SOCIAL DESIRABILITY AS A FUNCTION OF SCORING TECHNIQUE: A COMPARATIVE ANALYSIS

By

KENNETH WAYNE YOUNGER II

Bachelor of Arts
Carson-Newman College
Jefferson City, Tennessee
1972

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1974

Submitted to the Faculty of the Graduate College of
the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1975

SOCIAL DESIRABILITY AS A FUNCTION

OF SCORING TECHNIQUE: A

COMPARATIVE ANALYSIS

Thesis Approved:

Barbara J. Weiner
Thesis Adviser

Eugene Bailey

[illegible]

V. Barnes

N. N. Durham
Dean of the Graduate School

PREFACE

The investigation of the error variance in an individual's test score is an important area of investigation in the behavioral sciences. If such influences can be reduced, a more accurate estimate of an individual's true score will be made available. As measurement techniques are refined, the assessment of personality will become a more exact science. The purpose of the present study was to investigate the effects of different scoring techniques on the reduction of error variance in an individual's test score.

I wish to express my appreciation to Dr. Barbara Weiner, who served as Chairperson of my committee, for her encouragement and assistance in the writing of the manuscript. I should also like to thank her for instilling in me an interest in the area of measurement.

Appreciation is also extended to my other committee members, Dr. W. E. Jaynes, Dr. E. A. Weiner, and Dr. T. E. Bailey for their interest and critical reading of the manuscript.

Finally, I would like to thank my parents for their support and encouragement. Most of all, however, I would like to thank Nancy Lynn for being there.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION AND STATEMENT OF THE PROBLEM

It is unfortunate that many psychologists feel that measurement is a branch of psychology, rather than an underlying foundation relevant to all of psychology. As is the case with any discipline calling itself a science, major strides forward are usually contingent upon the advancement and refinement of measurement techniques used in that field. This is especially true in the behavioral sciences. For example, early attempts at assessing an individual's personality were subjective. Earlier assessments were made on the basis of an intuitive process applied by the evaluator. Now, the techniques used are to a great extent more objective and strongly associated with mathematical procedures. The application of these mathematical techniques to the process of evaluation marked a tremendous leap forward in the realm of personality assessment. Hence, advances in the use and application of the mathematical sciences have aided in the making of a more rigorous psychological science.

In 1940, a very significant move forward came out of the University of Minnesota (Hathaway & McKinley, 1940). With the introduction of the Minnesota Multiphasic Personality Inventory (MMPI), objective personality assessment became widespread. One of the major advantages of the MMPI over other inventories being used at the time (e.g.,

Strong Vocational Interest Blank) is that the scales were more "clinically" oriented. Like other objective tests of personality, the MMPI was developed using the empirical technique. The scales were developed by contrasting normal groups with selected clinically diagnosed subgroups. The MMPI has become a model for modern self-report inventories. One of the greatest assets of the inventory is the rich and diversified item pool. The 566 items offer the researcher limitless possibilities for research and further investigation.

As the MMPI grew in use, awareness of the limitations of the measurements made with this inventory also grew. There have been numerous attempts to identify the components which produce error variance in the scores of individuals. Although some progress has been made, there still exist some obscure areas. The chief concern here is that of response bias, or those influences which cause a person to present a distorted picture of himself. The person may be aware or unaware that these distortions are being produced. Cronbach (1946, 1950) makes a further distinction by dividing response bias into two components: response style and response set. Response style refers to a predisposition of the individual to answer in a certain manner without respect to item content. On the other hand, response set is the error variance produced by individuals presenting a false picture of themselves as they respond to item content. In other words, response style is error variance brought in by the individual and unassociated with the test and response set is error variance produced by an interaction of the attitudes of the individual and his responses to item content. Four specific types of response bias have been identified in the literature: (1) social desirability (Edwards, 1957),

(2) acquiescence (Couch & Keniston, 1961), (3) defensiveness (Rosen, 1956) and (4) lying (Hathaway & McKinley, 1951). Acquiescence is a response style, while the other three are response sets.

Receiving the most investigation have been social desirability and acquiescence. Social desirability is the endorsement of items in the socially approved direction as experienced by the individual. The concept comes mostly from Edwards (1957). It is based on the assumption that there is a positive relationship between the judged desirability of an item and the probability that that item will be endorsed (Edwards, 1953a). While social desirability has a relatively short history, acquiescence dates back to 1927 (Fritz, 1927). Acquiescence is the tendency for an individual to be a "yea-sayer" (e.g., respond "true" or agree to items). Several researchers (Peabody, 1961; Husek, 1961; Banta, 1961) have found that acquiescence appears to function in relation to item content. As items become more difficult, ambiguous, and unfamiliar to the subject, there is a definite increase in the probability that the individual will agree with the item. Diers (1964) points out that acquiescence can only be found when social desirability is not influencing the response of the individual. Thus, it appears that social desirability is a greater influence with respect to error variance than is acquiescence.

Although there have been numerous attempts to remove the social desirability component from an individual's test score (Edwards, 1957; Meehl & Hathaway, 1946; Hanley, 1956), a suitable solution has yet to be reached. Most of the solutions have centered around the investigation of the item's content. These solutions have been criticized as inadequate by numerous writers (Horst & Wright, 1959; Comrey, 1958;

Feldman & Corah, 1960; Block, 1965; Rorer, 1965). Hence, another approach or solution is needed. It is the purpose of this paper to investigate several recent developments and integrate them into a feasible solution to the social desirability problem.

CHAPTER II

THE ISSUE OF SOCIAL DESIRABILITY

Introduction

Before dealing explicitly with the issue of social desirability, one is wise to discuss some of the mathematical and historical foundations of the area of personality assessment. This will enable the reader to realistically appraise the present situation.

Although numerous theories of personality appear in the early Greek writings, it was the early humoral theory of Hippocrates that was one of the major theories of personality and bodily function for several centuries to follow (Watson, 1971, p. 13). However, with the exception of Sheldon's theory of constitutionality (Sheldon, 1940), the psychologist no longer looks to bodily function or appearance to assess an individual's personality. There is an ever-growing effort to assess characteristics of an individual's personality through measurement techniques.

What, then, is measurement? S. S. Stevens (1951, p. 2) points out that "in its broadest sense measurement is the assignment of numerals to objects or events according to rules." However, this definition still has its limitations. Using an even broader perspective, measurement can be regarded as the delimitation and fixation of one's own ideas of things. Hence, the problems of measurement merge with

the problems of prediction. It must not be overlooked that a numerical evaluation of things is not the only way of making evaluations, although, at present, it seems by far the best.

Numerical evaluation is the best, because, in addition to the fact that numbers are a universally recognized language, they make possible a refinement of analysis without the loss of clarity and their emotional neutrality permits a symbolic presentations of invariant relations in a domain where there are a manifold of changing qualities. Hempel (1952, pp. 56-57) delineates four advantages of numerical classification over verbal classification. His list is as follows:

> a. By means of ordering or metrical concepts, it is possible to differentiate among instances which are lumped together in a given classification; in this sense a system of quantitative terms provides a greater descriptive flexibility and subtlety.
> b. A characterization of several items by means of a quantitative concept shows their relative position in the order represented by that concept.
> c. Greater descriptive flexibility also makes for greater flexibility in the formation of general laws.
> d. The introduction of metrical terms makes possible an extensive application of the concepts and theories of higher mathematics. General laws can be expressed in the form of functional relationships between different quantities.

Thus, as Torgerson acknowledges (1958, p. 12), measurement, especially in the numeric sense, is an important aspect in the foundation of a science.

The historical development of the application of measurement techniques in personality assessment is of interest, if there is to be a grasp of its present status. As it was previously noted, personality has not always been assessed by questionnaires or self-report inventories. Some of these early techniques, such as astrology and

palmistry, are being revived and enjoying an increasing following. However, there is little, if any, scientific evidence in their favor for them to receive any more mention.

An important distinction must be made between idiographic and nomothetic approaches to personality assessment. An investigator may wish to study a single individual to acquire a detailed knowledge of the specificity and uniqueness of this particular person. This type of research approach is known as idiographic. On the other hand, the researcher may wish to approach the problem in terms of universal principles found in large numbers of persons (nomothetic). Each approach is a viable alternative. However, there is a strong tendency for the researcher to adopt one technique and ignore the advantages of the other. The present author suggests a compromise in which both procedures are used to give a broader foundation to personality assessment. Allowing the individual to respond in a more extended format provides greater consideration of his uniqueness as well as maintaining a nomothetic characteristic in measurement.

The initial application of objective techniques of measurement to psychological research was in the area of skills and abilities. This grew out of early academic interest in individual differences. Oddly enough, this was spurred by astronomers and their interest in the "personal equation." The phenomenon that is responsible for the development of the personal equation is the personal difference that was found among various astronomers in the observation and recording of stellar events. This discovery of individual variation formed an important impetus for later exploration into the use of measurement. This can be seen in the development of the correlation coefficient.

It was Francis Galton's keen interest to uncover the principles of the inheritance of manifest characteristics (Burt, 1962). Also, from his work in inheritance, Galton became familiar with scatterplots and what he termed as the principle of "regression toward mediocrity." In 1869, he formalized that notion into the "index of co-relation" (Galton, 1869). Karl Pearson, fascinated by Galton's attempts to mathematize biological and psychological principles, took this index and developed it into what is known today as the product moment correlation coefficient.

During World War I, the two major movements in personality testing, self-report inventories and projective techniques, had their beginnings. R. S. Woodworth was asked by the United States government to construct an instrument that could be used to weed out men who were not suitably fit to serve in the armed forces. The inventory developed was the Woodworth Personal Data Sheet. At the same time, Hermann Rorschach, a Swiss psychologist, was developing a markedly different approach to personality assessment. The Rorschach test consists of several ink-blots which the individual is to describe and interpret (Beck, 1944). There have been numerous other projective techniques introduced since Rorschach's initial work. These tests are very popular in the realm of clinical psychology, but have received a great deal of criticism because of the difficulties involved in standardizing and validating them.

As previously mentioned, the MMPI produced a significant forward movement in personality assessment. The forerunner of the MMPI was the Strong Vocational Interest Blank (SUIB). This was developed using the empirical method of test construction. Students working

under Strong later migrated to the University of Minnesota. At Minnesota, the MMPI movement began. It borrowed the empirical method of test construction from the SUIB, but changed the criterion groups to clinical populations. Thus, the first clinically-oriented objective measure of personality was developed. However, Cronbach and Meehl (1955) changed the direction of the construction of personality inventories with the introduction of construct validity. As this concept became popular, deficiencies in the MMPI became apparent. The diagnostic classifications were no longer feasible for several reasons. First of all, the groups and the scales corresponding to diagnostic classifications have never been cross validated. A second major complaint emerging at this time was that the normals used came from a very selected population. Finally, the within group variability is as great as the between group variability. There is a considerable overlap between scores of members of the various groups. This would seem to indicate very little predictive ability.

It is evident that a great deal of progress has been made toward the development of an accurate and objective assessment of an individual's personality. However, as the tests themselves improve, a greater awareness of the limitations of the measurements made with these instruments has become apparent. There have been numerous attempts to identify the components which produce error variance in the scores of individuals.

Although acquiescence was being considered as early as 1927, it was not until Cronbach (1946) published the first of his two classic articles that a great deal of interest was generated in what he called "response sets." For him, response set was "any tendency causing a

person consistently to make different responses to test items than he would have made had the same content been presented in a different form" (Cronbach, 1946, p. 491). He felt that response sets should be removed because their effect could not be predicted. For example, he pointed out that response sets could lower or raise the reliability and the validity of the measurement instrument. At this early stage, he was predicting that response sets were in part a function of item ambiguity which later research (Banta, 1961) proved to be true, especially with respect to acquiescence.

Furthering his arguments from 1946, Cronbach (1950) became even more pronounced on the issue. He suggested that these response sets might be variables of personality and not just sources of error variance. However, his most emphatic suggestion with respect to removing such sources from tests has virtually gone unheeded. He recommended that the use of multiple choice item formats could be used as a control for response sets. Most major inventories being used at present are the True/False variety. The most notable exception is the Comrey Personality Inventory (Comrey, 1969), which grew out of a series of factor analyses of the various scales of the MMPI. Cronbach's two articles served the purpose of bringing to light the possible consequences that the meaning of test results might have if response sets go unheeded. In this sense, his articles represent two of the classics in the area of response bias, and in the general discipline of psychology.

Since the appearance of Cronbach's articles, the study of response bias has taken several directions. One of the first distinctions to be made was between the influence of the testee's behavior

and the influence of item content. Meehl and Hathaway (1946) deal only with the former. It is their contention that the subject brings his own tendencies into the testing situation. They feel that the subject has the potential either to "fake good" or "fake bad," depending on the situational circumstances at that time. Their research indicates that a correction can be made in an individual's score by introducing the K scale. The K factor acts as a suppressor variable and can be used as a corrective measure. This will be explored in more detail later.

On the other hand, Edwards (1957) feels that it is the content of the items that elicits the error variance. His suggestion is that a forced choice format should be set up and the social desirability scale value of the two choices matched. Much has been said both in favor and against this procedure, as later arguments will show. A third alternative to response bias was suggested by Rorer (1965). It is his intention to show that the entire issue was without foundation. Although he argues convincingly, the evidence points in the other direction. For example, Jackson's (1968) model of responding to items is being considered as an effective approach to measurement (Rogers, 1971). The process oriented theory of Jackson is shedding light on the theoretical issue of response bias. Many of Rorer's arguments are falling by the wayside with the theory offered by Jackson.

As the study of response bias grew, social desirability became the center of controversy. It was given its impetus by Allen Edwards. Studies of social desirability have produced different methods of test construction, various scoring strategies, and a great deal of controversy. It is the purpose of the next section to sort out the varied

positions and give the reader some understanding of the social desirability issue. From that discussion, the material will be integrated and a solution proposed to the problem of removing the social desirability component from a person's score.

Summary of the Literature

Social Desirability Scale Value Approach (Edwards)

In a series of articles and books, Edwards (1953a, 1953b, 1955, 1957, 1958) creates a movement in psychological testing. Edwards' approach asserts that all statements about people can be placed along a continuum of social desirability. Thus, a social desirability scale value (SDSV) of a personality statement is "its location on the continuum of social desirability as determined by one of the various scaling techniques" (Edwards, 1957, p. 5). The research of Edwards generated three main areas of investigation: (1) the probability of the endorsement of an item and its SDSV, (2) the independence of the Edwards' Social Desirability Scale (ESD Scale) and personality characteristics, and (3) the suitability of the forced choice method as a testing procedure which controls for social desirability.

Numerous diverse results surround the first area of investigation. Edwards (1953a) used 140 items which were judged by 152 judges to determine the SDSV of each statement. He, then, administered the items in an inventory to 140 different individuals. From these two independent groups, a correlation was found between the SDSV and the probability of endorsement (Pearson $r = 0.87$). Several other studies

(Navran & Stauffacher, 1954; Hanley, 1956; Kenny, 1956; Wright, 1957), using the same technique with slight variations, report correlations from 0.82 to 0.92. More recent evidence confirms the high level of correlation obtained at the group level (Messick, 1963; Hanley, 1967). A problem arises when interpretation is made at the individual level. Norman (1967) points out that correlations between group derived SDSV's and the endorsement proportions of another group cannot be applied to inferences concerning individuals.

In order to alleviate this situation, biserial correlations between group SDSV's and an individual's endorsement of the items have been obtained (Taylor, 1959; Boe & Kogan, 1963, 1964). Boe and Kogan (1963) report a much lower correlation, 0.27, than the 0.87 reported by Edwards. Norman (1967) makes two assertions about the discrepancy found in the two procedures. The individual correlations are affected by the items chosen. If a large number of items are neutral with respect to social desirability, then this would minimize the chance of observing a relation between the two. Secondly, the differences can be attributed to the fact that the correlations are measuring different relationships and should not be expected to be equal. Messick (1963) suggests that the individual correlation be indicative of the influence of social desirability on an individual when responding to test items.

The two preceding procedures of correlation still fail to take into account individual differences in what is desirable. Although biserial correlations between an individual's judgements and responses tend to be higher than the correlation of group judgements with individual responses (Messick, 1963; Boe and Kogan, 1963), there exists some evidence that there is little difference in what is

considered desirable (Wiggins, 1966; Messick & Jackson, 1961). Wiggins (1966) shows that a factor analysis of desirability judgements yields one excessively large factor. This would seem to indicate very little disagreement within an individual in the judgement of what is socially desirable.

The second major area of controversy generated by Edwards' research is the ESD scale and its independence from personality influences. The ESD scale was developed under the assumption that social desirability varies among individuals and needs to be determined. The original scale consisted of 79 items taken from the MMPI. These were reduced to 39 on the basis of the greatest differentiation between high and low scores. It should be noted that articles before 1957 (e.g., Edwards, 1953b; Fordyce, 1956) use the longer version of the ESD scale in their research.

TABLE I

A COMPARISON BETWEEN ESD CORRELATIONS WITH THE MMPI SCALES AND A PSYCHOTIC FACTOR

MMPI Scale	ESD Correlations	Psychotic Factor
F	-0.82	-0.67
K	0.69	0.70
Hs	-0.70	-0.53
D	-0.69	-0.48
Pd	-0.57	-0.70
Mf	-0.21	-0.38
Pa	-0.60	-0.63
Pt	-0.86	-0.94
Sc	-0.91	-0.92
Ma	-0.50	-0.62

Source: Fordyce (1956)

Edwards (1957, p. 31) considers social desirability as measured by the ESD scale a personality trait independent of other personality traits. A question has been raised in the literature as to the feasibility of such a statement. DeSoto, Keuthe, and Bosley (1959) conclude that the ESD scale is a measure of social well-being rather than social approval. Fordyce (1956) has noted a marked similarity between loadings on the largest MMPI factor from Wheeler, Little, and Lehners (1951) and the correlations of MMPI scales and the ESD scale (refer to Table I). Wheeler et al. identify two major factors, one neurotic and the other psychotic. It is the latter that is referred to by Fordyce.

Crowne and Marlowe (1960) point out that when items are drawn from the MMPI, it is obvious they have scalability for social desirability, but the items also have pathological implications. It is this point that Block (1965) further develops. He suggests that a substantive interpretation of the ESD scale be made. Making such an interpretation, Block feels that the ESD scale is reflecting what he calles "ego-resiliency." Edwards (1970) offers evidence which complicates Block's position. Edwards reports the construction of several rational SD scales from an item pool of more than 2,500 non-pathelogical items. A factor analysis of these scales with the original ESD scale taken from the MMPI items yielded one common factor. Thus, if Block's assertion is correct, these other scales must also be relevant to ego-resiliency; however, as Edwards points out, these scales have a low correlation with Block's scale.

Although Edwards substantially defends his point of view, there is still the question of item overlap. Crowne and Marlowe (1960), constructed a social desirability scale (MC-SD) with the intent to

eliminate the presence of pathology in the items. Basic to their approach is the delineation of items which are socially acceptable, but indicate behaviors which are unlikely to occur. Hence, social desirability refers to the need of an individual to obtain approval by responding in a socially acceptable manner. Comparing the MC-SD scale with the ESD scale on the basis of correlations with the MMPI scales, one finds that higher correlations are found with the ESD scale. This would seem to indicate that the Edwards approach does have a problem of overlap. According to Crowne and Marlowe, their results suggest a theory that social desirability accounts for some of the variance in the MMPI, but not all of it.

Does pathology account for the differences in these two measures of social desirability or are they actually measuring different aspects of a single construct? Messick (1960) found several dimensions of desirability in a factor analytic study. He concludes that ESD leaves much of the consistent variance which is attributed to individual differences in the perception of desirability uncontrolled. Rosen (1956) said there must be a distinction between personal and social desirability. Furthering Rosen's contention, Wiggins (1959) described the differentiation between the tendency to endorse certain desirable items which exhibit large shifts under desirability instructions and the tendency to endorse other desirable items which presumably reflect a group norm.

In a factor analytic study by Wiggins (1964), the two SD scales loaded on separate factors which indicates different dimensions as proposed by Messick. Edwards' SD scale loaded on the factor resembling the first factor of the MMPI (Wheeler et al., 1951), while MC-SD

loaded on what Wiggins called the "role-playing" factor which was first identified by Edwards, Diers, and Walker (1962). It should be concluded that there is a difference among stylistic response measures of social desirability. Whether the differences in social desirability scales can be attributed to a difference in pathology content of the items remains to be seen.

The final controversy stirred up by Edwards centers around the Edwards Personal Preference Schedule or EPPS (Edwards, 1953b). The statements appearing in the inventory were paired on the basis of their SDSVs. The individual chooses which statement of the pair is more descriptive of himself. Edwards (1957) cites numerous studies reporting a favorable reduction of social desirability when using the forced choice format. However, Feldman and Corah (1960) present findings which suggest that social desirability is not minimized by the forced choice technique. They even suggest that the individual's sensitivity to social desirability is heightened in order to discriminate between the two items.

Horst and Wright (1959) in a study of the reliability of the forced choice and ipsative rating procedure found that the forced choice technique reduced reliability. Additionally, Heilburn and Goodstein (1959) found a greater increase in the influence of personal desirability in the forced choice format. On the basis of these findings and those of Feldman and Corah, one can conclude that personal desirability is not reduced in the forced choice format. This would account for a loss of reliability, since an individual's personal desirability can be a function of experience. On the other hand, Levonian, Comrey, Levy, and Proctor (1959) conclude that social

desirability is controlled, but there is a definite loss of validity and reliability when using the forced choice format.

Several conclusions are apparent from the discussion of the Edwards approach to social desirability. The concept of social desirability is multi-faceted. The evidence is very conclusive that the various measures of social desirability do not have convergent validity. This indicates that a single measure cannot be used to make broad inferences. In addition to this there are two specific problems created in the way Edwards handles the problem. The scale construction used by Edwards is on weak foundations due to the conflicting evidence concerning the probability of endorsement and its correlation with the SDSVs. This does not imply that psychometrically Edwards is wrong. It simply means that one of his basic assumptions underlying the construction of the scale is not tenable. The second problem is the way Edwards controls for social desirability in the testing situation. It is evident that the forced choice technique accomplishes very little, while reducing the validity and reliability of the scores.

Correction for Response Style (Meehl and Hathaway)

Although Edwards is the most prominent figure in the social desirability issue, two other approaches have been suggested. The suggestions of Wiener (1948) and Hanley (1956) were to develop inventories in which the statements are subtle and neutral to social desirability. The other approach has been emphasized by Meehl and Hathaway (1946) which involves the use of a correction factor to account for the

deviation of the scores due to response styles. Of the two, Meehl and Hathaway's approach seems to be the most popular, and frequently MMPI scores are reported and interpreted in the light of this correction factor.

Meehl (1945) presents an algebraic foundation for the existence of suppressor variables. Following these guidelines, Meehl and Hathaway (1946) derived the K scale as a correction factor for improving the discriminability of the MMPI scales. The K scale was developed by an item analysis of responses by normals obtaining normal scores and responses of deviates with normal scores. Items discriminating between these two groups are assumed to measure the defensiveness of the individual taking the test. Correlations of the K scale with the other scales indicated a differential effect of the K scale in the other scales. The actual correction figures are given by McKinley, Hathaway, and Meehl (1948).

There is a controversy over the exact nature of the K scale. Edwards (1953b) found that K correlated with the ESD scale at 0.89. Hence, he concluded that the K scale was a measure of social desirability and not a measure of defensiveness. An important distinction should be made at this point. Meehl and Hathaway are more interested in the testee's behavior rather than the specific item content. There are only five questions overlapping ESD and K (Edwards, 1957, p. 32), thus the item content may be substantially different. Hanley (1956) notes that the K scale has over twice as many neutral items than the Schizophrenic (Sc) scale. Edwards (1957) feels that test taking attitude consists largely of the tendency to endorse socially desirable items and reject socially undesirable ones. He concludes that the ESD scale should be used as a correction factor instead of the K scale.

However, it appears that both Edwards and Meehl and Hathaway may be wrong in their respective viewpoints.

Supporting still another more recent view, Fricke (1956) found that psychotic items were keyed true most of the time and neurotic items false most of the time. This would indicate that there is some relationship between personality and response style. In a follow-up study on the nature of the K scale, Smith (1959) found that individuals with low insight were more defensive; however, the K scale did not reflect this relationship. He concluded that the K scale was inadequate when using normal individuals. There is a definite trend in the literature for this position. On the same line as Smith, Heilburn (1961) presented some very important results and recommendations. He found that with maladjusted individuals the K scale measured defensiveness, but with normals there is no such evidence. Also, he found a significant sex difference. Females' K scores were less valid than their male counterparts. From these results, he makes three recommendations with respect to the use of K-corrected MMPI scores. When testing maladjusted males, the K-corrected scores are advantageous, but in testing adjusted females they have no significance. For adjusted males and maladjusted females, the K-corrected scores are less useful, but still worth some consideration. One of the most significant studies with respect to the K scale is Comrey's (1958) factor analysis of the items comprising that scale and certain marker variables. He concluded that the K-corrected scores are useless and that the K scale does not measure defensiveness. The major factors appearing in the study were cynicism, euphoria, and shyness. The evidence for the use of the K scale is clouded by these studies. The main

conclusion reached by the present author is, while it is almost automatic to report K-corrected scores to aid in the removal of response style variance, the validity of these scores is very questionable.

Subtle/Obvious Approach to Social Desirability (Wiener)

It is apparent that the Edwards forced choice method and the Meehl and Hathaway K correction factor do not suitably solve the problems raised by the social desirability issue. The final method used to deal with this controversy was introduced by Wiener (1948). It was his intention to develop inventories which are subtle or neutral with respect to response style. Hanley (1956) and Fricke (1957) concur with the Wiener approach. The evidence presented in these three articles gives strong support for the use of subtlety as a procedure to account for response style influences. Seeman (1952) points out that subtlety is a viable construct in the construction of personality inventories. However, as Wiener points out, there are several undesirable properties inherent in the use of subtle items which must be taken into account. He shows that the K scale has a number of subtle items comprising it. Therefore a question arises because the K scale does not represent a clear solution to the problem of social desirability. Another dilemma pointed out by Wiener is that if all items were subtle, there could be no discrimination in the items with respect to abnormal populations. Obvious items are needed. Thus, a problem is encountered. One can use subtle items to eliminate response bias variance and reduce the ability of the items to discriminate between normal and abnormal populations or include obvious items

and increase response bias influence. Although the use of subtle items seems to be the most feasible approach to social desirability, there still remains some concern as to its effectiveness.

Need for Approval (Marlowe-Crowne)

Recalling the MC-SD, a fourth approach to social desirability can be presented. Crowne and Marlowe (1960) point out that when items are drawn from the MMPI, although they might be scaled for social desirability, the items are related to pathology. This is the result of the items also belonging to the various clinical scales. Crowne and Marlowe offer a different approach. Their approach was the delineation of items which are socially acceptable, but indicate behaviors which are unlikely to occur. Hence, social desirability refers to the need of an individual to obtain approval by responding in a socially acceptable manner. Mosher (1965) and Schill (1966) report evidence to support the approval motive as developed by Crowne and Marlowe. However, a question still remains. Does the approach of Crowne and Marlowe relate to pathology, even though the items are derived rationally?

Stone (1966a) offered evidence that supports a contention that the Marlowe-Crowne SD scale shares a higher relationship with MMPI items than was initially believed to be the case. In a follow-up study (Stone, 1966b), he showed that a restriction of range had been producing low correlations of the MMPI and MC-SD scale. Additional proof for the overlap with pathological content of the MMPI and MC-SD has been offered by Spilka, Horn and Langerderfer (1966). In a factor analysis of several measures of social desirability, they found that there was no single attribute of social desirability. In other words,

there are several different components of social desirability. The factor they labelled "self-sentiment" did not contain either the ESD or the MC-SD. From these results, they concluded that both the ESD and the MC-SD have some overlap with the clinical scales of the MMPI.

Threshold Theory of Responding (Jackson)

These approaches deal mainly with item content as opposed to response technique to resolve the social desirability issue. With the introduction of the threshold theory of responding (Jackson, 1968), a new alternative seems more feasible. Jackson's theory is a combination of the classical psychophysical approach and modern scaling techniques. The threshold theory focuses on the delineation of salient parameters of responding and the relationship these parameters have to the probability of a given response. Fiske (1968) suggests the application of characteristic curves for the analysis of individual responses, but does not provide the elaboration proposed by Jackson. These curves must not be confused with the traditional item characteristic operating curves introduced by Lord (1952). Subject operating curves represent the relation of one subject between the salient item responding parameters and the probability of response. Instead of giving one item to numerous individuals (Lord, 1952), the curves suggested by Jackson are constructed by giving one individual numerous items. The threshold represents the crucial level on a certain parameter at which there is a change in the individual's response from "True" to "False" or "False" to "True."

Damarin (1970) suggests that item characteristic curves may be adapted to personality items. Similarly, Rogers (1971) demonstrates

that subject characteristic operating curves can be applied to personality items and can account for the appearance of response bias in an individual's score. While the threshold theory remains only a theoretical perspective, there is some evidence that it can account for the results obtained by others with respect to the social desirability issue (Rogers, 1971, p. 56).

In a similar approach, Kuncel (1973) suggests that the occurrence of inappropriate response processes should be related to the nearness of the subject and the item. When the individual and various items are scaled on the same continuum, the typical subject seems to have a threshold where his reasons for responding "True" counterbalance his reasons for responding "False." More exactly, the probability of a true response is equal to the probability of a false response.

It would seem that the best approach to the social desirability issue is no longer found in the analysis of item content, but in the analysis of the response process of an individual. If an individual's threshold can be identified, then the exact nature of the individual can be pinpointed. It is on this problem that the rest of the paper will be focused. Can social desirability influences be reduced by something other than item content?

A Proposed Solution

In developing a solution to the social desirability problem, the various evidence supporting the diverse approaches to response format and item content and subsequent scoring of items must be accounted for by the final result. Such a solution does seem possible. That

solution is the Differential Weighting System DWS). The logic of that system will be discussed in this section with later chapters relating an evaluation of its effectiveness.

The DWS is founded on two practical suggestions and a theoretical development. The two practical suggestions are made by Cronbach (1950) and Fricke (1956) and the theoretical development used in the development of DWS is proposed by Jackson (1968). Cronbach recommended the use of multiple choice tests for the control of response set. This is based on the fact that the subject's actual response might be more likely in a set of five alternatives than in a set of two. The multiple choice format seems to be consistent with Jackson's threshold theory. Fricke suggested that subtlety is a key factor in reducing response bias, especially the social desirable influences.

Upon using the threshold theory as one's model, there is an obligation to do away with the two choice system. The use of true/false becomes ineffective because the response is then based on other parameters, if neither alternative is feasible. Two choices greatly restrict the probability of response. For example, if the individual's probability of response for each of the two alternatives were 0.1 respectively, that would leave 0.8 as the probability of response for an alternative not available. In such a circumstance, the controlling factor becomes an outside parameter, very probably social desirability (Rogers, 1971). If in this example the most socially desirable response were "True" for the individual, then that would probably be the response given. Hence, social desirability becomes the important feature of the response procedure.

Using multiple choice or expanding the alternatives alleviates the influences of response set to some extent (Cronbach, 1950). Continuing with the same example, but expanding to a five choice response format, the probabilities could be distributed as follows: never true, 0.1; seldom true, 0.5; sometimes true, 0.1; often true, 0.1; and always true, 0.1, with 0.1 being left for other influences. Since seldom true now has the highest probability of response, it will most likely be the individual's response. Thus, seldom true is quite different from the true response made with the two choice response format when the influences of other factors played a greater role. It has been shown that five to seven alternatives is the most optimal range (Torgerson, 1954). Thus, the use of expanded choices allows for a more accurate identification of an individual's real response.

Expanding the alternatives creates another problem, while solving one. The problem of keying these expanded alternatives is introduced. According to Fricke (1956), subtlety is one of the best solutions for response influences, but Wiener (1948) emphasizes that obvious questions are needed to discriminate the abnormal population from the normal population. However, if the subtlety is in the keying, the obvious questions can be retained. Ryan (1972) pointed out that by using a dichotomous keying technique with expanded alternatives, response biases can be eliminated. He felt that there would still be some deviation, but the influences would rarely cause a move across a cutting point. He offered evidence as to the effectiveness of this technique, but Burt (1956) offers mathematical evidence that differentially weighted items increase subtlety and validity in a test.

Hence, it would seem feasible that differentially weighting response alternatives would work just as well.

Integrating the preceeding arguments, DWS uses a five choice response format with decreased weighting of the alternatives as they move away from the keyed direction. This method is to be tested against the traditional true/false scoring procedure and a dichotomous procedure on the expanded alternatives. The dichotomous procedure will use the five alternatives but will reduce them to a two point scoring procedure. The test instrument to be used is the MMPI, since a great deal of information is available in order to check how well social desirability is being controlled.

Hypotheses

The problems with which this study is concerned are embodied in the following questions: Does scoring technique affect the influence of the response bias known as social desirability? If so, which technique has the greatest effect on reducing the influences of social desirability? In answering these questions, three general hypotheses and several specific hypotheses have been developed.

General Hypotheses

I. Differential weighting, dichotomous splitting, and traditional keying differ in the amount of social desirability found in the scores produced by each scoring technique.

Factor analysis was used to test this hypothesis. Two different procedures were used: (1) a principal component solution with the

first three factors rotated to simple structure using the procedure, (2) a multiple group factor solution based on a hypothesized solution.

II. Traditional scores, K-corrected and non K-corrected, DWS scores, and dichotomous scores differ.

This hypothesis was tested by a repeated measures analysis of variance. It was expected that a significant difference between the scoring techniques would occur.

III. DWS when compared to traditional keying and dichotomous scoring normalizes the shape of the distribution.

This hypothesis was tested based on information obtained from the second, third, and fourth moments of the distributions for each scoring technique. From this information, tests of skewedness and kurtosis were made.

Specific Hypothesis

It was predicted that (1) DWS would remove social desirability from the first factor of the MMPI, (2) DWS scores would be slightly higher than the other techniques, (3) social desirability would be equivalent for males and females, and (4) scores found through DWS would have less skewedness and kurtosis than scores obtained by the other scoring techniques.

CHAPTER III

METHOD AND PROCEDURE

Subjects

All subjects were students at Oklahoma State University enrolled in courses in Introductory Psychology, Elementary Lab Psychology, Introduction to Research Methods in Psychology, or Individual Differences during the Fall of 1974. Although partial data were obtained on 102 subjects, complete data were available on only 84. The sample consisted of 42 females and 42 males. Due to the different MMPI norms for females and males, sex had to be accounted for in this study.

Tests

Minnesota Multiphasic Personality Inventory

The 399 item version of the MMPI questionnaire requiring the individual to respond to items which he feels describe his feelings and attributes was used. The scales scored for this study were the following: L, F, K, the validity scales; Hy, D, Hs, Pd, Mf, Pa, Pt, Sc, Ma, the clinical scales. (See Appendix A for information on acquiring a copy of the items.)

Edwards Social Desirability Scale

The ESD is a 39 item questionnaire requiring the individual to

respond to items which he feels describe his feelings and attitudes. It was included in this study to serve as a marker variable in the factor analysis. This scale measures social desirability on the basis of endorsing statements with a high SDSV. Edwards (1957) presents evidence for the validity of the test. A listing of the items appears in Appendix A.

Marlowe-Crowne Social Desirability Scale

The MC-SD is a 33 item questionnaire requiring the individual to respond to items which he feels describe his attitudes. It was used as a marker variable in the factor analysis and served a similar function as the ESD scale. This scale measures social desirability on the basis of a need for social approval. Crowne and Strickland (1961) present evidence for its construct validity. A listing of the items appears in Appendix A.

Administration of the Tests

The tests were administered as a single inventory with the items of the three tests being mixed. This inventory was administered during one testing session for each subject. The inventory was group administered to 11 to 16 subjects at each testing session. Groups were counter-balanced so that half of the subjects received the five alternative response format first and the other half received the two choice response format first. The five choice response format was (a) Always True, (b) Often True, (c) Equally True and False, (d) Often False, and (e) Always False. The two choice response format were true and false. The instructions for each group are in Appendix B.

Answers were marked on a special answer sheet with both sets of alternatives listed by each item number. A copy of the answer sheet also is given in Appendix B. The subjects were given two hours to complete the task, but could leave whenever they had finished.

Scoring of the Tests

Four different techniques were used in scoring the data. Two traditional scoring procedures (with and without K-correction) were used for the two choice alternatives, while DWS and dichotomous splitting were applied to the five choice set of alternatives.

Traditional Keying

The norms used in this study were the norms presented for college students by Dahlstrom, Welsh and Dahlstrom (1971). Five of the clinical scales are K-corrected: Hs, Pd, Pt, Sc, and Ma. These scales have two scores reported, one without K-correction and the other with K-correction. The remaining scales have only one score reported when using the traditional keying technique (non K-corrected).

Dichotomous Splitting

The major difference between this procedure and the traditional method is found in the number of alternatives to which a subject responds. For dichotomous splitting, five alternatives were available to the subject. The two alternatives on each side of the middle were considered the same response when scoring. The middle alternative was randomly keyed in either the true or false direction. This randomization was restricted such that 50 percent of the items comprising a

scale had the middle alternative of the five response alternatives placed in the keyed direction. Hence, the five choices were condensed to two categories: false and true. The scoring then proceeded in the traditional manner; however, no K-correction was made.

Differential Weighting System

This was the final scoring procedure used to score the data. It used the five alternative format with each alternative being weighted in some manner. The weights of the alternatives were dependent on the direction in which the item was keyed. The weights were four, three, two, one, and zero. They were assigned in such a manner that there is a decrease in the weighting as the alternatives move away from the keyed response. The final score for each scale was then divided by four in order to transform the DWS scores into a raw score numerically comparable to the other scoring methods' raw scores. All four types of scores were then referred to a norm table (Dahlstrom, Welsh & Dahlstrom, 1971) for conversion to T scores.

Analysis of the Data

Several statistical techniques were used to analyze the data. Prior to the implementation of these techniques, simple descriptive statistics were obtained. These were obtained for two reasons. One was to provide some initial insight into the relationships in the data and the other was to provide a basis upon which further computations could be performed.

Factor Analysis

Two different factor analytic techniques were applied to the matrices of intercorrelation obtained in the initial computations: (1) a principal components extraction rotated to simple structure using the Varimax procedure, and (2) a multiple group extraction based on a hypothesized factor pattern.

Hotelling (1933, 1936) developed the principal components method of factor analysis. The major intention of Hotelling was to propose a method which did not have to deal with the problem of communality. The analysis of a set of variables into their principal components employs the following model:

$$
\begin{aligned}
X_1 &= a_{11}S_1 + a_{12}S_2 + \ldots + a_{1p}S_p \\
X_2 &= a_{21}S_1 + a_{22}S_2 + \ldots + a_{2p}S_p \\
&\vdots \\
X_m &= a_{m1}S_1 + a_{m2}S_2 + \ldots + a_{mp}S_p
\end{aligned}
$$

where S_p, $p = 1, 2, \ldots, p$, are the first p principal components and the coefficients a_{ij}, $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, p$, are scalars which represent the weighting of the variables on each component. It is important to note that m components are required to reproduce the correlation matrix of m variables exactly. This technique extracts the components in such a manner that each new component accounts for the greatest possible portion of the total variance of the variables unaccounted for by the preceeding components.

The number of components that are reported was based on the systematic factor variance. The criterion for inclusion was that the

systematic factor variance had to be greater than 1.00. This was a necessary but not sufficient condition. These components were then rotated to simple structure via the varimax criteria (Kaiser, 1958).

A second factor analytic technique used to analyze the data was the multiple group method (Thurstone, 1947). The multiple group procedure is a special case of the rank reduction theorem. A binary matrix, Q, was arbitrarily constructed based on the hypothesized relationship of the variables and the information obtained from the correlation matrices. Younger (1974b) provides a program for computing the multiple group solution based on the input of the Q matrix.

Analysis of Variance

Nine 2 by 3 repeated measures analyses of variance were computed in order to compare the scores obtained using the different scoring techniques. Sex (male vs. female) was the A factor and scoring technique (traditional non K-corrected vs. dichotomous splitting vs. differential weighting) was the B factor and represented the repeated measure. The nine dependent variables were MC-SD, L, F, K, D, Hy, Mf, Pa, and ESD.

Five 2 by 4 repeated measures analyses of variance were computed in order to compare the scores obtained using the different scoring techniques. Sex (male vs. female) was the A factor and scoring technique (traditional non K-corrected vs. traditional K-corrected vs. dichotomous splitting vs. differential weighting) was the B factor and represented the repeated measure. The five dependent variables used in these ANOVA's were the MMPI scales which can be K-corrected (Hs, Pd, Pt, Sc, and Ma).

Distribution Comparisons

The final method of analysis was the description of the distributions formed for each scale in the analysis by the various scoring techniques. This was accomplished by computing the second, third, and fourth moments of each distribution. These were then used to obtain measures of skewedness and kurtosis for each distribution (Snedecor, 1956). Tests for divergence from normality for each measure were computed (Snedecor, 1956).

CHAPTER IV

RESULTS

The results of the statistical analysis of the data are presented in this chapter. This material is divided into four principal sections: (1) simple descriptive statistics, (2) factor analyses, (3) analyses of variance, and (4) distribution comparisons.

Simple Descriptive Statistics

In this section the means and standard deviations of the variables for each scoring technique (traditional non K-corrected, dichotomous splitting, differential weighting) are given. This information is broken down by sex. Tables II and III contain the means and standard deviations for each scoring technique of the men and women respectively. It should be noted that the traditional K-corrected technique is not included in this section since only five scales are scored in this manner.

The intercorrelations of the variables by sex and scoring technique are given in Tables IV through IX. Since sex differences were observed in the intercorrelations, intercorrelations for the total sample were obtained only for the traditional scoring technique. These intercorrelations are presented in Table X and their use is limited to comparison with previously published results, since this is what is typically reported in the literature.

TABLE II

MEANS AND STANDARD DEVIATIONS OF MEN
BY SCORING TECHNIQUE
(N = 42)

	Traditional		Dichotomous Splitting		Differential Weighting		K Correction	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
MC-SD	38.57	8.07	53.21	10.60	46.21	8.99		
L	47.36	10.79	55.26	12.91	59.90	11.75		
F	62.78	17.83	68.33	18.20	69.52	15.44		
K	41.88	9.67	46.59	8.34	44.90	7.75		
Hs	39.45	13.22	46.21	15.28	55.31	14.37	50.35	13.27
D	52.95	13.30	57.43	11.61	55.40	10.39		
Hy	40.19	10.07	44.24	9.68	42.90	9.02		
Pd	39.29	10.94	43.83	10.83	46.76	9.34	44.75	10.45
Mf	50.62	10.75	52.26	9.50	49.19	8.49		
Pa	59.64	13.02	61.33	10.77	62.59	11.59		
Pt	36.86	14.89	36.88	14.82	47.71	12.32	45.00	11.26
Sc	42.59	21.02	45.76	19.22	54.52	16.82	51.72	17.83
Ma	47.24	11.21	47.26	10.83	48.81	10.06	49.00	10.32
ESD	42.69	8.98	45.81	8.86	42.78	8.76		

TABLE III

MEANS AND STANDARD DEVIATIONS OF WOMEN
BY SCORING TECHNIQUE
(N = 42)

	Traditional		Dichotomous Splitting		Differential Weighting		K Correction	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
MC-SD	43.57	7.61	52.47	9.58	51.40	9.03		
L	51.33	13.76	58.23	13.42	59.71	11.06		
F	54.31	8.31	57.09	10.09	56.45	8.19		
K	36.07	20.15	42.14	19.26	38.09	16.82		
Hs	38.26	11.75	44.36	12.21	51.92	12.40	47.55	11.33
D	44.74	9.87	46.90	11.11	45.50	9.79		
Hy	45.36	10.03	47.86	9.22	46.07	8.47		
Pd	37.31	11.43	38.76	12.49	45.36	10.56	42.67	10.26
Mf	51.78	7.33	51.21	7.65	57.52	6.53		
Pa	51.62	11.01	51.45	9.51	52.47	10.50		
Pt	28.14	13.95	31.76	13.37	42.07	9.97	38.19	9.76
Sc	32.28	11.34	38.35	10.18	46.33	8.22	41.89	9.34
Ma	52.12	9.71	54.38	8.47	54.90	8.42	54.99	8.63
ESD	50.26	6.85	51.50	6.25	51.21	5.29		

TABLE IV

INTERCORRELATION OF MEN'S SCORES
USING TRADITIONAL SCORING
(N = 42)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		20	-23	22	-06	-02	12	08	-30	21	03	01	04	02
L	10		22	15	30	-05	01	07	-19	06	-04	13	01	-05
F	-12	-15		-14	24	04	00	01	26	54	26	52	04	-37
K	-17	-04	08		-11	-25	-18	-19	-34	-05	-63	-51	-35	30
Hs	-11	19	11	01		42	-01	23	12	26	37	35	13	05
D	04	-11	-05	-05	-06		13	51	38	24	57	34	-08	01
Hy	-01	00	12	10	-31	-24		29	12	15	-05	02	08	12
Pd	08	-08	-18	-04	-24	-07	-07		34	16	43	40	19	-07
Mf	-03	-29	06	-11	-10	06	01	03		23	18	14	01	04
Pa	-04	-17	14	11	-01	04	09	-12	15		44	52	11	-44
Pt	07	05	-02	-08	01	15	-18	02	-04	-04		78	33	-52
Sc	03	08	08	01	01	-01	-02	02	-08	-06	01		42	-65
Ma	-05	03	-20	-03	-01	-18	12	07	-02	-30	-16	-12		-34
ESD	-14	-05	03	-21	10	02	-08	-03	15	-06	01	-04	09	

Decimal points have been omitted. Correlations are above the diagonal and residuals for the multiple group solution are below the diagonal.

TABLE V

INTERCORRELATION OF MEN'S SCORES
USING DICHOTOMOUS SCORING
(N = 42)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		09	-01	45	-06	-21	08	-32	-24	-06	-39	-31	-30	48
L	00		07	25	26	-07	04	12	-11	07	-12	05	05	05
F	07	-32		-13	32	23	20	13	37	57	24	43	-09	-17
K	-14	05	-05		-06	-35	06	-09	-34	-21	-75	-51	-33	47
Hs	-01	12	04	01		37	08	29	06	41	28	40	08	-04
D	03	-14	-01	-14	-10		22	40	37	31	48	47	-06	-02
Hy	02	06	10	08	-30	-17		22	28	29	16	27	08	-05
Pd	-03	-04	-13	04	-13	-15	-10		23	04	27	36	05	-08
Mf	01	-19	18	-13	-16	08	13	-05		21	36	22	09	-07
Pa	04	-12	17	03	07	01	08	-17	00		38	58	-06	-37
Pt	03	00	01	-12	-03	09	-08	02	09	-09		72	48	-53
Sc	03	-01	04	02	-02	02	-02	03	-09	-01	-05		43	-62
Ma	-11	13	-24	06	-01	-12	03	10	-01	-41	-04	-08		-40
ESD	-10	-06	-01	-18	00	11	-10	-01	12	-07	09	-06	05	

Decimal points have been omitted. Correlations are above the diagonal and residuals for the multiple group solution are below the diagonal.

TABLE VI

INTERCORRELATION OF MEN'S SCORES
USING DWS SCORING
(N = 42)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		11	-19	32	-09	-15	10	-04	-13	11	-26	-08	-12	36
L	01		17	41	43	-10	-11	03	-35	-02	-28	-05	-15	21
F	02	-16		-13	24	17	19	01	37	41	33	46	-06	-23
K	-18	06	-01		01	-31	05	-01	-41	-18	-56	-41	-12	37
Hs	-07	26	-01	02		34	11	25	00	28	29	30	12	07
D	04	-11	00	-11	-09		33	36	28	26	56	35	-13	-02
Hy	00	-06	18	06	-28	-13		39	19	14	22	30	26	03
Pd	03	-08	-16	02	-20	-18	-07		17	-04	25	34	17	-01
Mf	06	-31	24	-20	-13	06	07	00		25	29	16	-09	-04
Pa	02	-09	05	-07	04	12	-01	-15	16		44	52	-04	-16
Pt	01	00	05	-03	-01	14	-09	-03	04	-01		72	29	-50
Sc	09	06	07	01	-03	-02	00	06	-04	-04	-03		32	-56
Ma	-14	03	-18	10	01	-24	11	12	-17	-39	-14	-12		-22
ESD	-13	-07	-01	-21	05	07	-05	-06	14	04	02	-10	03	

Decimal points have been omitted. Correlations are above the diagonal and residuals for the multiple group solution are below the diagonal.

TABLE VII

INTERCORRELATION OF WOMEN'S SCORES
USING TRADITIONAL SCORING
(N = 42)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		19	02	32	11	-21	-04	-16	-13	10	-20	-24	-08	17
L	16		68	-10	10	-15	01	26	11	14	37	32	26	-28
F	06	07		-18	04	-24	06	38	34	21	50	49	42	-38
K	-20	-09	-01		-26	-12	19	-33	-14	-05	-65	-59	-23	49
Hs	19	18	11	-13		26	39	47	-19	27	37	47	-11	-34
D	-08	-17	-28	08	-19		24	11	-38	45	32	34	01	-33
Hy	-13	01	12	08	-17	-14		34	-20	34	02	26	-07	-05
Pd	02	-03	04	-04	00	-30	-07		06	16	56	67	39	-48
Mf	-07	03	22	-05	03	-24	04	16		-29	12	07	16	11
Pa	-10	-10	-09	-01	-02	19	04	-21	-21		31	36	22	-09
Pt	12	12	07	-03	07	-02	-08	03	08	-10		83	46	-75
Sc	04	09	11	-02	03	-09	00	05	09	-11	02		46	-68
Ma	-06	-11	-09	07	-07	-07	03	12	03	-20	-10	05		-37
ESD	-26	-06	-05	-09	-06	-01	05	01	13	11	-08	-08	-01	

Decimal points have been omitted. Correlations are above the diagonal and residuals for the multiple group solution are below the diagonal.

TABLE VIII

INTERCORRELATION OF WOMEN'S SCORES
USING DICHOTOMOUS SCORING
(N = 42)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		10	11	30	13	11	-10	-13	23	12	-16	-10	-04	30
L	10		61	-26	18	17	25	38	20	27	45	39	29	-25
F	09	-01		-26	24	22	32	46	34	44	54	55	32	-37
K	-19	-10	-03		-33	11	02	-31	-06	-10	-68	-54	-36	54
Hs	16	14	10	-16		15	38	38	-21	37	40	44	-10	-28
D	-05	-18	-17	09	-18		20	22	-19	56	30	30	-09	-15
Hy	-08	08	10	06	-14	-18		33	-10	33	16	28	-04	-17
Pd	-01	-05	-04	00	-06	-20	-12		-14	33	59	66	25	-44
Mf	10	09	26	-17	04	-14	10	-01		-23	-01	-01	14	22
Pa	-11	-11	-05	01	01	12	00	-14	-21		43	42	11	-11
Pt	09	08	04	-04	03	03	-08	01	08	-05		84	39	-72
Sc	03	05	06	01	-01	-02	-03	07	11	-14	01		47	-65
Ma	-01	-01	-05	03	-03	-13	11	05	01	-20	-13	-02		-36
ESD	-18	-01	-05	-12	00	-04	02	01	07	10	-04	-04	-01	

Decimal points have been omitted. Correlations are above the diagonal and residuals for the multiple group solution are below the diagonal.

TABLE IX

INTERCORRELATION OF WOMEN'S SCORES
USING DWS SCORING
(N = 42)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		05	09	19	09	01	-20	-01	13	21	-10	03	10	38
L	05		57	-22	18	13	25	35	07	26	44	40	27	-26
F	06	-01		-20	00	04	21	29	26	31	46	51	47	-44
K	-27	-05	06		07	01	22	-01	-12	-01	-41	-25	-15	57
Hs	12	12	04	-09		23	28	45	-31	26	18	35	-05	-12
D	-02	-16	-20	-01	-14		17	22	-39	58	38	40	04	-15
Hy	-10	05	15	11	-22	-19		45	-26	15	03	34	02	-14
Pd	01	-01	01	-01	-05	-26	-05		-27	17	45	60	38	-24
Mf	06	05	19	-07	06	-15	05	03		-29	-15	-16	04	12
Pa	-02	-09	-10	-02	04	21	-01	-24	-16		37	42	26	-09
Pt	03	12	03	00	00	04	-09	05	01	-09		73	34	-61
Sc	06	07	11	-01	-02	-06	05	04	12	-14	-01		49	-53
Ma	-06	-01	-03	03	-01	-18	06	13	02	-22	-17	-03		-22
ESD	-14	00	-12	-07	-03	03	-01	01	00	04	-02	-04	02	

Decimal points have been omitted. Correlations are above the diagonal and residuals for the multiple group solution are below the diagonal.

TABLE X

INTERCORRELATION OF ALL SCORES
USING TRADITIONAL SCORING
(N = 84)

Variable	MC	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	ESD
MC		23	-23	19	01	-19	11	-06	-20	05	-16	-15	06	21
L	23		28	-06	18	-15	05	16	-03	05	12	14	17	-07
F	-23	28		-06	18	07	-05	14	24	49	37	56	07	-44
K	19	-06	-06		-18	-08	12	-25	-20	01	-51	-37	-29	25
Hs	01	18	18	-18		35	16	35	00	26	37	38	02	-12
D	-19	-15	07	-08	35		08	34	09	39	52	40	-11	-24
Hy	11	05	-05	12	16	08		28	01	14	-09	02	07	14
Pd	-06	16	14	-25	35	34	28		21	18	50	48	26	-26
Mf	-20	-03	24	-20	00	09	01	21		02	13	09	08	08
Pa	05	05	49	01	26	39	14	18	02		44	51	07	-40
Pt	-16	12	37	-51	37	52	-09	50	13	44		79	29	-66
Sc	-15	14	56	-37	38	40	02	48	09	51	79		32	-69
Ma	06	17	07	-29	02	-11	07	26	08	07	29	32		-21
ESD	21	-07	-44	25	-12	-24	14	-26	08	-40	-66	-69	-21	

Decimal points have been omitted.

Factor Analyses

The initial factor analytic solution consisted of a principal components extraction and varimax rotation of the correlation matrix presented in Table X. Table XI contains both the principal component solution and varimax rotation. It should be pointed out that only loadings having an absolute value greater than 0.29 are reported for the varimax solution.

TABLE XI

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF TRADITIONAL SCORING FOR TOTAL SAMPLE

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-23	67	16		61	36
L	19	50	48			67
F	59	-06	03	49		
K	-46	29	-43		57	-36
Hs	49	38	-22	62		
D	53	-03	-62	69		-43
Hy	04	62	-22		58	
Pd	60	27	03	60		
Mf	20	-26	09		-33	
Pa	60	22	-35	72		
Pt	89	-10	04	75	-45	
Sc	91	-01	04	79	-37	
Ma	33	14	64			70
ESD	-72	21	-07	-56	48	
Systematic Variance Of Factors	4.19	1.57	1.47	4.39	1.94	1.19

The decimal points have been omitted except at the bottom.

The solutions presented in Table XI closely resemble those found in the literature. Block (1965) presented the most representative solution to that date. His solution may be found in Appendix C. The replication of the present findings indicates that the sample used in this study does not appear extremely biased. For several reasons only three factors are extracted and rotated in the initial solution. First, only the factors having latent roots greater than one are considered as important. Cattell has suggested that this method is the most reliable for designs with 15 to 50 variables. A second reason is based on the literature. Most factor analyses of the MMPI report only two or three factors.

Principal component solutions and varimax rotations were also applied to the correlation matrices found in Tables IV through IX. Again, only three factors are reported for each solution. It should be noted that three of the six solutions have four factors with latent roots greater than one. In order to be consistent, only three factors are reported and rotated.

Tables XII through XVII present the principal components and varimax rotations. For the principal component solution all factor loadings are reported, but it should be noted that only loadings having an absolute value greater than 0.29 are reported for the varimax rotation.

The varimax solution reveals several interesting changes in patterns for the different scoring techniques. The most intriguing altercation is the loading of the ESD on the first factor. One can note that there is a small decrease from the traditional to the dichotomous with the differential weighting completely removing ESD from the first factor.

TABLE XII

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF TRADITIONAL SCORING FOR THE MEN

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-05	20	61			65
L	07	32	58			65
F	53	32	03	61		
K	-60	08	55	-52		62
Hs	47	-17	34		52	
D	55	-62	16		83	
Hy	06	-31	46		44	30
Pd	52	-48	27		75	
Mf	39	-56	-23		55	-47
Pa	62	22	32	60		31
Pt	88	-02	-09	73	46	
Sc	90	24	01	88	32	
Ma	41	25	-13	50		
ESD	-59	-58	-19	-83		
Systematic Variance Of Factors	4.11	1.82	1.69	3.89	2.76	1.84

The decimal points have been omitted except at the bottom.

The traditional solutions for both the men and the women have comparable factor patterns. They are also similar to the solution by Block (1965) presented in Appendix C. Although the solutions are comparable, some distinctive differences do appear. For the women, MC loads highly negative on the first factor. A second major difference is found on the third factor with the appearance of Pt and Sc in the women's solution. The final difference can be seen in the difference of the Mf loading on the second factor for both the men's and women's solution.

TABLE XIII

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF TRADITIONAL SCORING FOR THE WOMEN

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-22	-01	67	-55		39
L	46	-40	54			81
F	60	-53	42			85
K	-62	13	51	-81		
Hs	49	47	17		64	
D	32	70	-24	37	58	-43
Hy	23	53	38		69	
Pd	73	04	07	57	32	34
Mf	06	-70	-13		-62	33
Pa	39	45	38		68	
Pt	91	-06	-15	85		32
Sc	93	05	-07	82	34	31
Ma	54	-34	03	45		44
ESD	-78	-03	17	-75		
Systematic Variance Of Factors	4.71	2.29	1.66	4.25	2.53	1.76

The decimal points have been omitted except at the bottom.

The differences in the Mf loading on the second factor would be expected since the scale is supposedly a measure of masculinity and feminity. The higher loadings of Sc and Pt on the third factor are somewhat surprising due to the men's higher means and greater variability. The higher correlations of the women could possibly be the result of a restriction of range. The increased loading of MC on the women's first factor adds to confirming the premise that social desirability has an influence in the response process for a two-choice format.

TABLE XIV

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF DICHOTOMOUS SCORING FOR THE MEN

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-52	45	09	-64		
L	-04	38	59		52	-47
F	48	52	02		64	
K	-69	45	19	-73		-42
Hs	44	48	28		71	
D	59	29	-43		34	70
Hy	29	40	-12		40	31
Pd	43	22	-18		30	40
Mf	49	12	-55			72
Pa	59	45	22		73	
Pt	86	-23	-03	76		40
Sc	87	05	25	66	59	
Ma	41	-50	36	72		
ESD	-63	40	-37	-81		
Systematic Variance Of Factors	4.47	2.04	1.39	3.98	2.87	1.56

The decimal points have been omitted except at the bottom.

The solution using dichotomous splitting has some major differences from the previous solution. Tables XIV and XV present the factor patterns for men and women respectively. These solutions include the high negative loading of MC on the first factor. This is the major difference obtained between these solutions and the solutions based on the traditional scoring technique. There is little, if any, reduction in the high negative loading of ESD. A final difference is found in comparing the two men's solutions. For the traditional scoring MC has

a high positive loading on the third factor which is not the case for the dichotomous splitting solution.

TABLE XV

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF DICHOTOMOUS SCORING FOR THE WOMEN

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-12	38	61	-46		53
L	58	02	47	33		60
F	70	06	51	40	38	67
K	-61	52	17	-82		
Hs	53	35	-19		61	
D	36	63	-03		72	
Hy	40	44	-07		59	
Pd	73	07	-10	56	48	
Mf	-04	-32	79		-37	76
Pa	55	59	04		78	
Pt	90	-16	-07	82	37	
Sc	90	-08	-03	76	43	
Ma	45	-49	23	58		37
ESD	-71	29	33	-81		
Systematic Variance Of Factors	4.97	1.95	1.73	3.56	2.47	1.87

The decimal points have been omitted except at the bottom.

Again, there are sex differences found in the factor patterns of a single scoring technique. As in the traditional solution, the Mf scale loads on the second factor negatively for women and positively for men. This similarly appears in the L scale except the signs are

reversed; positive for women and negative for men. This is probably related to the lack of a high positive loading of MC on the third factor for the men. The third factor also exhibits differences in the interpretability of that factor for men and women. For men, the factor represents a more clinical nature as evident by the high loading on D and Pa, while the women's factor pattern indicates that the third factor is largely determined by sex identity and response bias as seen in the high loadings of Mf, L and F.

TABLE XVI

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF DWS SCORING FOR THE MEN

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-33	34	12		48	
L	-26	71	-32	46	65	
F	52	25	-49	72		
K	-62	50	09		79	
Hs	35	66	-11	67		
D	60	25	11	44		44
Hy	38	34	55			73
Pd	38	36	58			76
Mf	47	-16	-07		-45	
Pa	55	25	-40	68		
Pt	88	-08	02	43	-66	39
Sc	83	13	-01	56	-48	41
Ma	28	-11	53			53
ESD	-53	42	14		68	
Systematic Variance Of Factors	3.98	1.98	1.51	3.68	2.19	1.56

The decimal points have been omitted except at the bottom.

TABLE XVII

PRINCIPAL COMPONENT AND VARIMAX ROTATION OF DWS SCORING FOR THE WOMEN

Variable	Components			Varimax		
	I	II	III	I	II	III
MC	-05	13	79	31		74
L	61	-18	26	67		
F	64	-42	36	84		
K	-33	60	36	-31		68
Hs	39	53	-03		65	
D	47	47	-06		66	
Hy	39	42	-10		58	
Pd	67	26	02	41	59	
Mf	-24	-65	39		-74	
Pa	55	34	29	43	52	
Pt	81	-19	-13	62	32	-47
Sc	88	01	03	68	49	
Ma	52	-27	36	68		
ESD	-64	37	50	-38		80
Systematic Variance Of Factors	4.37	2.13	1.62	4.06	2.68	1.33

The decimal points have been omitted except at the bottom.

The results of the factor analysis on the DWS scores are very different from the solutions for traditional and dichotomous scoring in that, for the men the social desirability scales, MC, K, and ESD, are removed from the first factor and now load highly on the second factor. The third factor of the men's solution still remains clinical in nature with the high loadings on D and Pd supplemented by moderate loadings on Ma, Sc, and Hy. The Mf loading for men is no longer in the positive direction, loading moderately negative on the second

factor. This change does not appear to be as significant when looking at the high negative loading of Mf for women. It appears that the DWS

TABLE XVIII

HYPOTHESIS MATRICES FOR MEN AND WOMEN
FOR THE MULTIPLE GROUP SOLUTION

Variable	Men				Women			
	I	II	III	IV	I	II	III	IV
MC				+				+
L	+				+			
F	+				+			
K				+				+
Hs	+	+			+	+		
D	+	+			+			
Hy	+	+			+	+		
Pd	+	+			+	+		
Mf		+				-		
Pa	+		+		+		+	
Pt	+		+		+		+	
Sc	+		+		+		+	
Ma	+		+		+	-		
ESD				+				+

scoring technique simply shifted the Mf loadings toward the negative pole.

Major differences exhibited in the women's factor solution of the DWS scores are not as obvious as the men's. Although the degree to which the social desirability scales load on the first factor is greatly reduced, all three still load on the first factor. The third factor is a social desirability factor, similar in nature to the third factor of the women's dichotomous factor solution. The sex identity is no longer a part of the third factor. This could possibly be due to the high negative loading appearing on the second factor.

The second factoring technique used to look at the relationships of the various scales using different scoring techniques was the multiple group technique. One of the requirements of this technique is the construction of a binary hypothesis matrix to be used in the computational algorithm. Since sex differences were observed in the varimax solutions previously discussed, two hypothesis matrices were constructed. The hypothesized relationship of the scales using DWS scoring is given in the hypothesis matrices appearing in Table VIII. The '+' sign represents a moderate or high positive loading for that variable on the factor and the '-' sign represents a moderate or high negative loading for that variable on the factor. Four factors were hypothesized because in some of the principal components solutions four components had latent roots greater than 1.00. The absence of a '+' or '-' indicates that the variable is hypothesized to have a zero or near zero loading.

The development of the hypothesis matrix was based on the intercorrelations and the hypothesized relationship based on the literature review. It was also felt that the hypothesis matrices should be based on the DWS scoring technique and then used to extract factors for all

three scoring techniques. If the factor patterns obtained were the same using this one hypothesis matrix for each sex, then one could state that there were no differences evident in the factor structures using the different techniques of scoring. The extraction matrix is taken from the hypothesis matrix.

The multiple group solutions appear in Tables XIX and XX. The men's solutions in Table XIX do show differences among the three scoring techniques. The dichotomous and DWS more closely resemble the hypothesized structure than the traditional factor structure. Similarly, the women's solutions using DWS and dichotomous scoring produce factor structures that are more representative of the hypothesis matrix than the traditional scoring factor structure. Since the factor structures for each sex are based on different hypothesis matrices, a discussion of the differences between sexes would not be necessary. It is more important to examine differences in scoring technique of each sex. A factor by factor comparison would be the most useful.

Examining the men's differences first, one finds that the first factor in each of the solutions is almost as hypothesized except for the appearance of a moderately positive loading of ESD in the traditional and dichotomous techniques. It is important to note that the traditional technique has a higher loading than the dichotomous technique. Ma does not load on the first factor as was hypothesized for the dichotomous and the DWS scoring techniques. Other than these differences the first factor is as hypothesized for all three techniques.

The second factor of the traditional technique is surprising in that ESD loads moderately positive. This has not appeared in the

TABLE XIX

MULTIPLE GROUP SOLUTIONS OF THE MEN'S SCORES
USING THREE DIFFERENT SCORING TECHNIQUES

Variable	Traditional				Dichotomous				DWS			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
MC				82				79				81
L	41		-53		41		-48		32		-64	
F	58	-40			62				60	-35		
K			-50	54			-58	48			-50	57
Hs	50	37			57	34			58			
D	52	57			58	46			55	49		
Hy		53			34	48			42	56		
Pd	61	47			56	44			54	48		
Mf	32			-34	37							
Pa	67				63				58		32	
Pt	63		61		56		68		56		65	
Sc	73		50		75		53		69		55	
Ma	39		44				55				49	
ESD	-38	43	-45		-33		-55	48			-50	58
Systematic Variance Of Factors	3.34	1.50	1.63	1.36	3.41	1.06	2.11	1.31	2.85	1.19	2.06	1.56

The decimal points have been omitted except at the bottom.

TABLE XX

MULTIPLE GROUP SOLUTIONS OF THE WOMEN'S SCORES
USING THREE DIFFERENT SCORING TECHNIQUES

Variable	Traditional				Dichotomous				DWS			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
MC				69				74				74
L	62	-31	-39		70				68		-33	
F	69	-35			78				70	-32		
K			-55	56			-54	53				69
Hs		73			32	68				69		
D	33	48			54	30			53	33		
Hy		68			38	64			33	62		
Pd	65	40			68	33			63	48		
Mf		-32				-37				-45		
Pa	62			40	70				68			
Pt	71		48		72		54		64		51	
Sc	74		44		75		55		73		50	
Ma	65	-34	30		48	-46	43		63	-34		
ESD	-55			48	-43		-45	53	-31		-30	74
Systematic Variance Of Factors	3.80	1.96	1.20	1.41	4.10	1.53	1.54	1.36	3.80	1.78	0.96	1.74

The decimal points have been omitted except at the bottom.

literature nor was it expected on the basis of the varimax solutions. The second factor for the dichotomous technique is as hypothesized except for the absence of the Mf scale. Both the traditional and DWS scoring techniques have F loading moderately negatively on the second factor.

All three scoring techniques produce identical third factors. These factors are as hypothesized except for the moderately negative loadings of L, K, and ESD. It would be desirable to have the loadings of K and ESD not present, but for this factor the variance they are contributing may not be social desirability response biases in the individual's response.

The fourth factor is as hypothesized for the DWS and dichotomous scoring techniques. Again, it is the traditional scoring technique that produces surprising results. ESD does not load on the final factor and Mf does moderately negative.

The women's results do not produce as clear factor structures as do the men's with respect to the hypothesis matrix. The first factor is as hypothesized except for the inclusion of ESD on all three first factors. An important difference does exist among scoring techniques when looking specifically at that loading. The amount ESD loads on the first factor decreases remarkably. It goes from a -55 for traditional scoring to a -31 for DWS scoring.

The second factor is as hypothesized for all three scoring techniques. It should be pointed out that the traditional scoring technique does not have ESD loading on the second factor for women as it does for men. Also, L and F appear on the second factor but not with much influence.

The third factor has the most obvious differences. ESD loads negatively on the third factor for the dichotomous and DWS scoring techniques, but not for the traditional technique. K loads highly negative for the traditional and dichotomous techniques but not at all for the DWS. F loads moderately negative for the DWS technique, while L loads moderately negative for the traditional technique. The third factor is not a stable factor across scoring techniques.

The fourth factor is as hypothesized for the DWS and dichotomous techniques. For the traditional technique, it is as hypothesized except for one additional loading. The Pa scale loads moderately positive on this factor. It should be noted, however, that it loads the least of the scales loading on the factor for the traditional technique.

Analyses of Variance

Fourteen repeated measures analyses of variance were computed. The overall design for each ANOVA was a 2 by 3 (or for Hs, Pd, Pt, Sc, and Ma a 2 by 4) analysis of variance with subjects nested in the sex factor, A, and repeatedly measured on the scoring technique factor, B. The ANOVA tables are given in Appendix D.

A summary of the significant differences resulting from the ANOVA's is presented in Table XXI. On only two of the dependent variables, L and Pd, were there no significant sex differences. Females had higher scores on both social desirability scales, MC-SD and ESD, as well as on the MMPI scales of Hy, Mf and Ma. Males scored higher on the MMPI scales of F, K, Hs, D, Pa, Pt, and Sc.

In general, scoring technique produced significant differences on all five of the MMPI scales with K correction (Hs, Pd, Pt, Sc, and Ma).

TABLE XXI

SUMMARY OF ANOVA RESULTS

Scale	A(Sex)	B(Scoring Technique)	A x B
MC	F > M ($p < .01$)	Dichotomous > Traditional ($p < .0001$)	N.S.
L	N.S.	Dichotomous > Traditional ($p < .01$)	N.S.
F	M > F ($p < .01$)	N.S.	N.S.
K	M > F ($p < .01$)	N.S.	N.S.
Hs	M > F ($p < .01$)	Trad + K > DWS > D.S. > Trad ($p < .0001$)	N.S.
D	M > F ($p < .01$)	N.S.	N.S.
Hy	F > M ($p < .05$)	N.S.	N.S.
Pd	N.S.	Trad + K > DWS > D.S., Trad ($p < .0001$)	N.S.
Mf	F > M ($p < .01$)	N.S.	Men: DWS > Trad Women: Trad > DWS ($p < .0001$)
Pa	M > F ($p < .01$)	N.S.	N.S.
Pt	M > F ($p < .01$)	Trad + K > DWS > D.S., Trad ($p < .0001$)	N.S.
Sc	M > F ($p < .01$)	Trad + K > DWS > D.S., Trad ($p < .0001$)	N.S.
Ma	F > M ($p < .01$)	Trad + K > DWS > D.S., Trad ($p < .0001$)	N.S.
ESD	F > M ($p < .01$)	N.S.	N.S.

However, on only two of the remaining nine scales (MC-SD and L) were there any differences due to scoring technique. On all five scales with K correction, the traditional K-corrected scores were significantly higher than those from the traditional and dichotomous splitting procedures. Similarly, on four of these five scales (Hs, Pd, Pt, and Sc) the DWS scoring produced higher means than either the traditional or dichotomous techniques. On the Hs scale dichotomous scores were higher than traditional scores. For both the other analyses where there was a significant main effect due to scoring technique (MC-SD and L), the significance was due to the dichotomous scoring producing higher means than the traditional technique.

The only significant interaction appeared in the ANOVA of the Mf dependent variable. Examiniation of the means indicated that both the significant main effect due to sex and the significant interation effect were the result of the DWS scores. There were no differences in scores for men and women on the traditional and dichotomous scoring techniques. There were, however, significant changes in the men's scores and the women's scores when using the DWS scoring technique. Men's scores significantly increased when using DWS, while women's scores decreased significantly when using DWS scoring.

Distribution Comparisons

Measures of skewedness and kurtosis were computed for each variable in each scoring technique except traditional K-corrected. Those measures were then tested for divergence from normality. Table XXII presents the results of the men when examined for kurtosis and skewedness, while Table XXIII presents the same results for women.

TABLE XXII

MEASURES OF KURTOSIS AND SKEWEDNESS FOR THE THREE SCORING TECHNIQUES USING MEN'S SCORES

Variable	Traditional		Dichotomous		DWS	
	Skew	Kurt	Skew	Kurt	Skew	Kurt
MC	-0.104	-0.622	-0.095	-0.107	-0.006	-0.195
L	0.172	-1.051**	0.013	-0.789	0.702*	0.205
F	0.709*	-0.367	0.212	-0.869*	0.214	-0.581
K	0.619*	-0.812	0.616*	-0.773	0.869*	0.134
Hs	0.374	-0.334	0.183	-0.164	0.067	-0.601
D	0.421	-0.412	0.204	-0.032	0.226	-0.259
Hy	0.223	-0.568	0.061	-0.648	0.559	-0.434
Pd	0.033	-0.840	0.664*	-0.101	0.190	-0.744
Mf	0.278	-0.559	0.139	-0.804	0.039	-0.863*
Pa	0.995**	1.098*	1.112**	1.280*	1.095**	0.922*
Pt	0.523	-0.566	0.411	-0.728	0.435	-0.752
Sc	0.745*	-0.410	0.645*	-0.309	0.734*	-0.186
Ma	0.155	-0.374	-0.355	-0.114	-0.009	-0.639
ESD	-0.006	0.299	0.194	0.295	0.215	0.536

*Significantly different from normality ($p < .05$)
**Significantly different from normality ($p < .01$)

Since there is no technique available statistically to compare the various measures across scales, the procedure used to draw out differences was simply to count the number of scales for each scoring technique which significantly differ from normality. For the men, traditional had four scales that were significantly skewed and two whose kurtosis were different from normality. The dichotomous and DWS scoring techniques also had four skewed scales and two whose kurtosis were different from normality. Thus, for the men, there appears to be no tendency for one of the scoring techniques to normalize the scores more than any of the others. On the other hand, the women do exhibit

differences across the scoring techniques. Six scales were skewed using traditional scoring, while only four were skewed for dichotomous and DWS scoring techniques. Six scales are different from normality with respect to kurtosis when using the traditional scoring technique. This is greatly reduced when using the dichotomous and DWS scoring techniques. Dichotomous scoring has two scales which have kurtosis significantly different from normality and DWS has three scales which are different from normality with respect to kurtosis.

TABLE XXIII

MEASURES OF KURTOSIS AND SKEWEDNESS FOR THE THREE SCORING TECHNIQUES USING WOMEN'S SCORES

Variable	Traditional		Dichotomous		DWS	
	Skew	Kurt	Skew	Kurt	Skew	Kurt
MC	-0.196	0.084	0.132	-0.594	0.183	-0.592
L	0.449	-0.530	0.383	-0.259	0.786*	0.234
F	1.307**	2.245**	1.052**	0.582	1.434**	1.995**
K	0.320	-1.164**	0.208	-0.936*	0.103	-1.021*
Hs	0.226	-0.989*	0.226	-0.354	0.183	-0.927
D	0.623*	-0.634	1.131**	0.682	0.976**	0.593
Hy	0.360	-0.010	0.353	0.285	0.449	0.376
Pd	0.595*	-0.277	0.586*	-0.228	0.154	-0.348
Mf	0.665*	0.479	-0.064	0.268	0.254	-0.391
Pa	1.634**	4.061**	1.057**	1.551*	1.065**	1.296*
Pt	0.324	-0.953*	0.283	-0.804	0.089	-0.762
Sc	0.619*	-0.263	0.584	-0.330	0.355	-0.359
Ma	0.013	-0.943*	0.450	-0.294	0.041	-0.322
ESD	-0.062	-0.692	0.055	-0.675	0.180	-0.482

*Significantly different from normality ($p < .05$)
**Significantly different from normality ($p < .01$)

CHAPTER V

DISCUSSION AND SUMMARY

Discussion

The major result of the present study is the finding that a differential weighting system of scoring is more advantageous than the traditional True/False scoring technique employed by the MMPI. Several specific arguments can be developed in support of this statement in light of the results obtained in the previous chapter. First the factor analyses, both the varimax solution and multiple group solution, provide evidence that a differential weight system reduces the influence of social desirability in the responses of an individual to a personality inventory. A second argument can be developed on the basis of the findings of the analyses of variance. These results indicate that the scores obtained using the DWS are not significantly different from the scores of the traditional scores with respect to interpretability. A final argument can be developed on the basis of the overall results and the response theory of Jackson.

The factor analytic results indicate two important aspects. First, the data produce the same factor structure of the MMPI when traditionally scored as reported in the literature. This is important in order to establish that the data are not anomolous. If this were not the case, then the other factor analytic results would not be

reliable. Some of the important similarities should be noted. The high negative loading of the ESD scale on the first factor has been used by Edwards as the foundation of his argument against the use of True/False items in a personality inventory. Both the men's and women's factor analytic results produced a high negative loading of ESD on the first factor.

A second similarity is found in the loading of the K scale on the first factor. The present results concurred with the findings obtained by Meehl and Hathaway (1946). Again, a high negative loading on the first factor was obtained for both men and women. The final similarity is found only in the men's factor analytic solution. The MC scale has been found in several studies (Crowne & Marlowe, 1960; Hanley, 1967) to load on a separate factor.

Although the social desirability scales load in a similar manner for the present study and those studies in the literature, the factor structure of the clinical scales should be replicated. Based on the factor structure obtained by Wheeler, Little and Lehner (1951), the present findings replicate their two factor solution. The first factor in the present study is similar to the psychotic factor and the second factor is similar to the neurotic factor.

Now that it has been established that the factor structures obtained in the present study duplicate those of previous studies, one can look at the differences in the factor structures of the different scoring techniques to assess their similarities and differences. The most significant finding is the decrease of the ESD loading in both the men's and women's factor structures. Both the varimax and multiple group solutions support this finding. The reduction of the high

negative loading of ESD on the first factor is one of the goals of Edwards.

The reduction of the ESD loading on the first factor in the DWS solution is coupled with the absence of K and MC loadings on that factor. Since this factor can no longer be identified as a social desirability component, how should this dimension be construed in light of the present results? Block (1965) points out that behavioral correlates of the first factor as identified in previous research have indicated that this factor is related to differences along an adaptability-vulnerability personality dimension. This may no longer be the case due to the shift of factor loadings. Also, most solutions are in terms of simple structure; whereas the present multiple group solution is a hierarchial approach to the factor structure of the MMPI. Tyler (1951), in a similar solution, names the first factor "general maladjustment." This may be an appropriate term to apply to the first factor due to the hierarchial nature of the solution. On the other hand, Younger (1974) in a factor analysis of several measures of personality identified the factor containing the majority of MMPI scales as "intra-individual stability." This is more appropriate in terms of Block's findings with respect to the behavioral correlates than Tyler's label. In addition, it allows for a more suitable position within a theoretical framework from which predictions can be made.

The second factor in the present study closely resembles the traditional second factor, Beta. Kassebaum, Couch and Slater (1959) interpret this factor in terms of "introversion-extraversion." In other factor analyses of the MMPI this component has been identified as "neuroticism" (Wheeler, et al., 1951) or "acquiescence" (Couch &

Keniston, 1961) or "ego-control" (Block, 1965). Only the label applied by Wheeler et al. appears to be appropriate. Since this factor is a minor group factor of the general factor, "neuroticism" as a label has theoretical implications. It also is related to the behavior correlates found by Block for the Beta dimension. Block points out that there are sex differences with respect to this factor which is supported by the factor solutions and the ANOVAs in the present study.

Again, the findings of Wheeler et al. can be used to label the third factor or the second group factor. Since the most serious deviations are represented by the scales loading on this factor, "psychoticism" is the most appropriate label. The Pt and Sc scales are typically referred to as psychotic scales. What one has in the first three factors is a hierarchial arrangement. The general factor is labelled as "inter-individual stability" with two group factors associated with it, neuroticism and psychoticism. This has several theoretical implications which are pertinent to discussions centering around personality structure. Initially, it was felt that psychoticism and neuroticism were poles of the same continuum. The present results indiciate that this is not the case. They are two different dimensions which affect the internal stability of the individual. A second implication is centered on the psychoticism factor and the loading of ESD on it. Does this imply that the psychoticism factor is closely related to social desirability? Wikoff (1965) found that there are various personality correlates to social desirability which could possibly account for the loading of ESD on the third factor. On the other hand, Block (1965) develops an extensive argument that a social desirability interpretation is inappropriate. On the basis of more

recent research Block's arguments may be substantiated to some extent. Boe, Gocka, Edward, and Kogan (1966) factor analyze individual social desirability scale values and indicate personal desirability is more influential than social desirability in biasing an individual's test response. In addition, Messick (1960) points out that there are several dimensions of social desirability. Thus, as Block argues, the overlap of the ESD with the psychoticism factor is due to content overlap rather than the influence of social desirability on the response process. This is further made evident by the fact that MC does not load on this factor at all.

The final factor is labeled "social desirability" for obvious reasons. All three social desirability scales involved in the study load highly on the fourth factor, while all other scales do not appear on that factor. Not only has social desirability been removed from the first factor by the DWS scoring technique, but also it is independent of the stability hierarchy obtained from the MMPI scales.

A major question can be raised concerning the scores of the various techniques. Are the scores of the traditional scoring technique equivalent to the other methods or are they measuring the same thing? In order to answer this question, the repeated analyses of variance were computed for each scale. These results indicated several important relationships among the scoring techniques.

The most important aspect revealed by the ANOVAs was that the scores of the DWS were similar to those of the traditional, and when appropriate to the traditional K-corrected scores. For example, there were only significant main effects for scoring technique when K-corrected scores were available. This indicates that the DWS scores

produce K-corrected scores automatically and reduce social desirability influence at the same time. Thus for the DWS scoring technique, no K-corrections need be made. The subtlety incorporated in the scoring technique appeared to pick up the individual's true response more readily. In visual examination of the scores of individuals across the scoring techniques, one finds a great deal of individual variability. In other words, an individual might have the DWS score equivalent to the K-corrected score for one scale but not for another, for example, Pt and Sc. This indicates that DWS is not just automatically correcting for social desirability influences, but is producing a more accurate estimate of the individual's true score.

The results of the tests for skewedness and kurtosis indicated that for men there were no major changes in the shapes of the distributions, while for women there were reductions in divergences from normality. This could possibly be attributed to the nature of the men's scores. For most of the ANOVAs sex differences were present and the direction of difference was for the men to be more deviant from the mean than the women. The extreme nature of the men's scores could be the reason for the differences in the effect of scoring technique on the normalization of the distributions. Therefore, at this point it would be presumptuous to state that DWS scoring produces score distributions which are more normal than traditional scoring techniques.

To summarize these results, one can look at the hypotheses stated in the second chapter. The first hypothesis is supported in that differential weighting, dichotomous splitting and traditional scoring techniques produce different factor patterns with respect to the social desirability. These patterns lead one to conclude that the

amount of social desirability influence in the DWS scores is greatly reduced. The expectation of the second hypothesis was marginally supported. Scoring techniques did produce significant differences but only for the five scales which require K-correction. Significant differences were obtained for the MC scale, but this is possibly due to the fact that it is a social desirability scale. The third hypothesis was not supported, but some evidence was obtained from the scores of women that a tendency for normalization existed when the DWS technique was used. It can be concluded that the DWS scoring technique is better than traditional procedures. However, further research needs to be done to ascertain the exact effects of using DWS on the interpretation of the scores obtained from that procedure. Several suggestions can be made. First of all more subjects should be used. This may not change the results of the factor analyses and ANOVAs, but the shape of the distribution will be more sensitive to the actual shape of the distribution and will pick up changes more readily. A second procedure that is suggested is to use a between subjects design and match subjects by previous clinical classification. This would be similar to renorming the present test. A final consideration would be to use a less homogeneous group in order to get a greater variability in the scores.

Theoretical Implications

The findings of the present study offer support for the threshold theory of responding. Since the differential weighting was developed using the logic of the theory, it is not surprising that the results support the threshold response process. The threshold theory of

responding postulates that the desirability of an item interacts with two subject parameters, salience and threshold, to produce a response. The threshold is the critical level at which the subject makes the transition from a true to a false endorsement of the item. On the other hand, the salience parameter indicates the sensitivity of the subject to the desirability dimension.

Rogers (1971) points out that the threshold theory accounts for the factor structure of the MMPI based on several Monte Carlo studies. The threshold theory is a stylistic theory and does not deal with item content. Similarly, the DWS scoring technique does not use item content to reduce social desirability. It appears that focusing on item content has not been a suitable solution to the problem as the literature indicates, and the research both in the present study and the study by Rogers offers a feasible alternative to the social desirability issue.

The major implication of these results is to demonstrate that the process orientation has the potential to explain the test responses of an individual. These results de-emphasize item content. Concentration on the responding process has several advantages. Not only does it aid in the understanding of the measurement process, but it could be important in the development of advances in cognitive research. A second advantage of the process orientation is that it offers a framework for a more humanistic approach to the measurement process. In the past, most researchers have emphasized the item's content and neglected the individuals' process of responding.

Given the results of the present study, the process orientation is an effective approach to measurement. As the research progresses,

the role of the individual in the testing situation will become more clear. The error variance in the test score will be reduced and a more accurate picture of the individual will be made. Hopefully, these results will aid in resolving the differences between the S-R and R-R approaches to psychology. Cronbach (1957) suggests that the S-R orientation focuses on between treatments' differences, while the R-R approach places emphasis on between subjects' variance. Rogers (1971) points out that the threshold theory of responding offers an opportunity for treatment-subject interactions. The development of such a compromise could aid in moving the psychological sciences out of their present quagmire.

Summary

The purpose of the present study was to investigate the effects of different scoring techniques on the reduction of error variance in an individual's test score. Specifically, the study investigated the reduction of social desirability. The study emphasized the response process rather than the item's content.

Four scoring techniques were used: traditional true/false, traditional true/false with K-correction dichotomous splitting based on five response alternatives, and differential weighting of five response alternatives. The instrument used to test the various scoring techniques was the Minnesota Multiphasic Personality Inventory. This was chosen for several reasons, but primarily because of the voluminous research which has been published on the instrument and the fact that it is the most frequently used personality instrument in the clinical setting.

A within-subjects design was used with each subject being scored on fourteen variables using all four techniques. Factor analyses were used to determine the reduction of social desirability by investigating the loading of the Edwards' Social Desirability scale on the first factor of the MMPI. Repeated measures analysis of variance investigated sex and scoring technique differences for each variable included in the study in order to determine the equivalence of scores for each scoring technique. A final statistical analysis was performed to test the hypothesis that the differential weighting system would have a more normal distribution than the other scoring techniques. Measures of skewedness and kurtosis were employed to test this hypothesis.

The results indicated that social desirability had less influence when using the differential weighting technique. A hierarchial factor solution was obtained as the factor structure of the MMPI. A general "intra-individual" stability factor was initially extracted. Two minor group factors were related to this general factor, psychoticism and neuroticism. Independent of this hierarchy was a factor of social desirability. The differential weighting system not only reduced social desirability, but produced a novel factor solution to the MMPI.

The repeated measures analysis of variance revealed that the differential weighting system produced scores equivalent to the traditional except when scales required K-correction. In the case of those scales, the differential weighting system yielded scores which were similar to the K-corrected scores. Thus, two important findings resulted. On the one hand, differential weighting reduced social desirability influence in the response process; while, on the other

hand, it produced mean scores which were similar to those which would have been produced using the traditional method of scoring.

The results of the skewedness and kurtosis of the distributions produced by each scoring technique were not as clear cut. For women, there was a tendency for the differential weighting scores to be somewhat more normal than those of the traditional scoring technique. However, the men did not produce such results. There was no difference across the scoring techniques for the men's scores. Therefore, it cannot be concluded that at present there is substantial evidence for the differential weighting scores to produce distributions which are more normal in shape than those produced by traditional scoring.

Two major points can be made from the results of the present study. First, the differential weighting system is a better scoring method than traditional true/false. Second, the development of a process orientation has beneficial implications for theoretical arguments. The movement from a content orientation to a responding orientation can be important in the development of a more humanistic approach to measurement.

In conclusion, the use of differential weighting as an alternative method of scoring personality inventories has been supported. In fact, this technique has several advantages over the traditional true/false strategy. The use of this process oriented analysis also supports the threshold theory of responding and its implications for the development of psychological tests. Although broader researcher needs to be implemented, especially with regard to other types of tests, the effectiveness of the differential weighting system in the present study in reducing the influence of error variance in an individual's

test response indicates that a process orientation is a more viable alternative than a content orientation. It is hoped that a suitable compromise can be reached between the two points of view in order to move the psychological sciences further in their quest of the understanding of man.

SELECTED BIBLIOGRAPHY

Banta, T. J. Social attitudes and response styles. Educational and Psychological Measurement, 1961, 21, 543-557.

Beck, S. J. Rorschach's Test. Vol. 1. Basic processes. New York: Grune and Stratton, 1944.

Block, J. The challenge of response sets. New York: Appleton-Century-Crofts, 1965.

Boe, E. E. and Kogan, W. S. Social desirability response set in the individual. Journal of Consulting Psychology, 1963, 27, 369.

Boe, E. E. and Kogan, W. S. Effects of social desirability instructions on several MMPI measures of social desirability. Journal of Consulting Psychology, 1964, 28, 248-251.

Boe, E. E., Gocka, E. F. and Kogan, W. S. A factor analysis of individual social desirability scale values. Multivariate Behavioral Research, 1966, 1, 287-292.

Burt, C. C. Weighted summation applied to pair comparisons. British Journal of Mathematical and Statistical Psychology, 1956, 9, 63-65.

Burt, C. C. Francis Galton and his contribution to psychology. British Journal of Mathematical and Statistical Psychology, 1962, 15, 1-49.

Comrey, A. L. A factor analysis of items on the Minnesota Multiphasic Personality Inventory K scale. Educational and Psychological Measurement, 1958, 18, 633-639.

Comrey, A. L. Construct validation of the Comrey Personality Scales. Multivariate Behavioral Research, 1970, 5, 469-477.

Couch, A. and Keniston, K. Agreeing response set and social desirability. Journal of Abnormal and Social Psychology, 1961, 62, 175-179.

Cronbach, L. J. Response sets and test validity. Educational and Psychological Measurement, 1946, 6, 475-494.

Cronbach, L. J. Further evidence on response sets and test design. Educational and Psychological Measurement, 1950, 10, 3-31.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J. and Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Crowne, D. P. and Marlowe, D. A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 1960, 24, 349-354.

Dahlstrom, W. G., Welsh, G. S. and Dahlstrom, S. MMPI Handbook, Vol. 1: Clinical Interpretation. Minneapolis: University of Minnesota Press, 1972.

Damarin, F. A latent-structure model for answering personal questions. Psychological Bulletin, 1970, 73, 23-40.

DeSoto, C. B., Keuthe, J. L., and Bosley, J. J. A redefinition of social desirability. Journal of Abnormal and Social Psychology, 1959, 58, 273-275.

Diers, C. J. Social desirability and acquiescence in response to personality items. Journal of Consulting Psychology, 1964, 28, 71-77.

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, 1953a, 37, 90-93.

Edwards, A. L. Manual for the Edwards Personal Preference Schedule. New York: Psychological Corporation, 1953b.

Edwards, A. L. Social desirability and Q sorts. Journal of Consulting Psychology, 1955, 19, 462.

Edwards, A. L. The social desirability variable in personality assessment and research. New York: The Dryden Press, 1957.

Edwards, A. L. Social desirability and the probability of endorsement of items in the Interpersonal Check List. Journal of Abnormal and Social Psychology, 1958, 57.

Edwards, A. L. Measurement of personality traits by scales and inventories. New York: Holt, Rinehart and Winston, 1970.

Edwards, A. L., Diers, C. J. and Walker, J. N. Response set and factor loadings on 61 personality scales. Journal of Applied Psychology, 1962, 46, 220-225.

Feldman, M. J. and Corah, N. L. Social desirability and the forced choice method. Journal of Consulting Psychology, 1960, 24, 480-482.

Fiske, D. W. Items and persons: Formal duals and psychological differences. Multivariate Behavioral Research, 1968, 3, 393-401.

Fordyce, W. E. Social desirability in the MMPI. Journal of Consulting Psychology, 1956, 20, 171-175.

Fricke, B. G. Response set as a suppressor variable in the OAIS and MMPI. Journal of Consulting Psychology, 1956, 20, 161-169.

Galton, F. Hereditary genius, an inquiry into its laws and consequences. London: University of London Press, 1869.

Hanley, C. Social desirability and responses to items from three MMPI scales D, Sc, K. Journal of Applied Psychology, 1956, 40, 324-328.

Hanley, C. Deriving a measure of test-taking defensiveness. Journal of Consulting Psychology, 1957, 21, 391-397.

Hathaway, S. R. and McKinley, J. C. Manual for the MMPI. New York: Psychological Corporation, 1940.

Hathaway, S. R. and McKinley, J. C. Manual for the MMPI (Rev. Ed.). New York: Psychological Corporation, 1951.

Heilbrun, A. B. The psychological significance of the MMPI K scale in a normal population. Journal of Consulting Psychology, 1961, 25, 486-491.

Heilbrun, A. B. and Goodstein, L. D. Relations between personal and social desirability sets and performance on the Edwards Personal Preference Schedule. Journal of Applied Psychology, 1959, 43, 302-305.

Hempal, C. G. Fundamentals of concept formation in empirical science. In O. Neurath (Ed.), International Encyclopedia of Unified Science. Chicago: University of Chicago Press, 1952.

Horst, P. and Wright, C. E. The comparative reliability of two techniques of personal appraisal. Journal of Clinical Psychology, 1959, 15, 388-391.

Hotelling, H. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 1933, 24, 417-441, 498-520.

Hotelling, H. Simplified calculations of principal components. Psychometrika, 1936, 1, 27-35.

Husek, T. R. Acquiescence as a response set and as a personality characteristic. Educational and Psychological Measurement, 1961, 21, 295-308.

Jackson, D. N. A treshold model for stylistic responding. Paper read at the American Psychological Association, San Francisco, September, 1968.

Kaiser, H. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Kassebaum, C. G., Couch, A. S., and Slater, P. E. The factorial dimensions of the MMPI. Journal of Consulting Psychology, 1959, 23, 226-236.

Kenny, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. Journal of Consulting Psychology, 1956, 20, 315-318.

Kuncel, R. B. Response processes and relative location of subject and item. Educational and Psychological Measurement, 1973, 33, 545-563.

Levonian, E., Comrey, A. L., Levy, W. and Proctor, D. A statistical evaluation of the Edwards Personal Preference Schedule. Journal of Applied Psychology, 1959, 43, 355-359.

Lord, F. M. A theory of test scores. Psychometric Monographs, 1952, No. 7.

McKinley, J. C., Hathaway, S. R., and Meehl, P. E. The MMPI: VI. The K Scale. Journal of Consulting Psychology, 1948, 12, 20-31.

Marlowe, D. and Crowne, D. P. Social desirability and response to perceived situational demands. Journal of Consulting Psychology, 1961, 25, 109-115.

Meehl, P. E. An investigation of a general normality or control factor in personality testing. Psychological Monographs, 1959, 59, No. 4.

Meehl, P. E. and Hathaway, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. Journal of Applied Psychology, 1946, 36, 525-564.

Messick, S. Dimensions of social desirability. Journal of Consulting Psychology, 1960, 24, 279-287.

Messick, S. Desirability judgements and inventory responses in the assessment of personality. Paper read at the American Psychological Association, Philadelphia, September, 1963.

Messick, S., and Jackson, D. N. Acquiescence and the factorial interpretation of the MMPI. Psychological Bulletin, 1961, 58, 299-304.

Mosher, D. L. Approval motive and acceptance of "fake" personality test interpretations which differ in favorability. Psychological Reports, 1965, 17, 395-402.

Navran, L. and Stauffacher, J. C. Social desirability as a factor in the Edwards Personal Preference Schedule performance. Journal of Consulting Psychology, 1954, 18, 442.

Norman, W. T. On estimating psychological relationships: Social desirability and self report. Psychological Bulletin, 1967, 67, 273-293.

Peabody, D. Attitude content and agreement set in scales of authoritarianism, dogmatism and economic conservatism. Journal of Abnormal and Social Psychology, 1961, 63, 1-11.

Rogers, T. B. The process of responding to personality items: Some issues, a theory and some research. Multivariate Behavioral Research Monographs, 1971, 6, No. 2.

Rorer, L. G. The great response-style myth. Psychological Bulletin, 1965, 63, 129-156.

Rosen, E. Self appraisal, personal desirability, and perceived desirability of personality traits. Journal of Abnormal and Social Psychology, 1956, 52, 151-158.

Ryan, L. R. The FIRO-B: Its advantages and uses in the clinical setting. Colloquim given at Oklahoma State University, November, 1972.

Scedecor, G. W. Statistical Methods. Ames, Iowa: Iowa State University Press, 1956.

Schill, T. R. Effects of approval motivation and varying conditions of verbal reinforcement on incidental memory for faces. Psychological Reports, 1966, 19, 55-60.

Seeman, S. "Subtlety" in structured personality tests. Journal of Consulting Psychology, 1952, 16, 278-283.

Sheldon, W. H. The varieties of human physique: An introduction to constitutional psychology. New York: Harper, 1940.

Smith, E. E. Defensiveness, insights and the K scale. Journal of Consulting Psychology, 1959, 23, 275-277.

Spilka, B., Horn, J. and Langerderfer, L. Social desirabilities among measures of social desirability. Educational and Psychological Measurement, 1966, 26, 111-120.

Stevens, S. S. Mathematics, measurement and psychophisics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: John Wiley and Sons, 1951.

Stone, L. A. Social desirability and order of item presentation in the MMPI. Psychological Reports, 1965, 17, 518.

Stone, L. A. Corrected (for curtailment) correlations between the Marlowe-Crowne Social Desirability Scale and the MMPI. Psychological Reports, 1966, 19, 103-106.

Taylor, J. B. Social desirability and MMPI performance: The individual case. Journal of Consulting Psychology, 1959, 23, 514-517.

Thurstone, L. L. Multiple Factor Analysis. Chicago: University of Chicago Press, 1947.

Torgerson, W. S. Theory and methods of scaling. New York: John Wiley and Sons, 1958.

Tyler, F. T. A factorial analysis of fifteen MMPI scales. Journal of Consulting Psychology, 1951, 15, 303-320.

Watson, R. I. The great psychologists. New York: J. B. Lippincott Co., 1971.

Wheeler, W. M., Little, K. B. and Lehner, G. F. The internal structure of the MMPI. Journal of Consulting Psychology, 1951, 15, 134-141.

Wiener, D. Subtlety and obvious keys for the MMPI. Journal of Consulting Psychology, 1948, 12, 164-170.

Wiggins, J. S. Intercorrelations among MMPI measures of dissimulation under standard and social desirability instructions. Journal of Consulting Psychology, 1959, 23, 419-427.

Wiggins, J. S. Convergences among stylistic response measures from objective personality measures. Educational and Psychological Measurement, 1964, 24, 551-562.

Wiggins, N. Individual viewpoints of social desirability. Psychological Bulletin, 1966, 66, 68-77.

Wikoff, R. L. Personality correlates of psychometric response styles. Unpublished doctoral dissertation, Oklahoma State University, 1965.

Wright, C. E. Relations between normative and ipsative measures of personality. Unpublished doctoral dissertaion, University of Washington, 1957.

Younger, K. W. Selective factor rotation: A rotational procedure yielding hierarchial factor patterns. Unpublished master's thesis, Oklahoma State University, 1974a.

Younger, K. W. A revised program for multiple group factor solutions. Unpublished manuscript, Oklahoma State University, 1974b.

APPENDIXES

APPENDIX A

SCALES

The scales used in this investigation are presented in this section. The items for the Minnesota Multiphasic Personality Inventory can be obtained from:

The Psychological Corporation
304 East 45th Street
New York, New York 10017

Edwards' Social Desirability Scale

1. Criticism or scolding hurts me terribly. (False)
2. When in a group of people I have trouble thinking of the right things to talk about. (False)
3. It does not bother me particularly to see animals suffer. (False)
4. I am never happier than when alone. (False)
5. My family does not like the work I have chosen. (False)
6. I am not afraid to handle money. (True)
7. No one cares much what happens to you. (False)
8. I am liked by most people who know me. (True)
9. Most of the time I would rather sit and daydream than do anything else. (False)
10. I can easily make other people afraid of me, and sometimes do for the fun of it. (False)
11. I usually expect to succeed in things I do. (True)
12. It makes me uncomfortable to put on a stunt at a party even when others are doing the same sort of thing. (False)

13. I have had periods in which I carried on activities without knowing later what I had been doing. (False)

14. I have reason for feeling jealous of one or more members of my family. (False)

15. It makes me impatient to have people ask my advice or otherwise interrupt me when I am working on something important. (False)

16. My parents and family find more fault with me than they should. (False)

17. People often disappoint me. (False)

18. I dream frequently about things that are best kept to myself. (False)

19. My hands and feet are usually warm. (True)

20. I am very seldom troubled by constipation. (True)

21. I find it hard to keep my mind on a task or job. (False)

22. My sleep is fitful and disturbed. (False)

23. I am happy most of the time. (True)

24. I cry easily. (False)

25. I do not tire quickly. (True)

26. I frequently notice my hand shakes when I try to do something. (False)

27. I sweat very easily on cool days. (False)

28. Life is a strain for me much of the time. (False)

29. I am easily embarrassed. (False)

30. I cannot keep my mind on one thing. (False)

31. I feel anxiety about something or someone almost all of the time. (False)

32. I have been afraid of things or people that I knew could not hurt me. (False)

33. I am not usually self-conscious. (True)

34. I am hungry almost all of the time. (False)

35. I worry quite a bit over possible misfortunes. (False)

36. It makes me nervous to have to wait. (False)

37. I blush more often than others. (True)

38. I sometimes feel that I am about to go to pieces. (False)

39. I shrink from facing a crisis or difficulty. (False)

Marlowe-Crowne Social Desirability Scale

1. I have almost never felt the urge to tell someone off. (True)

2. I am sometimes irritated by people who ask favors of me. (False)

3. I am always courteous, even to people who are disagreeable. (True)

4. There have been many occasions when I have felt like smashing things. (False)

5. I always try to practice what I preach. (True)

6. I don't find it particularly difficult to get along with loud mouthed obnoxious people. (True)

7. It is sometimes hard for me to go on with my work if I am not encouraged. (True)

8. I have never deliberately said something that hurt someone's feelings. (True)

9. I sometimes feel resentful when I don't get my way. (False)

10. At times I have really insisted on having things my own way. (False)

11. I would never think of letting someone else be punished for my wrongdoings. (True)

12. I have never felt that I was punished without cause. (True)

13. I never make a long trip without checking the safety of my car. (True)

14. I never resent being asked to return a favor. (True)

15. I am always careful about my manner of dress. (True)

16. No matter who I'm talking to, I'm always a good listener. (True)

17. I can remember "playing sick" to get out of something. (False)

18. I'm always willing to admit it when I make a mistake. (True)

19. There are times when I have been quite jealous of the good fortune of others. (False)

20. On occasion I have had doubts about my ability to succeed in life. (False)

21. When I don't know something I don't mind admitting it. (True)

22. There have been times when I have felt like rebelling against people in authority even though I knew they were right. (False)

23. I have been irked when people expressed ideas different from my own. (True)

24. I have never intensely disliked anyone. (True)

25. Before voting I thoroughly investigate the qualifications of all the candidates. (True)

26. I sometimes think when people have a misfortune they only got what they deserved. (False)

27. I never hesitate to go out of my way to help someone in trouble. (True)

28. On a few occasions, I have given up doing something because I thought too little of my ability. (False)

29. There have been occasions when I took advantage of someone. (False)

30. My table manners at home are as good as when I eat in a restaurant. (True)

31. There have been occasions when I felt like smashing things. (False)

32. If I could get into a movie without paying and be sure I was not seen I would probably do it. (False)

APPENDIX B

INSTRUCTIONS

The purpose of this experiment is to study the mathematical foundations of test construction. Your task will be to read each statement in the test booklet and decide how that statement best applies to you.

You are to mark your answers on the answer sheet you have. Look at the example of a correctly marked answer sheet to the right. First you are to decide which of the following alternatives is most appropriately applied to you with respect to that statement (a - ALWAYS TRUE, b - OFTEN TRUE, c - EQUALLY TRUE AND FALSE, d - OFTEN FALSE, e - ALWAYS FALSE). Then you are to decide whether each statement is true or false as applied to you. Do NOT circle your answers, MARK THEM OUT WITH AN 'X'.

Section of an answer sheet correctly marked

1.	X	b	c	d	e	T	F
2.	a	b	c	X	e	T	X
3.	a	b	X	d	e	T	X

Remember to give your own opinion of your self. Give your first response to the statements. If some are unclear, try to answer them as best you can. Do not leave any statements unanswered. In marking your answers on the answer sheet be sure that the last two digits of the statement number correspond to the item number on the answer sheet.

Do NOT write your name on either the test booklet or the answer sheets. In order to ensure complete anonymity, place your answer sheets in the box labelled 'ANSWER SHEETS' and return your test booklet. When you have done this, please sign the extra credit sheet making sure you indicate which class the credit is to be reported.

The questions are designed to cover many aspects of personal thoughts, feelings and behavior. They were also designed to be used with a wide variety of people. Some may seem trivial, while others quite serious. However all are important. Therefore, it is important that you answer each question. Since complete anonymity is guaranteed, it is hoped that you will finish the questions so that your data may be used in the analysis and thus make the study more meaningful. However, if you feel that you cannot finish answering the questions, you may turn in your answer sheet.

In about three weeks you will receive from your class a brief summary discussing this research area and stating the specific hypothesis this study is investigating. YOUR COOPERATION AND INTEREST IS VERY MUCH APPRECIATED.

INSTRUCTIONS

The purpose of this experiment is to study the mathematical foundations of test construction. Your task will be to read each statement in the test booklet and decide how that statement best applies to you.

You are to mark your answers on the answer sheet you have. Look at the example of a correctly marked answer sheet to the right. First you are to decide whether each statement is true or false as applied to you.

Section of an answer sheet correctly marked

1. T F a b c d e
2. T F a b c d e
3. T F a b c d e

Then you are to decide which of the following alternatives is most appropriately applied to you with respect to that statement (a - ALWAYS TRUE, b - OFTEN TRUE, c - EQUALLY TRUE AND FALSE, d - OFTEN FALSE, e - ALWAYS FALSE). Do NOT circle your answers, MARK THEM OUT WITH AN 'X'.

Remember to give your own opinion of yourself. Give your first response to the statements. If some are unclear, try to answer them as best you can. Do not leave any statements unanswered. In marking your answers on the answer sheet be sure that the last two digits of the statement number correspond to the item number on the answer sheet.

Do NOT write your name on either the test booklet or the answer sheets. In order to ensure complete anonymity, place your answer sheets in the box labelled 'ANSWER SHEETS' and return your test booklet. When you have done this, please sign the extra credit sheet making sure you indicate which class the credit is to be reported.

The questions are designed to cover many aspects of personal thoughts, feelings and behavior. They were also designed to be used with a wide variety of people. Some may seem trivial, while others quite serious. However all are important. Therefore, it is important that you answer each question. Since complete anonymity is guaranteed, it is hoped that you will finish the questions so that your data may be used in the analysis and thus make the study more meaningful. However, if you feel that you cannot finish answering the questions, you may turn in your answer sheet.

In about three weeks you will receive from your class a brief summary discussing this research area and stating the specific hypothesis this study is investigating. YOUR COOPERATION AND INTEREST IS VERY MUCH APPRECIATED.

APPENDIX C

TABLE XXIV

THE FIRST THREE FACTORS TRADITIONALLY DERIVED BY PRINCIPAL COMPONENTS WITH VARIMAX ROTATION

Variable	Factor 1	Factor 2	Factor 3
F	-.55	.07	-.30
Hs	-.52	.18	.46
D	-.58	.45	.15
Hy	-.21	.09	.77
Pd	-.61	-.13	.24
Mf	-.45	-.12	.03
Pa	.14	-.09	.56
Pt	-.83	.25	.10
Sc	-.47	.21	-.05
Ma	-.35	-.25	.35
ESD	.66	-.26	.24
Si	-.30	.68	-.56

Source: Block (1965, p. 51)

APPENDIX D

TABLE XXV

ANALYSIS OF VARIANCE SUMMARY TABLE FOR MC

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	625.43	7.65	0.0066
Subj. W. Groups	82	81.77		
Within Subjects				
Scoring Technique (B)	2	3007.02	36.92	0.0001
A x B	2	238.38	2.91	0.0663
B x Subj. W. Groups	164	81.43		

TABLE XXVI

ANALYSIS OF VARIANCE SUMMARY TABLE FOR L

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	320.06	2.20	0.1366
Subj. W. Groups	82	159.04		
Within Subjects				
Scoring Technique (B)	2	2431.69	15.29	0.0001
A x B	2	99.36	0.68	0.5114
B x Subj. W. Groups	164			

TABLE XXVII

ANALYSIS OF VARIANCE SUMMARY TABLE FOR F

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	7524.32	41.24	0.0001
Subj. W. Groups	82	182.42		
Between Subjects				
Scoring Technique (B)	2	520.15	2.69	0.0696
A x B	2	112.73	0.61	0.5468
B x Subj. W. Groups	164	192.90		

TABLE XXVIII

ANALYSIS OF VARIANCE SUMMARY TABLE FOR K

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	2040.03	9.30	0.0032
Subj. W. Groups	82	219.28		
Within Subjects				
Scoring Technique (B)	2	611.57	2.93	0.0553
A x B	2	29.39	0.13	0.8748
B x Subj. W. Groups	164	208.48		

TABLE XXIX

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Hs

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	3018.00	36.74	0.0001
Subj. W. Groups	82	82.14		
Within Subjects				
Scoring Technique (B)	3	1978.74	138.46	0.0001
A x B	3	17.69	1.19	0.3126
B x Subj. W. Groups	246	14.29		

TABLE XXX

ANALYSIS OF VARIANCE SUMMARY TABLE FOR D

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	5742.89	16.86	0.0003
Subj. W. Groups	82	340.60		
Within Groups				
Scoring Technique (B)	2	231.75	1.82	0.1637
A x B	2	30.01	0.25	0.7800
B x Subj. W. Groups	164	118.48		

TABLE XXXI

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Hy

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	1000.01	4.20	0.0410
Subj. W. Groups	82	238.07		
Within Subjects				
Scoring Technique (B)	2	225.24	2.61	0.0752
A x B	2	23.09	0.25	0.7814
B x Subj. W. Groups	164	91.90		

TABLE XXXII

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Pd

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	500.09	4.05	0.0437
Subj. W. Groups	82	123.55		
Within Subjects				
Scoring Technique (B)	3	1978.74	16.77	0.0001
A x B	3	100.49	0.81	0.5124
B x Subj. W. Groups	246	117.94		

TABLE XXXIII

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Mf

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	497.28	6.82	0.0098
Subj. W. Groups	82	66.36		
Within Subjects				
Scoring Technique (B)	2	36.21	1.98	0.0946
A x B	2	180.16	9.89	0.0001
B x Subj. W. Groups	164	18.42		

TABLE XXXIV

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Pa

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	5497.34	67.01	0.0001
Subj. W. Groups	82	82.04		
Within Subjects				
Scoring Technique (B)	2	77.27	0.46	0.6338
A x B	2	27.63	0.33	0.7197
B x Subj. W. Groups	164	165.31		

TABLE XXXV

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Pt

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	3895.05	6.70	0.0110
Subj. W. Groups	82	581.16		
Within Subjects				
Scoring Technique (B)	3	8886.31	363.29	0.0001
A x B	3	61.28	2.51	0.0585
B x Subj. W. Groups	246	24.46		

TABLE XXXVI

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Sc

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	8122.33	9.87	0.0027
Subj. W. Groups	82	822.51		
Within Subjects				
Scoring Technique (B)	3	6533.13	313.60	0.0001
A x B	3	40.64	1.93	0.0672
B x Subj. W. Groups	246	20.83		

TABLE XXXVII

ANALYSIS OF VARIANCE SUMMARY TABLE FOR Ma

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	3018.00	8.39	0.0050
Subj. W. Groups	82	359.64		
Within Subjects				
Scoring Technique (B)	3	559.50	55.54	0.0001
A x B	3	17.69	1.75	0.1546
B x Subj. W. Groups	246	10.07		

TABLE XXXVIII

ANALYSIS OF VARIANCE SUMMARY TABLE FOR ESD

Source	DF	MS	F	Observed Significance Level
Between Subjects				
Sex (A)	1	3293.34	20.01	0.0001
Subj. W. Groups	82	164.55		
Within Subjects				
Scoring Technique (B)	2	108.62	1.52	0.2137
A x B	2	41.19	0.62	0.5638
B x Subj. W. Groups	164	70.39		

VITA

Kenneth Wayne Younger

Candidate for the Degree of

Doctor of Philosophy

Thesis: SOCIAL DESIRABILITY AS A FUNCTION OF SCORING TECHNIQUE: A COMPARATIVE ANALYSIS

Major Field: Psychology

Biographical:

Personal Data: Born in Jacksonville, Florida, October 26, 1951, the son of Mr. and Mrs. Ralph B. Younger, Jr. Married the former Nancy Lynn Ford, June 5, 1971.

Education: Graduated from Englewood High School, Jacksonville, Florida in June of 1969; received the Bachelor of Arts degree from Carson-Newman College in 1972, with majors in mathematics and psychology; received the Master of Science degree from Oklahoma State University in 1974 in general psychology; received the Degree of Doctor of Philosophy from Oklahoma State University in July, 1975.

Professional Experience: Served as a Teaching Assistant, Department of Psychology, Oklahoma State University, 1972-1975.

Professional and Honorary Organizations: Phi Kappa Phi, Sigma Xi, Psi Chi, Kappa Mu Epsilon, Midwestern Psychological Association.