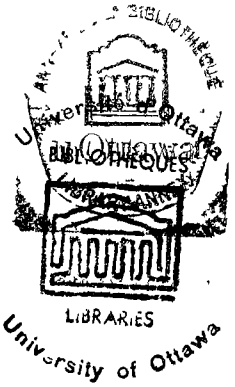
A COMPARISON OF SELF-REPORTED AND OBSERVED PERSONALITY TRAITS IN A CUSTODIAL SETTING

by Jacqueline A. Marshall

Thesis presented to the School of Psychology and Education of the University of Ottawa as partial fulfillment of the requirements for the degree of Master of Arts in Psychology



Ottawa, Canada, 1964

ACKNOWLEDGEMENTS

This thesis was prepared under the direction of Professor Maurice Chagnon, Ph.D., of the School of Psychology and Education of the University of Ottawa.

The writer wishes to thank Mr. H.M. Hooper, Superintendent of the Ontario Training Centre, Brampton, for his cooperation in allowing this study to be carried out at Brampton. The author also wishes to express her gratitude to Mr. Frank Blum, Resident Psychologist at the Ontario Training Centre, Brampton, for his assistance in administering the test and the rating scale at the institution.

The writer is indebted to Mr. Michael Maloney for help received in the wording of the rating scale. The assistance of those students of the Training Centre who helped with the scoring and preparation of the data is gratefully acknowledged.

CURRICULUM STUDIORUM

Jacqueline Agnes Marshall was born September 26, 1930, in Hamilton, Ontario. She received the Bachelor of Arts degree from the University of Ottawa, in 1958.

TABLE OF CONTENTS

LIST OF TABLES

Table page

INTRODUCTION

In the allocation of young, male, first offenders in the Province of Ontario psychological tests are frequently used to assist in selection. In the Province of Ontario, most of the selection takes place at the Ontario Reformatory, Guelph. One of the psychological tools used in this process is the Guilford-Zimmerman Temperament Survey.[1]

Guilford considers that a score on the Survey represents a "confirmed dimension of personality and a dependable descriptive category".[2] It is the purpose of the present study to examine the latter claim in a correctional setting for a selected group of young first offenders.

The first chapter of this thesis will be concerned with a review of the pertinent literature. It will include some notions on the concept of validity particularly as it applies to personality inventories. Two studies in which Guilford Personality Factors are validated by means of rating scales will be reviewed and discussed. This chapter will end by indicating the lack of validation studies in the field of corrections.

---

1 J.P. Guilford and Wayne S. Zimmerman, The Guilford-Zimmerman Temperament Survey, Beverly Hills, California, Sheridan Supply Co., 1949, 12 p. Hereafter referred to as the Survey.

2 Ibid., p. 6-8.

The statement of the hypothesis to be investigated will be followed by a description of the experimental design of the study. This will include a description of the sample and of the independent and dependent variables. The chapter will end with a description of the statistical treatment of the data used in testing the hypothesis.

The results of this study are then presented, interpreted, and discussed. Some possible weaknesses in the experimental design and in the general theoretical framework underlying it are considered. The implication of the findings for an Ontario Reformatory is indicated.

In conclusion, some suggestions for further research will be made. Supplementary information relevant to this study will be found in the appendix.

CHAPTER I

REVIEW OF THE LITERATURE

In this chapter the concept of test validity will be briefly discussed with specific reference to personality inventories. This discussion will be limited to predictive—concurrent validity and specifically to the use of rating techniques as criterion. The two validation studies of Guilford Personality Inventory Factors where rating techniques were employed as criteria will be reviewed, and the lack of validation studies in the field of corrections will be noted. The chapter will conclude with the statement of the hypothesis to be investigated in this study.

1. Notions on Test Validity.

Anastasi[1] quotes Guilford[2] as saying, "In a very general sense a test is valid for anything with which it is correlated". Validity, Anastasi emphasizes, is only meaningful with reference to specific criterion, and the validity of a test should not be confused with the validity of the "will-o'-the wisp" psychological processes underlying the

---

1 Anne Anastasi, "The Concept of Validity in the Interpretation of Test Scores", Educational and Psychological Measurement, Vol. 10, 1950, p. 74.

2 J.P. Guilford, "New Standards for Test Evaluation", Educational and Psychological Measurement, Vol. 6, (no number), 1946, p. 429, quoted by Anastasi, Op. Cit., p. 74.

test.[3] Cronbach and Meehl[4] analyzed the concept of validity as presented in the recommendations of the American Psychological Association[5] and found that Anastasi's notion of validity referred to only one type of validity - that was to criterion-oriented validation procedures, or to predictive and concurrent validity. In the aforementioned types of validity the criterion behaviour is of primary concern to the tester.

Validity of Personality Inventories.- A typical personality inventory asks the examinee to answer questions about himself and his opinions. Anastasi's definition of a psychological test - "it is a device for determining within a relatively short period of time what could be otherwise discovered only by means of a prolonged follow up"[6] - is applicable to personality inventories. In the case of the personality inventory, the "prolonged follow up" consists of the observations and opinions of others who have been familiar with the examinee for a long period of time.

---

3 Anastasi, Op. Cit., p. 76.

4 Lee J. Cronbach and Paul E. Meehl, "Construct Validity in Psychological Tests", Psychological Bulletin, 1955, Vol. 52, No. 4, p. 281-300.

5 "Technical Recommendations for Psychological Tests and Diagnostic Techniques", Psychological Bulletin Supplement, Vol. 5, No. 2, Part 2, 1954, p. 1-38, cited by Cronbach and Meehl, Op. Cit., p. 174.

6 Anastasi, Op. Cit., p. 67.

In analyzing the relationship between self reported and observed behaviour from the point of view of personality, Guilford and Martin state:

> It is said that personality is at least three things. It is first of all what it is apart from any human observer. It is secondly what a person observes. Finally, it is what other people observe. The three have many points of correspondence, naturally, because the second and third depend upon the first. Because of the points of correspondence we can translate findings from one view to another.[7]

Guilford maintains that the key to a successful validity measure is a good criterion measure.[8] Ellis, however, points out some of the deficiencies of ratings by others as criterion measures for the validation of personality inventories.[9] Nevertheless he finds that many studies use some sort of rating technique to validate personality questionnaires - in forty-four studies that he reviewed he found positive results in twelve cases, questionably positive results in ten, and negative results in twenty-two cases.[10] Ellis therefore concluded that "in no case does it seem that any one of the most popular questionnaires have shown to

---

7 J.P. Guilford and Howard Martin, "Age Differences and Sex Differences in Some Introversive and Emotional Traits", Journal of General Psychology, 1944, Vol. 31, Second Half, p. 226.

8 J.P. Guilford, Psychometric Methods, New York, McGraw-Hill, 1954, p. 402.

9 A. Ellis, "The Validity of Personality Questionnaires", Psychological Bulletin, 1946, Vol. 43, p. 391, 406.

10 Ibid., p. 411.

have indisputably high validity in any particular area".[11] He attributed some of this low validity to the inadequacies of ratings as a criterion; however, it is unlikely that he would disagree with Guilford[12] that for purposes of validation the rating method with all its weaknesses will remain in use until a better procedure becomes available. Guilford states that we are familiar with many of the sources of bias and error where rating scales are vulnerable, but he also states that these are to some degree controllable by the application of correction procedures.[13]

2. The Validity of the Guilford Factorial Personality Inventories.

A survey of the literature disclosed two studies where Guilford Personality Inventory Factors were validated against ratings by others. These two studies pertain to the Inventory of Factors STDCR.[14] Guilford and Zimmerman state that the findings of the two studies can be applied with confidence to the scores of the Guilford-Zimmerman Temperament

---

11 Ellis, Op. Cit., p. 423.

12 Guilford, Psychometric Methods, Op. Cit., p. 297.

13 Ibid., p. 298.

14 --------, An Inventory of Factors STDCR, Beverly Hills, California, Sheridan Supply Co., 1942. Hereafter referred to as Inventory.

Survey.[15] Factors S, T, and R of the Inventory were taken over directly by the Survey, the only exception being a reversal in the direction of the scoring on Factor R. An equal number of items from Factors C and D, from the Inventory were combined to give Factor E of the Survey.[16] The basis for the choice of the specific items for this factor was not indicated.

The first study was carried out by Guilford at the University of Southern California in the summers of 1938 and 1939.[17] The subjects, fifty and fifty-one respectively, were a group of Guilford's psychology students. A nine point bipolar rating scale of the adverbial type was constructed to examine twenty-six traits.Included in the aforementioned rating scale were Factors STDCR. The experiment consisted of each student rating himself on the rating scale, being rated on the same scale by five classmates or "peers" with whom he was well acquainted, and completing the Inventory. (The exact order of the steps in the experiment was not reported.) The "peer" rating for each student consisted of the median rating

---

15 J.P. Guilford and Wayne S. Zimmerman, The Guilford-Zimmerman Temperament Survey, Manual of Instructions and Interpretations, Beverly Hills, California, Sheridan Supply Co., 1949, p. 8. Hereafter referred to as the Survey.

16 Ibid., p. 4.

17 Guilford and Martin, Op. Cit., p. 228.

obtained from the five classmates. No experimental reliability of either the Inventory or the rating scales was reported as having been established. Inter-rater reliabilities were reported by the authors for Factors STDCR respectively. These reliability coefficients were obtained on a similar set of ratings in a study conducted on twenty-two Nebraska students. With respect to the Southern California study the relevant findings are as follows:(no ratings for Factors T and C were obtained in 1938); significant relationships were obtained for both years for Factors S, D and R, with coefficients of correlation ranging from r = .39 to r = .70; no significant correlations were obtained for Factors T and C. Guilford and Martin considered the relationships for Factors S, D and R substantial; for Factors T and C much less satisfactory.[18]

Considering the low reliability of the latter traits, the authors suggest that the highly implicit nature of the behaviour connected with these traits makes their evaluation by an external observer yield doubtful validity. The authors conclude that, based on the "slender thread of evidence" provided by these data, it would seem that ratings by others are to be trusted for some traits, but not for others.[19]

---

18 Guilford and Martin, Op. Cit., p. 227.

19 Ibid., p. 228.

One certainly feels that the qualification "slender thread of evidence" is justified. Three errors in the study are noteworthy: i) no experimental reliabilities of the ratings were obtained; ii) by exclusively accepting the median score of the five available ratings as the "peer" rating, much information was lost and the possibility of constant errors in ratings thereby increased; and iii) the small samples may partly account for the fluctuation of some of the validity coefficients between the two years.

The second study on the relationship between the Inventory, self ratings, and peer ratings was carried out by Carroll.[20] The sample consisted of 125 men who had a median age of about twenty-five years. Most of these 125 men had had some previous college training and were enrolled in the Army Specialized Training Program curriculum in Personnel Psychology. Carroll states that "conditions appeared particularly favourable to obtain reliable interpersonal ratings" inasmuch as these men were housed together for four months, usually six to a room, and took identical courses.[21]

Based on the description of the factors given in the Manual of the Inventory, an eleven point graphic rating scale

---

20 John B. Carroll, "Ratings on Traits Measured by a Factored Personality Inventory", Journal of Abnormal and Social Psychology, 1952, Vol. 47, p. 626-632.

21 Ibid., p. 626.

was constructed as a class exercise. The author points out that, while it was crucially important that the rating scales be based on the same traits as the personality inventory, there was no way of knowing whether or not this ideal was achieved. Difficulty was experienced in constructing the scale because only one pole of the bipolar scale was described in the Manual of the Inventory. By examining the scoring of the items that made up the individual factor, the author decided what was likely to be at the opposite ends of the scales.

As in the previous study, each man rated himself, rated each of his five room-mates and was, in turn, rated by them, and filled out the Inventory in the aforementioned order. Complete data were available for 110 men. Scores were assigned to the ratings by graphic interpolation of the norms of the 388 University of Nebraska students, (the standardizing population of the Inventory). The sample showed "minor divergences" from the norm group.[22] The means of the sample corresponded to the respective 53rd, 49th, 43rd, 37th, and 33rd centile rank value of the standardizing population, but the author does not comment about these differences apart from noting them.

The coefficients of internal consistency ranging from r = .39 to r = .91, as reported in the Manual were accepted,

---

22 Carroll, Op. Cit., p. 627.

and no further estimates of the reliability of the independent variable were made. Inter-rater reliability ranged from r = .664 for Factor T, to r = .806 for Factor S. The findings most relevant for this study, namely the correlation between the Inventory scores and ratings by others, were as follows: for Factor S, r = .37; for Factor T, r = .32; For Factor D, r = .27; for Factor C, r = .26; and for Factor R, r = .45. Carroll regards these correlations as indicating only very moderate validity for the personality inventory. (All correlations significant at P = .01.)[23]

In discussing the results, he reports that the tests were taken seriously, carefully, and honestly by the testees; but because of this, he states that generalizations based on the findings should be limited to situations where falsifications are unlikely to occur.[24]

This study was carried out and reported more carefully than the previously described Guilford-Martin study. Nevertheless, three criticisms can be made, two of which have been previously mentioned in connection with the Guilford-Martin study: i) failure to present a measure of the experimental reliability of the independent variable; ii) selection of the median rating as the criterion score, rather than the

---

23 Carroll, Op. Cit., p. 630.

24 Ibid., p. 631.

mean of the five ratings. It is not explained why the advantages offered by linear transformations were not utilized; iii) assumptions underlying Carroll's use of graphic interpolation were not spelled out.

Personality Test Validation in Correctional Settings.- Doppelt and Seashore report that there exists a dearth of literature in the area of testing in correctional settings.[25] A survey of the literature confirms their statement. Specifically, no study pertaining to the validation of a personality inventory in a correctional setting was found.

This study is an attempt to diminish the paucity of information in this area. The subject of this study is the validation of a personality inventory, specifically the Guilford-Zimmerman Temperament Survey in a correctional institution. The experimental hypothesis stated in the null form is as follows: There is no statistically significant relationship between self ratings on the Guilford-Zimmerman Temperament Survey by individuals under penal custody, and ratings by custodial treatment staff on the same traits using a scale derived from the Survey.

---

25 Jerome D. Doppelt and Harold C. Seashore, "Psychological Testing in Correctional Institutions", Journal of Counseling Psychology, 1959, Vol. 6, No. 1, p. 81.

In this chapter some notions of validity which have direct application to this study were presented. Two studies in which Guilford Personality Inventory Factors were validated by rating techniques were evaluated, and mention was made of the lack of test validation studies in the field of corrections. The chapter concluded with a statement of the hypothesis to be tested.

The next chapter describes the experimental design of this study.

CHAPTER II

EXPERIMENTAL DESIGN

Included in this chapter will be the procedures involved in conducting the study to test the hypothesis proposed in the preceding chapter.

The chapter begins with the description of the experimental sample, and the independent and dependent variables, the Guilford-Zimmerman Temperament Survey and the Rating Scale respectively. This is followed by the presentation of the procedures used in the administration and scoring of the instruments. The chapter concludes with the description of the statistical treatment of the data.

1. Sample.

The Ontario Training Centre, Brampton, is an "open institution" by which is meant that the inmates are not confined to cells, and the doors are unlocked. A selected group of young male, first offenders between the ages of 16-24 years, with an Otis I.Q. or its equivalent of 85 or more, are transferred from the Ontario Reformatory, Guelph, to the Ontario Training Centre, Brampton,[1] to serve the

---

1 For selection standards for Brampton, see Appendix 1. For general description of Brampton, see H.M. Hooper, S. Kean, et al., "The Brampton Story", Canadian Journal of Corrections, Vol. 4, No. 4, October 1962, p. 226-227. Hereafter referred to as O.T.C. Brampton.

remainder of their sentence. These inmates receive training in certain trades and they may in addition follow an academic curriculum in the institution. Physical training is emphasized. The inmates are referred to as "students". They are housed in five dormitories or halls.

The rating scale, the dependent variable in this study demanded familiarity on the part of the staff with the students. Thus, the criterion for inclusion in the sample was at least three months residency at the Ontario Training Centre, Brampton. On July 1, 1962, fifty-six students drawn from the five dormitories in the institution met this criterion. Five of the fifty-six students had to be excluded due to an insufficient knowledge of English, and one student was unavailable at the time of testing. Therefore, the experimental sample consisted of fifty students. For a description of the experimental sample, see Table I.

2. Instruments Used.

a) Guilford-Zimmerman Temperament Survey.- Since the Guilford-Zimmerman Temperament Survey[2] is the independent variable in the present study, a short description of this instrument will be presented.

---

2 J.P. Guilford and Wayne S. Zimmerman, The Guilford-Zimmerman Temperament Survey, Beverly Hills, California, Sheridan Supply Co., 1949. Hereafter referred to as the Survey.

Table I.-

Age, Education and Intelligence[a] of Experimental Sample.

	M	σ	Range
Age	18.46	2.30	16-23
Education	8.68	1.42	6-12
Intelligence	99.54	11.29	81-124

a As measured by Otis Group Intelligence Scale, Advanced Examination.

Buros[3] indicated that this instrument is appropriate for individuals who have achieved at least a Grade IX educational level. The Survey consists of a test booklet containing three hundred questions representing ten factors. For each one of the ten factors, the Survey contains thirty items to which the individual answers "yes", "no", or "?", with instructions to avoid using the latter if possible. The Survey also includes a standard IBM answer sheet with thirty item spaces per column. Five columns appear on each side of the answer sheet. The items in the booklet have been rotated in such a manner that the answers for each trait fall in the same column. It is possible to score the answer sheets by hand or by machine. Two stencils are provided for hand scoring, one for the five traits on the front page, and the other for the five traits on the back page. Scores of zero and one are assigned to the responses according to the scoring stencil. The scores for each factor range from zero to thirty, and with the help of a table, they can be converted into scores on a C scale, a centile rank scale, and a T scale.

In this study only the raw scores obtained by hand scoring were used.

---

3 Oscar K. Buros, (ed.), The Fifth Mental Measurement Yearbook, Highland Park, New Jersey, The Gryphon Press, 1959, p. 65.

The Manual reports internal consistency coefficients between r = .75 to r = .87 for each of the ten factors.

On data collected by Blum,[4] test-retest reliabilities were computed by this writer for nine of the factors of the Survey. The stability of Factor R was reported by Blum.[5] The test-retest sample consisted of seventy students at the O.T.C. Brampton, where approximately two to three weeks had elapsed between test-retest. The obtained reliability coefficients ranged from .010 for Factor O to .760 for Factor E. With the exception of Factors O and F which were not significantly different from zero at P = .05, the others were significant at P = .01. See Appendix 2.

These values are not directly applicable to the present study because: i) the samples are different; ii) the reliability coefficients are based on tests administered at the receptions at the Ontario Reformatory, Guelph, and O.T.C. Brampton respectively. The aforementioned coefficients were listed since no other reports of the stability of the instrument were found.

---

4 Frank Blum, personal communication with the author.

5 Frank J. Blum, An Investigation of the Relationship of a Measure of Extraversion-Introversion and Subsequent Recidivism for a Selected Group of Young Offenders, unpublished master's thesis presented to the School of Psychology and Education of the University of Ottawa, Ontario, May 1963, p. 33.

b) Rating Scale Booklet: Construction.- In current use at the O.T.C. Brampton is a regular monthly Review Board Report[6] which is circulated among the staff and which claims to assess the students' progress at the institution. This report is in the form of a five point adverbial rating scale. Many staff members have privately expressed dissatisfaction with the Review Board Report for the following reasons: i) its "complicated"and ambiguous descriptive terms; ii) the work involved in completing the forms; and iii) doubts concerning the independence of each man's report - in the dormitories one rater, or one rating could, and occasionally did influence another. In constructing the rating scale used in this study, a conscious effort was therefore made to avoid the aforementioned complaints raised against the institutional scale.

Three psychologists, including the writer, and a high school English teacher participated in the construction of the behavioural rating scale for each of the ten Survey factors. The English teacher assisted by indicating possible changes in the wording of the behavioural descriptions of the traits, with the purpose of restricting the reading level of the scale within the Grade VIII limits. This was recommended by the resident psychologist at the O.T.C. Brampton to

---

6 The Review Board Report is a summary rating prepared monthly on each student to measure his degree of adjustment in the institution.

ensure the goodwill and cooperation of all the prospective raters so that the proposed task would not be looked upon with disfavour due to reading and comprehension difficulties.

The construction of the scale was guided by the description of the ten traits given in the Manual of the Survey.[7] This was the basis upon which the behavioural descriptions of the ten traits was made. Criterion for inclusion of any part of the descriptive paragraphs for each pole of each trait called for unanimous agreement by the three participating psychologists. Appendix 3 contains an illustration of the derived rating scale.

c) Rating Scale Booklet: Description.- The rating scale constructed for this study was eleven pages in length. The scale contained the same traits as measured by the Survey. The first page was an introduction page consisting of instructions to the raters and a completed example of the rating procedure. The following ten pages of the rating scale were made up of the ten traits to be rated. Each of the ten pages contained the ratings of a single rater for a maximum of ten students on the same trait. A line drawn across the top centre of each page contained the name of the rater, the date he commenced to rate, and the date he completed his rating. Directly below this line was a large

---

7 Guilford-Zimmerman, Manual of Instructions and Interpretations, Op. Cit., p. 2-3.

rectangular block consisting of specific instructions to the rater cautioning him to consider only that behaviour which he had personally witnessed and reminding him to put only one mark under each number. To the far left hand margin of the page, four blocks were vertically aligned. The first block contained a short description of both the positive and negative poles (that is the socially desirable and less desirable aspects) of the trait under consideration. For the third block the following instructions were given "(...) about half the time like the person in block 2, and about half the time like the person in block 4". Two arrows each extend from blocks two and four. These arrows connote that each of these blocks represents two rating categories, depending upon whether the student manifests the behaviour described in the block "nearly all the time", or "a lot of the time". From block three there extends only one arrow.

The body of the rating scale is composed of ten numbered columns of five boxes each corresponding to the five ways the trait can be manifested. All ten trait pages are of identical format. The ten "trait pages" were put together in a random order, i.e., the numbers on to ten were pulled from a hat in an attempt to control for the constant errors of sequence effect and fatigue. The ten randomized "trait pages", faced with an introduction page, were then stapled.

A unique feature of this rating scale was the fact that the names of the individuals to be rated did not appear on the rating scale pages. To identify and yet maintain the anonymity of both raters and ratees, an individual rater code was printed on the introduction page of each rater's booklet. Each rater was then provided with a cardboard insert containing the names of the students to be rated. When this cardboard insert was fitted over the ten numbered columns, each of the names on the insert was aligned with a numbered column on the rating scale.

This method of anonymity was considered advisable so that student help could be utilized in the clerical preparation of the raw data for analysis.

3. Administration and Scoring of Instruments.

a) Administration of the Survey.- Having obtained the superintendent's permission to collect the experimental data within the institution, the chairman of the student's council introduced the students to the study. It was felt that an introduction by a fellow student would increase the motivation and cooperation of the entire student body. The introduction was made at the Monday morning assembly. The student council chairman announced that those students who had been at the Training Centre for a certain period of time

would be asked to participate in a scientific study to be conducted by the resident psychologist.

The student council chairman implied in his address that by participating in this scientific experiment, the group would not only be contributing to science but would probably be contributing to a practical application of the findings of the study, namely, the reduction of the lengthy testing program at the Ontario Reformatory, Guelph, into a shorter and simpler one. The chairman concluded by explaining that the experiment consists of taking a short test and a longer test, and he further explained that each of the participants would be approached individually concerning the most suitable date for their test.

On the evening following the chairman's address, the resident psychologish accompanied by the chairman visited each of the five dormitories. Each student in the experimental sample was individually approached and asked if he would be free to take the test on Friday morning. Thirty-six of the fifty were free to come on Friday; the other fourteen students received individual call slips to report to the resident psychologist's office any afternoon from Monday to Thursday of the following week.

The test was administered by the resident psychologist to the group of thirty-six students in a classroom. The remaining group of fourteen students in the experimental

sample were administered the test in groups of 2-5 in the resident psychologist's office. The instructions for the Survey were read aloud by the resident psychologist in both cases.

Before the Survey was administered to the experimental sample, each student filled out a self-rating scale. This rating scale was identical to that filled out by the staff with the exception that the "he's" were changed to "I's", and only the responses of one individual were requested for each trait on the scale. The self-rating scale was introduced to increase the face validity of the experiment. The data obtained from this scale was not used in this study.

b) Administration of the Staff Rating Scale.- A total of seventeen staff members participated in this study as raters. Fifteen of these raters were dormitory supervisors, and the remaining two were physical training instructors. Each student in the study was rated by five staff members.[8] These five raters consisted of the two physical training instructors, and the students' three dormitory supervisors.

The cooperation of the staff members participating in the study was encouraged by the resident psychologist. This

---

8 This was true in all but ten cases. Rater D was found to be uncooperative, so in the aforementioned cases only four raters in toto rated these students.

was accomplished by personally approaching each prospective rater approximately two weeks prior to the commencement of the study. At this time it was explained to the raters that the experiment involved rating a group of students known to the rater for at least three months on the basis of ten personality traits. All of the raters appeared willing to participate.

On the morning the experiment was started, seven staff members were approached individually and each was given the rating scale with the appropriate cardboard insert containing the names of the students he was to rate. The raters were directed to read the instructions on the cover page and to proceed from there. A check on the same afternoon revealed that only one of the aforementioned seven raters had completed the ratings; the other six raters claimed that they could not proceed with the ratings as they failed to comprehend the instructions. It was discovered that the confusion experienced by the raters was due to the lack of instructions concerning the use of the cardboard insert of names. As a consequence, the raters failed to comprehend that the cardboard inserts were employed as a substitute for the printed names of the subjects on each individual rating scale.

Due to the aforementioned confusion covering the rating scale instructions, it was decided that individual instructions be given in the use of the rating scale to each

rater by the resident psychologist. This instruction consisted of the psychologist reading a copy of the cover page while the rater followed it on his own sheet. The cardboard insert containing the names of the students to be rated was then placed over the numbers from one to ten on the first trait of the individual scale. It was then pointed out to the rater that this procedure was to be followed for the other nine traits. Any questions that the rater had were then answered.

At this time, it was indicated to the rater that secrecy and anonymity dictated the need for placing the students' names on cardboard inserts rather than printing them on the individual scale. The raters were requested to treat their cardboard inserts as privileged information; and they were also asked not to discuss their ratings with other staff members. The rating booklets were to be completed as soon as possible and deposited in the resident psychologist's locked mail box.

As a result of this explanation, the staff members were able to proceed with the experiment. Within three days all the ratings were completed and accounted for.

c) Reliability Study of the Rating Scale.- The "reliability phase" of the rating scale commenced eight days following the return of the last rating booklet. The raters were individually contacted as in the "rate phase" procedure,

and asked to repeat the ratings a second time. All the raters consented to repeat the rating procedure and, on the following day, the rating booklets were re-distributed. Again, within three days, all booklets were returned.

d) Scoring of the Survey.- The answer sheets were hand scored using two stencils, one for the five traits on the front page and the other for the five traits on the back page. Scores of 0 and 1 were assigned to the responses according to the scoring stencils.

e) Scoring of the Rating Scale.- The raw data from the rating scale was scored, checked, and tabulated by four O.T.C. Brampton students chosen for their maturity, intelligence, arithemetical ability, and accuracy. These four students were experienced in this type of work as they had participated in the scoring of data from another project. A periodic spot-check of their work was made by the resident psychologist. To maintain anonymity, the dormitory and rater code was printed on each sheet of the rating scale by the resident psychologist.

The scorers assigned raw scores from 1 - 5 to the five rating scale categories, with a score of one being assigned to the most negative category, and five to the most positive category.

4. Statistical Treatment of the Data.

a) Rating Scales.- In instances where different raters rate different parts of the sample, Guilford suggests weighing the raw scores of each rater.[9] This method is designed to reduce the constant errors of rating and allows the combination of ratings. Therefore the following steps were taken:

i) using the raw scores, the mean and standard deviation were calculated for each rater on each trait of the rating scale;

ii) a new mean of 2.5 and a new standard deviation of .8 were assigned to the ratings of each rater on each trait;

iii) each raw score for each rater on each trait was equated against a new weighted (standard) score by means of the following formula:[10]

$$X_2 = M_2 + \frac{\sigma_2}{\sigma_1}\left(X_1 - M_1\right)$$

where: X_2 - new score to be found
M_2 - assigned mean (2.5)
σ_2 - assigned standard deviation (.8)
σ_1 - calculated standard deviation
X_1 - raw score
M_1 - calculated mean

---

9 J.P. Guilford, Psychometric Methods, New York, McGraw-Hill, 1954, p. 289.

10 "Hull's Method", quoted by David Wechsler, The Measurement of Adult Intelligence, Third Edition, Baltimore, Williams & Wilkins Co., 1944, p. 219.

iv) the weighted scores for each rater on each trait were then contracted into a five point scale in the following manner:

a) all weighted scores of -.5 and below became 0,
b) all weighted scores of -.4 to 1.4 became 1,
c) all weighted scores of 1.5 to 2.4 became 2,
d) all weighted scores of 2.5 to 3.4 became 3,
e) all weighted scores of 3.5 to 4.4 became 4, and
f) all weighted scores of 4.5 and above became 5;

v) for each member of the sample, the average of the weighted ratings of the five raters was calculated for each trait.[11]

The same procedure as described above was repeated for the "reliability phase" of the rating scale.

b) Relationship between Survey and Rating Scale.- To justify using a Pearson Product Moment Formula to test the hypothesis, a graphic check of the independent and dependent variable on each of the ten traits indicated that no curvilinear relationship existed between them.

The following formula was used in the calculation of the coefficients of correlation between the independent and dependent variable on the ten personality traits:

$$r_{xy} = \frac{N\Sigma XY - [(\Sigma X)(\Sigma Y)]}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

where: N = 50
X = Guilford-Zimmerman Factor Scores
Y = Average weighted trait scores.

---

11 This applied for all except the ten students previously mentioned. See footnote 8.

In this chapter the tools and procedure used to test the hypothesis were described.

The next chapter will deal with the findings of this study, their interpretation, and their discussion.

CHAPTER III

FINDINGS AND DISCUSSION

In this chapter the findings of the experiment will be presented, interpreted and discussed. This will be followed by a consideration of some of the shortcomings of the design of the present experiment that may have contributed to the acceptance of the null hypothesis for nine of the ten traits. A short discussion will be presented of some of the theoretical issues involved in the problems of cross media validation of personality inventories. In concluding the chapter, the application of the Guilford-Zimmerman Temperament Survey[1] in an Ontario Reformatory will be considered as an example of one use of psychological tests in a correctional setting.

1. Findings: Presentation and Interpretation.

a) Split Half Reliability of the Independent Variable.- First-half-second-half internal consistency coefficients were calculated for the experimental sample on each of the ten factors measured by the Survey. These findings are presented in Table Ia. All are significant

---

1 J.P. Guilford and Wayne S. Zimmerman, The Guilford-Zimmerman Temperament Survey, Beverly Hills, California, Sheridan Supply Co., 1949. Hereafter referred to as Survey.

Table Ia-

First Half-Second Half Internal Consistency of Guilford-Zimmerman Temperament Survey Factors (N-50).

Survey Factors	r^a
G	.657
R	.746
A	.848
S	.891
E	.849
O	.805
F	.818
T	.870
P	.834
M	.717

a All coefficients are significant at $P = .01$.

at P = .01, and they ranged from .657 for Factor G, to .891 for Factor S. Most of these coefficients are generally similar to the ones presented in the Manual of the Survey.[2]

b) Stability of the Dependent Variable.- The rate-rerate reliability coefficients for the ten traits measured by the rating scale are presented in Table II.[2a] These coefficients range from .178 for Factor O, to .872 for Factor T. The reliability coefficient for Factor O was not significantly different from zero (P = .05). This is probably due to the instability of the trait rather than the unreliability of the raters. (See Appendix 2.) All the other factors were significant at the .01 level of probability.

c) Relationship between the Survey and the Rating Scale.- The obtained coefficients of correlation between the factors measured by the Survey and the same traits measured by the rating scale are presented in Table III. The aforementioned coefficients range from .025 for Factor T, to .283 for Factor R. With the exception of Factor R which is significant at the .05 level of probability, none of the other obtained coefficients are significantly different from zero. Thus the null hypothesis was rejected for Factor R and accepted for the other nine factors.

---

2 Guilford and Zimmerman, The Guilford-Zimmerman Temperament Survey, Manual of Instructions and Interpretations, Op. Cit., p. 6. Hereafter referred to as the Manual.

2a The six students omitted from the experimental sample were included in the "reliability sample". This was done in order to increase the size of the "reliability sample" even though in this way the experimental and the "reliability" samples were no longer identical.

Table II.-

Test-Retest Reliability Coefficients of the Rating Scale Traits (N-56).

Rating Scale Traits	r^a
G	.600
R	.830
A	.807
S	.648
E	.618
O	.178
F	.726
T	.872
P	.643
M	.539

a All coefficients except that for Factor O significant at P = .01. Coefficient for Factor O not significantly different from zero. (P = .05).

Table III.-

Obtained Coefficients of Correlation between Guilford Personality Traits and Ratings on the Same Traits as Measured by the Rating Scale (N-50).

Guilford Personality Traits	r
G	.235
R	.283[a]
A	.249
S	.189
E	.040
O	.034
F	.058
H	.025
P	.152
M	.192

a Significant at P = .05.

If the findings of the previous studies of Guilford-Martin,[3] and Carroll[4] are compared with the present study, it appears that for a certain universe the validity of the Guilford Personality Inventory Traits is minimal. To facilitate comparison of the present findings with those of the two previous studies of Guilford-Martin and Carroll, Table IV presents the pertinent data of the three studies.

An explanation of why the findings of the present study are considerably more restricted than the findings of Guilford-Martin and Carroll will be attempted in terms of three kinds of judgments called for in the execution of the present project: i) With respect to the rating scale, there is no assurance that it is measuring the identically same phenomena as that measured in the Survey - a problem also shared by Guilford-Martin and Carroll. In addition to the aforementioned, a conscious bias was present in the construction of the rating scale for this study - only that phenomena readily amenable to observation was included. ii) Consideration must be given to the question of whether staff reports

---

3 J.P. Guilford and Howard Martin, "Age Differences and Sex Differences in Some Introversive and Emotional Traits", Journal of General Psychology, 1944, Vol. 31, Second Half, p. 219-229.

4 John B. Carroll, "Ratings on Traits Measured by a Factored Personality Inventory", Journal of Abnormal and Social Psychology, 1952, Vol. 47, p. 626-632.

Table IV.-

Correlations Obtained on Three Studies Using Guilford Personality Factors and Ratings by Others, on the Same Traits.

Author	Types of Raters/ratees	(N)	r Coefficients S	T	D	C	R
Guilford and Martin(1938)	Peer ratings/ Psychology Stu	50	.70	.08	.60	.20	.48
Guilford and Martin(1939)	Peer ratings/ Psychology Stu	51	.43		.39		.46
Carroll(1952)	Peer ratings/ Psychology Stu	110	.37	.32	.27	.26	.45
Marshall(1964)	Custodial Staff/ Inmates	50	.19	.03		.04	.28

Single underline: Significant at .01 level of probability.

Double underline: Factor E (Marshall's study) constructed from a combination of Factors C and D.

Triple underline: Significant at .05 level of probability.

and peer reports with respect to the existence of the same phenomena would differ. By virtue of their position as authority figures in the institution, staff do not see everything that goes on. Also Allport's observation based upon various research findings that:

> It is (...) that members of the (...) same age (...) are ordinarily the best judges of one another. Students are as a rule better judges of other students than of professors - and vice versa.[5]

iii) The validity of the self reports by students - the question here is not whether the students deliberately falsified their answers, but whether they possess the reflective ability to correctly assess themselves. This point becomes meaningful when considered in relation to the one positive finding of the study - Factor R. The mean score of Factor R for the experimental group was 13.2, with a standard deviation of 5.25. The corresponding statistics for the standardizing sample reported in the Manual were 16.9 and 4.94 respectively.[6] The items making up this factor suggest that by asking the person with a low score on this factor to fill out a questionnaire about himself, one asks him to do something which he is not used to doing, and which he furthermor

---

5 Gordon W. Allport, Pattern and Growth in Personality, New York, Holt, 1963, p. 507.

6 Guilford-Zimmerman, Manual, Op. Cit., p. 7.

does not like doing. It is possible that this fact could reduce the validity of the self inventory.

2. Shortcomings in the Design of the Present Study.

Some shortcomings in the design of the present study may in part account for the predominantly negative findings. Some of these shortcomings will be grouped with respect to the independent variable, and some with respect to the dependent variable. The nature of these shortcomings is possibly such that they would tend to reduce the coefficients of correlation between the test and the rating scale.

With respect to the independent variable: i) No test-retest reliability measure was obtained, therefore, it is difficult to differentiate theoretically between the stability of the behaviour being observed, and the stability of the ratings. Thus two different sources of instability may be contributing to a decrease in the reliability of the ratings. ii) The suggested educational level of the Survey as indicated by Buros[7] is Grade IX, yet the educational level of the experimental sample was Grade 8.7. Blum[8] indicated that the

---

7 Oscar K. Buros, (ed.), The Fifth Mental Measurement Yearbook, Highland Park, New Jersey, The Gryphon Press, 1959, p. 65.

8 Frank J. Blum, An Investigation of the Relationship of a Measure of Extroversion-Introversion, and Subsequent Recidivism for a Selected Group of Young Offenders, unpublished master's thesis, presented to the School of Psychology and Education of the University of Ottawa, Ontario, May 1963, p. 77.

students tend to experience difficulty understanding certain words in the Survey. Thus the discrepancy between the desirable and the actual reading-educational level of the sample could reduce the validity of the Survey.

With respect to the dependent variable: i) Failure to obtain the ratings of the shop instructors. In line with Guilford's comments that certain traits enter into the determination of certain areas of behaviour but do not affect all or nearly all behaviour[9] is the frequent finding of the institutional Review Board[10] that the behaviour of the students in the shops is markedly different from that observed in the dormitories.[11] Thus the rating scale disregarded situations where different aspects of the trait could be manifested by a failure to sample part of the student's day. ii) In the calculation of the average weighted score for each student on each trait of the rating scale, equal emphasis was assigned to the scores of each of the five raters even though certain raters have more contact with the students than others(dormitory

---

9 J.P. Guilford, Personality, New York, McGraw Hill, 1959, p. 74.

10 This is a board composed of senior staff which meets weekly to evaluate the progress of students in the institution.

11 Scott Kean, Social Worker, O.T.C., Brampton, personal communication.

supervisors). Thus different results may have been obtained if the importance placed on each score was in proportion to the amount of time the rater spent with the students.

The theoretical problem: With respect to the general problem of cross media validation of psychological measurements, this study has referred to the hopeful rationale of Guilford,[12] and the cautioning and sobering comments of Ellis[13] and Anastasi.[14] More recently Becker[15] presented a "counter-rationale" to the possibilities of substantial cross media matchings. He not only criticized Cattell's claims of almost one to one correspondence between the factorial composition of behaviour ratings and personality inventories on the ground of Cattell's method of factor analysis, but he also expressed doubts about the possibility of one to one matching on psychological grounds for the following reasons:

---

12 Guilford and Martin, Op. Cit., p. 226.

13 A. Ellis, "The Validity of Personality Questionnaires", Psychological Bulletin, 1946, Vol. 43, p. 385-440.

14 Anne Anastasi, "The Concept of Validity in the Interpretation of Test Scores", Educational and Psychological Measurement, 1950, Vol. 10, p. 67-78.

15 Wesley C. Becker, "The Matching of Behaviour Rating and Questionnaire Personality Factors", Psychological Bulletin, 1960, Vol. 57, No. 3, p. 201-212.

> (a) defenses of various sorts act to limit self-awareness, (b) the frame of reference (adaption level) for making self-judgments will vary from person to person, making such scores not comparable, and (c) the experiences to be judged of necessity constitute different samples of behaviour for self and other raters.[16]

3. Testing in Corrections: The Use of the Survey in an Ontario Reformatory.

The findings of the present study, and of earlier validation studies of the Guilford Personality Inventory Factors indicate that Guilford's statement, cited in the Introduction, that the factors represent "reliable, descriptive categories"[17] has to be considered with qualifications. Caution in interpreting test findings becomes especially warranted if test scores and their interpretations are to be used for selection-screening purposes. Such is the case in one Ontario reformatory.

When a first offender between the age of sixteen and twenty-four years and with a reformatory sentence of at least three months is selected-screened for placement in various institutions, the selection board, ordinarily consisting of the two superintendents of the two training centres for young male adult first offenders[18] and a psychologist from

---

16 Becker, Op. Cit., p. 210.

17 Guilford-Zimmerman, Manual, Op. Cit., p. 7.

18 These refer to the superintendents of O.T.C. Burtch, and O.T.C. Brampton.

the Ontario Reformatory, Guelph, decide on the basis of a majority vote to which institution the offender will be placed. Before any decision regarding allocation is made, a copy of the psychological report on each prisoner to be considered is given to each member of the board, and is read aloud. This report is based on various psychometrics and a single interview with the psychologist. While statements "he is unfriendly, impulsive and unstable" that did on occasion appear ten years ago no longer appear, statements like "our test findings indicate that he is unfriendly, impulsive, and unstable" are not infrequent. In spite of the qualifying phrase "test findings", and the presence of a psychologist on the board, one wonders how statements of this nature may contribute to the decision of a group where decision goes by majority vote, and two laymen can constitute a majority.

It is true that until a study similar to the present study is carried out at the Ontario Reformatory, Guelph, one could possibly claim that the Survey yields valid descriptive information at Guelph, but not at Brampton. However, if such claims were made, it would appear to be extremely tenuous. After all Guilford himself says that, "(...) instruments developed to assess some traits are not applicable to all segments of the population".[19]

---

19 Guilford, Personality, Op. Cit., p. 73-74.

In this chapter the results of the study were presented, interpreted and compared to two pertinent studies. An explanation of the predominantly negative findings of the present study was attempted in terms of the shortcomings in the design and general theoretical considerations. The chapter ended with a suggestion that caution be exercised in the interpretation of Survey findings in Ontario reformatories.

In the next chapter, a short resume of the results of this study together with some conclusions that can be drawn will be presented. The chapter will end with some suggestions for further research.

SUMMARY AND CONCLUSIONS

In examining the relationship between the ten traits as measured by the Guilford-Zimmerman Temperament Survey, and the same ten traits as measured by a rating scale constructed on the basis of the Survey, it was found that the null hypothesis was accepted for nine of the ten traits. In the case of Factor R, the hypothesis was rejected: the obtained Pearson-Product Moment coefficient of correlation was significant at the .05 level of probability.

The findings of the present study tend to support the previous findings of Guilford-Martin and Carroll, namely that the concurrent-predictive validity of the Guilford Personality Factors is moderate at best.

With respect to selection purposes in custodial settings, this study indicates the need for more caution and criticality in using the Guilford-Zimmerman Temperament Survey as a basis of personality description.

Suggestions for further research can be presented in three categories.

a) Using Available Data for a Pilot Study.- i) Relate the self-ratings which were utilized in this study in order to increase the face validity of the experiment, to Survey scores and ratings by others. This would provide another medium by which the present data could be compared

to the previous studies of Guilford-Martin and Carroll. If the self ratings were to be employed in a future study their stability would have to be established. ii) In arriving at an average weighted rating, assign emphasis to the scores of each rater in proportion to the amount of time he spends with the students, and see whether or not the obtained correlations are thereby changed.

b) Using New Data in Same Setting.- i) Employ a larger sample and utilize the ratings of shop instructors and possibly school teachers. ii) Utilize peer ratings, that is ratings by other students, as a criterion. By so doing this project would be even more comparable to the previously cited studies. iii) By using Survey scores as a criterion, a multiple coefficient of correlation could be calculated whereby it would be possible to evaluate the respective contributions of ratings by dormitory supervisors, shop instructors, academic school teachers, physical training instructors and other students.

c) Repeating the Study in Different Settings.- As the rating scale used in the study appears to be a reliable instrument in at least one setting, the present study with the improvements-extensions suggested above, could be repeated in other settings, both penal and non-penal, to determine if wider generalization of the findings is possible.

BIBLIOGRAPHY

Allport, Gordon W., Pattern and Growth in Personality, New York, Holt, 1963, p. 507.

A revision of the 1937 classic. His statement, based on the summary of the most recent relevant research about age similarity between rater and ratee making for improved judgments, was used as one possible explanation for the predominantly negative findings of this study.

Anastasi, Anne, "The Concept of Validity in the Interpretation of Test Scores", Educational and Psychological Measurement, 1950, Vol. 10, p.67-78.

In a well written and comprehensive article, the author elucidates what she means by validity. This study validates the Guilford-Zimmerman Temperament Survey in Anastasi's sense.

Becker, Wesley C., "The Matching of Behaviour Rating and Questionnaire Personality Factors", Psychological Bulletin, 1960, Vol. 57, No. 3, p. 201-212.

This article challenges assertions made in recent publications by Cattell and his co-workers to the effect that present research has shown in most cases a one-to-one matching of behaviour ratings and questionnaire Personality factors. A review of the evidence fails to support these assertions. Psychological factors involved in cross media matching are discussed.

Carroll, John B., "Ratings on Traits Measured by a Factored Personality Inventory", Journal of Abnormal and Social Psychology, 1952, Vol. 47, p. 626-632.

Carroll's study is similar to the present study except that he uses peer ratings and his factors are from the Inventory of Factors STDCR. This author cautions against being too optimistic about the possibility of obtaining valid inter-correlations between these two variables.

Ellis A., "The Validity of Personality Questionnaires", Psychological Bulletin, 1946, Vol. 43, p. 385-440.

The author reviews several studies in which personality questionnaires have been validated against different types of objective criteria among which have been rating techniques. None of these criteria have been found to be too satisfactory. The author wonders whether more validation studies, making possible a finer breakdown of the data, will indicate that specific personality questionnaires are valid for some situations and not for others.

Guilford, J.P., Personality, New York, McGraw Hill, 1959, p. 73-74.

A text based on Guilford's most recent systematic exposition of his conception of personality. He suggests, on the one hand, that instruments developed to assess some traits are not applicable to all segments of the population, and on the other hand that certain traits do not affect behaviour in certain areas. The former suggestion is the "raison d'etre" for this study, the latter points toward a possible deficiency of the design of the study.

-------- and Howard Martin, "Age Differences and Sex Differences in Some Introversive and Emotional Traits", Journal of General Psychology, 1944, Vol. 31, p. 219-229.

Generally this is a poorly executed study and one which does not seem worthy of the Guilford of today. This article's importance in terms of the present study lies in the fact that rating techniques were used to validate personality inventory factors. It also provides a theoretical framework inasmuch as Guilford spells out what he considers to be the components of personality.

-------- and Wayne S. Zimmerman, The Guilford-Zimmerman Temperament Survey: Manual of Instructions and Interpretations, Beverly Hills, California, Sheridan Supply Co., 1949, 12 p.

Apart from the description of the ten Guilford personality factors, measured by both the independent and dependent variable in this study, the claim about the factor scores giving dependable descriptive categories of personality initiated the present experiment.

APPENDIX 1

SELECTION STANDARDS FOR THE ONTARIO TRAINING CENTRE, BRAMPTON

APPENDIX 1

SELECTION STANDARDS FOR THE ONTARIO TRAINING CENTRE, BRAMPTON

AGE;	16 to 24 years inclusive. (i.e. up to their 25th birthday at time of admission to Brampton).
INTELLIGENCE:	A. An I.Q. of 85 or higher - less than 85 are sent to O.T.C., Burtch. B. Inmates with an I.Q. of 80 to 84 may be sent to either the O.T.C., Brampton or the O.T.C., Burtch.
SENTENCE:	Generally speaking an inmate should have not less than three months of his sentence left to serve at time of screening.
CUSTODIAL RISK:	The inmate should be considered to be sufficiently stable to adjust to an open institution.
CRIMINAL RECORD:	A. Usually the inmate who has already received training at the O.T.C., Brampton, on a previous conviction will not be considered for further training at that Centre. B. At time of screening the inmate should not have had more than three previous adult convictions or have served a sentence of more than six months.
MENTAL HEALTH:	An inmate should be rejected for training if there is evidence of: A. Psychosis B. Severe Neurosis C. Homosexuality D. Arson.

APPENDIX 2

TEST-RETEST RELIABILITY STUDY OF THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

APPENDIX 2

TEST-RETEST RELIABILITY STUDY OF THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

Table V .-

Test-Retest Reliability Study of the Guilford-Zimmerman Temperament Survey Carried Out at G.T.C. Brampton in 1961. (Guelph-Brampton) (N:70).

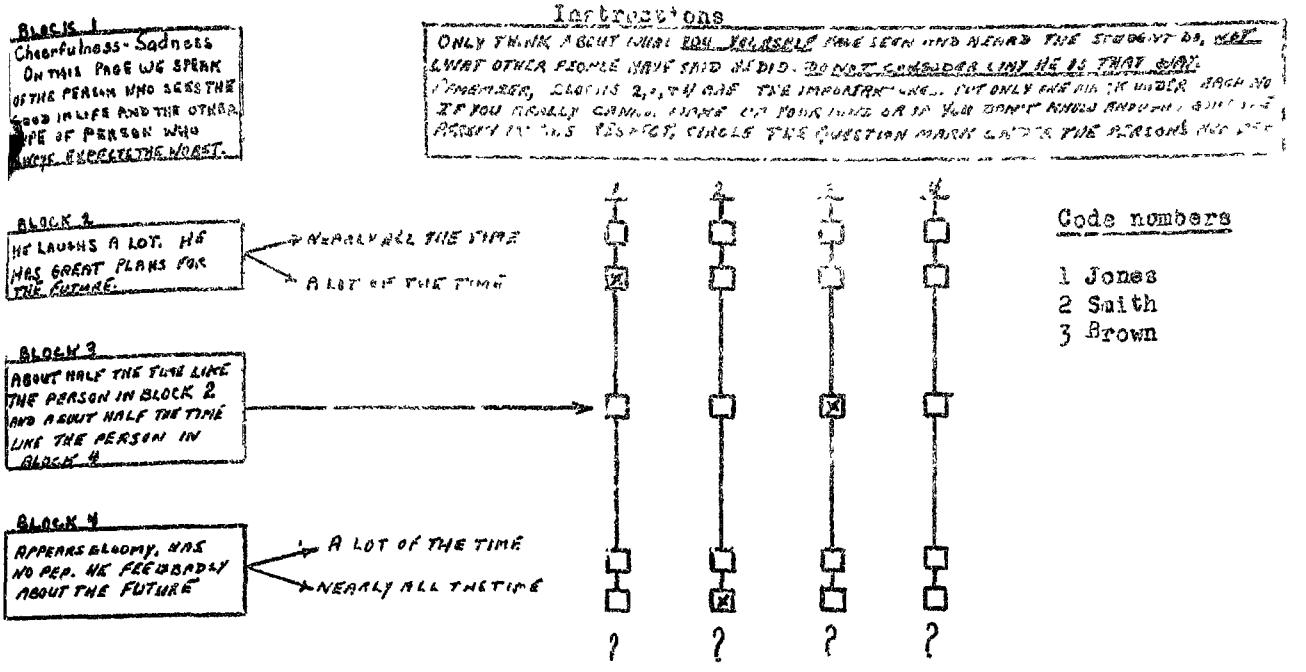
Survey Factors	r
G	.754
R	.576
A	.755
S	.685
E	.760
O	.010
F	.040
T	.534
P	.662
M	.682

All coefficients significant at P = .01 with the exception of Factors O and F which were not significantly different from zero.

APPENDIX 3

CONSTRUCTED RATING SCALE

EXAMPLE APPENDIX 3



EXPLANATION OF EXAMPLE

This paper talks about (x no.) of students. On each page you are supposed to say how much each student acts or talks in a certain way. We will talk about 10 of the many ways students can act and talk, and we will use one page for each of these ways. This gives us ten pages altogether.

There are five blocks of writing on each page. Read first the top block at the right of the page. This block gives you advice on what to do.

Read next, block 1 at the top left hand side of the page. This block tells you a little bit about what blocks 2, 3, & 4 talk about. Then you should read blocks 2, 3, & 4 at the left hand side of the page. Blocks 2, 3, & 4 are the important ones.

Block 2 tells about one kind of acting and talking a student can do in one of our ten ways.

Block 4 tells us about the kind of acting and talking that is the most different from block 2.

Block 3 is for talking and acting that is half like block 2 and half like block 4.

For example a student could be either cheerful or sad. Then cheefful acting and talking would be in the 2nd block and sad acting and talking would be in block 4.

We would also want to know whether the student acts and talks cheerful a lot of the time or whether he does it nearly all of the time. On the other hand, if he talks and acts sad, we would want to know whether he does it a lot of the time, or nearly all the time.

Suppose we fixed it so that we could mark the kind of talking and acting that a student does, and the number of times he does it. Our page would look like the example at the top of this page.

If #1 (Jones) were cheerful a lot of the time, and #2 (Smith) were sad nearly all the time, and #3 (Brown) was cheerful half of the time, and sad half the time, we would mark the page as above. If you really cannot make up your mind, or if you don't know enough about the person in any one of the ten different ways, circle the question mark

Make sure you put down a mark for every student on every page. Put your X - mark in one of the little boxes under the student's number, unless you have to choose to circle the question mark.

Now turn the page and start on page 1. Go on until you have completed page 10.

APPENDIX 3

RATER | DATE COMMENCED | DATE COMPLETED

Only think about what you yourself have seen and heard the student do, not what other people have said he did. Do not consider why he is that way. Remember blocks 2, 3, & 4 are the important ones. Put only one mark under each number. If you really cannot make up your mind or if you don't know enough about the person in this respect, circle the question mark under the person's number.

BLOCK 1

FAST, ENERGETIC-SLOWNESS - On this page we are talking about the person who has strong drives, is energetic and active and the other kind of person who is slow and not very active.

BLOCK 2

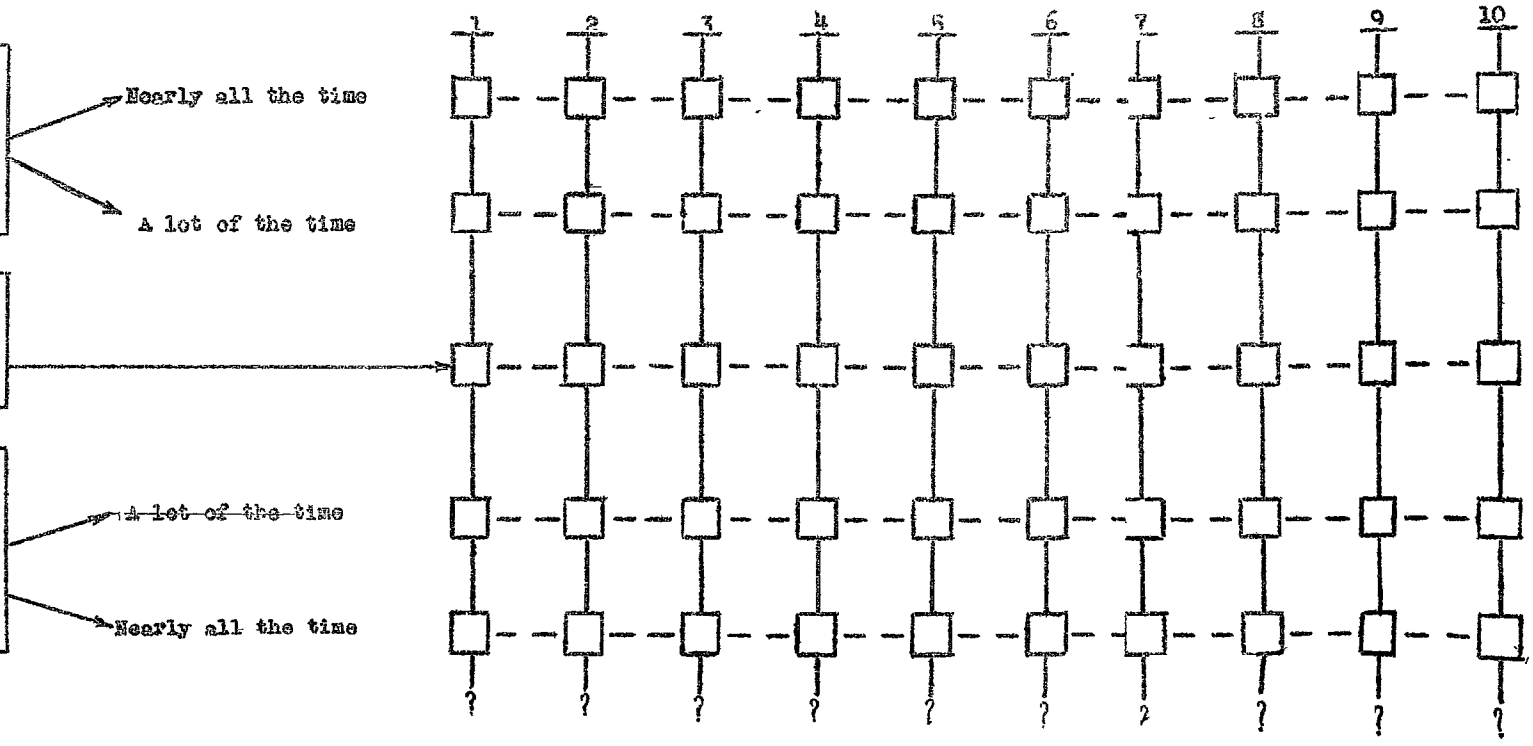
He is someone who is always on the go. He has lots of pep. He likes to do things, fast and well.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He takes his time doing things and rests a lot. He gets tired easily and he doesn't get much done.



BLOCK 1

SERIOUS MINDEDNESS - HAPPY GO LUCKY

On this page we are talking about the person who is over-restrained, over serious, and the other kind of person who is happy-go-lucky, carefree and impulsive.

BLOCK 2

He is a very serious minded person. He takes a long time to make up his mind and do things, but once he starts something he hardly ever gives up. He has a lot of self control.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He likes things if they are lively and exciting. He usually acts first and thinks later. He gets really interested in something and then gets tired of it before he finishes it. He acts as though he hasn't a care in the world.

BLOCK 1

TAKING CHARGE - FOLLOWING - On this page we are talking about the person who is a leader, and the other kind of person who is a follower.

BLOCK 2

He likes to run things and doesn't take anything from anybody that he thinks he shouldn't. He likes talking to people. He is a little pushy.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He lets people push him around. He hardly ever takes a stand. He lets somebody else run things.

BLOCK 1

LIKING FOR COMPANY - LIKING FOR BEING ALONE - On this page we are talking about the person who is at ease with others and enjoys their company, and the other kind of person who is reserved and withdrawn and hard to get to know.

BLOCK 2

He has lots of friends and likes getting around. He is a good talker and likes to be in on everything.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He doesn't have very many friends. He doesn't talk much. He doesn't like doing things with a bunch of others. He sticks to himself a lot.

BLOCK 1

EVEN TEMPERED - MIXED UPNESS - On this page we are talking about the person who is cheerful and able to look on the bright side of life, and the other kind of person who is gloomy and depressed.

BLOCK 2

He's not a moody person. He's usually cheerful and doesn't have too many ups and downs.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He feels really good one day, and lousy the next. He is gloomy. He usually thinks the worst thing of all will happen. He lets things get him down long after they've happened.

BLOCK 1

THICK SKINNED - TOUCHINESS - On this page we are talking about the person who is thick skinned and the other kind of person who is touchy and overly sensitive.

BLOCK 2

He is really thick skinned. He doesn't care when people talk about him.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He is very touchy and very easily hurt. He thinks only of himself. He thinks almost everybody is against him. Often he thinks people are talking about him even when they aren't.

BLOCK 1

FRIENDLINESS - HOSTILITY - On this page we speak about the person who desires to be liked and who likes to please others, and the other kind of person who has a fighting attitude.

BLOCK 2

Even if other people are mean and unfriendly, he is not mean in return. He doesn't mind being told what to do.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He doesn't like being told anything, and will fight for the smallest reason. He likes to have his own way with others.

BLOCK 1

THINKING - DOING THINGS - On this page we are talking about the person who tends to think a lot about things before doing something, and the other kind of person who seems to like acting before thinking.

BLOCK 2

He likes to think about why things happen in life. He seems to think a lot before he does things. He seems to turn it over in his mind why people do things.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He likes jobs where he doesn't have to think. He likes moving around more than thinking. He does not like planning things.

BLOCK 1

GETTING ALONG - FAULTFINDING - On this page we talk about the person who has the ability to get along with others, and the other kind of person who is fault-finding and critical.

BLOCK 2

He seems to feel that people are usually honest and good. He puts up with other people's faults. He does not mind if other people see things differently from him. He can get along well with others.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He feels most people are stupid, and follow around like sheep. He thinks there are too many stupid laws and rules that get in a person's way. He thinks people help themselves before anyone else, and he doesn't get along too well with others.

BLOCK 1

MANLINESS - On this page we are talking about the person who acts and has interests that are typical of men, and the other kind of person who is of a less rugged type.

BLOCK 2

He likes rugged work, like building things or hunting, or he might like studying scientific things. He seems to be hard boiled. He hardly ever shows his feelings. Things that might sicken others a little bit don't bother him at all.

BLOCK 3

About half the time like the person in block 2 and about half the time like the person in block 4.

BLOCK 4

He likes to help people in trouble. He might also like to take care of animals. He shows his feelings if he likes somebody. He might also show it if he feels sad. He likes clothes and styles. He might like music, reading or even flowers.

APPENDIX 4

RAW DATA OF THE RATING SCALE

Table VI.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores, For All Raters on Factor C.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-9.8[b]	0[c]	-5.7	0	-1.6	0	2.5	3	6.6	5	-10.5	0	-6.6	0	-2.7	0	1.2	1	5.0	5
B	-7.5	0	-4.3	0	-1.0	0	2.3	2	5.6	5	-8.3	0	-4.5	0	- .8	0	3.0	3	6.8	5
C	-6.0	0	-2.9	0	.3	1	3.5	4	6.7	5	-7.6	0	-3.8	0	0	1	3.8	4	7.6	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-7.3	0	-3.6	0	0	1	3.6	4	7.3	5	-7.3	0	-3.4	0	.4	1	4.2	4	8.0	5
F	-7.2	0	-4.1	0	- .9	0	2.2	2	5.3	5	-7.6	0	-4.3	0	-1.0	0	2.3	2	5.6	5
G	-8.8	0	-5.4	0	-2.0	0	1.3	2	4.7	5	-4.7	0	-1.6	0	1.6	2	4.7	5	7.8	5
H	-9.0	0	-5.5	0	-2.1	0	1.4	1	4.8	5	-10.0	0	-6.4	0	-2.9	0	.7	1	4.3	4
I	-8.8	0	-5.4	0	-2.0	0	1.3	1	4.7	5	-6.9	0	-3.1	0	.8	1	4.6	5	8.4	5
J	-5.5	0	-2.5	0	.6	1	3.7	4	6.8	5	-8.2	0	-5.0	0	-1.9	0	1.3	1	4.4	4
K	-5.9	0	-2.6	0	.7	1	3.9	4	7.2	5	-6.1	0	-3.1	0	0	1	3.1	3	6.1	5
L	-6.5	0	-3.4	0	- .3	1	2.8	3	5.9	5	-8.5	0	-4.8	0	-1.1	0	2.6	3	6.3	5
M	-10.7	0	-7.1	0	-3.6	0	0	1	3.6	4	-8.5	0	-5.3	0	-2.2	0	.9	1	4.1	4
N	-8.9	0	-5.6	0	-2.3	0	1.0	1	4.3	4	-8.9	0	-5.6	0	-2.3	0	1.0	1	4.3	4
O	-8.6	0	-5.0	0	-1.4	0	2.2	2	5.8	5	-9.8	0	-6.2	0	-2.5	0	1.1	1	4.7	5
P	-7.1	0	-3.9	0	- .6	0	2.6	3	5.8	5	-7.3	0	-3.8	0	- .3	0	3.1	3	6.6	5
Q	-7.8	0	-4.4	0	-1.0	0	2.4	2	5.8	5	-9.4	0	-5.8	0	-2.2	0	1.4	1	5.0	5

a Raw scores.
b Weighted scores.
c Contracted Weighted scores.

Table VII.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores For All Raters on Factor R.

	Test										Retest									
	1[a]		2		3		4		5		1		2		3		4		5	
A	-5.7[b]	0[c]	-2.5	0	.6	1	3.8	4	6.9	5	-6.6	0	-3.3	0	0	1	3.3	3	6.6	5
B	-6.8	0	-3.6	0	- .3	1	2.9	3	6.2	5	-7.3	0	-3.8	0	- .3	1	3.1	3	6.6	5
C	-6.2	0	-3.0	0	.3	1	3.6	4	6.9	5	-6.7	0	-3.0	0	.7	1	4.5	5	8.2	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-7.2	0	-4.1	0	- .9	0	2.2	2	5.3	5	-8.1	0	-4.8	0	-1.6	0	1.6	3	4.8	5
F	-6.3	0	-3.2	0	0	1	3.2	3	6.3	5	-6.7	0	-3.2	0	.4	1	3.9	4	7.4	5
G	-5.5	0	-2.3	0	1.0	1	4.2	4	7.4	5	-6.1	0	-3.1	0	0	1	3.1	3	6.1	5
H	-7.3	0	-4.3	0	-1.2	0	1.8	2	4.9	5	-7.5	0	-4.4	0	-1.3	0	1.9	2	5.0	5
I	-6.7	0	-3.4	0	0	1	3.4	3	6.7	5	-7.5	0	-3.8	0	0	1	3.8	4	7.5	5
J	-6.4	0	-3.4	0	- .3	1	2.7	3	5.8	5	-7.1	0	-4.0	0	- .9	0	2.2	2	5.3	5
K	-6.7	0	-3.5	0	- .3	1	2.9	3	6.0	5	-6.9	0	-3.8	0	- .6	0	2.5	3	5.7	5
L	-6.7	0	-3.6	0	- .6	0	1.8	2	4.9	5	-7.5	0	-4.4	0	-1.3	0	1.9	2	5.0	5
M	-8.8	0	-5.7	0	-2.7	0	.3	1	3.3	3	-7.1	0	-3.6	0	0	1	3.6	4	7.1	5
N	-7.4	0	-4.3	0	-1.2	0	1.8	2	4.9	5	-7.4	0	-4.3	0	-1.2	0	1.8	2	4.9	5
O	-10.7	0	-6.6	0	-2.5	0	1.7	2	5.8	5	-8.1	0	-4.7	0	-1.4	0	2.0	2	5.4	5
P	-6.3	0	-3.0	0	.3	1	3.6	4	7.0	5	-6.3	0	-3.0	0	.3	1	3.7	4	7.0	5
Q	-8.5	0	-5.1	0	-1.7	0	1.7	2	5.1	5	-7.2	0	-3.9	0	- .7	0	2.6	3	5.9	5

a Raw scores.
b Weighted scores.
c Contracted Weighted scores.

Table VIII.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores For All Raters on Factor A.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-8.2[b]	0[c]	-4.8	0	-1.4	0	2.1	2	5.5	5	-7.8	0	-4.4	0	-1.0	0	2.4	2	5.8	5
B	-7.1	0	-3.7	0	- .3	1	3.1	3	6.4	5	-7.0	0	-3.5	0	0	1	3.5	4	7.0	5
C	-7.1	0	-3.7	0	- .3	1	3.0	3	6.4	5	-6.5	0	-2.9	0	.7	1	4.3	4	7.9	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-10.0	0	-5.7	0	-1.3	0	3.1	3	7.4	5	-10.0	0	-5.8	0	-1.7	0	2.5	3	6.7	5
F	-6.2	0	-2.9	0	.3	1	3.6	4	6.8	5	-6.6	0	-3.3	0	0	1	3.3	3	6.6	5
G	-8.9	0	-5.2	0	-1.5	0	2.2	2	5.9	5	-8.9	0	-5.2	0	-1.5	0	2.2	2	5.9	5
H	-7.9	0	-4.6	0	-1.3	0	2.0	2	5.2	5	-6.5	0	-3.4	0	- .3	1	2.8	3	5.9	5
I	-7.6	0	-4.3	0	-1.0	0	2.3	2	5.6	5	-7.0	0	-3.7	0	- .3	1	3.0	3	6.4	5
J	-8.5	0	-5.3	0	-2.2	0	.9	1	4.1	4	-8.3	0	-5.1	0	-1.9	0	1.3	1	4.5	5
K	-9.7	0	-6.4	0	-3.0	0	.3	1	3.7	4	-8.3	0	-4.8	0	-1.4	0	2.1	2	5.5	5
L	-7.2	0	-3.9	0	- .7	0	2.6	3	5.9	5	-6.5	0	-3.1	0	.3	1	3.8	4	7.2	5
M	-8.8	0	-5.4	0	-2.0	0	1.4	1	4.7	5	-10.0	0	-6.6	0	-3.1	0	.3	1	3.8	4
N	-8.9	1	-5.3	0	-1.8	0	1.8	2	5.3	5	-7.7	0	-4.4	0	-1.0	0	2.3	2	5.7	5
O	-8.4	0	-4.7	0	-1.1	0	2.5	3	6.2	5	-8.6	0	-5.0	0	-1.4	0	2.2	2	5.8	5
P	-6.4	0	-3.2	0	0	1	3.2	3	6.4	5	-6.9	0	-3.5	0	0	1	3.5	4	6.9	5
Q	-7.9	0	-4.5	0	-1.0	0	2.4	2	5.8	5	-9.1	0	-5.4	0	-1.8	0	1.8	2	5.4	5

a Raw scores.
b Weighted scores.
c Contracted Weighted scores.

Table IX.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores
For all Raters on Factor S.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-7.6[b]	0[c]	-4.3	0	-1.0	0	2.3	2	5.6	5	-7.9	0	-4.5	0	-1.0	0	2.4	2	5.8	5
B	-7.5	0	-3.5	0	.4	1	4.3	4	8.3	5	-8.2	0	-3.6	0	.9	1	5.5	5	10.0	5
C	-7.0	0	-3.3	0	.4	1	4.1	4	7.8	5	-8.2	0	-3.6	0	.9	1	5.5	5	10.0	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-6.7	0	-2.5	0	1.7	2	5.8	5	10.0	5	-4.8	0	-1.1	0	3.0	3	6.7	5	10.4	5
F	-4.9	0	-1.8	0	1.2	1	4.3	4	7.4	5	-4.8	0	-1.6	0	1.6	2	4.8	5	8.0	5
G	-5.9	0	-2.8	0	.3	1	3.4	3	6.6	5	-4.8	0	-1.8	0	1.2	1	4.2	4	7.1	5
H	-8.7	0	-5.2	0	-1.7	0	1.7	2	5.2	5	-11.5	0	-7.6	0	-3.6	0	.4	1	4.4	4
I	-7.0	0	-3.8	0	- .6	0	2.5	3	5.7	5	-7.5	0	-4.4	0	-1.3	0	1.9	2	6.2	5
J	-7.3	0	-4.3	0	-1.2	0	1.8	2	4.9	5	-7.5	0	-4.4	0	-1.3	0	1.9	2	5.0	5
K	-7.5	0	-4.4	0	-1.3	0	1.9	2	5.0	5	-7.4	0	-4.2	0	-1.0	0	2.2	2	5.4	5
L	-9.7	0	-6.6	0	-3.4	0	- .3	1	2.8	3	-8.3	0	-5.1	0	-1.9	0	1.3	1	4.5	5
M	-6.9	0	-3.9	0	- .9	0	2.1	2	5.1	5	-8.2	0	-5.2	0	-2.1	0	.9	1	4.0	4
N	-8.2	0	-5.2	0	-2.1	0	.9	1	4.0	4	-7.6	0	-4.5	0	-1.3	0	1.9	2	5.1	5
O	-9.6	0	-6.3	0	-3.0	0	.3	1	3.6	4	-8.1	0	-4.7	0	-1.4	0	2.0	2	5.4	5
P	-8.1	0	-4.7	0	-1.3	0	2.0	2	5.4	5	-8.0	0	-4.5	0	-1.0	0	2.4	2	5.9	5
Q	-6.9	0	-3.8	0	- .6	0	2.5	3	5.7	5	-8.8	0	-5.3	0	-1.8	0	1.8	2	5.3	5

a Raw scores.
b Weighted scores.
c Contracted weighted scores.

Table X.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores
For All Raters on Factor E.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
	[b]	[c]																		
A	-8.9	0	-5.7	0	-2.5	0	.6	1	3.8	4	-10.5	0	-7.0	0	-3.5	0	0	1	3.5	4
B	-8.4	0	-4.9	0	-1.4	0	2.1	2	5.6	5	-6.9	0	-3.3	0	.4	1	4.0	4	7.6	5
C	-7.9	0	-4.5	0	-1.0	0	2.4	2	5.8	5	-6.5	0	-2.9	0	.7	1	4.3	4	7.9	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-9.3	0	-6.0	0	-2.7	0	.7	1	4.0	4	-10.4	0	-6.7	0	-3.0	0	.7	1	4.5	5
F	-8.2	0	-4.8	0	-1.4	0	2.1	2	5.5	5	-8.5	0	-4.8	0	-1.1	0	2.6	3	6.3	5
G	-6.9	0	-3.3	0	.4	1	4.0	4	7.6	5	-4.8	0	-1.8	0	1.2	1	4.2	4	7.2	5
H	-5.9	0	-2.8	0	.3	1	3.4	3	6.5	5	-6.5	0	-3.4	0	- .3	1	2.8	3	5.9	5
I	-7.0	0	-3.7	0	- .3	1	3.0	3	6.4	5	-9.5	0	-6.0	0	-2.5	0	1.1	1	4.6	5
J	-8.3	0	-5.2	0	-2.1	0	.9	1	4.0	4	-7.4	0	-4.3	0	-1.2	0	1.9	2	5.0	5
K	-6.7	0	-3.5	0	- .3	1	2.9	3	6.0	5	-6.9	0	-3.8	0	- .6	0	2.5	3	5.7	5
L	-9.4	0	-6.3	0	-3.1	0	0	1	3.1	3	-8.5	0	-5.3	0	-2.2	0	.9	1	4.1	4
M	-6.5	0	-3.4	0	- .3	1	2.8	3	5.9	5	-8.3	0	-5.1	0	-1.9	0	1.3	1	4.5	5
N	-7.0	0	-4.0	0	- .9	0	2.1	2	5.2	5	-9.3	0	-6.1	0	-2.9	0	.3	1	3.5	4
O	-9.6	0	-6.3	0	-3.0	0	.3	1	3.6	4	-8.9	0	-5.6	0	-2.3	0	1.0	1	4.3	4
P	-7.9	0	-4.6	0	-1.3	0	2.0	2	5.3	5	-7.1	0	-3.7	0	- .3	1	3.1	3	6.5	5
Q	-9.0	0	-5.7	0	-2.3	0	1.0	1	4.4	4	-9.0	0	-5.5	0	-2.1	0	1.4	1	4.8	5

a Raw scores.
b Weighted scores.
c Contracted weighted scores.

Table XI.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores
For All Raters on Factor O.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-7.5[b]	0[c]	-4.3	0	-1.0	0	2.3	2	5.6	5	-8.1	0	-4.6	0	-1.1	0	2.5	3	6.0	5
B	-7.3	0	-4.0	0	- .7	0	2.7	3	6.0	5	-7.5	0	-3.5	0	.4	1	4.3	4	8.3	5
C	-6.8	0	-3.4	0	0	1	3.4	3	6.8	5	-6.7	0	-3.0	0	.7	1	4.5	5	8.2	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L	-10.0	0	-5.8	0	-1.7	0	2.5	3	6.7	5	-8.3	0	-5.0	0	-1.7	0	1.7	2	5.0	5
P	-6.8	0	-3.4	0	0	1	3.4	3	6.8	5	-5.0	0	-1.7	0	-1.7	2	5.0	5	8.3	5
G	-6.7	0	-3.6	0	- .6	0	2.4	2	5.5	5	-5.8	0	-2.7	0	.3	1	3.4	3	6.4	5
H	-8.8	0	-5.2	0	-1.5	0	2.2	2	5.9	5	-7.1	0	-3.9	0	- .6	0	2.6	3	5.8	5
I	-6.2	0	-2.7	0	.7	1	4.1	4	7.5	5	-7.3	0	-3.8	0	- .3	1	3.1	3	6.6	5
J	-6.4	0	-3.4	0	- .3	1	2.7	3	5.8	5	-6.9	0	-3.8	0	- .6	0	2.5	3	5.7	5
K	-8.6	0	-5.1	0	-1.7	0	1.7	2	5.1	5	-6.7	0	-3.5	0	- .3	1	2.9	3	6.0	5
L	-9.0	0	-5.7	0	-2.3	0	1.0	1	4.4	4	-8.8	0	-5.2	0	-1.5	0	2.2	2	5.9	5
M	-9.9	0	-5.8	0	-1.7	0	2.5	3	6.6	5	-8.6	0	-5.4	0	-2.2	0	1.0	1	4.1	4
N	-5.4	0	-2.2	0	1.0	1	4.1	4	7.3	5	-6.1	0	-2.9	0	.3	1	3.5	4	6.7	5
O	-7.1	0	-3.6	0	0	1	3.6	4	7.1	5	-7.3	0	-3.0	0	1.3	1	5.6	5	9.8	5
P	-8.3	0	-4.7	0	-1.1	0	2.5	3	6.1	5	-7.9	0	-4.3	0	- .72	0	2.9	3	6.5	5
Q	-8.0	0	-4.0	0	- .7	0	2.9	3	6.6	5	-8.9	0	-5.2	0	-1.5	0	2.2	2	5.9	5

a Raw scores.
b Weighted scores.
c Contracted weighted scores.

Table XII.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores
For All Raters on Factor F.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-6.1[b]	0[c]	-2.9	0	.3	1	3.5	4	6.7	5	-7.9	0	-4.6	0	-1.3	0	2.0	2	5.3	5
B	-8.7	0	-5.1	0	-1.5	0	2.2	2	5.8	5	-6.9	0	-3.3	0	.4	1	4.0	4	7.6	5
C	-7.2	0	-3.9	0	- .7	0	2.6	3	5.9	5	-7.5	0	-3.9	0	- .4	1	3.2	3	6.8	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-9.6	0	-6.4	0	-3.2	0	0	1	3.2	3	-8.9	0	-5.7	0	-2.5	0	.6	1	3.8	4
F	-7.4	0	-3.9	0	- .4	1	3.2	3	6.7	5	-7.3	0	-3.6	0	0	1	3.6	4	7.3	5
G	-8.9	0	-5.6	0	-2.3	0	1.0	1	4.3	4	-6.8	0	-3.5	0	- .3	1	2.8	3	6.0	5
H	-10.2	0	-6.8	0	-3.4	0	0	1	3.4	3	-7.5	0	-4.4	0	-1.3	0	1.9	2	5.0	5
I	-9.9	0	-5.4	0	- .9	0	3.6	4	8.1	5	-6.9	0	-3.1	0	.8	1	4.6	5	8.4	5
J	-9.2	0	-6.0	0	-2.9	0	.3	1	3.5	4	-7.2	0	-3.9	0	- .7	0	2.6	3	5.9	5
K	-8.1	0	-5.0	0	-1.9	0	1.3	1	4.4	4	-7.2	0	-4.1	0	- .9	0	2.2	2	5.3	5
L	-8.8	0	-5.5	0	-2.3	0	1.0	1	4.2	4	-9.2	0	-5.9	0	-2.6	0	.7	1	3.9	4
M	-11.7	0	-8.5	0	-5.2	0	-2.0	0	1.3	1	-10.3	0	-7.0	0	-3.6	0	.3	1	2.0	3
N	-8.3	0	-5.6	0	-2.3	0	1.0	1	4.3	4	-9.0	0	-5.9	0	-2.8	0	.3	1	3.4	3
O	-11.5	0	-8.1	0	-4.7	0	-1.4	0	2.0	2	-10.9	0	-7.1	0	-3.4	0	.4	1	4.1	4
P	-7.5	0	-4.2	0	-1.0	0	2.3	2	5.5	5	-8.2	0	-4.8	0	-1.4	0	2.1	2	5.5	5
Q	-8.4	0	-4.9	0	-1.4	0	2.1	2	5.6	5	-8.5	0	-5.1	0	-1.7	0	1.7	2	5.1	5

a Raw scores.
b Weighted scores.
c Contracted Weighted scores.

Table XIII.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores
For All Raters on Factor T.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-5.1[b]	0[c]	-1.9	0	1.3	1	4.5	5	7.7	5	-7.3	0	-3.8	0	- .3	0	3.1	3	6.6	5
B	-7.9	0	-4.6	0	-1.3	0	2.0	2	5.3	5	-6.5	0	-3.1	0	.3	1	3.8	4	7.2	5
C	-6.7	0	-3.5	0	- .3	0	2.9	3	6.0	5	-6.8	0	-3.2	0	.4	1	3.9	4	7.5	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-8.2	0	-4.8	0	-1.4	0	2.1	2	5.5	5	-8.6	0	-5.1	0	-1.7	0	1.7	2	5.1	5
F	-6.5	0	-3.4	0	- .3	1	2.8	3	5.9	5	-6.8	0	-3.4	0	0	0	3.4	3	6.8	5
G	-4.6	0	-1.5	0	1.5	2	4.6	5	7.7	5	-6.6	0	-3.5	0	- .3	1	2.8	3	6.0	5
H	-7.3	0	-4.3	0	-1.2	0	1.8	2	4.9	5	-7.6	0	-4.5	0	-1.3	0	1.9	2	5.1	5
I	-6.6	0	-3.3	0	0	0	3.3	3	6.6	5	-5.9	0	-2.2	0	1.5	2	5.2	5	8.9	5
J	-5.9	0	-2.8	0	.3	1	3.4	3	6.6	5	-6.7	0	-3.5	0	- .3	1	2.9	3	6.0	5
K	-6.3	0	-3.3	0	- .3	1	2.7	3	5.7	5	-5.2	0	-2.1	0	.9	1	4.0	4	7.1	5
L	-6.6	0	-3.4	0	- .3	1	2.8	3	5.9	5	-6.6	0	-3.4	0	- .3	1	2.8	3	5.9	5
M	-9.6	0	-6.3	0	-3.0	0	.3	1	3.6	4	-9.2	0	-5.8	0	-2.4	0	1.0	1	4.4	4
N	-7.3	0	-3.8	0	- .3	1	3.1	3	6.6	5	-9.9	0	-5.6	0	-1.3	0	3.0	3	7.3	5
O	-8.4	0	-4.7	0	-1.1	0	2.5	3	6.2	5	-8.6	0	-5.0	0	-1.4	0	2.2	2	5.7	5
P	-5.9	0	-2.6	0	.7	1	3.9	4	7.2	5	-6.5	0	-3.1	0	.3	1	3.8	4	7.2	5
Q	-8.5	0	-5.0	0	-1.4	0	2.1	2	5.7	5	-8.3	0	-4.8	0	-1.4	0	2.1	2	5.5	5

a Raw scores.
b Weighted scores.
c Contracted Weighted scores.

Table XIV.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores For All Raters on Factor P.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
	[b]	[c]																		
A	-7.1	1	-3.9	1	- .6	1	2.6	3	5.8	5	-8.7	0	-5.4	0	-2.0	0	1.3	1	4.7	5
B	-7.5	1	-4.3	1	-1.0	1	2.3	2	5.6	5	-6.8	0	-3.0	0	.8	1	4.5	5	8.3	5
C	-8.6	1	-5.1	1	-1.7	1	1.7	2	5.1	5	-7.8	0	-4.1	0	- .4	1	3.3	3	7.0	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-9.5	1	-6.2	0	-2.9	0	.3	1	3.6	4	-7.3	0	-4.1	0	-1.0	0	2.2	2	5.4	5
F	-8.3	1	-5.0	0	-1.7	0	1.7	2	5.0	5	-9.1	0	-5.4	0	-1.8	0	1.8	2	5.4	5
G	-7.1	1	-3.9	0	- .6	0	2.6	3	5.8	5	-6.0	0	-3.0	0	0	1	3.0	3	6.0	5
H	-12.6	1	-8.4	0	-4.2	0	0	1	8.4	5	-9.8	0	-6.3	0	-2.8	0	.7	1	4.2	4
I	- 9.2	0	-5.8	0	-2.4	0	1.0	1	4.4	4	-8.1	0	-3.6	0	.9	1	5.4	5	9.9	5
J	-9.0	0	-5.8	0	-2.6	0	.6	1	3.9	4	-6.8	0	-3.7	0	- .6	0	2.5	3	5.6	5
K	-6.6	0	-3.4	0	- .3	1	2.8	3	5.9	5	-6.8	0	-3.7	0	- .6	0	-2.5	3	-5.6	5
L	-9.2	0	-5.9	0	-2.6	0	.7	1	3.9	4	-9.0	0	-5.7	0	-2.3	0	1.0	1	4.4	4
M	-11.1	0	-7.8	0	-4.6	0	-1.3	0	2.0	2	-9.8	0	-6.5	0	-3.3	0	0	1	3.3	3
N	-6.7	0	-3.4	0	0	1	3.4	3	6.7	5	-7.1	0	-3.4	0	.4	1	4.1	4	7.9	5
O	-10.9	0	-7.1	0	-3.4	0	.4	1	4.1	4	-10.9	0	-7.1	0	-3.4	0	.4	1	4.1	4
P	-8.0	0	-4.8	0	-1.6	0	1.6	2	4.8	5	-6.5	0	-3.3	0	0	1	3.3	3	6.5	5
Q	-9.9	0	-6.3	0	-2.6	0	1.1	1	4.8	5	-9.2	0	-5.7	0	-2.1	0	1.4	1	5.0	5

a Raw scores.
b Weighted scores.
c Contracted Weighted scores.

Table XV.-
Table of Raw Scores, Weighted Scores, and Contracted Weighted Scores For All Raters on Factor M.

	Test										Retest									
Rater	1[a]		2		3		4		5		1		2		3		4		5	
A	-7.9[b]	0[c]	-4.3	0	- .7	0	2.9	3	6.5	5	-7.8	0	-4.4	0	-1.0	0	2.4	2	5.8	5
B	-7.1	0	-3.7	0	- .3	1	3.0	3	6.4	5	-6.8	0	-3.2	0	.4	1	3.9	4	7.5	5
C	-8.1	0	-4.6	0	-1.1	0	2.5	3	6.0	5	-7.2	0	-3.2	0	.8	1	4.8	5	8.8	5
D	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
E	-6.4	0	-3.0	0	.3	1	3.7	4	7.1	5	-6.4	0	-3.2	0	0	1	3.2	3	6.4	5
F	-7.9	0	-4.6	0	-1.3	0	2.0	2	5.3	5	-8.5	0	-4.8	0	-1.1	0	2.6	3	6.3	5
G	-9.7	0	-6.4	0	-3.0	0	.3	1	3.7	4	-8.5	0	-5.2	0	-2.0	0	1.3	1	4.6	5
H	-7.9	0	-4.6	0	-1.3	0	2.0	2	5.2	5	-6.9	0	-3.9	0	- .9	0	2.1	2	5.1	5
I	-7.1	0	-3.5	0	0	0	3.5	4	7.1	5	-7.5	0	-3.8	0	0	1	3.8	4	7.5	5
J	-7.8	0	-4.6	0	-1.3	0	2.0	2	5.2	5	-9.3	0	-6.0	0	-2.7	0	.7	1	4.0	4
K	-10.4	0	-6.8	0	-3.2	0	.4	1	3.9	4	-6.6	0	-3.4	0	- .3	1	2.8	3	5.9	5
L	-7.4	0	-3.9	0	- .4	1	3.2	3	6.7	5	-10.3	0	-6.2	0	-2.1	0	2.1	2	6.2	5
M	-5.4	0	-2.2	0	1.0	1	4.1	4	7.3	5	-12.0	0	-8.4	0	-4.7	0	-1.1	0	2.5	3
N	-7.0	0	-3.6	0	- .3	1	3.0	3	6.3	5	-5.2	0	-2.0	0	1.3	1	4.6	5	7.8	5
O	-8.8	0	-5.4	0	-2.0	0	1.4	1	4.7	5	-7.8	0	-4.4	0	-1.0	0	2.4	2	5.8	5
P	-7.3	0	-3.8	0	- .3	1	3.1	3	6.6	5	-6.3	0	-2.8	0	.7	1	4.2	4	7.7	5
Q	-8.4	0	-5.0	0	-1.7	0	1.7	2	5.0	5	-10.1	0	-6.4	0	-2.6	0	1.1	1	4.9	5

a Raw scores.
b Weighted scores.
c Contracted weighted scores.

APPENDIX 5

ABSTRACT OF

A Comparison of Self-Reported and Observed Personality Traits in a Custodial Setting

APPENDIX 5

ABSTRACT OF

A Comparison of Self-Reported and Observed Personality Traits in a Custodial Setting.[1]

Guilford claims his Personality Factors are reliable descriptive categories of personality. This study was an attempt to evaluate his claim in a penal institution, using as a criterion a rating scale consisting of the same traits as measured by the Guilford-Zimmerman Temperament Survey.

The sample consisted of all those students who were between sixteen and twenty-four years old, with an I.Q. of 85 or more, and who had been at the Ontario Training Centre, Brampton, for at least three months prior to July 1, 1962. Each student in the sample was administered the Survey, and five staff members were asked to rate him on a rating scale made up of the same traits as measured by the Survey.

The hypothesis that no statistically significant relationship exists between the traits as measured by the Survey and the same traits as measured by the rating scale was accepted for nine of the ten traits. In the case of Factor R, the coefficient of correlation between the trait

---

1 Jacqueline A. Marshall, master's thesis presented to the School of Psychology and Education of the University of Ottawa, Ontario, May 1964, viii-69 p.

as measured by the Survey and the same trait as measured by the rating scale was significant at the .05 level of probability.

The findings suggest that for a certain universe the validity of the Guilford Personality Factors is minimal.

Possible weaknesses in the design of the study and the general theory underlying the study were pointed out, and suggestions for further research were made.