



1953

An Investigation of the Differences in Masculine-Feminine Interests Patterns of Three Groups of Women Students

Richard James Stanek
Loyola University Chicago

Recommended Citation

Stanek, Richard James, "An Investigation of the Differences in Masculine-Feminine Interests Patterns of Three Groups of Women Students" (1953). *Master's Theses*. Paper 1289.
http://ecommons.luc.edu/luc_theses/1289

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



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 1953 Richard James Stanek

AN INVESTIGATION OF THE DIFFERENCES IN MASCULINE-
FEMININE INTERESTS PATTERNS OF THREE
GROUPS OF WOMEN STUDENTS

by

Richard James Stanek

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University in Partial Fulfillment of
the Requirements for the Degree of
Master of Arts

February

1953

Life

Richard James Stanek was born in Oak Park, Illinois, September 13, 1927.

He was graduated from Oak Park River Forest Township High School, Oak Park, Illinois, June, 1945, and from Loyola University, Chicago, Illinois, February, 1950, with the degree of Bachelor of Science.

From 1950 to the present writing the author has attended Loyola University, Chicago, Illinois, as a member of the Graduate School in the Department of Psychology.

Since his entrance into the Graduate School the author has held an assistantship in the Department of Psychology and the Loyola Guidance Center. At present he is an assistant psychologist on the staff of the Loyola Guidance Center.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. REVIEW OF RELATED LITERATURE	5
III. DESCRIPTION OF TEST MATERIALS	21
IV. PROCEDURE--TEST ADMINISTRATION, SCORING, AND METHOD OF ANALYSIS	43
V. ANALYSIS OF RESULTS	54
VI. SUMMARY AND CONCLUSIONS	72
BIBLIOGRAPHY	76
APPENDIX I	80

LIST OF TABLES

Table	Page
I. MEANS, STANDARD DEVIATIONS, AND CRITICAL RATIOS OF SCORES FOR TWO SUB-GROUPINGS OF THE COEDUCATIONAL STUDENTS ON THREE MEASURES OF MASCULINITY-FEMININITY	45
II. DIFFERENCES BETWEEN MEANS AND CRITICAL RATIOS OF SCORES FOR THREE GROUPS OF STUDENTS ON THREE MEASUREMENTS OF MASCULINITY-FEMININITY	55
III. COMPARISON OF GROUP MEAN SCORES AND STANDARD DEVIATIONS WITH SPECIAL REFERENCE TO GROUP TRENDS IN MASCULINITY-FEMININITY	57
IV. COEFFICIENTS OF CORRELATION AND SIGNIFICANCE LEVELS OF CORRELATION ON SCORES AMONG FOUR MEASURES OF MASCULINITY-FEMININITY	60
V. MEAN SCORES, STANDARD DEVIATIONS, DIFFERENCES BETWEEN MEANS, AND CRITICAL RATIOS FOR THREE GROUPS OF STUDENTS ON SEVEN EXERCISES OF THE ATTITUDE-INTEREST ANALYSIS TEST	64
VI. MEAN SCORES AND STANDARD DEVIATIONS OF MMPI SCALES FOR THE WOMEN'S COLLEGE, COEDUCATIONAL, AND NURSING STUDENTS	66
VII. PERCENTAGE OF WOMEN STUDENTS WITH T SCORES ABOVE 70 ON THE SEPARATE SCALES OF THE MMPI	70

LIST OF FIGURES

Figure	Page
1. A GRAPHIC PRESENTATION OF THE PERFORMANCE OF WOMEN'S COLLEGE, COEDUCATIONAL, AND NURSING STUDENTS ON THE MMPI BASED ON MEAN T SCORES . . .	67

CHAPTER I

INTRODUCTION

Recent investigation in the fields of vocational and educational guidance has revealed the importance of factors in personality in addition to intelligence which are necessary for efficiency and success in various occupations. Increased interest has been shown in interests, values, and attitudes, which, other things being equal, give promise of success in a particular occupation or academic curriculum. One of these personality variables, masculinity-femininity, constitutes the subject matter for the present investigation.

The fact that men and women display characteristic sex differences, i.e., either masculine or feminine behavior, can hardly be questioned. These differences lend a distinctive character to the total personality structure for each sex. It is interesting to observe, however, that some differences, long assumed to separate the sexes, were found to be nonexistent. For instance, psychological science refuted the once prevalent view that women were considerably inferior to men in regard to intellectual endowment. Intelligence tests established the fact that intellectual ability was distributed normally throughout the

entire population regardless of sex. Psychometric measurement has shown the equality of the sexes in other areas as well, e.g., musical ability, mathematical ability, and even mechanical ability.

Although the growing tendency has been to concede equality or near equality to the sexes in regard to human abilities, the belief still remains that the sexes differ basically in their instinctive and emotional equipment. Measurement of interests, attitudes, and opinions, derivatives of this basic equipment, confirms the belief in the existence of typically masculine and typically feminine personality types.

Many studies on masculinity-femininity have been concerned with determining the typical interest pattern for each sex. However, little research has been conducted on differences found only within the female sex even though knowledge in this area could be profitably employed in vocational guidance. If it can be established that students in various academic curricula differ significantly in interest pattern and also display a characteristic interest pattern, then an individual's masculinity-femininity level may become an important factor for prediction of success or failure in school. It is the purpose of this investigation to explore differences among three particular groups of women students and to determine typical masculinity-femininity levels for each group.

The hypothesis simply stated is that women students in one type of academic institution differ significantly in typical interest pattern from women students in another type of academic institution. Three schools will be utilized in this project: a women's college, a coeducational university, and a nursing school. Furthermore, careful analysis of the data will reveal not only differences among the groups but also the amount and direction of these differences, i.e., which group is most feminine and which least feminine.

A secondary purpose of the study involves ascertaining the degree to which the direct method corresponds with the indirect method of measuring the personality variable, masculinity-femininity. The direct method is so described because it directly asks for an evaluation of masculinity-femininity in the form of a rating. The indirect method is so called since the person taking the test presumably has no knowledge of the trait it is measuring.

The first test used in this study is the Minnesota Multiphasic Personality Inventory (MMPI). This test yields measurements of many personality variables but it was used primarily for only one of its scales, masculinity-femininity. The second test is the Attitude-Interest Analysis Test which is designed exclusively to measure masculinity-femininity. It utilizes 450 items which touch upon various areas of interest, in-

formation, opinion, etc., relative to the sexes. The third instrument is a group rating scale especially constructed for this thesis. With this scale each student rated all of the women in her group according to the degree of masculinity or femininity she believed each one to possess.

Thus, the three measures of masculine-feminine interest patterns served the two ends of this investigation. Various statistical techniques were applied to the material in order first, to determine significant differences among the three groups and second, to ascertain the degree of agreement between the two methods of measuring masculinity-femininity.

Investigations pertinent to the present study will be reviewed in the following chapter. The tests used in the project will be described in a separate chapter as will the procedure of test administration and scoring. Analysis of results will then be considered. A final chapter will be devoted to summary and conclusions.

CHAPTER II

REVIEW OF RELATED LITERATURE

An extensive investigation of the literature on masculinity-femininity revealed that some studies were similar to this project. Many investigations were concerned with differences between the sexes in various areas of interests, abilities, or occupations. Very few of them treated differences exclusively within the female sex. None of the research explored the basic hypothesis of this study.

The present chapter is designed to review some of the more representative studies on the general problem involved in the thesis. A fuller analysis of the test instruments will be found in Chapter III. These two chapters, then, will serve to provide the reader with a knowledge of the more important aspects of the project undertaken in this investigation as these have been outlined in the literature.

The most pertinent research was done by Lough.¹ She conducted an experiment to determine significant differences on

¹ Orpha M. Lough, "Teacher's College Students and the MMPI," Journal of Applied Psychology, XXX, June, 1946, 241-247.

any of the scales of the MMPI between two groups of women students in two different teacher training curricula, school music and elementary grades. The mean chronological age for both groups was 18.8 which is very close to the mean age for the groups of this study. Lough did not find any significant differences on any of the scales. However, two trends were noted: the students as a whole were stable with a slight tendency toward hypomania and certain of the scales might be useful for selecting students for the teaching profession. The author concluded that more research was warranted before the last conclusion could be wholly accepted.

Lough's study differed from this one in that this project: (1) used different types of schools; (2) employed two additional personality measurements, a personality test and a rating scale; and (3) used only one scale of the MMPI, masculinity-femininity, to determine significant differences.

Lough again investigated the same hypothesis in a study conducted about a year later using the subjects from her earlier study and two additional groups, liberal arts and nursing cadet students.² Again the mean ages were similar to those in this project but the purpose and number of instruments em-

2 Orpha M. Lough, "Women Students in Liberal Arts, Nursing, and Teacher Training Curricula and the MMPI," Journal of Applied Psychology, XXXI, August, 1947, 437-445.

ployed differed as before. Lough found that there were no significant differences between any of the four groups enrolled in the different curricula. She confirmed her previous finding that the groups were relatively stable with a slight tendency toward hypomania. Another more interesting and relevant conclusion was the trend in the nursing group toward more masculine interests with greater stability and less emotionality than the other female groups. In part, this conclusion prompted the present investigation.

Cillis and Orbison conducted some recent research on the two personality tests used in the present investigation.³ Their hypothesis was that since there were marked discrepancies and low correlations among tests of masculinity-femininity it seemed that the Terman-Miles test and the MMPI were measuring different aspects of the trait. The subjects consisted of 129 male students in the School of Business Administration with a median age of twenty-four years and fifty female students in the College of Liberal Arts and Sciences with a median age of eighteen years. Besides the total score for the Terman-Miles test Cillis and Orbison compared several exercises of the test with the MMPI Mf scale and found correlations of $-.30$ for

³ Olga E. de Cillis and William D. Orbison, "A Comparison of the Terman-Miles M-F Test and the Mf Scale of the MMPI," Journal of Applied Psychology, XXXIV, October, 1950, 338-342.

exercise four for the women and $-.21$ and $-.32$ for exercises four and five, respectively, for the men. All of these correlations were at least significant at the 5 per cent level of confidence. The high correlations for the men were in keeping with expected results since the MMPI derived many of its items from exercises four and five of the Terman-Miles test.⁴ However, the authors of the research were at a loss to explain the high correlation on exercise three for the women since the MMPI derived only a few of its items from this exercise.

The Terman-Miles test and the MMPI Mf scale showed a significant but not high correlation. Cillis and Orbison concluded that the two tests did measure something in common but that they were testing different aspects of it. Furthermore, much of what the two tests had in common was ascribed to exercises four and five. The authors believed that until a factor analysis of the Terman-Miles test was made caution should be employed in interpreting the results of the test.

Heston made a comparison of four different masculinity-femininity tests to determine their degree of agreement and capacity to differentiate between the sexes.⁵ The four tests

4 Ibid.

5 Joseph C. Heston, "A Comparison of Four Masculinity-Femininity Scales," Educational and Psychological Measurement, VIII, Autumn, 1948, 375-387.

used were: (1) the Strong Vocational Interest Blank for Men, Form M; (2) the Kuder Preference Record, Form BM; (3) the Minnesota Multiphasic Personality Inventory, Group Form; and (4) the DePauw Adjustment Inventory. The population was composed of thirty-four male college freshmen with a mean age of 18.3 years and forty-five female college freshmen with a mean age of 17.9 years. The four Mf scales were found to be about 80 per cent satisfactory in their ability to place men above the mean score and women below the mean score for the combined sexes.

When one of two variables in a correlation problem is a dichotomy, such as male versus female, the point-biserial correlation is employed. In Heston's study the test scores of the four Mf scales constituted one variable and the men and women, the other variable. The MMPI was found to be slightly superior (r was $+.743$). A critical ratio demonstrated that all of the tests exhibited significant differences between mean scores earned by each sex. Again, the MMPI was somewhat higher than the other scales (t was 9.75). The two interest tests, the Kuder and the Strong tests, had the highest correlation ($+.726$). The MMPI and the Strong test had a positive correlation of $.686$. Thus, the MMPI appeared to be the most effective instrument in distinguishing between men and women for this particular study.

Another similar study by Nance compared the MMPI, the Strong Vocational Interest Blank, and the Guilford-Martin Inven-

tory of Factors GAMIN to see if they measured the same variable.⁶ The subjects used for this study were 102 college students, 51 of each sex. The median age was twenty-three years for the men and nineteen years for the women. The population was broken down into six subgroups in order to vary the sequence of test presentation.

Correlations for women were found to be low on all three tests, possibly because the trait is not so well defined with respect to its manifestations in females. Correlations were higher for the men on all of the tests than for the women. The MMPI and the Strong test correlated: $+0.51$ for the men; $+0.20$ for the women; and $+0.71$ for both sexes combined. The men in general received a less masculine score than the mean score for the normative population. The women tended to have a more feminine score than the general population of females. When compared for various teaching curricula groups, music education students (both sexes) scored at least consistently more feminine and secondary education students scored at least consistently more masculine for all of the tests. Nance concluded that Mf inventories might prove useful for individual selection of students for various kinds of teaching curricula.

⁶ R. D. Nance, "Masculinity-Femininity in Prospective Teachers," Journal of Educational Research, XLII, May, 1949, 658-666.

Hampton emphasized the importance of an accurate method for identifying student difficulties in college.⁷ He administered the MMPI to 407 college girls in an attempt to identify those students with personality difficulties. The median age was seventeen years. Ninety-six of these students showed significant elevations on one or more of the personality scales (T scores of 70 or more). Elevated scales standing out by themselves were: masculinity-femininity, hypomania, depression, and paranoia. Hampton concludes that those individuals with high Mf scores had a tendency toward homosexuality.

Very little research has been done on the behavioral meaning of a high score on the Mf scale of the MMPI when made by a woman.⁸ The authors of the test in describing the Mf scale state that "[a]mong females high scores cannot yet be safely assumed to have similar clinical significance [to men], and the interpretation must be limited to measurement of the general trait."⁹ Thus, it seems unjustified to interpret an abnormal

7 Peter G. Hampton, "The Minnesota Multiphasic Personality Inventory as a Psychometric Tool for Discovering Personality Disorders Among College Students," Journal of Social Psychology, XXVI, August, 1947, 99-108.

8 Willie M. Verniaud, "Occupational Differences in the MMPI," Journal of Applied Psychology, XXX, December, 1946, 609.

9 S. R. Hathaway and J. C. McKinley, Manual for the Minnesota Multiphasic Personality Inventory, Revised, New York, 1951, 20.

score for women in this scale as a tendency toward homosexuality.

Terman and Miles have investigated the area of masculinity-femininity extensively. They have presented their findings in a single volume entitled Sex and Personality.¹⁰ However, some of the authors' statements contradict certain findings of other studies.

For instance, Terman and Miles stated that student nurses were less masculine than any college group of the same age.¹¹ This statement contradicts Lough's finding that student nurses have more masculine interest patterns than other college populations.

Terman and Miles also commented on the unreliability of masculinity-femininity ratings. They believed that ratings of masculinity-femininity were less reliable than ratings of any other personality variable, "as we have shown by several experiments."¹² Two studies indicating very low reliabilities were presented. Composite teacher's ratings consisting of one to six teachers were correlated with a preliminary series of masculinity-femininity test scores in one study.¹³ Correlations

10 Lewis M. Terman and Catherine C. Miles, Sex and Personality, New York, 1936.

11 Ibid., 154.

12 Terman and Miles, Sex and Personality, 64.

13 Ibid., 68, 69.

of test scores with teacher's ratings of the various areas measured by the test were: information, +.10; interest, +.15; introvertive response, +.02; and word association, +.03.

A second experiment was reported in which eighty-two male Stanford students took Form A of the Attitude-Interest Analysis Test and were then asked to rate themselves on masculinity-femininity.¹⁴ Correlations of the test scores with the self-rated areas were: childhood interest, +.08; occupational interest, +.06; leisure time, +.22; emotionality, +.21; general personality makeup, +.13; and average of the five self-ratings, +.19.

However, there is contrary evidence which points to at least a moderate correlations between tests and ratings of masculinity-femininity. Gilkinson considered a rating scale of masculinity-femininity a criterion for validating the Attitude-Interest Analysis Test.¹⁵ The Terman-Miles test was administered first in his study. Then, ratings for three traits, Masculinity-Femininity, Crudeness-Refinement, and Shyness-Boldness, were made by 110 men and 84 women. Numerical values were assigned to scale locations and the score for each person was

14 Ibid., 69.

15 Howard Gilkinson, "Masculine Temperament and Secondary Sex Characteristics: A Study of the Relationships Between Psychological and Physical Measures of Masculinity," Genetic Psychology Monographs, XIX, November, 1937, 105-154.

computed by averaging all of the ratings given to him by both male and female judges. The Masculinity-Femininity and the Crudeness-Refinement rating scales were found to differentiate significantly between men and women. The difference between averages for men and women on the Masculinity-Femininity scale was equal to 7.63 times the standard error of the difference.

Reliability of the scales was determined by correlating the average ratings by one half of the judges with the average ratings by the other half of the judges. These correlations were $+.73 \pm .03$ for the Masculinity-Femininity scale and $+.63 \pm .04$ for the Crudeness-Refinement scale.

Validity was determined by correlating the Attitude-Interest Analysis Test scores with the two traits on the rating scales. Correlations of the test scores with the rating scale scores for Masculinity-Femininity and Crudeness-Refinement for the 110 men were: $+.32 \pm .06$ and $+.40 \pm .05$, respectively; and for the 84 women, $+.40 \pm .06$ and $+.30 \pm .07$, respectively. These correlations are significant and clearly higher than those quoted by Terman and Miles, even though the judges in Gilkinson's study were less familiar with the task of evaluating behavior.¹⁶

¹⁶ The judges in the first experiment by Terman and Miles were teachers while the judges in Gilkinson's study were university students.

Another study by Smith using the Goodenough Speed-of-Association Test and a rating scale also indicated moderate correlations for masculinity-femininity rating scales and test scores.¹⁷ The ratings were made in ten areas, including such traits as: leadership, tomboyishness, aggressiveness, etc. The rating scales were given to the college women and their parents. Additional ratings were made by associates living in the house. The members of the sorority also ranked each girl in regard to total masculinity-femininity. Positive but moderate correlations were found between test scores and total masculinity-femininity rankings.¹⁸ The ratings by the associates who were least intimate with the persons rated yielded the highest correlations.

Disher administered the Terman-Miles test to determine regional differences in masculinity-femininity.¹⁹ The test was given to 556 Florida State College female students, 492 University of Florida male students, and 485 male and female Florida high school students. Testing data for four groups of students

17 Jane H. Smith, "The Relation of Masculinity-Femininity of Scores of Sorority Girls on a Free Association Test to Those of Their Parents," Journal of Social Psychology, XXII, August, 1945, 79-85.

18 Cf. *infra*, 36.

19 Dorothy R. Disher, "Regional Differences in Masculinity-Femininity Responses," Journal of Social Psychology, XV, February, 1942, 53-61.

from Stanford University, Oregon State University, Washington State University, and University of Utah were also used in the study for purposes of regional comparison.

The conclusions following upon an analysis of the data were: (1) in general, the results of the research supported Terman and Miles' finding in that the test differentiated between the sexes for both high school and college populations; (2) Florida populations rated more feminine than the western populations; (3) the northern born of northern born parentage were more masculine than the Florida born of Florida born parentage; (4) the longer the Florida influence had operated, within two generations, the more femininizing it was; and (5) the most outstanding regional differences, according to this test, centered around emotional and ethical reactions and interests.

Burger, Nemzek, and Vaughan attempted to determine the personalities of different types of sex offenders.²⁰ Their population was composed of 120 male prison inmates charged with sex offenses. Of seventeen factors studied in relationship to the Terman-Miles test only five were found to be significantly related to it, i.e., age, intelligence, recidivism, car owner-

²⁰ F. E. Burger, Claude L. Nemzek, and Charles L. Vaughan, "The Relationship of Certain Factors to Scores on the Terman-Miles Attitude-Interest Analysis Test," Journal of Social Psychology, XXVI, August, 1942, 39-50.

ship, and broken home background. Specifically, these conclusions were drawn from the data: (1) the older criminal tended to be more feminine; (2) the more intelligent criminal was more masculine; (3) the recidivist tended to be more masculine; (4) the car-owning criminal was more masculine; and (5) the broken home criminal tended to be more feminine.

The first two conclusions corroborated Terman and Miles' findings. In regard to age they found an initial rise in youth in masculinity for both sexes. With increasing age the masculine peak was followed by a decline throughout maturity toward femininity. In males the peak of masculinity was reached in the high school period while the most feminine scores were found in old age. In females the peak of femininity was attained in eighth grade while the most masculine scores were found in the college period.²¹

Terman and Miles also stated that intelligence was probably positively correlated with masculinity-femininity at certain age levels.²² In childhood the correlation was more pronounced for males than for females. Thereafter, the relationship was more marked for female than for male populations. The authors added a caution that these relationships between age,

21 Terman and Miles, Sex and Personality, 155.

22 Ibid., 156.

intelligence, and masculinity-femininity only held for large general-population groups. Selected groups did not always conform to the expectations postulated for the larger populations, e.g., extremely feminine male religious groups or relatively masculine professional female nurses.

The last conclusion in the study by Burger, Nemzek, and Vaughan concerning the feminine trend in the broken home criminal agreed with a finding of another investigation. Ferguson found that pleasant or desirable childhood experiences enabled the child to accept appropriate models of the cultural pattern,²³ Conversely, unpleasant or undesirable childhood experiences caused the child to reject the normal models and acquire behavior patterns which were atypical, i.e., femininity in men and masculinity in women.

The preceding review of literature on sex differences indicates that very little research has been devoted to the behavioral meaning of masculinity-femininity in women. It was seen that some workers erroneously interpreted extreme scores for women on tests of masculinity-femininity as indicative of homosexuality (Hampton). Most of the workers realized, however, that only the general trait should be involved in its inter-

²³ Leonard W. Ferguson, "The Cultural Genesis of Masculinity-Femininity," Psychological Bulletin, XXXVIII, July, 1941, 584-585.

pretation, at least until more extensive work on it has been accomplished.

Many of the studies reviewed concentrated on female subjects exclusively while others emphasized differences between the sexes. Only two of the studies (both of Lough) made any intra-sex comparisons. No significant differences were discovered among the women students in these studies.

A few inconsistencies were discovered in the course of reviewing the literature on masculinity-femininity. Lough, for example, found that student nurses were more masculine than three other college groups. On the other hand, Terman and Miles stated that student nurses were more feminine than any other college group of equal age. As will be evident in the chapter on analysis of results the apparent contradiction is resolved if the low correlation between the two tests used in the respective studies is considered.²⁴

A rather important point of disagreement involved the reliability of ratings of masculinity-femininity. Terman and Miles believed that ratings of this particular personality trait were very unreliable. Some low correlations between ratings and test scores were then reported. Self ratings were found to have higher correlations with test scores than judge's ratings.

24 Cf. *infra*, 60.

Both Smith and Gilkinson found higher correlations between ratings and test scores than those cited by Terman and Miles. Furthermore, Smith concluded that self ratings were no better than mere chance.²⁵

Finally, review of the literature revealed the importance of the role of masculinity-femininity in vocational guidance. Two authors suggested the use of an individual's level of masculinity-femininity as an aid in the selection of students for the teaching curricula. It is the purpose of this project to investigate the relationship between masculinity-femininity and the type of school attended by women students. Thus, the thesis may serve a dual purpose by: (1) contributing to the knowledge of intra-sex differences and (2) indicating possible applications of masculinity-femininity to vocational guidance.

²⁵ Compare with the findings of the present study, 64.

CHAPTER III

DESCRIPTION OF TEST MATERIALS

The present chapter is concerned with the reliability, validity, and standardization of the tests used in the study. Thus, the reader will gain some idea of the adequacy of the three different measures of masculinity-femininity in accomplishing the end of this research. The physical makeup of the testing instruments will also be described.

The Minnesota Multiphasic Personality Inventory by Hathaway and McKinley represents the most adequate development of the paper-and-pencil type of personality test. Here is a brief description of the test as found in the testing manual:

The Minnesota Multiphasic Personality Inventory is a psychometric instrument designed ultimately to provide, in a single test, scores on all the more important phases of personality. The point of view determining the importance of a trait in this case is that of the clinical or personnel worker who wishes to assay those traits that are commonly characteristic of disabling psychological personality. The instrument itself comprises 550 statements covering a wide range of subject matter - from the physical condition to the morale and the social attitudes of the individual being tested.¹

1 S. R. Hathaway and J. C. McKinley, Manual for the Minnesota Multiphasic Personality Inventory, 5.

Originally, personality characteristics could be assessed on nine clinical scales: Hypochondriasis, Depression, Hysteria, Psychopathy, Masculinity-Femininity, Paranoia, Psychasthenia, Schizophrenia, and Hypomania.² In addition, the authors tried to circumvent the usual criticisms of personality inventories by devising four validity scales: Lie, Cannot Say, F, and K. The authors also stated that the use of the scales was not strictly confined to clinical diagnosis since, "they have all been shown to have meaning within the normal range."³

The test has two forms, Individual or Card Form and Group or Booklet Form. As the titles imply the difference between the two forms is in the manner of administration. The Group or Booklet MMPI is the form used in this study since it is designed to meet the need for group administration. This form can be scored by IBM scoring machine or by hand. The items in the Group Form are identical to those found in the Individual Form, with the addition of sixteen duplicated statements. However, since all of the validation data for the test were based on the Individual Form, users of the Group Form were cautioned that the results might not correspond exactly to those of the Individual Form. The authors still encourage the use of the

2 Ibid.

3 Ibid., 6.

Individual Form in testing small groups or individual cases. Nevertheless, two authors have shown, through separately conducted research, that college, high school, or professional subjects who are accustomed to reading and writing obtain almost identical test results for both forms.⁴

The test materials for the Group Form include test booklets, separate answer sheets, and electrographic pencils. The administrator is provided with scoring keys and profile forms. The booklets, answer sheets, and pencils are distributed to every individual in the group, instructions are read, and the test is begun. After the test has been completed it is collected, scored, and the results are recorded on the profile forms.

With regard to the particular scale employed in this study the authors did not believe that a high score for women had any clinical significance but claimed that an interpretation based on the general aspects of masculinity-femininity was legitimate.⁵ For the purpose of this thesis the limitation for female populations is not too important since only differences in

⁴ William C. Cottle, "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, XXXIV, August, 1950, 255-259; Daniel N. Wiener, "Differences Between the Individual and Group Forms of the MMPI," Journal of Consulting Psychology, XI, March-April, 1947, 104-106.

⁵ Hathaway and McKinley, Manual for the Minnesota Multiphasic Personality Inventory, 20.

the general trait, masculinity-femininity, were sought.

The Mf scale is composed of sixty items. Heston presented a brief description of the type of item appearing on the scale in a study previously cited.⁶ He reported that twenty-seven of the sixty items in the scale were concerned with likes and dislikes of an occupational or avocational nature. Thus, males prefer science, hunting, and military life and dislike nursing, poetry, and dramatics. The remaining thirty-three items were said to comprise personality reactions referring to: (1) emotional feelings; (2) people; (3) sexual inversion; and (4) miscellaneous activities.

Cillis and Orbison gave a summary of the derivation of the items appearing in the scale in a study also previously reviewed.⁷ They stated that about thirty-one of the items originated from items appearing in exercises four through seven of the Terman-Miles' Attitude-Interest Analysis Test. Examples of similar items appearing on the respective tests were: (1) Terman-Miles, "Are your feelings often badly hurt?" MMPI, "My feelings are not easily hurt"; (2) Terman-Miles, "There is plenty of proof that life continues after death" MMPI, "I believe in

⁶ Heston, "A Comparison of Four Masculinity-Femininity Scales," Educational and Psychological Measurement, VIII, 377-378.

⁷ Cillis and Orbison, "Comparison of the Terman-Miles M-F Test and the Mf Scale of the MMPI," Journal of Applied Psychology, XXXIV, 340-341.

a life hereafter"; and (3) Terman-Miles, "Were you ever fond of playing with snakes?" MMPI, "I do not have a great fear of snakes."

The norms for the MMPI were based on about seven hundred men and women representing a cross-section of the general Minnesota population in age and education. The normal subjects consisted of the friends and relatives of the patients in the university hospital. The criterion for establishing normality was the answer given to one question: whether or not the subject was under a doctor's care. If an individual declared that he was not undergoing medical attention he was presumed normal.⁸ Thus, individuals included in the normalizing population could conceivably be disturbed, have not requested doctor's care, and yet be accepted as normal by these standards. This type of standardization method is subject to considerable criticism. However, it is not within the realm of this discussion to criticize the particular standardizing techniques utilized in the construction of the test. If a test is effective in measuring what it is supposed to measure, and does this reliably, then its use can be recommended.

The scales were constructed by contrasting normal

⁸ S. R. Hathaway and J. C. McKinley, "The Minnesota Multiphasic Personality Inventory: V Hysteria, Hypomania, and Psychopathic Deviate," Journal of Applied Psychology, XXVIII, April, 1944, 155.

groups with carefully scrutinized clinical cases. Selection of items to be included in the inventory came from various sources: the Humm-Wadsworth Temperament Scale, the Bernreuter Personality Inventory, the Bell Adjustment Inventory, psychiatric manuals, and clinical experience. Items were assigned to scales on the basis of the extent to which they discriminated 221 classified psychiatric patients from 724 normal persons, 265 college entrance applicants at the University of Minnesota, and other individuals assumed to be normal. Each clinical group was comprised of approximately fifty patients.

Test-retest reliability coefficients of the scale seem to be about as high as that expected for most personality inventories. The authors reported reliability coefficients ranging from $+0.71$ to $+0.83$.⁹ Cottle, using the two forms of the test within a period of one week, obtained coefficients ranging from $+0.46$ to $+0.91$. The Mf scale had the highest reliability coefficient in this study. Holzberg and Alessi administered the complete Individual Form and a shortened Individual Form within a period of three days.¹⁰ They found test-retest reliabilities ranging from $+0.52$ to $+0.93$.

⁹ Cottle, "Card Versus Booklet Form of the MMPI," Journal of Applied Psychology, XXXIV, 255-259.

¹⁰ Jules D. Holzberg and Salvatore Alessi, "Reliability of the Shortened MMPI," Journal of Consulting Psychology, XIII, August, 1949, 288-292.

From 50 to 80 per cent of each of the psychiatrically diagnosed groups were differentiated from the normal group and also from each other.¹¹ The scales which differentiated between each other were: hysteria, hypomania, psychopathic deviate, hypochondriasis, psychasthenia, and depression.¹² Subsequent research confirmed Hathaway and McKinley's claim that the scales differentiated between themselves.

Some of the studies reviewed in Chapter II provided evidence for the validity of the test.¹³ A study by Benton deserves special mention at this point since one of the purposes of this study was the validation of the Mf scale for males.¹⁴ The MMPI was given to eighty-five male patients of known disorders with no doubt as to diagnosis: (1) ten schizophrenics; (2) nine hysterics; (3) sixteen delinquents or psychopathic deviates; (4) ten homosexuals; and (5) forty organically diseased patients. The age range extended from seventeen to sixty years with a median age of twenty-one years. Results were

11 Hathaway and McKinley, Manual for the Minnesota Multiphasic Personality Inventory, 6.

12 Ibid.

13 Cf. Supra studies by Lough, Cillis and Orbison, Heston, and Hampton.

14 Arthur L. Benton, "The Minnesota Multiphasic Personality Inventory in Clinical Practice," Journal of Nervous and Mental Disease, CII, October, 1945, 416-420.

considered positive if at least one of the following two criteria were met: (1) the trend in question showed a T score of 70 or more, regardless of the relative strengths of the other trends elicited in the test or (2) the trend in question showed a T score of 65-69 and was the highest score on the test. Test papers with Cannot Say or Lie scores above 66 were rejected as being of questionable validity.

The homosexual patients were assured that their performance on the test would not affect the disposition of their cases and were encouraged to be honest. Nine out of the ten patients gave positive results on the masculinity-femininity scale. These nine patients were given the test again and told to conceal the fact of their homosexuality. Six of these nine men gave negative results on the femininity scale. Thus, two-thirds of the positive scoring homosexuals had enough insight to conceal their psychosexual trends. Kelly, Miles, and Terman reached a similar conclusion with a group of normal men and women, e.g., their subjects were able to shift their masculinity-femininity scores at will.¹⁵

It seems, then, that the MMPI will elicit the feminine response for homosexual males only if: (1) the men are frank

¹⁵ E. Lowell Kelly, Catherine C. Miles, and Lewis M. Terman, "Ability to Influence One's Score on a Typical Paper-and-Pencil Test of Personality," Character and Personality, IV, March 1936, 206-215.

and honest or (2) assuming they are not frank and honest they lack sufficient insight to hide the fact of their homosexuality. A significant proportion of this group did lack this insight. Thus, the MMPI was said to identify confessed male homosexuals with reasonable accuracy.

Burton conducted a study on thirty-four inverts, twenty rapists, and eighty-seven delinquents.¹⁶ The mean ages for the rapists, inverts, and delinquents were: 17.70, 16.86, and 17.19, respectively. The mean T scores on the Mf scale were 53 for the rapists, 60 for the inverts, and 50 for the delinquents. The difference between the invert and the rapist was significant since the critical ratio was 5.31. Retests were made on the invert group not less than three months nor more than four months after the original test. The product moment correlation for the retest group was $+ .70 \pm .09$ which corresponds rather closely to the reliability coefficients found by Holzberg and Alessi ($+ .76$)¹⁷ but is quite a bit less than that found by Cottle ($+ .91$).¹⁸

16 Arthur Burton, "The Use of the Masculinity-Femininity Scale of the MMPI as an Aid in the Diagnosis of Sexual Inversion," Journal of Psychology, XXIV, July, 1947, 161-164.

17 Holzberg and Alessi, "Reliability of the Shortened MMPI," Journal of Consulting Psychology, XIII, 289.

18 Cottle, "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, XXXIV, 257-258.

Burton concluded that because of the relatively large variability of the Mf scale it is limited as a measure of sexual inversion and cannot be used for individual clinical application. He further believed that even though the test were extended its reliability would not thereby be improved.

The Attitude-Interest Analysis Test was developed exclusively as a measure of masculinity-femininity. The authors, Terman and Miles, wanted a quantitative estimation of the degree and direction of an individual's deviation from his own sex in regard to interests, attitudes, and thought trends. In constructing the test the authors avoided any theoretical discussion of the trait, masculinity-femininity. They proceeded on an empirical basis and disregarded any general theory of sex differences. The authors do not believe that a theoretical approach is necessary for the construction of a personality test, for as they state in their manual:

The M-F test rests upon no assumption as to the factors which determine an individual's sex temperament. The causes may be either physiological and biochemical, or psychological and cultural; or they may involve both of these types of influence. The aim has been merely to devise a test which would measure existing M-F differences, however caused. It then becomes possible to investigate the influence of numerous physical, social, and psychological factors that may affect a subject's rating.¹⁹

¹⁹ Lewis M. Terman and Catherine C. Miles, Manual of Information and Directions for Use of Attitude-Interest Analysis Test, New York and London, 1938, 1.

Applications of the test include the relationships of masculinity-femininity of temperament to homo- and heterosexuality, to body build, to metabolic and other physiological factors, and to excess or deficiency of gonadal hormone stimulation. It has also been related to environmental situations such as number and sex of children, parent-child relationships, sex of teachers, type of education, and choice of friends or occupation.

The test is comprised of seven exercises: Word Association, Ink-Blot Association, Information, Emotional and Ethical Response, Interests, Personalities and Opinions, and Introvertive Response. The test has two equivalent forms, A and B. The authors have established the fact that the forms measure the same trait reliably. The tests are responded to by checking one of four, three, or two multiple responses. Separate stencils are provided for each page of the test and each response carries a weight of either plus or minus, i.e., masculine or feminine. The algebraic sum of the weighted items is the individual's score.

The test is administered without time limit and may be given either to an individual or to a group. It is not applicable to subjects of less than seventh-grade education and ability. The separate exercises were intended to sample a wide variety of sex differences in numerous areas. The total score in the general adult population ranges as follows: +200 to

-100, with a mean of +52 and a standard deviation of 50 for males and +100 to -200, with a mean of -70 and a standard deviation of 47 for females. The administration of a single form is adequate for a comparison of population groups and for approximate ratings of individual subjects. If a very accurate evaluation is necessary both forms should be given and the average of the two scores computed.

The test was constructed on an empirical basis. First, the experimental literature dealing with sex differences was perused for suggestions of test items. Then, items of many different types were formulated and administered to male and female groups. Several thousand items were tested and those which did not discriminate between the sexes were discarded, those that did discriminate between the sexes were retained. The final form contained 910 items for both forms, A and B.

The test was administered to more than six thousand subjects of both sexes, all ages, and many occupations. The authors studied the relationship of M-F scores to sex, age, education, scholarship, intelligence, occupation, interest, domestic milieu, physique, tendencies toward homosexuality, clinical case histories, and many psychological tests.

Reliability coefficients for one form computed by the split-half technique were +.92 for combined sex groups and +.78

for single sex groups.²⁰ The reliability coefficients for both forms were $+ .96$ for the combined sexes and $+ .88$ for either sex separately. The reliabilities for the seven exercises separately vary a great deal, ranging from $+ .24$ to $+ .89$ for single sex groups and from $+ .32$ to $+ .90$ for combined sex groups. If the exercises are employed separately, the only subtest reliable enough to estimate an individual's M-F level is exercise four, Emotional Attitudes. Exercise two, Ink-Blot Associations, and exercise seven, Introvertive Response, are so unreliable that they can be used by themselves only in comparison of extremely large populations.

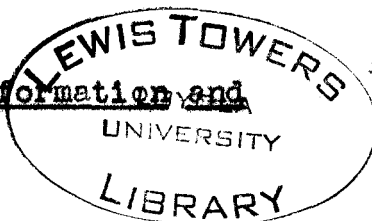
Commenting on the adequacy of the test in measuring what it is supposed to, the authors remarked that:

The validity of the M-F test is a necessary consequence of the method by which it was derived. Each item is in fact discriminative as between the sexes resident in a given culture and locale - the United States of the 1930's. The validity of the scale in discriminating sex temperaments naturally diminishes with remoteness of the culture patterns from our own.²¹

A typical procedure of validating a test is to select two criterion groups on the basis of some outside measure and ascertain how well the two groups are differentiated by the test in question. However, it was a relatively simple task for the

²⁰ Terman and Miles, Manual of Information and Directions, 6.

²¹ Ibid., 10.



authors to find clear-cut criterion groups since boys and girls, men and women were everywhere at hand and the maleness or femaleness of either group was known with 100 per cent accuracy. It then only remained to administer the test to comparable sex groups and note the amount of overlap in the distributions for the two sexes.²²

The average overlap for the two sex groups was about 8.02 per cent for the total score. The average index of overlap ranged from 8.84 per cent on exercise five to 30.89 per cent on exercise six.²³ However, the very method of selecting the items made the overlap on the total scores small. If all possible items of the type found in a given exercise had been included, the overlap would have been much greater.

The authors have correlated the test with many variables. The M-F test correlated about $+.20$ with mental age for single grade groups. When chronological age was held constant this correlation diminished to almost zero. The correlation of scholarship of college men was slightly negative indicating a tendency for men of higher scholarship to be less masculine than the average. College women of higher scholarship had a tendency

22 Terman and Miles, Sex and Personality, 63-64.

23 Ibid., 67.

to be a little more masculine than the average.²⁴

The total scores correlated $+.20$ to $+.30$ with the Stenquist mechanical ability test and about $+.13$ with the McQuarrie mechanical ability test. Correlation with the Cady-Raubenheimer character tests was approximately zero. The test correlated about $+.30$ with Allport ascendance, about $-.24$ to $-.52$ with Conklin introversion, and near zero with both the Neyman-Kohlstedt introversion and the Watson fairmindedness tests. For college women a multiple correlation of $+.36 \pm .06$ was found between the M-F test and a number of physical measurements. Exercise four correlated significantly for women with the Terman group IQ test, $+.28 \pm .06$, and with the Thorndike intelligence test, $+.36 \pm .06$.²⁵

The author of this research found no published rating scales which exclusively measured the trait, masculinity-femininity. Terman and Miles constructed a rating scale which sampled nineteen different areas of behavior, all allegedly related to masculinity-femininity.²⁶ Aside from the poor reliability of the instrument reported by the authors, the use of the scale would not have been practicable for the purpose of this research.

²⁴ Terman and Miles, Manual of Information and Directions, 10-11.

²⁵ Ibid., 11.

²⁶ Terman and Miles, Sex and Personality, 555-563.

since every individual had to rate every other person in her class a rating scale of such length would have been too unwieldy and complicated for group administration. However, some of the areas of behavior in the Terman-Miles rating scale were inserted in the instructions to the subjects for the rating scale constructed for this thesis.²⁷

A study by Smith, already reviewed in Chapter II, deserves consideration again at this point since her conclusions are particularly relevant to the findings of the present study.²⁸ In her study significant correlations were discovered between the Goodenough Speed-of-Association Test and a rating scale of ten traits. Correlations between test scores and total masculinity-femininity rankings were: $+.40 \pm .17$ for impersonal associates; $+.45 \pm .16$ for women living on another floor of the sorority house; and $+.36 \pm .18$ for women living on the same floor of the sorority house.

The author drew the following conclusions from the above relationships: (1) if correlation with M-F score was a measure of the validity of these ratings, accuracy of judgement changed with the judge and the trait to be judged; (2) self-

²⁷ Cf. *infra*, 50.

²⁸ Smith, "The Relation of Masculinity-Femininity of Scores of Sorority Girls on a Free Association Test to Those of Their Parents," Journal of Social Psychology, XXII, 79-85.

ratings were no better than chance judgement on all traits except popularity with boys and typical interests; (3) moderate correlations were found between total M-F scores and total M-F rankings; and (4) highest correlations were found with the ratings by the associates who were least intimate with the persons rated.²⁹ In regard to the last conclusion, the author suggested that close associates were less inclined than those with a more impersonal attitude to assign ratings which carry an implication of undesirability to their friends. Thus, the author inferred that judgements obtained from a number of persons, both familiar and unfamiliar, gave a truer evaluation than those obtained from any single individual.

After a survey of the literature on the topic it was discovered that most authorities preferred the graphic rating scale to other rating methods because of its greater reliability and accuracy in rating behavior. Freyd advocated the use of the graphic rating scale and pointed out that it: (1) was simple and easily grasped; (2) was interesting and required little motivation; (3) was quick; (4) freed the rater from direct quantitative terms; (5) enabled the rater to make as fine a discrimination as he desired; (6) was universal; and (7) allowed the administrator to alter the scoring method at will, e.g., 1 to 5

29 Ibid.

or 1 to 100.³⁰

Garrett listed three general principles to be followed in the construction of a rating scale: (1) the qualities evaluated must be valuable; (2) the qualities evaluated must admit of exact definition; and (3) the qualities evaluated must be capable of objective evaluation and measurement.³¹ The second principle posed one of the most difficult phases in the construction of the scale. Masculinity-femininity is a term which has come to have ambiguous meanings. Need for a clear and precise definition of the term was therefore made all the more important.

Garrett concluded that there were certain variables which, if left uncontrolled, affected reliability and validity adversely. These six factors were that: (1) close associates rated more reliably but there was little correlation between degree of acquaintance and competence as a rater; (2) judges tended to rate their friends too high on desirable traits and too low on undesirable ones; (3) individuals differed in ability to judge and ratings in which the judge expressed himself as very sure were more reliable than ordinary ratings; (4) characteristics exhibited in one's reactions to assigned tasks were better rated

³⁰ Max Freyd, "The Graphic Rating Scale," Journal of Educational Psychology, XIV, February, 1923, 94.

³¹ Henry E. Garrett and Mathew R. Schneck, Psychological Tests, Methods, and Results, New York and London, 1933, 110.

than social or personal traits; (5) judges over-rated themselves on desirable traits and under-rated themselves on undesirable traits; and (6) the halo effect was remedied by rating one trait at a time.³²

Freeman devoted a section of his text on psychological testing to the construction of rating scales.³³ In speaking of the consistency of ratings he stated that reliability was due primarily to the extent of variation of the judges' ratings. Since judges did not always agree on the trait allegedly possessed by a person, the average of all the judgements made by five or seven judges was taken as representative of the true rating. The most dependable judges can be determined by noting the extent to which each of their ratings agreed with the mean of all the ratings of a particular trait. It is also possible to determine which subjects have been most reliably rated by noting agreement of ratings among judges.

Freeman had little to say on the validity of rating scales except that it was very difficult to determine. Oftentimes, the validity of the rating scale rested on the judge's understanding of the meaning of the trait and his accuracy in

32 Ibid.

33 Frank S. Freeman, Theory and Practice of Psychological Testing, New York, 1950, 362-382.

rating it. Freeman concluded that the most effective means of securing validity was the careful selection of judges and a conscientious rating of traits.³⁴

In constructing the scale for the present study the writer attempted to incorporate the fundamental principles involved in the construction and use of graphic rating scales. The following paragraphs describe the manner in which the scale was constructed and the principles which were utilized in its construction. Brief mention is made of the scoring. A more detailed analysis of the administration, scoring, and interpretation of the scale is contained in the next chapter.

Smith found that judgements obtained from a number of subjects, both familiar and unfamiliar, yielded a more accurate evaluation of masculinity-femininity than those from only one person. Freeman specifically stated that from five to seven judges were needed for a valid appraisal of an individual. The present study employed at least thirty judges in each of the three groups, involving both familiar and unfamiliar associates.

The graphic rating scale was chosen since most authorities in the field agreed that it is more reliable and accurate than other rating devices. Freyd outlined many advantages of

34 Ibid.

the graphic rating method. Langlie³⁵ and Cronbach agreed that this type of rating method was more reliable and valid than other methods.

Garrett stated that the divisions in a rating scale varied from two to seven, five being the number most often employed. Cronbach believed that from five to seven divisions were adequate. The rating scale in this thesis was composed of five main divisions: Extremely Masculine, Somewhat Masculine, Average Femininity, Quite Feminine, and Extremely Feminine.

The halo effect was cited by Freeman as the most serious cause for unreliability of ratings. Garrett suggested two ways in which the halo effect may be eliminated; (1) rate one trait at a time and (2) define the trait clearly. The halo effect was lessened if not absent in this study since only one trait was rated and the trait was clearly defined in the instructions to the subjects.

Another device used to increase reliability is to instruct the judges to state how sure they are in each of their judgements.³⁶ Thus another graphic rating scale followed the scale of masculinity-femininity, making it possible for every

³⁵ T. A. Langlie, "Personality Ratings: I Reliability of Teacher's Ratings," Pedagogical Seminary and Journal of Genetic Psychology, L, June, 1937, 339.

³⁶ Both Garrett and Freeman advocate this practice.

judge to indicate his degree of certainty in each evaluation. The five dimensions of this scale were: Very Certain, Certain, Fairly Certain, Uncertain, and Very Uncertain. Therefore, each judge made two ratings for every subject, one referring to masculinity-femininity and the other referring to the degree of certainty of his judgement.

Freeman also commented on the scoring of the ratings. Since judges do not always agree as to the trait possessed by a person the mean of all judgements is taken as representative of the true rating. The scale divisions are usually assigned numerical values, e.g., one to five or one to one hundred. The check marks of the judges are then converted to scores by noting the positions of the ratings on the scale. This method of scoring was used in the present study.

Thus, the physical appearance of the rating form was a list of the names of classmates down the left-hand margin of the page. The characteristics which were mentioned in the instructions appeared at the top of each page. Two rating scales followed each subject's name, one describing the trait and the other the degree of certainty.

CHAPTER IV

PROCEDURE--TEST ADMINISTRATION, SCORING, AND METHOD OF ANALYSIS

The procedure of the research involved four stages:

(1) administration of the MMPI; (2) administration of the Attitude-Interest Analysis Test; (3) administration of the group form rating scale of masculinity-femininity; and (4) scoring, analysis, and interpretation of the test data. It is the purpose of this chapter to acquaint the reader with these various stages so that he will be able to adequately evaluate the findings of the study contained in the next chapter.

The subjects consisted of three groups of women students in attendance at three different types of educational institutions. Of the 132 students in the study, 57 attended a women's college, 34 attended a coeducational university, and 41 attended a nursing school. The mean age for the women's college group was 20.8 years with a standard deviation of .8 and an age range of 19-1 to 23-7 years. The mean age for the coeducational group was 20.3 years with a standard deviation of 1.4 and an age range of 17-10 to 24-11 years. The mean age for the nursing group was 20.2 years with a standard deviation of .8 and an age

range of 18-9 to 22-5 years. Examination of these data indicates that none of the mean ages for any of the groups differed to a significant degree.

As far as possible then chronological age was controlled by using only those students of one class, juniors, on the assumption that their ages would be relatively homogeneous. This method had the added advantage of equating the educational level for all three groups.

The selection of groups presented another problem. An attempt was made to procure distinct, clear-cut groupings since student A, a hypothetical student, who is a member of a coeducational group could have conceivably attended a women's college in the past. However, because of the limited number of subjects in the coeducational group, elimination of individuals like student A would have diminished the size of this group to a crucial degree.

Instead, critical ratios for determining the significance of difference were computed between those students whose background included another type of school and those students whose background did not. The findings of this analysis will be found in Table I.¹ No significant differences were discovered on any of the three tests between the two sub-groupings of the

1 Table I, 45.

TABLE I

MEANS, STANDARD DEVIATIONS, AND CRITICAL RATIOS OF SCORES
FOR TWO SUB-GROUPINGS OF THE COEDUCATIONAL
STUDENTS ON THREE MEASURES OF
MASCULINITY-FEMININITY

Test	Students With Attendance In Another Type Of School N = 14		Students Without Attendance In Another Type Of School N = 20		
	Mean	SD	Mean	SD	CR*
Attitude-Interest Analysis Test	-47.14	39.90	-35.00	52.10	0.87
MMPI Mf Scale	36.36	4.31	36.20	3.91	0.11
Rating Scale	2.78	.35	2.70	.45	0.59

*Refers to critical ratio. The formula for the critical ratio (t) for the difference between the means of small samples is:

$$t = \frac{M_1 - M_2}{\sqrt{\frac{\frac{N_1 \sigma_1^2 + N_2 \sigma_2^2}{N_1 + N_2 - 2}}{\frac{N_1 N_2}{N_1 + N_2}}}}$$

Source: Lindquist, E. F., A First Course in Statistics, Houghton Mifflin Company, Revised ed., 1942, 138.

coeducational group. It was concluded that the factor of previous attendance at a different type of school had no significant influence on the level of masculinity-femininity for this particular group of coeducational students.

The administration of the MMPI followed the procedure outlined in the manual for the Group Form of the test.² The materials were distributed to the subjects in this order: an electrographic pencil; an IBM answer sheet on which the subject printed her name, date, and birthdate; and a test booklet which was left closed on the desk. The instructions were then read aloud by the administrator while the subjects read them silently, any questions were answered, and the subjects were instructed to open the booklets and begin the test. One circuit was made among the subjects to make certain that the procedure was understood and followed. No time limits were imposed but the subjects were encouraged to work rapidly.

The tests were scored by the IBM scoring machine since it was a more economical and accurate method than hand scoring. Hathaway and McKinley present instructions in their manual for setting up the scoring machine.³ These were followed in scoring the tests of this investigation. The obtained raw scores were

2 Hathaway and McKinley, Manual for the Minnesota Multiphasic Personality Inventory, 9-10.

3 Ibid., 10-11.

then transferred to coded cards and converted to T scores from the tables provided in the manual.⁴ No T scores above 70 were found on any of the three validity scales, Cannot Say, Lie, or F.

Since masculinity is considered an undesirable personality trait to ascribe to a woman the subjects were assured that the rating scales would remain anonymous. Thus, it was hoped that the judges would be more truthful in their ratings. However, so that comparisons between the self-ratings and the three M-F estimations could be made it was necessary to know what rating each student assigned to herself. The manner of meeting this problem is described below.

Since Kelly, Miles, and Terman have demonstrated that subjects can influence their scores if they know the purpose of the test,⁵ the Attitude-Interest Analysis Test was administered previous to the rating scale. This procedure was made necessary by the fact that some of the subjects may have discovered the purpose of the test from the nature of the rating scale. This sequence also made identification of the rating scales possible by the following method. The Terman-Miles test was not collected until everyone in the group had finished the test.

4 Ibid., 14-16.

5 Kelly, Miles, and Terman, "Ability to Influence One's Score on a Typical Paper-and-Pencil Test of Personality," Character and Personality, IV, 205-215.

Particular care was taken to collect the tests in a certain order so that when the ratings scales were gathered, in the same order, the sequence of the two sets of materials would be identical. The names on the tests were then transcribed to the rating scales. Later, the names of the subjects were coded numerically so that the judges actually would remain anonymous.

Instructions for the administration of Form A of the Terman-Miles test were quite simple.⁶ The subject was given a test booklet, requested to read the instructions at the top of the front page, fill in the desired information, read the instructions at the bottom of the page, and begin the test proper. Any undue levity was discouraged and talking was not permitted. The administrator made himself as inconspicuous as possible throughout the test. Subjects who were excessively slow were urged to work more rapidly.

The Terman-Miles test can only be scored manually. The response to each item was scored as either plus or minus, according to the stencils for the test. Each plus counted one point toward masculinity and each minus as one point toward femininity. For each of the seven exercises the pluses and minuses were totaled separately and their algebraic sum recorded on the cover of the test. These separate scores were then

⁶ Terman and Miles, Manual of Information and Directions, 8.

weighted by various values given in the manual.⁷ The weighted scores were totaled algebraically, yielding the total weighted score. These total scores were transferred to coded cards and standard scores were determined by a method given in Garrett's text.⁸ The following paragraph describes the manner in which these raw scores were converted to standard scores.

The conversion of raw scores to standard scores might best be explained by way of an illustration. The mean of the distribution for the women's college group was -58.27 and the standard deviation, 40.53 . Each raw score in this distribution was expressed as a sigma deviation from the mean. For example, a score of -17.8 is subtracted from the mean and divided by the standard deviation resulting in a standard score of -1.00 (a standard score called Z). Each raw score was calculated in the same manner. These sigma scores were then transferred into a new distribution with the mean at 50 and the standard deviation at 10 . Thus, the sigma score of -1.00 is multiplied by 10 , added to the mean of 50 and then becomes 40 . All of the sigma scores were converted in the same way. High standard scores denoted masculinity while low standard scores denoted femininity.

The administration of the group rating scale im-

7 Ibid., 4.

8 Henry E. Garrett, Statistics in Psychology and Education, 3rd ed., New York, 1948, 156-157.

mediately followed the Terman-Miles test. Each subject was given a hectographed list of all the women in her class with two rating scales following each name: one scale consisting of five dimensions, ranging from extremely masculine to extremely feminine and the other scale consisting also of five dimensions, ranging from very certain to very uncertain. The rating scale was introduced with these instructions:

I would like to have each one of you rate each girl on the list which has been given to you. This rating will remain entirely anonymous since you will not identify the sheets with your own name. No one will know how anyone has rated her companions. Now I will explain the procedure you are to follow in filling out this form. All people possess both masculine and feminine characteristics to a greater or lesser degree. You are to rate each girl on the scale after her name by checking the point which best describes her according to these characteristics. In making these judgments you may keep the following ideas in mind although I am most interested in your own general impression of each girl. Sometimes the following points are indicative of femininity: one's speech (pause), one's mannerisms (pause), one's hobbies (pause), one's interests (pause), one's habits (pause), the clothes one wears (pause), physical characteristics (pause), the things one reads (pause), the ideas one expresses (pause), and probably a number of more subtle clues of which you are not fully aware but which make a person appear more or less feminine or masculine.

Therefore, considering these things and your own general impression of each girl I would like you to rate every girl on this list according to the degree of femininity or masculinity you believe her to possess. Place a check mark, thus (✓), at the appropriate point on the scale following her name. When you have completed this rating place another check mark on the scale following the one you have just finished to indicate how certain you are in your estimate of each girl. Then complete the list repeating the procedure for each girl. A list of some of the characteristics I mentioned is reproduced on each page for your convenience. Don't forget to rate yourself when you come to your own name. Are there any questions? Then begin.

A blackboard demonstration of the procedure accompanied the verbal explanation. Most of the subjects completed the rating scale within ten minutes. However, all of the subjects waited until everyone had completed the ratings before the scales were collected in order to make identification of the scales possible.

The five dimensions, Extremely Masculine, Somewhat Masculine, Average Femininity, Quite Feminine, and Extremely Feminine, were assigned numerical values of 5, 4, 3, 2, and 1, respectively. Any ratings of masculinity-femininity for which the individual judged herself to be uncertain or very uncertain were omitted from the tabulations.

In order to arrive at a mean rating score for each student the following procedure was employed. Tabulations of the ratings for each subject were made within the numerical categories, 5, 4, 3, 2, and 1. By way of example, assume that thirty judges have rated student A in the following manner: five ratings, Somewhat Masculine; twenty ratings, Average Femininity; and five ratings, Quite Feminine. Since the dimension, Somewhat Masculine, was assigned a value of 4 the five ratings in this category were multiplied by 4, yielding a sub-total of 20. The next dimension, Average Femininity, was assigned a value of 3 which when multiplied by the twenty ratings in this category gave a sub-total of 60. Finally, the dimension, Quite Feminine,

was assigned a value of 2. Since only five ratings were checked in this category the sub-total equaled 10. The sum of these three sub-totals was 90. Then, dividing this total score by the number of ratings, thirty, yielded the mean score of 3.0. Thus, student A has a mean rating scale score of 3.0 or Average Femininity. These raw scores were then converted to standard scores in the same manner as the Terman-Miles test scores.⁹ Thus, each subject had a rating scale standard score. Again the mean standard score was set at 50 and the standard deviation at 10. A high standard score indicated masculinity while a low standard score indicated femininity.

The analysis and interpretation of results constituted the last phase of the procedure. It will be remembered that the primary purpose of the study was to attempt to establish the fact that significant differences in masculinity-femininity existed among three groups of women students enrolled in different types of schools. The means, standard deviations, and critical ratios were calculated in order to ascertain the significance of difference between the three groups. A further analysis of the test data revealed those groups that were most feminine or most masculine. Nine tests of significance of difference were made for the three groups of students, using the three test instruments.

9 Ibid.

The secondary purpose of the study was to determine the degree to which the direct method, the rating scale, agreed with the indirect method, the personality tests. Three correlations were computed between: (1) the rating scale and the MMPI; (2) the rating scale and the Attitude-Interest Analysis Test; and, in addition, (3) the MMPI and the Attitude-Interest Analysis Test.

The following chapter presents the statistical treatment of the data and the findings of the study.

CHAPTER V

ANALYSIS OF RESULTS

The findings concerning the three measurements of masculinity-femininity in relation to women students are set forth in this chapter. In accord with the statistical procedure outlined in Chapter IV the results have been analyzed to determine significant differences among the three groups of women students. Correlations between the rating scale and the two personality tests have also been computed.

The critical ratios and actual differences between group mean scores for the three scales will be found listed in Table II.¹ From the data it is apparent that there is only one significant difference between the means of the coeducational and women's college groups. Since the behavioral meaning of femininity on the MMPI has not been thoroughly investigated the masculinity-femininity scale might not be expected to discriminate as well as the Attitude-Interest Analysis Test which has been quite well standardized on both men and women. Thus, if a significant difference were to occur it would more likely appear on the

1 Table II, 55.

TABLE II

DIFFERENCES BETWEEN MEANS AND CRITICAL RATIOS OF SCORES FOR
THREE GROUPS OF STUDENTS ON THREE MEASUREMENTS
OF MASCULINITY-FEMININITY

Test	Women's College And Nursing Groups		Coeducational And Women's College Groups		Coeducational And Nursing Groups	
	Differ- ence	Critical Ratio	Differ- ence	Critical Ratio	Differ- ence	Critical Ratio
Attitude- Interest Analysis Test	11.33	1.38	19.36	2.06*	8.03	0.81
MMPI Mf Scale	0.05	0.06	1.04	0.49	1.09	0.63
Rating Scale	0.12	1.20	0.17	1.76	0.05	0.50

*Significant at the .05 level of confidence.

latter test. In line with this analysis a difference was found on the Attitude-Interest Analysis Test which was significant at the .05 level of confidence.² Evaluation of the other critical ratios revealed no further significant differences. It seems then that the basic hypothesis of the research was not generally

² In all of the analyses of differences Fisher's t test of significance was utilized since the N in each of the three groups was below 100. For formula cf. supra, 45.

borne out since only one significant difference at the .05 level of confidence was found.

Therefore, with the exception of one difference it may be concluded that women students in three different schools, women's college, coeducational university, and nursing school, do not differ significantly in typical interest pattern. Any differences which did arise were due only to chance. This finding agrees with Lough's study in which four college curricula and all nine scales of the MMPI were employed.³ She found no significant differences among the four groups on any of the scales of the MMPI.

Although the critical ratios as reported in Table II show only one statistically significant difference some trends in masculinity-femininity may be noted among the groups. A summary of these trends will be found in Table III.⁴ On the MMPI masculinity-femininity scale the nursing group had the most feminine score while the coeducational group had the least feminine score. This finding does not agree with Lough's study in which the nursing students were the most masculine group when

³ Lough, "Women Students in Liberal Arts, Nursing, and Teacher Training Curricula and the MMPI," Journal of Applied Psychology, XXXI, 437-445.

⁴ Table III, 57.

TABLE III

COMPARISON OF GROUP MEAN SCORES AND STANDARD DEVIATIONS
WITH SPECIAL REFERENCE TO GROUP TRENDS IN
MASCULINITY-FEMININITY

Degree of Femininity	MMPI Mf Scale	Attitude-Interest Analysis Test	Rating Scale
Most Feminine Group	Nursing	Women's College	Coeducational
Mean	49.15	-58.27	2.73
SD	6.42	40.53	0.41
Next Most Feminine Group	Women's College	Nursing	Nursing
Mean	49.20	-46.94	2.79
SD	9.31	36.57	0.49
Least Feminine Group	Coeducational	Coeducational	Women's College
Mean	50.24	-38.81	2.90
SD	8.39	46.84	0.46

compared with three other college groups.⁵

The group mean scores on the Attitude-Interest Analysis

⁵ Lough, "Women Students in Liberal Arts, Nursing, and Teachers Training Curricula and the MMPI," Journal of Applied Psychology, XXXI, 437-445.

Test revealed that the women's college students gave the most characteristically feminine response and the coeducational students the least characteristically feminine response. Terman and Miles stated that nursing students were less masculine than any other college group of equal age.⁶ In this study the nursing group fell between the most and least feminine positions. Compared to the general population of females all of the groups in this study were less feminine since the mean score for women in the general population is -70 with a standard deviation of 47. However, this trend toward masculinity is in conformity with a finding by Terman and Miles who report that women become masculinized, i.e., acquire masculine interest patterns, during the college years.

The group mean scores on the rating scale suggest that the coeducational students were the most feminine group and the women's college students the least feminine group. Again the nursing group assumed a middle position between the other two groups.

Thus, the three measures do not generally agree in ranking the three groups of students according to level of masculinity-femininity when group mean scores are used as criteria. Two similarities were noted, however, in ranking the

6 Terman and Miles, Sex and Personality, 154.

three groups from most to least feminine: (1) the MMPI masculinity-femininity scale and the Attitude-Interest Analysis Test agreed in describing the coeducational students as the least feminine group and (2) the rating scale and the Attitude-Interest Analysis Test agreed in placing the nursing group between the two extremes of the most and least feminine groups.

Two observations may be made concerning the trends of the student's interest patterns when subjected to this analysis. Since practically no significant differences were discovered among the three groups of students the group trends as indicated by the mean scores cannot be considered very reliable. The obtained differences between the group mean scores have already been explained on the basis of chance fluctuation. Any group trends in terms of mean scores should be similarly interpreted, i.e., due to chance factors.

The lack of agreement found among the tests in ranking the three groups according to level of masculinity-femininity may be better understood if the correlations among the three measurements are examined.⁷ The low but positive correlations indicate that the three tests were measuring something in common. However, because of the low correlations the tests could hardly be expected to yield a point by point correspondence in ranking the

7 Table IV, 60.

TABLE IV

COEFFICIENTS OF CORRELATION AND SIGNIFICANCE LEVELS OF
CORRELATION ON SCORES AMONG FOUR MEASURES
OF MASCULINITY-FEMININITY*

Test	Self Rating Scale		Attitude-Interest Analysis Test			MMPI Mf Scale		
	r	t	r	t	r	t		
Group Rating Scale	.20	2.20 S.	.18	2.07 S.	.23	2.58 S.		
MMPI Mf Scale	.09	.97 N.S.	.17	1.86 N.S.				
Attitude-Interest Analysis Test	.07	.76 N.S.						

*The abbreviations "N.S." and "S." following the t values refer to levels of confidence: Not Significant and Significant, respectively. Since N was relatively small the PEr was not used. Instead, Fisher's test of significance for r was employed. "Significant" means that the probability a given coefficient will arise by chance is .05, or 5 chances out of a 100. The formula for the test of significance of r in small samples is:

$$t = r \sqrt{\frac{N - 2}{1 - r^2}}$$

Cf. J. P. Guilford, Fundamental Statistics in Psychology and Education, 2nd ed., New York, 1950, 227.

groups from most to least feminine.

Since neither the ratings nor the test scores were considered a criterion against which the validity of the other measure could be assessed, the comparison made in this study was not a validation of either measure. Its purpose was merely to note the relative agreement or disagreement of the two methods in measuring the trait, masculinity-femininity. The coefficients and the reliabilities of the correlations between the rating scale and the personality tests will be found in Table IV. The correlation between the rating scale and the Attitude-Interest Analysis Test was $+.18$ and between the rating scale and the MMPI masculinity-femininity scale, $+.23$. Both of these correlations proved to be significant at the $.05$ level of confidence. The sizes of the correlations were admittedly small, however, indicating only a very slight degree of agreement between the direct and indirect methods of measuring masculinity-femininity. In spite of the fact that the correlations in this study were lower than those of Gilkinson⁸ and Smith⁹ the findings for the three studies are in essential agreement.

8 Gilkinson, "Masculine Temperament and Secondary Sex Characteristics: A Study of the Relationships Between Psychological and Physical Measures of Masculinity," Genetic Psychology Monographs, XIX, 124.

9 Smith, "The Relation of Masculinity-Femininity of Scores of Sorority Girls on a Free Association Test to Those of Their Parents," Journal of Social Psychology, XXII, 82.

The coefficients and the reliabilities of the correlations between the self ratings of masculinity-femininity and the three test instruments will also be found in Table IV. The self ratings did not correlate significantly with the MMPI masculinity-femininity scale or the Attitude-Interest Analysis Test. Stated in another way, the correlations between the self ratings and the two tests of masculinity-femininity were no better than chance. Smith arrived at the same conclusion in her study.¹⁰ However, the self rating scale did correlate significantly at the .05 level of confidence with the group rating scale.

Table IV also presents the coefficient of correlation between the MMPI masculinity-femininity scale and the Attitude-Interest Analysis Test. The correlation of $+.17$ was not significant at either the .01 or .05 level of confidence. It may be concluded that the two tests were not measuring the same trait. This finding is particularly surprising since thirty-one of the sixty items comprising the MMPI masculinity-femininity scale were derived from the Attitude-Interest Analysis Test. Low but at least significant correlations were discovered by Cillis and Orbison in a comparison study of the same two tests.¹¹ They

10 Ibid.

11 Cillis and Orbison, "A Comparison of the Terman-Miles M-F Test and the Mf Scale of the MMPI," Journal of Applied Psychology, XXXIV, 340.

found a correlation of $-.30$ for men and a slightly higher correlation of $-.37$ for women.¹²

Since the Attitude-Interest Analysis Test was the only scale which indicated any significant differences the separate exercises were analyzed to determine which ones contributed most to distinguishing the differences found by the entire test. Table V contains the means, standard deviations, differences between means, and critical ratios for the three groups on the seven exercises of the test.¹³ From this analysis only one significant difference was found on exercise 7, Introvertive Response. The group mean scores for exercise 7 between the co-educational and women's college students were significantly different at the .01 level of confidence. No further significant differences between group means were found on the separate exercises. It is in keeping with expected results that the signi-

12 The correlations in the study by Gillis and Orbison were negative since a high raw score on the Attitude-Interest Analysis Test indicates masculinity while a high raw score on the MMPI masculinity-femininity scale indicates femininity. Thus, any correlations based on raw scores would have to be a negative value. However, in the present study the raw scores of the Attitude-Interest Analysis Test were converted to standard scores, a high standard score representing masculinity. The raw scores of the MMPI masculinity-femininity scale were converted to T scores from the tables provided in the manual of the test. Here a high T score for women also indicates masculinity. Thus, a correlation based on these two sets of scores would yield a positive value.

13 Table V, 64.

TABLE V

MEAN SCORES, STANDARD DEVIATIONS, DIFFERENCES BETWEEN MEANS, AND CRITICAL RATIOS FOR THREE GROUPS OF STUDENTS ON SEVEN EXERCISES OF THE ATTITUDE-INTEREST ANALYSIS TEST*

Exercise	Women's College Group N = 57		Coeducational Group N = 34		Nursing Group N = 39		Women's College and Nursing Groups		Coeducational and Women's College Groups		Coeducational and Nursing Groups	
	Mean	SD	Mean	SD	Mean	SD	Difference	Critical Ratio	Difference	Critical Ratio	Difference	Critical Ratio
1	84.32	6.81	82.52	5.81	84.46	7.59	0.14	0.13	1.80	1.06	1.94	1.05
2	99.77	0.97	99.61	1.16	99.39	1.35	0.38	1.62	0.16	0.84	0.22	0.64
3	90.90	7.62	92.91	6.44	93.51	8.86	2.61	1.52	2.01	1.27	0.60	0.32
4	115.01	18.35	121.38	20.54	114.99	22.90	0.02	0.02	6.37	1.56	6.39	1.23
5	57.00	27.30	64.36	36.08	65.08	24.54	8.08	1.47	7.36	1.09	0.72	0.10
6	95.04	5.60	95.85	6.39	95.76	5.62	0.72	0.68	0.81	0.70	0.09	0.07
7	97.44	2.04	98.56	1.77	97.90	1.61	0.46	1.33	1.12	2.63**	0.66	1.83

*Since most of the scores on the separate exercises were of a minus value a constant of 100 was added to each score to facilitate computation.

**Significant at the .01 level of confidence. Cf. Guilford, Fundamental Statistics in Psychology and Education, 610.

ficant difference found on exercise 7 occurred between the same two groups, coeducational and women's college, which were found to be significantly different by the total test.

Table VI contains the group mean scores and standard deviations for all of the scales of the MMPI.¹⁴ However, Figure 1 represents these same scores in a much more meaningful manner in the form of group profiles.¹⁵ The three group profiles approximate a fairly straight line at the T score mean level of 55. None of the group mean scores fell below a T score of 45 or exceeded a T score of 60. The profiles in general tended to be similar although some differences may be noted. Generally speaking, the profiles for the coeducational students deviated most and the profiles for the nursing students deviated least from the mean level of 50. The coeducational students as a whole manifested the most irregular profile, i.e., extreme shifts from high to low mean scores.

It will be remembered that the coeducational group was also found to be the least feminine as measured by the indirect method, the two personality tests. Nevertheless, when the coeducational students rated each other using the direct method, the group rating scale, they were found to be the most feminine

14 Table VI, 66.

15 Figure 1, 67.

TABLE VI

MEAN SCORES AND STANDARD DEVIATIONS OF MMPI SCALES FOR
THE WOMEN'S COLLEGE, COEDUCATIONAL, AND
NURSING STUDENTS

Scale*	Women's College Group		Coeduca- tional Group		Nursing Group		Combined Group	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
?	51.94	2.39	51.62	0.68	51.60	0.62	51.72	1.23
L	53.94	3.95	52.32	2.87	52.38	2.25	52.88	3.02
F	55.59	5.62	54.09	3.98	53.35	3.86	54.34	4.49
K	55.32	8.14	58.32	7.67	55.11	11.32	56.25	9.04
Hs	52.71	7.86	53.50	9.37	50.82	6.89	52.34	8.04
D	51.67	8.59	48.44	9.26	49.35	8.06	49.82	8.64
Hy	55.76	8.06	55.97	10.62	53.65	7.02	55.13	8.57
Pd	55.40	9.00	58.56	8.72	54.21	8.30	56.06	8.67
Mf	49.20	9.31	50.24	8.39	49.15	6.42	49.53	8.04
Pa	51.76	9.81	52.68	8.32	52.38	7.88	52.27	8.67
Pt	55.42	9.38	54.68	8.61	53.55	8.25	53.55	8.75
Sc	56.98	9.18	56.56	7.72	52.38	7.73	55.31	8.21
Ma	55.76	10.88	57.74	10.00	55.89	10.06	56.46	10.31

*(?) Question

(L) Lie

(F) Validity

(K) K

(Hs) Hypochondriasis

(D) Depression

(Hy) Hysteria

(Pd) Psychopathic

Deviate

(Mf) Masculinity-
Femininity

(Pa) Paranoia

(Pt) Psychasthenia

(Sc) Schizophrenia

(Ma) Hypomania

T Scores

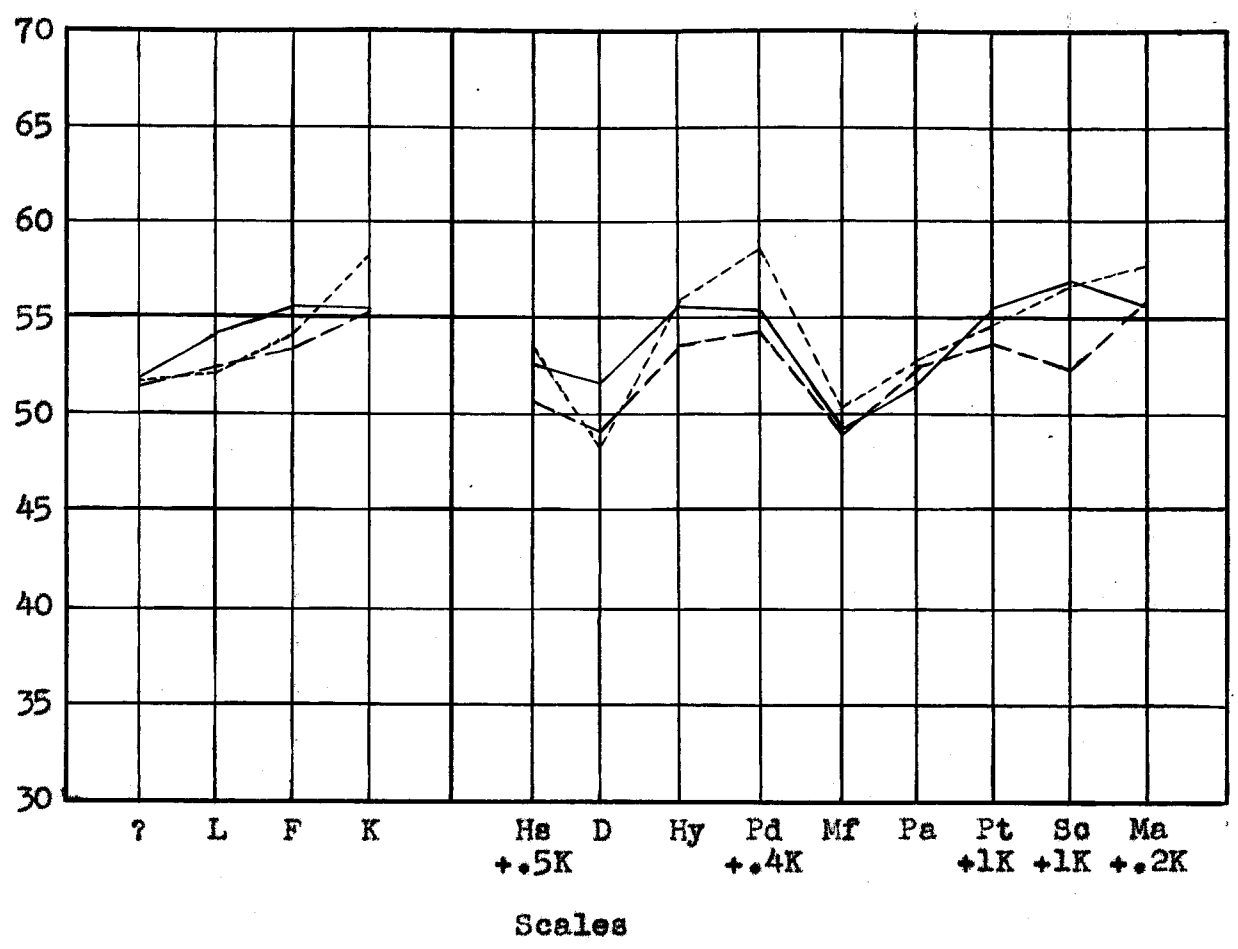


Figure 1

A GRAPHIC PRESENTATION OF THE PERFORMANCE OF WOMEN'S COLLEGE, COEDUCATIONAL, AND NURSING STUDENTS ON THE MMPI BASED ON MEAN T SCORES*

*Solid line (————) represents 46 Women's College students. Interrupted line (— — —) represents 41 Nursing students. Broken line (-----) represents 34 Coeducational students.

group. This opposite trend toward femininity as measured by the direct method might be attributed to a compensating over-evaluation of femininity by the coeducational group. In other words, the high feminine score might have been due to an effort on the part of the coeducational students to conceal the masculine tendencies of the group which were suggested by the two personality tests. This last speculation may be related to the fact that the coeducational students also achieved the highest mean score of the three groups on the psychopathic deviate scale of the MMPI.

Since the validity scores for the three groups were very close to the normal T score level of 50 it was concluded that the remaining clinical scales were accurate measures of the students in these areas. The two highest mean scores for the entire population were found on the psychopathic deviate and K scales for the coeducational students. Both of these mean scores were one and a half standard deviations above the mean level of 50. The lowest point on the profile for the coeducational group appeared on the depression scale. The women's college students scored highest on the schizophrenia scale and lowest on the masculinity-femininity scale. The nursing group had the lowest mean scores for the entire population on five of the nine scales: hypochondriasis, hysteria, psychopathic deviate, masculinity-femininity, and schizophrenia. The nursing students

scored highest on the hypomania scale.

From a group profile analysis of the MMPI it would seem that the students from the women's college, the coeducational university, and the nursing school were on the whole normal and stable. The most significant trend was demonstrated by the nursing students who appeared to be more stable and less emotional than the other two college groups. Lough found very similar results.¹⁶ The nursing students in her study were also the most well-adjusted group when compared to three other college groups.

Normal persons do not often score above 70 on any of the scales of the MMPI. However, if environmental pressure is small or other factors in the personality are favorable a person may score over 70 and yet escape the need for special attention. Table VII shows the percentage of students in this study with T scores above 70 on each scale of the Inventory.¹⁷

On the basis of these percentages it would seem that there may be some personality differences among the three groups of women students. The extremely high scores among the women college students were on the psychasthenia-schizophrenia-hypo-

16 Lough, "Women Students in Liberal Arts, Nursing, and Teacher training Curricula and the MMPI," Journal of Applied Psychology, XXXI, 443.

17 Table VII, 70.

TABLE VII

PERCENTAGE OF WOMEN STUDENTS WITH T SCORES ABOVE
70 ON THE SEPARATE SCALES OF THE MMPI

Scale*	Per Cent of 46 Women's College Students with T Scores Above 70	Per Cent of 34 Coeducational Students with T Scores Above 70	Per Cent of 41 Nursing Stud- ents with T Scores Above 70	Per Cent of 121 Students Above 70
?	0.00	0.00	0.00	0.00
L	0.00	0.00	0.00	0.00
F	0.00	0.00	0.00	0.00
K	2.17	5.88	0.00	2.48
Ha	2.17	5.88	2.44	3.31
D	2.17	2.94	2.44	2.48
Hy	2.17	11.76	0.00	4.13
Pd	2.17	11.76	2.44	4.96
Mf	0.00	2.94	0.00	0.83
Pa	2.17	0.00	0.00	0.83
Pt	8.70	2.94	4.88	5.79
Sc	10.87	5.88	0.00	5.79
Ma	6.52	14.71	7.32	9.09

*Cf. Supra, 66 for the names of the scales.

mania scales; among the coeducational students on the hysteria-psychopathic deviate-hypomania scales; and among the nursing students on the psychasthenia-hypomania scales. None of the nursing students scored above 70 on four of the nine diagnostic scales and, compared to the other two groups, the nursing group had only small percentages above 70 on the remaining scales.

When the three groups were combined the highest percentages of T scores over 70 were on the hypomania scale and in decreasing order of percentage, on the schizophrenia, psychasthenia, psychopathic deviate, and hysteria scales. Very similar results were discovered in Lough's study in which the same analysis of T scores was made.¹⁸ When the two groups of teacher's college students were combined the highest percentage of T scores above 70 were found to be on the hypomania scale and in decreasing value of percentage, hysteria, paranoia, psychopathic deviate, and schizophrenia scales.

In general, the coeducational students had the largest percentage of T scores above 70 while the nursing students had the smallest percentage of T scores above 70.

¹⁸ Lough, "Teacher's College Students and the MMPI," Journal of Applied Psychology, XXX, 245.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The purpose of the study was to determine significant differences in interest patterns among three groups of women students in attendance at different types of educational institutions: a women's college, a coeducational university, and a nursing school. A secondary purpose involved ascertaining the degree of agreement between the direct and indirect methods of measuring masculinity-femininity, i.e., a rating scale and two personality tests. The instruments of measurement were the Minnesota Multiphasic Personality Inventory, the Attitude-Interest Analysis Test, and a rating scale of masculinity-femininity especially constructed for this thesis.

A review of the literature on sex differences demonstrated that little research has been done on the behavioral meaning of masculinity-femininity in women. Many studies considered differences between the sexes while others emphasized differences within only the female sex. A discrepancy in the literature was noted concerning the reliability of ratings of masculinity-femininity. Some authors regarded ratings of this trait as very unreliable. Others found significant correlations

between ratings and test scores of masculinity-femininity. A general review of the research disclosed the importance of masculinity-femininity in vocational guidance. In particular it was proposed that an individual's level of masculinity-femininity be utilized in the selection of students for the teaching curricula. No study was discovered which explored the basic hypothesis of this thesis.

When the tests had been administered to the three groups various statistical techniques were applied to the testing data. Means, standard deviations, critical ratios, and correlations were calculated in order to discover the exact nature of the relationships involved in the hypothesis of the thesis.

On the basis of the findings the primary hypothesis of the study was not substantiated. Women students in attendance at three different types of schools do not differ significantly in masculinity-femininity. One exception was found on the Attitude-Interest Analysis Test. The difference in group mean scores between the coeducational and women's college students was significant at the .05 level of confidence. Some trends in interest patterns were noted among the students but there was no consistency among the tests in ranking the three groups according to level of masculinity-femininity.

The direct and indirect methods of measuring masculinity-femininity as represented by the rating scale and personality

tests, respectively, correlated positively but only to a moderate degree. The self ratings of masculinity-femininity did not correlate significantly with either of the two personality tests.

When the separate exercises of the Attitude-Interest Analysis Test were analyzed one significant difference was found on exercise 7, Introvertive Response, between the coeducational and women's college students.

The MMPI profiles for the three groups approached a fairly straight line at the mean T score level of 50. The co-educational group deviated most from the mean of 50 and the nursing group least.

Analysis of the test results obtained from 132 women students suggests that:

- (1) In general, women students in a women's college, a coeducational university, and a nursing school do not manifest significantly different interest patterns. However, since the behavioral meaning of masculinity-femininity is not yet clearly defined with regard to women, any conclusion in which no significant differences are found cannot be considered definitive until further research has been conducted in this area.
- (2) Positive but moderate correlation exists between the direct and indirect methods of measuring masculinity-femininity. Self ratings of masculinity-femininity are no better than chance.
- (3) The nursing students are more stable and less emotional than either the women's college or coeducational students. The coeducational students are the least well-adjusted and the least feminine of the three groups.
- (4) More research on the Attitude-Interest Analysis Test

with special reference to the separate exercises might prove fruitful.

BIBLIOGRAPHY

A. Books

- Cronbach, Lee J., Essentials of Psychological Testing, New York, 1949.
- Freeman, Frank S., Theory and Practice of Psychological Testing, New York, 1950.
- Garrett, Henry E., Statistics in Psychology and Education, 3rd ed., New York, 1948.
- Garrett, Henry E. and Schneek, Mathew R., Psychological Tests, Methods, and Results, New York and London, 1933.
- Guilford, J. P., Fundamental Statistics in Psychology and Education, 2nd ed., New York, 1950.
- Hathaway, Starke R. and McKinley, J. C., Manual for the Minnesota Multiphasic Personality Inventory, Revised, New York, 1951.
- Lindquist, E. F., A First Course in Statistics, Revised ed., Boston, 1942.
- Terman, Lewis M. and Miles, Catherine C., Sex and Personality, New York, 1936.
- Terman, Lewis M. and Miles, Catherine C., Manual of Information and Directions for Use of Attitude-Interest Analysis Test, New York and London, 1938.

B. Articles

- Benton, Arthur L., "The Minnesota Multiphasic Personality Inventory in Clinical Practice," Journal of Nervous and Mental Disease, CII, October, 1945, 416-420.
- Benton, Arthur L. and Probst, Kathryn A. "A Comparison of Psychi-

atric Ratings with MMPI Personality Inventory Scores," Journal of Abnormal and Social Psychology, XLI, January, 1946, 75-78.

Burger, F. E., Nemzek, Claude L., and Vaughan, Charles L., "The Relationship of Certain Factors to Scores on the Terman-Miles Attitude-Interest Analysis Test," Journal of Social Psychology, XXVI, August, 1942, 39-50.

Burton, Arthur, "The Use of the Masculinity-Femininity Scale of the MMPI as an Aid in Diagnosis of Sexual Inversion," Journal of Psychology, XXIV, July, 1947, 161-164.

Cillis, Olga E. de and Orbison, William D., "A Comparison of the Terman-Miles M-F Test and the Mf Scale of the MMPI," Journal of Applied Psychology, XXXIV, October, 1950, 338-342.

Cottle, William C., "Card Versus Booklet Forms of the MMPI," Journal of Applied Psychology, XXXIV, March-April, 1950, 255-259.

Disher, Dorothy R., "Regional Differences in Masculinity-Femininity Responses," Journal of Social Psychology, XV, February, 1942, 53-61.

Ferguson, Leonard W., "The Cultural Genesis of Masculinity-Femininity," Psychological Bulletin, XXXVIII, July, 1941, 584-585.

Freyd, Max, "The Graphic Rating Scale," Journal of Educational Psychology, XIV, February, 1923, 83-102.

Gilkinson, Howard, "Masculine Temperament and Secondary Sex Characteristics: A Study of the Relationships Between Psychological and Physical Measures of Masculinity," Genetic Psychology Monographs, XIX, November, 1937, 105-154.

Guilford, J. P. and Martin, Howard, "Age Differences and Sex Differences in Some Introvertive and Emotional Traits," Journal of General Psychology, XXXI, October, 1944, 219-229.

Hampton, Peter G., "The Minnesota Multiphasic Personality Inventory as a Psychometric Tool for Discovering Personality Disorders Among College Students," Journal of Social Psychology, XXVI, August, 1947, 99-108.

- Hathaway, Starke R. and McKinley, J. C., "A Multiphasic Personality Schedule (Minnesota): I Construction of the Schedule," Journal of Psychology, X, October, 1940, 249-254.
- Hathaway, Starke R. and McKinley, J. C., "The Minnesota Multiphasic Personality Inventory: V Hysteria, Hypomania, and Psychopathic Deviate," Journal of Applied Psychology, XXVIII, April, 1944, 153-174.
- Heston, Joseph C., "A Comparison of Four Masculinity-Femininity Scales," Educational and Psychological Measurement, VIII Autumn, 1948, 375-387.
- Holzberg, Jules D. and Alessi, Salvatore, "Reliability of the Shortened MMPI," Journal of Consulting Psychology, XIII, August, 1949, 288-292.
- Kelly, E. Lowell, Miles, Catherine C., and Terman, Lewis M., "Ability to Influence One's Score on a Typical Paper-and-Pencil Test of Personality," Character and Personality, IV, March, 1936, 206-215.
- Langlie, T. A., "Personality Ratings: I Reliability of Teacher's Ratings," Pedagogical Seminary and Journal of Genetic Psychology, L, June, 1936, 339-359.
- Lough, Orpha M., "Teacher's College Students and the MMPI," Journal of Applied Psychology, XXX, June, 1946, 241-247.
- Lough, Orpha M., "Women Students in Liberal Arts, Nursing, and Teacher Training Curricula and the MMPI," Journal of Applied Psychology, XXXI, August, 1947, 437-445.
- Nance, R. Dale, "Masculinity-Femininity in Prospective Teachers," Journal of Educational Research, XLII, May, 1949, 658-666.
- Smith, Jane H., "The Relation of Masculinity-Femininity of Scores of Sorority Girls on a Free Association Test to Those of Their Parents," Journal of Social Psychology, XXII, August, 1945, 79-85.
- Verniaud, Willie M., "Occupational Differences in the MMPI," Journal of Applied Psychology, XXX, December, 1946, 604-613.
- Weiner, Daniel N., "Differences Between the Individual and Group

Forms of the MMPI," Journal of Consulting Psychology, XI, March-April, 1947, 104-106.

Wolf, R. and Murray, Henry A., "An Experiment in Judging Personalities," Journal of Psychology, III, April 1937, 345-365.

APPENDIX I

THE GROUP FORM RATING SCALE OF
MASCULINITY-FEMININITY

Approval Sheet

The thesis submitted by Richard James Stanek, has been approved by three members of the Department of Psychology.

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

The thesis is therefore accepted in partial fulfillment of the requirements for the Degree of Master of Arts.

January 7, 1953
Date

Frank J. Kobler
Signature of Adviser