or that one's own unique report was unimportant (Wallack & Kogan, 1965; Wilee & Davis, 1976). Since the individual testing failed to reduce any validity scores, it may be that either the scales do not reflect this kind of distortion of that a moderate range of scores on L and K reflect a relatively undistorted report. Since the administrator gender variable may be confounded, the distortion hypothesis discussed earlier connected with anxiety when identification possibilities between examinee and examiner is reduced cannot be evaluated. Regardless, none of the treatment conditions, whether aimed at conscious or unconscious distortion, affected validity scoring. One reasonable explanation for this is that moderate K and L scores do not reflect distortion, but some other internal process or appropriate levels of characteristics on these scales.

It should also be mentioned that the finding that drug abuser use of alcohol may be associated with lower K scores is consistent with the idea that K measures a type of defensiveness essential to ego-strength (Campbell, Clarkson & Sensabaugh, 1977). Persons who have little defense against the vicissitudes of life might find that the use of heroin or other drugs would not provide them enough relief and so might also turn to alcohol. These same persons, according to the "ego-defense" explanation of the K scale, would obtain low K scores. Therefore, the present finding that alcohol use in drug abusers was grossly associated with lower K scores is consistent with the ego defense explanation of the K scale, even though the present finding did not directly support that explanation.

Among alcoholics, a significant positive correlation between K and denial and a significant negative correlation between K and acknowledgement of dependency has been found (Mozdzierz, Macchitelli, Conway, & Krauss, 1973). However, elevated K scores were associated with leaving treatment against medical advice, an apparently unhealthy level of defensiveness. The group that stayed in treatment had a mean K score of 11.6, about midway between the alcohol and non-alcohol users in the present study. Further examination of K elevations as they relate to alcohol use needs to occur before the relationship can be explained, although these two indicate that such a relationship may exist.

An experiment which employed drug abusers who use alcohol and alcoholics would clarify the matter. Ego-strength and defensiveness could first be assessed independently through some test, such as the Rorschach, or through clinical judgments made by independent raters. Trait defensiveness could further be isolated from state defensiveness by telling some subjects that the MMPI measured how "mentally disturbed" they were. The subjects in this special instruction group could be expected to be more defensive while taking the MMPI, though they might not generally be defensive. The alcoholic group could be compared with the drug-abuser group to discover whether one relationship between K and alcohol use was generalizable to both groups or whether the relationship was more complex. Other designs could also be employed, but some further research would be worthwhile which further examines the relationship that approached significance in the present study between K and use of alcohol by drug abusers.

What the K and L scores represent cannot be directly deter-

mined from this study, even though some of the findings are more consistent with the "clinical" rather than the dissimulation interpretation of the scales at extreme levels. A more full delineation of the types of distortion or other clinical meanings of the K and L scales awaits further systematic research.

More can be said about the elevated F score obtained as an average of the entire sample. As Table 8 shows, the T score for the present population is well above that of the normal population. A comparison with the average F score obtained from many samples of drug abusers ($T = 63$) reveals that the average F of the present sample is even unlike that of the general drug abuser population (Craig, 1979b). However, one study compared drug abusers volunteering for treatment with those not volunteering and found that the F scores of the former averaged at $T = 76$ while those of the latter averaged significantly different at $T = 59$ (Penk & Robinowitz, 1976). Since the subjects of the present study were all drug abusers who volunteered for treatment, the findings of the present study were consistent with those of the earlier study.

Still, why weren't the deviate F scores reduced by any of the treatment conditions? Penk and Robinowitz (1976) described their volunteers as more disturbed and as having committed more infractions of social standards than the non-volunteers. This description fits well with the hypothesis discussed earlier that an elevated F represents a characterological behavior disorder making F appropriate for clinical interpretation in some groups. Furthermore, if an elevated F with some populations represents a character disorder, the elevated F may not represent dissimulation, either conscious or unconscious and

so there might be little reason to expect F scores to be reduced with either special instructions or other special environmental conditions.

Dahlstrom, Welsh and Dahlstrom say of cases like these:

In these instances, elevated F scores are part and parcel of the behavior disorder generating the clinical scale configurations . . . not reflecting adversely upon the dependability of the MMPI protocol itself. (1972, p. 156)

As with the other scales, further research will be needed to clarify the meaning of the F score, even for the drug abuser population. However, the findings of the present study were consistent with a growing body of findings concerning elevated F scores obtained from certain populations.

In summary, the lack of findings in the present study did not challenge the traditional understanding of the MMPI validity scales as measures of internal, rather than situational influences. The results did not clearly support any hypothesis about the meaning of the validity scales, but they were more consistent with clinical, rather than dissimulation explanations of the scales. In particular, the hypotheses that F reflects a behavioral disorder and that K reflects ego defenses received indirect support.

REFERENCES

- Apeldorf, M. & Hunley, P. J. Exclusion of subjects with F scores at or above 16 in MMPI research on alcoholism. Journal of Clinical Psychology, 1976, 32, 498-500.
- Back, R. & Dana, R. H. Examiner sex bias and Wechsler Intelligence Scale for Children scores. Journal of Consulting and Clinical Psychology, 1977, 45, 500.
- Bespalec, D. A. The psychodiagnostic test battery: the effects of sequence of administration on Rorschach variables, Unpublished doctoral dissertation, Loyola University of Chicago, 1977.
- Black, F. W. & Heald, A. MMPI characteristics of alcohol- and illicit drug-abusers enrolled in a rehabilitation program. Journal of Clinical Psychology, 1975, 31, 572-575.
- Boe, E. E. & Kogan, W. S. Effect of social desirability instructions on several measures of social desirability. Journal of Consulting Psychology, 1964, 28, 248-251.
- Campbell, H. G., Clarkson, Q. D., & Sensabaugh, L. L. MMPI identification of non-rehabilitants among disabled veterans. Journal of Personality Assessment, 1977, 41, 266-269.
- Carlson, R. Where is the person in personality research? Psychological Bulletin, 1971, 75, 2-3-219.

- Chambers, N. S. An investigation into group and individual assessment of creativity (Doctoral dissertation, University of South Carolina, 1970). Dissertation Abstracts International, 1970, 31, 1614A. (University Microfilms No. 70-19, 591)
- Childress, R. & Gillis, J. S. A study of pre-therapy role induction as an influence process. Journal of Clinical Psychology, 1977, 33, 528-535.
- Cofer, C. N., Chance, J. E., & Judson, A. J. A study of malingering on the MMPI. Journal of Psychology, 1949, 27, 491-499.
- Collins, H. A., Burger, G. K., & Taylor, G. A. Personality patterns of drug abusers as shown by MMPI profiles. Journal of Clinical Psychology, 1977, 31, 897-900.
- Cowan, M. A., Walkins, B. A., & Davis, E. F. Level of education, diagnosis, and race-related differences in MMPI performance. Journal of Clinical Psychology, 1975, 31, 442-444.
- Coleman, J. C. Abnormal psychology and modern life (3rd ed.). Glenview, Illinois: Scott, Foresman & Co., 1964.
- Craig, R. J. Personality characteristics of heroin addicts: A review of the empirical literature with critique - Part I. International Journal of the Addictions, 1979a, 14, in press.
- Craig, R. J. Personality characteristics of heroin addicts: A review of the empirical literature with critique, Part III. International Journal of the Addictions, 1979b, 14, in press.

- Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. An MMPI Handbook
Volume 2: Research Applications. (Revised edition) Minneapolis:
 The University of Minnesota Press, 1975.
- Davis, W. W. Race and the differential "power" of the MMPI. Journal
of Personality Assessment, 1975, 39, 132-140.
- Dreger, R. M. & Miller, K. S. Comparative psychological studies of
 Negroes and whites in the United States. Psychological Bulle-
tin Monograph, 1969, 70 (part 2).
- Fink, A. M. & Butcher, J. N. Reducing objections to personality in-
 ventories with special instructions. Journal of Educational
Psychology, 1963, 54, 118-122.
- Garfield, S. L., Blek, L., & Melker, F. The influence of method of
 administration and sex differences on selected aspects of TAT
 stories. Journal of Consulting Psychology, 1952, 16, 140-146.
- Gough, H. G. The F minus K dissimulation index for the MMPI. Journal
of Consulting Psychology, 1950, 14, 408-413.
- Gough, H. G. Manual for the California Personality Inventory. Palo Alto,
 California: Consulting Psychologists' Press, Inc., 1975.
- Gynther, M. D. The clinical utility of "invalid" MMPI F scores. Jour-
nal of Consulting Psychology, 1961, 25, 543-548.
- Gynther, M. D. MMPI. In O. K. Buros (Ed.), The Seventh Mental Measure-
ments Yearbook, Vol. 1. Highland Park, New Jersey: The Gryphon

Press, 1972.

- Hagen, R. L., Durham, T., & Shannon, D. Administration of digit span on the Wechsler and Binet: differences that matter. Journal of Clinical Psychology, 1977, 33, 480-481.
- Harvey, M. A. & Sipprelle, C. N. Demand characteristics on the subtle and obvious subscales of the MMPI. Journal of Personality Assessment, 1978, 42, 512.
- Hunt, H. F. The effect of deliberate deception on MMPI performance. Journal of Consulting Psychology, 1948, 12, 396-402.
- Klinge, V. & Straus, M. E. Effects of scoring norms on adolescent psychiatric patients' MMPI norms. Journal of Personality Assessment, 1976, 40, 13-17.
- Klopfer, B., Ainsworth, M. D., Klopfer, W. G., & Holt, R. R. Developments in the Rorschach technique, Vol. 1. New York: Harcourt, Brace & World, Inc., 1954.
- Ligon, Administration of group tests. Educational Psychological Measurement, 1942, 2, 387-400.
- Marks, P. A. & Seeman, W. The actuarial description of personality: an atlas for use with the MMPI, Baltimore: Williams & Wilkins, 1963.
- Masling, J. The influence of situational and interpersonal variables in projective testing. Psychological Bulletin, 1960, 57, 65-85.

- McDonald, R. K. & Dewolfe, A. S. Behavioral and physical effects of experimenter sex. Journal of Clinical Psychology, 1976, 32, 225-227.
- Mikesell, R. H. & Calhoun, L. G. Faking on the Zung self-rating depression scale. Psychological Reports, 1969, 25, 173-174.
- Mikesell, R. H., Calhoun, L. G., & Lottman, T. J. Instructional set and the Cooper-Smith Self Esteem Inventory. Psychological Reports, 1970, 26, 317-318.
- Milgram, R. M. & Milgram, N. A. Group versus individual administration in the measurement of creative thinking in gifted and non-gifted children. Child Development, 1976, 47, 563-565.
- Mozdzierz, G. J., Macchitelli, F. J., Conway, J. A., and Krauss, H. H. Personality characteristic differences between alcoholics who leave treatment against medical advice and those who don't. Journal of Clinical Psychology, 1973, 29, 78-82.
- Newark, C. S., Ziff, D. R., Finch, H. J., & Kendall, J. C. Comparing the empirical validity of the standard form with two abbreviated MMPIs. Journal of Consulting and Clinical Psychology, 1978, 46, 53-61.
- Penk, W. E. & Robinowitz, R. Personality differences of volunteer heroin and nonheroin drug users. Journal of Abnormal Psychology, 1976, 85, 91-100.

- Rapaport, G. M. "Ideal self" instructions, MMPI profile changes and the prediction of clinical improvement. Journal of Consulting Psychology, 1958, 22, 459-563.
- Rice, D. G. Rorschach responses and aggressive characteristics of MMPI F<16 scores. Journal of Projective Techniques, 1968, 32, 253-261.
- Rosenthal, R. Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.
- Sarason, I. G., Smith, R. E., & Diener, E. Personality research: components of variance attributable to the person and the situation. Journal of Personality and Social Psychology, 1975, 32, 199-204.
- Saunders, R. & Cleveland, S. E. The relationship between certain examiner personality variables and subject's Rorschach scores. In B. I. Murstein (Ed.), Handbook of projective techniques, New York: Basic Books, Inc., 1965, 333-353.
- Shapiro, A. & Swenson, C. H. Self-disclosure as a function of self-concept and sex. Journal of Personality Assessment, 1977, 41, 144-149.
- Shomer, R. W. & Centers, R. Differences in attitudinal responses under conditions of implicitly manipulated group salience. Journal of Personality and Social Psychology, 1970, 15, 125-132.

- Shore, R. E. A statistical note on "differential misdiagnosis" of blacks and whites by the MMPI. Journal of Personality Assessment, 1976, 40, 21-23.
- Siegal, M. D. Psychological testing and the detection of drug abusers. The International Journal of the Addictions, 1976, 11, 1031-1043.
- Stevenson, H. W. Social reinforcement with children as a function of CA, sex of E and sex of S. Journal of Abnormal and Social Psychology, 1961, 63, 147-154.
- Stevenson, H. W. & Allen, S. Adult performance as a function of sex of experimenter and sex of subject. Journal of Abnormal and Social Psychology, 1964, 68, 214-216.
- Super, D. E. Appraising vocational fitness. New York: Harper & Brothers, 1949.
- Wallach, M. A. & Kogan, N. Modes of thinking in young children and a study of the creativity, intelligence distinction. New York: Holt, Rinehart & Winston, 1965.
- Wildman, R. W. & Wildman, R. W. II. An investigation into the comparative validity of several diagnostic tests and test batteries. Journal of Clinical Psychology, 1975, 31, 455-458.
- Wilee, C. T. & Davis, S. F. Level of self-esteem and size of human figure drawing as a function of group vs. individual administration. Psychological Reports, 1976, 39, 150.

Zubin, J., Eron, L. D. & Schumer, F. An experimental approach to projective techniques. New York: John Wiley & Sons, Inc., 1965.

Auckerman, M., Sola, S., Masterson, J. & Angelone, J. V. MMPI patterns in drug abusers before and after treatment in therapeutic communities. Journal of Consulting and Clinical Psychology, 1975, 43, 286-296.

REFERENCE NOTES

1. Dahlstrom, W. G. Personal communication, December 28, 1978.

APPROVAL SHEET

The dissertation submitted by Barry Joseph Weber has been read and approved by the following committee:

Dr. Alan DeWolfe, Director
Professor, Psychology, Loyola

Dr. James Johnson
Professor, Psychology, Loyola

Dr. Daniel Barnes
Adjunct Professor, Psychology, Loyola
Director of Loyola Counseling Center

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date

8/14/79

Director's Signature

Alan S. DeWolfe