A METHOD FOR TESTING HYPOTHESES AGAINST A STANDARDIZED PERSONALITY SCALE WITH CONTRIBUTIONS TO THE THEORY OF TEST CONSTRUCTION

A METHOD FOR TESTING HYPOTHESES AGAINST A STANDARDIZED PERSONALITY SCALE WITH CONTRIBUTIONS TO THE THEORY OF TEST CONSTRUCTION

By

THOMAS PERSHING HELMS

Bachelor of Arts

Oklahoma Agricultural and Mechanical College

Stillwater, Oklahoma

1949

Submitted to the Department of Psychology

Oklahoma Agricultural and Mechanical College

In Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

1950

APPROVED BY:

Melvin S Rigg
Chairman, Thesis Committee

N. K. Boggs
Member of the Thesis Committee

A L Reed
Head of the Department

M. E. Harlan
Dean of the Graduate School

PREFACE

In recent years a number of well constructed and well standardized personality tests have made their appearance upon the psychological scene in the United States. Interest has attached itself to those tests which aim at measurement of neurotic and psychotic personality deviations. This is doubtless due to the increasing popular interest in the abnormal and is an attempt to provide a rational and conventional method for understanding and diagnosing such deviations in personality. It is our opinion, however, that these attempts have only approached the validity of the many good intelligence tests available. The foremost reason for this lack of high validity probably arises from the fact that the tests have been standardized against an arbitrarily established criterion. In most cases the psychologist has worked in conjunction with a psychiatrist or a staff of psychiatrists and has accepted the psychiatric diagnosis of the subjects upon whom the test was standardized as the criterion. Although we are not prepared to suggest a better method, we recognize the possibility of error. The student of abnormal psychology knows that many syndromes are not easily categorized by diagnosis and that a mental disease may pass through several categories before it achieves its final character.

One suspects that underlying these attempts at measurement exists the motive to bring psychoanalytic, psychiatric, and

psychologic theory in line, each with the other. In this study a personality test is utilized as the criterion against which various psychoanalytic and psychologic theories can be validated. The test itself, the Terman, Miles Attitude-Interest Analysis Test, which is really a masculinity-femininity test, has the advantage that its validity and reliability compare favorably with the better intelligence tests. There is little that is arbitrary about the test. The criterion groups against which the test was standardized are the male and female population. The criterion is an objective one and meets the requirements of scientific procedure. The criticism that what it measures is not basic but superfluous can be met by pointing out that it measures the ego-involved traits which are just as important to the total personality as the unconscious determinants. As for the success of the study, the reader is referred to the body of the thesis.

I wish to express my thanks to the various members of the Psychology Department who guided and assisted me in the formulation of the problem, and the collection of the data. Dr. S. L. Reed made department facilities available and offered encouragement which helped me formulate the problem. I respect the thesis committee for their unbiased yet critical evaluation of the problem. Again, my thanks to Dr. S. L. Reed, and to Dr. M. G. Rigg and Mr. J. E. Farris, because without their assistance it would have been impossible for me to have gotten sufficient data. Lastly, my thanks to Professor H. K. Brobst for the discussions I had with him on statistical methods.

Thomas P. Helms

TABLE OF CONTENTS

	Page
STATEMENT OF THE PROBLEM.	1
THE SAMPLE.	3
MATERIALS USED.	3
METHOD OF COLLECTING DATA	6
ANALYSIS OF THE DATA.	6
THE CONSTRUCTION OF AN M-F TEST FROM THE EXPERIMENTAL ITEMS.	16
AN EXPERIMENT IN ITEM WEIGHTING	19
THEORETICAL CONSIDERATIONS.	22
DETECTION OF THE AFFECTED RESPONSE.	24
SUMMARY OF M-F FINDINGS	26

A METHOD FOR TESTING HYPOTHESES AGAINST A STANDARDIZED PERSONALITY SCALE WITH CONTRIBUTIONS TO THE THEORY OF TEST CONSTRUCTION

Statement of the Problem

The problem is a general one. There are too many controversies and too few ways of settling them, especially in the field of personality. The reason for this is disagreement on basic definitions and the lack of absolute criteria for making comparisons. In this study a non-arbitrary criterion of the personality trait masculinity-femininity has been selected and then various other personality phenomena said to be related to masculinity-femininity have been correlated with the criterion to determine agreement and to determine the value of the method as a scientific method of testing hypotheses about personality. This is the main problem. As will be noted in the pages which follow, several other phenomena appeared which merited special consideration.

The criterion is the Terman, Miles Attitude-Interest Analysis Test. The title is a disguise to decrease the possibility of manipulation. The test is generally recognized as the best of the available masculinity-femininity tests. A complete description of the test can be found in Terman and Miles book Sex and Personality, so only the most important features will be given here.[1] The items

---

[1] Lewis M. Terman and Catharine C. Miles, Sex and Personality.

were selected on the basis of their ability to differentiate the males and females and are therefore not arbitrary. Complete standardization includes more than six thousand subjects. The reliability computed by the split-half method is .92 for combined sex groups and .72 for single sex groups. Its validity derives as a consequence from the way it was constructed. The average overlap for comparable sex groups is 8 per cent.

Form A of the test was used. In order to diminish the time required for administration, subtest 7 was eliminated. This was a necessary step. Conditions under which the tests were administered imposed a time limit of 50 minutes. Normal administration time for the test is 40 to 50 minutes, and as 29 additional items were added as experimental material, time requirements would have been prohibitive. Actually, elimination of subtest 7 has little effect upon total score. The total score is the algebraic sum of the weighted subtest scores. There are 42 items in subtest 7, but the weight is one-third of total so the greatest possible change subtest 7 can produce is 14. However, inspection of the distribution reveals that the range between the fifth and ninety-fifth centiles is but 6 for both sexes. Actually the mean for males on subtest 7 is 0 and for females -1. This means that only one time in ten would the score exceed plus or minus 3 in the case of the males, or plus 2 and minus 4 in the case of the females. Further, it is the least reliable of all the subtests, its coefficient of reliability being but .24. Because of these facts it is felt that little was lost by this step. It should be noted that all responses are scored either plus or minus. The total score is the algebraic sum of the separate responses. The general population mean for females is -70, and for

males, +52. The standard deviations are: for females, 47, and for males, 50.

The Sample

The sample consists of 65 males and 36 females. All subjects were enrolled in psychology classes during the summer session at Oklahoma A. and M. College. The age range for the males was from 18 to 44 with the mean at 25. The range for females was 17 to 46 with the mean at 26.5. The masculinity-femininity scores (hereafter noted as M-F scores) are influenced by age, generally decreasing with advance in age, which means a trend towards femininity in both sexes. Actually, this was desirable as it was intended that the sample should resemble the general population as nearly as possible. The greatest contributor to bias is noted in the fact that the majority of the subjects were enrolled in either the Arts and Science School or the School of Education.

Materials Used

Form A of the Terman, Miles Attitude-Interest Analysis Test was used. The following items which constitute the experimental material were substituted for subtest 7 of the M-F test. The reasons for choosing them will be given in the section on analysis of the data.

Name_____Sex_____Age_____

(check one)

1. I consider it more important that people think well of:

   a. My character, honesty, integrity, etc. . . . . _____
   b. My ability to do a job . . . . . _____

2. I feel the greater self confidence in:
   a. My ability to succeed in my social affairs. . . _____
   b. My ability to succeed in the type of work I wish to do. . . . . _____

3. Supposing you are in a situation which involves doing something which is not serious but which is against your principles and which your friends want you to do and which if not done will disappoint your friends. Which are you most likely to do:
   a. Act to please your friends. . . . . _____
   b. Act in accordance with your principles. . . . . _____

4. Which of the following do you consider the most serious aspect of sexual relations outside of marriage:
   a. The danger of the loss of the respect of friends and associates. . . . . _____
   b. The danger of the loss of self respect. . . . . _____

5. Religion could be replaced by a "common code of ethics."
   a. Yes . . . . . _____
   b. No. . . . . _____

6. When I attempt to understand persons of the opposite sex:
   a. I try to imagine myself in their situation. . . _____
   b. I try to determine their motives by observing their actions . . . . . _____

7. I believe that men and women:
   a. Have basically the same emotional reactions . . _____
   b. Differ greatly in their basic emotional reactions . . . . . _____

8. I believe that men and women:
   a. Have basically the same interests . . . . . _____
   b. Differ greatly in interests . . . . . _____

9. When a woman falls in love the most important thing for her is:
   a. Whether or not the man is her ideal type. . . . _____
   b. Whether or not the man is tender, considerate, and a good provider regardless of type. . . . . _____

10. When a man falls in love the most important thing for him is:
    a. Whether or not the woman is his ideal type. . . _____
    b. Whether or not the woman is tender, considerate, and a good helpmate regardless of type . . . . . _____

11. I have had a dream in which I could fly or sail through the air without being drawn down by gravity:
    a. In my childhood. . . . . _____
    b. Within the last year . . . . . _____
    c. Never. . . . . _____

12. I have had a dream in which I was falling:
    a. In my childhood. . . . . _____
    b. Within the last year . . . . . _____
    c. Never. . . . . _____

13. I have had a dream in which I wanted to run away to escape some danger but could not because of a feeling of paralysis:
    a. In my childhood. . . . . _____
    b. Within the last year . . . . . _____
    c. Never. . . . . _____

14. I have had a habit of counting things that are not important such as bulbs on electric signs and so forth:
    a. In my childhood. . . . . _____
    b. Within the last year . . . . . _____
    c. Never. . . . . _____

15. I frequently find myself worrying about something:
    a. Yes. . . . . _____
    b. No . . . . . _____

16. Much of the time my head seems to hurt all over:
    a. Yes. . . . . _____
    b. No . . . . . _____

17. I frequently notice my hand shakes when I try to do something:
    a. Yes . . . . . _____
    b. No. . . . . _____

18. I think that I feel more intensely than most people do:
    a. Yes. . . . . _____
    b. No . . . . . _____

19. I have certainly had more than my share of things to worry about:
    a. Yes. . . . . _____
    b. No . . . . . _____

20. I have had an intense fear of: (Check ones which apply)

	Ever	Within past year	Never
a. Some type of animal	____	____	____
b. High places	____	____	____
c. Closed places	____	____	____
d. Open places	____	____	____
e. Fire	____	____	____
f. Water	____	____	____
g. The dead	____	____	____
h. The dark	____	____	____
i. Being alone	____	____	____

Method of Collecting Data

The M-F test was administered in accord with the directions of its authors to four groups of approximately twenty-five persons each. The experimental material was presented as a part of the test. The subjects were told that it was a research problem but were not told the nature of the problem. They were told their scores would be kept confidential and they were posted opposite a code number or given to the subjects personally by the test administrator.

Analysis of the Data

All of the experimental items were chosen because of some supposed relationship to masculinity-femininity. All of the M-F scores were then tabulated according to the sex groups and the responses on the experimental items. The tabulated data was then inspected and those items which apparently differentiated the subjects on M-F score were subjected to statistical analysis by the familiar bi-serial correlation technique. It must be understood that the bi-serial r's are in each case for one sex group only. This proves to be a refined analysis because generally the bi-serial r's obtained on single sex groups increase in value when the data

from both sex groups is combined. There is one important esception which will be encountered later.

The critical ratio (CR) of the bi-serial r's obtained was utilized as an indication of the significance of this differentiation. The CR of the bi-serial r is of course the obtained bi-serial r divided by its standard error. The method of computation was taken from Garrett.[2] Those items with significant CR's were understood as being closely related to masculinity-femininity as measured by the M-F test. As will be seen later, it was possible to construct two original M-F tests, one for males and one for females utilizing precisely these items.

The first two items are closely related and derive from the same source. In an investigation of sex differences in introversion-extroversion, Heidbreder[3] found no significant sex differences but found another trait which cut across introversion-extroversion and seemed to bear no relation to it. It was concerned with social relations. The introvert traits marked most often by men were those which would interfere with social adjustments. On the other hand, the women marked those most often which would interfere with efficient work. It was decided to carry this a step further and learn if the same relationship holds for M-F score as well as sex. If such were true, the subjects making high M-F scores would choose response b on experimental item 1, i.e. they would consider it more

---

[2] Henry E. Garrett, Statistics in Psychology and Education, p. 349.

[3] Edna Heidbreder, "Introversion and Extroversion in Men and Women," Journal of Abnormal and Social Psychology, XXII (1927), 52-61.

important that people think of them as efficient workers rather than as good "morally." It is assumed that character, honesty, and integrity are variables in social adjustments, or at least it is important that one is thought of as possessing these attributes if he is to make a social adjustment. Conversely, one would expect the subjects making high M-F scores to choose response b on item 2, i.e. they would express the most confidence in their ability to succeed in the vocation of their choice and the least confidence in their ability to succeed in their social affairs. The results were in the expected direction on item 1. The CR for the males was 6.25, but only 0.32 for the females. Expectations were reversed on item 2 and the male CR was 2.90, the female, 1.35.

Interpretation: There is a significant trend for males making high M-F scores to consider it more important that people think well of their ability to do a job rather than their character, honesty, and integrity. This trend is not significant in the female group. This is probably due to the moral standards imposed upon women. There is a significant trend for the high scoring, and hence more masculine, men to express greater self confidence in their ability to succeed in their social affairs than in the type of work they wish to do. While the trend is in the same direction in the female group the CR of 1.35 is not convincing. A rationalization of this trend is found in the fact that the socially aggressive man is spoken of as masculine and that the socially aggressive woman falls heir to this attribute through cultural determinants. It is proper to indicate at this point that we are dealing with mental masculinity-femininity and that the physical attributes are of

minor importance. (Refer to page 3 for the experimental items.)

Item 3 was calculated to measure the strength of conscience. Freud[4] has stated that the superego is more highly developed in men than in women. It is not at all certain that this single item can be considered valid for indicating the strength of the conscience, or superego. It was originally assumed that subjects with the stronger conscience would choose response b. According to Freud's assertion, one would expect more men than women to choose response b and also for the high M-F scores to correlate with it. Actually, 60 per cent of the men chose response b and 78 per cent of the women chose the same response. Also, the trend for high M-F scores was towards response a. The CR's were 1.50 for the males and 0.512 for the females and are not impressive. Interpretation: Due to the small CR's and ambiguous nature of the item, no conclusions are drawn.

Item 4 was constructed in an attempt to touch upon attitude towards promiscuous sexual relations and its relation to M-F score. The only pertinent feature is that whereas 22 per cent of the males chose response a, only one of the females chose that response. There was no apparent relationship to M-F score except that the one female who chose that response scored very high, in fact two standard deviations above the general population mean for females. Interpretation: It seems safe to strongly suspect that women consistently regard the loss of self respect, of becoming a "fallen woman," as the most serious aspect of promiscuous sexual relations.

Item 5 was included because of a suspicion that attitude towards

---

[4] Sigmund Freud, New Introductory Lectures on Psychoanalysis, pp. 176-177.

religion would reflect itself in the M-F score. Inspection of the data revealed that the bi-serial r would not be significant. The males were about evenly divided on their choice of response but only four of the females indicated that they thought religion could be replaced by a "common code of ethics." However, of these four, all were above the general population mean and two were the highest scores of the female group. Interpretation: There is a sex difference but it does not correlate with M-F score.

Item 6 was constructed on the basis of two considerations. First, there is the popular idea that people attempt to understand one another by imagining themselves to be in one another's circumstance. Also, the understanding type of person is sometimes described as passive and passivity is thought by some to be an attribute of femininity. Secondly, the item inquires if this method of understanding other persons is utilized regarding persons of the opposite sex and aims at detecting an identification with the opposite sex which might have an influence on the M-F score. The only apparent significance was that all of the extremely low scoring females chose response b. No interpretation is justified.

Items 7 and 8 belong to the same context. As Terman and Miles[5] have indicated that the most important sex differences in personality relate to interests and emotional and ethical attitudes it was conjectured that there might be a relation between the subjects' knowledge of this and their M-F scores. It was apparent that the bi-serial r would be extremely small, so none was computed.

---

5 Terman and Miles, op. cit., pp. 447-448.

Items 9 and 10 produced more interesting results. It was supposed that the choice of a mate on the basis of an ideal type might bear a relationship to the deeper stratas of the psyche. There was an inverted relationship between the sexes, i.e. the a response on both items correlated with high M-F scores in women and low M-F scores in men. The CR's are, however, too small to justify any conclusions, being only .886 for females and .587 for males on item 9, and only .512 for females and 1.46 for males on item 10. In a situation like this, bi-serial r computed for the sex groups combined will be lower than the bi-serial r's obtained from the separate sex groups. Later an interpretation of this phenomenon will be ventured.

The next three items treat very common psychic phenomena which have seldom if ever been investigated with a refined statistical technique, namely, dreams. The three dreams chosen were selected because of their universally common occurrence and because psychoanalysts have had something to say about their meaning.

According to Freud[6] the dream of flying may have various interpretations depending upon the dreamer but that quite frequently it is a dream of the erection of the penis. By displacement the body is substituted for the penis and the representation of being able to fly unaided, to overcome the law of gravity, which is in a sense exactly what the male organ can do, and the erotic sensations in the dream are the result of this psychic gratification of the sex impulse; the wish fulfillment. Women with unconscious penis envy might then be expected to have the flying dream as an erection dream

---

[6] Sigmund Freud, The Interpretation of Dreams, pp. 370-371.

and it might be that an unconscious wish of such a nature would effect an influence on M-F score. If these considerations are correct, one would expect a greater occurrence of this dream in adult men than adult women and also a correlation with high M-F score. The results were in the expected direction. Only one female reported the occurrence of the flying dream within the last year, whereas four of the men reported such. The one female was one of the higher scorers, being almost two SD's above the mean for general population females. The bi-serial r's are in the same direction but are not large, being .625 for females, but only .34 for males. These bi-serial r's were computed by taking all subjects who had ever had a dream of flying against all subjects who had never had such a dream (see experimental item 11). Interpretation: Although results were in the expected direction, in order to keep a clear conscience statistically, results are qualified.

In item 12, dealing with the dream of falling, results are more significant. Freud[7] and other psychoanalysts have said the dream of falling occurring in women nearly always represents the wish to experience a moral fall. For this reason and the fact that it is a commonly occurring dream determined its selection. The results were both significant and unaccountable. The occurrence of this dream in females within the last year of their lives correlates with low, highly feminine, M-F score. Conversely, the occurrence within the last year of the falling dream in males correlates with high M-F score, and hence increased masculinity. The CR for females was 3.54 and for males, 1.72. Due to this divergent result on the

---

[7] Ibid., pp. 371-372.

same response, a bi-serial r computed for the male and female groups combined would be reduced to non-significance. Here is an item which is valid for measuring masculinity in males and valid for measuring femininity in females. Interpretation of this phenomenon will be postponed until later.

Item 13, the dream of motor paralysis, the representation usually developing around an attempt to flee from some danger accompanied by an inability to do so because of the feeling of motor inhibition, was included because of its common occurrence and Freud's assertion[8] that the sensation of motor paralysis is the dream's way of representing a strong psychic no against some unconscious wish striving for expression via the dream. Considering the rather popular idea that psychic inhibition bears a relationship to femininity, found in the expression that women are more inhibited, there arises the expectation that this dream will correlate with M-F scores. The bi-serial r's are larger in both sexes than in the preceding item, and the divergence is again observed. The dream, occurring within the past year, correlated with low M-F score in females and high M-F score in males. It is pointed out that it would be impossible to detect such a phenomenon by any other method of item analysis. Such items may be the key to improved test construction techniques, but again interpretation is postponed until a later section.

Items 14 through 19 are taken directly from the Minnesota Multiphasic Personality Inventory. These particular items were chosen because the scoring depends upon the positive response and

---

[8] Ibid., pp. 320-322.

because there is no overlapping with items from the other scales. Items 14 and 15 are from the psycasthenic scale, 16 and 17 from the hysteria scale, and 18 and 19 from the paranoia scale. Freud[9] states that the obsessional neurosis has an affinity for masculinity and that hysteria has an affinity for femininity. Farnham[10] states that the compulsion to count, which item 14 taps, is a universal symptom in the syndrome of the obsessional neurosis. Paranoia was said by Freud[11] to spring from unconscious homosexuality.

Of all these latter items, only 14 produced statistically reliable results. The compulsion for counting, if it occurred within the past year, correlates with masculinity in both sexes. The CR's were: 5.5 for males and 1.77 for females.

Interpretation: Because of the frequency of occurrence of the counting compulsion in the syndrome of the obsessional neurosis and the value of the CR's, it is felt that the statement, "The obsessional neurosis has an affinity for masculinity," is correct and demonstrable by this technique.

CR's for items 16, 17, 18, and 19 were not computed because it was obvious from inspection of the data that no significant values would be obtained.

Interpretation: It is not sure that these items are capable of detecting hysteria and paranoia. The sampling is not large enough. This was anticipated and that is why two from each scale were

---

9 Sigmund Freud, The Problem of Anxiety, p. 84.

10 Ferdinand Lundberg and M. F. Farnham, Modern Woman: The Lost Sex, p. 255.

11 Sigmund Freud, A General Introduction to Psychoanalysis, p. 367.

utilized. It is also well-known that true hysteria is rarely found in a sample such as this and paranoia is rarely ever found in the general population. It is believed that these items are "dead wood" as far as this study is concerned.

Item 20 includes nine common fears or phobias. The results exceeded all expectations. Only one, the fear of high places, was included because of some suspected relationship to masculinity-femininity. Freud[12] has stated that the fear of high places is a manifestation of the death instinct aiming at self-destruction and that the ego's retreat in the face of the destructive impulse has a hidden feminine significance. The CR's, though not large, are in the feminine direction in both sexes. The others were included as a matter of curiousity. There were fewer subjects who confessed a phobia "within the last year" but the data was arranged on the basis of "ever had" versus "never had" and bi-serial r computed with this arrangement. The results are tabulated below. The minus sign indicates that the occurrence of the phobia correlates with femininity.

		Females CR	Males CR
a.	Fear of some animal	Nil	Nil
b.	" " high places	-1.15	-1.63
c.	" " closed places	-5.45	Nil
d.	" " open places	Nil	Nil
e.	" " fires	-2.36	Nil
f.	" " water	-4.00	-2.99
g.	" " the dead	-4.20	Nil
h.	" " the dark	-1.20	Nil
i.	" " being alone	-1.46	-1.47

First, it should be noted that all of the CR's are in the feminine direction. The confession of a phobia in all of these cases

---

[12] Sigmund Freud, The Problem of Anxiety, p. 131.

correlates with low M-F score. A bi-serial *r* utilizing the data from both sex groups was computed for phobia *f*, the fear of water, and the CR was -4.50, which is larger than either of the single sex group CR's. This may be taken as an indication that a true mental masculinity-femininity does exist regardless of sex although, of course, ultimately the concept must be traced to its origin in an association to one of the sexes. Only the bi-serial method could have brought this to light.

Interpretation: Because it is impossible to determine from the data whether or not these responses constitute a phobia in the psychoanalytic sense, one is justified in saying no more than that a confession of them correlates with mental femininity.

The Construction of an M-F Test From the Experimental Items

Proof that the experimental items are measuring the same thing as the criterion is demonstrated in the construction of an alternate test made of the best experimental items. The correlation obtained between the alternate and the criterion was highly significant, far beyond the .01 level.

On the construction of the alternate test, separate forms were made for each sex. This was necessary because there were not enough items which were significant for both sexes and partly because some of the best items must be scored according to the sex of the testee. An example of this latter is in items 12 and 13, the dreams. On the test, the occurrence within the past year of either of these dreams is scored plus in the males, but minus in the females.

The form for the females is given first. Each response is scored plus or minus one, depending upon the sign following the

response blank. To avoid confusion on the items with three responses, instructions to score only one response per item must be given.

The test follows.

12. I have had a dream in which I was falling:
    a. In my childhood. . . . . +
    b. Within the last year . . . . . −
    c. Never. . . . . +

13. I have had a dream in which I wanted to run away to escape some danger but could not because of a feeling of paralysis:
    a. In my childhood. . . . . +
    b. Within the last year . . . . . −
    c. Never. . . . . +

14. I have had a habit of counting things that are not important such as bulbs on electric signs and so forth:
    a. In my childhood. . . . . −
    b. Within the last year . . . . . +
    c. Never. . . . . −

20. I have had an intense fear of: (Check ones which apply)

	Ever	Within past year	Never
b. High places	−	−	+
c. Closed places	−	−	+
e. Fire	−	−	+
f. Water	−	−	+
g. The dead	−	−	+
h. The dark	−	−	+
i. Being alone	−	−	+

The items were standardized on all 36 of the females. Eight of them had left one or more of the items unchecked, so only 28 were available for correlation with the criterion. The correlation coefficient obtained for alternate test versus the criterion was .684. From the tables the $\underline{r}$ for N = 28, P = 1% level, is .478. The results are highly significant. The standard error of estimate when predicting the criterion scores from the alternate test scores is 34.2 or about three-fourths of one standard deviation of the criterion.

Considering that there are only ten items in the alternate, these results are impressive. It is felt that due to the nature of the items the alternate would possess good reliability. The test for the males follows:

(check one)

1. I consider it more important that people think well of:
   a. My character, honesty, integrity, etc. . . . . _____ −
   b. My ability to do a job . . . . . _____ +

2. I feel the greater self confidence in:
   a. My ability to succeed in my social affairs . . . _____ +
   b. My ability to succeed in the type of work I wish to do . . . . . _____ −

3. Supposing you are in a situation which involves doing something which is not serious but which is against your principles and which your friends want you to do and which if not done will disappoint your friends. Which are you most likely to do:
   a. Act to please your friends. . . . . _____ +
   b. Act in accordance with your principles. . . . . _____ −

10. When a man falls in love the most important thing for him is:
   a. Whether or not the woman is his ideal type. . . _____ −
   b. Whether or not the woman is tender, considerate and a good helpmate regardless of type . . . . . _____ +

12. I have had a dream in which I was falling:
   a. In my childhood . . . . . _____ −
   b. Within the last year. . . . . _____ +
   c. Never . . . . . _____ −

13. I have had a dream in which I wanted to run away to escape some danger but could not because of a feeling of paralysis:
   a. In my childhood . . . . . _____ −
   b. Within the past year. . . . . _____ +
   c. Never . . . . . _____ −

14. I have had a habit of counting things that are not important such as bulbs on electric signs and so forth:
   a. In my childhood . . . . . _____ −
   b. Within the last year. . . . . _____ +
   c. Never . . . . . _____ −

20. I have had an abnormal fear of: (Check ones which apply)

	Ever	Within past year	Never
b. High places	——— –	——— –	——— +
f. Water	——— –	——— –	——— +
i. Being alone	——— –	——— –	——— +

The items were standardized on 65 males, but due to omissions on some of the items only 56 were available for the correlation of the alternate against the criterion. The computed r was .6152. From the tables, for N = 56, r at the 1% level is .325. Obviously, results are highly significant. The standard error of estimate in predicting criterion scores from the alternate is 40.7 or about three-fourths of one standard deviation of the criterion. In light of the fact that there are only ten items in the alternate test, these results are impressive.

It is not meant for the alternates to be accepted as a qualified test, although it might be used for group comparisons. Rather, the results are presented as evidence that the procedure is valid.

An Experiment In Item Weighting

In the literature on personality test construction, one finds attempts to improve validity and reliability by weighting the items. Guilford[13] summarizes the methods and shows that little if anything is gained. The results obtained with the weighted items and the method of uniform scoring, plus one and minus one, are in all cases cited by Guilford, almost identical. In this study, several methods of weighting were tried but in no case were results superior to the uniform scoring of the items selected by the bi-serial method.

---

[13] Joy P. Guilford, Psychometric Methods .

In fact, all methods except one proved inferior. As a matter of formality, the methods and results are given here.

In order to reduce computation, the female group, N = 28, was used. The uniform weighting of items selected by the bi-serial method, 10 items with an r of .68 with the criterion, was utilized as a criterion of comparison. To answer the first question, what is the relationship between length of the test, the CR of the items, and the validity of the test, a test of five items with CR's all greater than 3 was constructed. An r was computed and compared with the r of the alternate which includes four items with CR's smaller than 2 but greater than 1. The r for the five-item test was .604, which is considerably lower than the .68 of the alternate, best test. If there is a shortage of items it would seem advisable to add items even though their CR's are less than 2 but greater than 1. When the CR is too low, validity is apparently reduced. The alternate was lengthened to 12 items by the inclusion of two items with CR's less than 1. The r was only .568.

A novel method of weighting was tried. The responses were unevenly divided on all of the items. It was thought that the response occurring less frequently could be considered more difficult and thus more diagnostic of the trait, and therefore should be given a greater weight. The weight was computed as follows: For example, the weight for a given response a was computed as the fraction obtained by placing the number making response a in the denominator and the number making response b in the numerator. The weight for b was taken as the reciprocal of this value. The r computed by this method utilizing the 10 best items was .573 which is inferior to the alternate.

A modified chi-square method of weighting was tried. The subjects were divided into two groups on the basis of having scored above or below the mean of the general population. They were then divided on the basis of response chosen on the experimental items. This arrangement is adapted to analysis by the chi-square test of independence.[14] The chi-square values were computed and those items having a P value of less than .02 were scored plus and minus 2. Those items having P values greater than .02 but less than .16 were weighted plus and minus 1. The r obtained by this method was .618. This method is effective but inferior to the bi-serial technique.

It seemed that those items which differentiated the sexes significantly and also correlated with M-F scores should be the most valid and hence merit a greater scoring weight. All scores were divided on the basis of sex and then again by response on the items. The chi-square test of independence was again utilized. The chi-squares were computed and of the ten items in the alternate test, those having P values of less than .10 were weighted plus and minus 2. The correlation with the criterion was .550.

Since the results were disappointing with the latter method, a sort of inversion of the latter was utilized. The reasoning is as follows: An item which differentiates the sexes and also correlates with M-F score is easier and less diagnostic in an abstract sense than one which only correlates with M-F score. Therefore, only items with a chi-square P value greater than .20 and a bi-serial CR greater than 3 were weighted plus and minus 2. All others were weighted plus and minus 1. This method proved the best and the r obtained was .681, comparable to the .684 of the alternate which utilized

---

[14] George W. Snedecor, Statistical Methods, pp. 194-197.

only bi-serial items without weighting. The items are the same in both, only in the former weighting was on the basis of chi-square values.

Finally, to answer the question is the bi-serial technique superior to a straight chi-square analysis, items 12, 13, 14, and phobias c, e, h, and i, all of which are in the alternate and have P values computed from the chi square test of independence of .15 or less. The obtained r was .478, which is highly significant but not comparable to the r of .68 obtained by the bi-serial technique.

Interpretation: The bi-serial method has again proven to be the superior method of item analysis. There is still the criticism that by this method it is possible to invent traits which do not exist. It is felt that this criticism is not valid if the items are examined individually for psychological validity.

Theoretical Considerations

It has been demonstrated that by use of the bi-serial method it is possible to discover valid items which would otherwise escape detection. Items such as these may be diagnostic of the trait, but not diagnostic of the sample upon which the test is standardized. As an example, the phobia, fear of water, has a low chi-square value and does not differentiate the sexes significantly, but does correlate significantly with M-F scores. Fear of water is characteristic of femininity. One can scarcely imagine a completely masculine man as being afraid of water without some intellectual rationalization to justify the relation. Also, psychoanalysts have something to say about the significance water plays in the unconscious thought

processes underlying neurotic symptoms.[15] Water is utilized symbolically by the thought processes for several purposes. It plays an important part in the symbolization of birth. Recall the Biblical injunction, "You must be born again," by baptism, and how Moses gained a new mother in the rescue from the water. The Oedipus complex is often disposed of in the formation of a phobia.[16] Farnham[17] explains that an extreme Oedipal fixation is often the cause of homosexuality in men. Piecing these considerations together, it becomes clear that a fear of water is a feminine attribute which springs from obscure associations, probably unconscious. It is known also that women are more susceptible to Oedipal fixations than are men.

It is only meant to demonstrate the mechanics of this technique. It is granted that results are only evidence and are not to be considered final.

Items 12 and 13, the dreams of falling and motor paralysis, are unique because the same response is scored differently depending upon the sex of the testee. It is suggested that whatever happens in these two dreams is basic for the development of the trait masculinity-femininity. It may be that after the occurrence of one or both of these dreams the constitutional factor then determines whether the character will develop into a masculine or feminine one, or it may be that environmental factors will be decisive. Only a case study technique could reveal which.

---

[15] Sigmund Freud, A General Introduction to Psychoanalysis, p. 137.

[16] Sigmund Freud, The Problem of Anxiety, pp. 29-41.

[17] Lundberg and Farnham, op. cit., pp. 349-350.

Detection of the Affected Response

It is recognized, especially in the field of industrial psychology, that screening tests almost always eliminate some of the subjects who qualify in terms of the criterion, and admit others who do not. If some correction could be made for this divergence, the validity of screening tests could be greatly improved. In the field of personality testing, this same problem is encountered in a slightly different form. One is never sure that the testee is being psychologically honest. Probably the most important development of this study is the invention of a technique for detecting subjects making these "affected" responses.

In the section on weighting of items, two fundamentally different methods of item selection were used, both of which produced highly significant results. The superior one was the bi-serial method. The other method was the chi-square test of independence. The principle of the latter is demonstrated below.

Suppose 25 males and 25 females make responses to a test item, and the results are tabulated as follows:

	Response No. 1	Response No. 2	
Number of males. . .	23	2	25
Number of females. .	3	22	25
	26	24	50 = N

It is apparent that this item is diagnostic of the sexes. The salient feature is that this method "counts noses" as it were of the subjects making the different responses. If a test is constructed of items selected in this manner, results will be valid as was demonstrated in the earlier section. If a test is constructed by the bi-serial method alone, the criterion has to be an arbitrary

one. Obviously, the chi-square principle, in this case, utilizes an absolute criterion. However, if a test is first constructed using the absolute criterion and then items are validated against it as was done in this study, the bi-serial items are then validated against the absolute criterion. The thesis that by this method the continuous variable found throughout both sex groups is isolated is put forth here. As evidence, consider the dreams, items 12 and 13. The chi-square principle establishes the occurrence of these dreams within the past year to be feminine determinants, the P values being .15 and .085 respectively, whereas the bi-serial method produces the divergent result, i. e. it is scored masculine if occurring in the males and feminine if occurring in the females. In the phobia, fear of high places, its occurrence is scored feminine by the bi-serial method and masculine by the chi-square principle for both sexes. The fact remains that the bi-serial method produced the highest correlation with the criterion. It seems that these abstracted traits are really superior. Now if a subject wants to manipulate the test he calls upon his memory, recalls from his experiences which group the item diagnoses, decides whether or not he wishes to identify himself with that group, and then makes the appropriate response. This is exactly what analysis of the item by the chi-square principle has accomplished. The subject has solved the test. Now suppose he is confronted by the phobia, fear of high places. If he chooses his response the same way the bi-serial scoring will frustrate his attempt. The subject who is psychologically honest will simply read the item, wait for his associations, and put down the appropriate response. The theory is this: The ideal test

should be composed of equal numbers of items chosen by both methods. Items selected by the chi-square principle are necessary for the standardization against the absolute criterion. Items selected by the bi-serial principle are necessary to get the abstract indicators of the trait. The two sets of items should be grouped into two subtests, both measuring the same trait. If responses have been "unaffected" the scores on the two subtests should be similar, and if manipulation has occurred, the scores should be dissimilar.

The only evidence in support of the theory comes from the section on item weighting. It will be recalled that when items having large bi-serial CR values and low chi-square values were weighted most heavily, best results were obtained. All other items were weighted plus and minus 1. This can only mean that the abstract bi-serial items were more diagnostic. The theory is developed, but testing it is beyond the scope of this study.

It is not fair to say that all people giving the affected response have lied. Anxiety and inhibitions can conceivably produce the same result.

Summary of M-F Findings

It seems appropriate to conclude by reviewing the indications of mental masculinity-femininity as revealed by the experimental items. The following statements are generalizations and there are exceptions. Only items utilized in the construction of the alternate tests are considered.

It is masculine to evaluate ability to do a job above moral character. The masculine person is more certain of social than of vocational success. Acting to please friends is indicative of

masculinity, whereas acting in accord with subjective principles correlates with femininity. The dreams of falling and motor paralysis if occurring within the recent life of an individual are indicative of masculinity in males and femininity in females. The compulsion to count things, on a senseless level, is a masculine trait. Although it depends somewhat upon the type, the confession of a phobia, even within the childhood period, indicates a probable trend towards femininity. The number of objects feared is also of importance. These conclusions are only as accurate as the sample is representative.

BIBLIOGRAPHY

Freud, Sigmund. A General Introduction to Psychoanalysis. New York: Liveright Publishing Co., 1920.

Freud, Sigmund. New Introductory Lectures on Psychoanalysis. New York: W. W. Norton and Co., Inc., 1933.

Freud, Sigmund. The Interpretation of Dreams. New York: McMillan Co., 1942.

Freud, Sigmund. The Problem of Anxiety. New York: W. W. Norton and Co., Inc., 1936.

Garrett, H. E. Statistics in Psychology and Education. New York: Longmans, Green and Co., 1947.

Guilford, J. P. Psychometric Methods. New York: McGraw Hill Book Co., 1936.

Lundberg, Ferdinand, and Farnham, M. F. Modern Woman: The Lost Sex. New York: Harper and Brothers, Publishers, 1947.

Snedecor, G. W. Statistical Methods. Ames, Iowa: The Iowa State College Press, 1946.

Terman, L. M., and Miles, C. C. Sex and Personality: Studies in Masculinity and Femininity. New York: McGraw Hill Book Co., 1936.

Typist: Archie C. Thomas