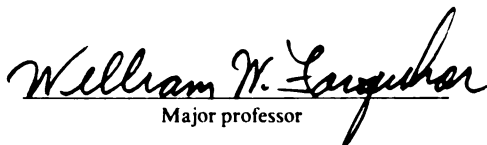This is to certify that the

dissertation entitled

The Relationship of Rating Error to Personality Characteristics of the Myers-Briggs Type Indicator

presented by

Thomas R. Holmes

has been accepted towards fulfillment
of the requirements for

_____Ph.D._____ degree in __Counseling__ Psychology

William W. Farquhar
Major professor

Date October 26, 1983

0-12771

THE RELATIONSHIP OF RATING ERROR TO PERSONALITY
CHARACTERISTICS OF THE MYERS-BRIGGS TYPE INDICATOR

By

Thomas Holmes

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

College of Education

1983

ABSTRACT

THE RELATIONSHIP OF RATING ERROR TO PERSONALITY

CHARACTERISTICS OF THE MYERS-BRIGGS TYPE INDICATOR

by

Thomas Holmes

This study explored the relationship between the
personality of raters and the type of rating errors they
make. The personality typology of Carl Jung, as
operationalized in the Myers-Briggs Type Indicator (MBTI)
was explored and several personality types were selected as
most related to factors involved in rating error.

The history of the study of rating identified the
rating errors: Leniency/Severity, Range Restriction, Halo
and low Interrater Reliability. The personality types were
then used to predict the nature and degree of rating error
expected.

A sample of fifty-six raters, undergraduate students,
rated six therapist - client interactions, and three
speeches. There was a total of seventy-two ratings from
each rater. The raters, who had been tested on the MBTI
prior to making their ratings, were categorized according to
personality types. Their patterns of rating and the nature
of their rating errors were then analyzed to see if there
were significant differences between type.

The results yielded a number of significant results.
It was found that, as predicted, the Sensing/Judging

managerial style persons made ratings which were consistently more severe than the Intuitive/Feeling managerial style persons. In addition it was found that the Sensing/Judging type was less accurate in their ratings of Unconditional Positive Regard than the Intuitive/Feeling types, while there was no difference in their accuracy in rating Accurate Empathy. This same result was found for the MBTI Judging type versus the Perceiving type.

The implications are two-fold: 1) that certain personality types will have predictably different levels of accuracy in their ratings; and 2) that these errors tend to vary according to the task they are rating.

ACKNOWLEDGEMENTS

I would like to express my appreciation to the following people: to William Farquhar for his patient and skilled guidance throughout the dissertation process; Ralph Kron for his intuitive insight and caring support; Marcia Carlyn for sharing her enthusiasm and expertise in the Myers-Briggs Type Indicator; and William Mehrens for assisting with my dissertation. I want also to acknowledge the computer expertise of my son, Dan, who innovated programs when SPSS was inadequate; and the editorial help and support of my wife, Lauri. In addition I want to thank Tim Trichler for his diligence and patience in the preparation of the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The task of having one person judge another person's performance is a common activity in the fields of counseling research, applied psychology, and in clinical settings. Rating scales are the most popular device used in this task, and considerable effort has gone into developing and improving the accuracy of these scales. Researchers have identified specific patterns of error, causing further efforts to develop scales which are less vulnerable to error patterns. Extensive research has been conducted on various methods of construction and analysis of rating scales.

That the source of some patterns of rating error lies beyond the rating scales and in the personality of the raters themselves is acknowledged by writers and researchers but has not been directly investigated. Mehrens and Lehman see the personality of the rater as one of four sources of error: "Error may be due to the scale itself (ambiguity), the personality of the rater, the nature of the traits being rated, and the opportunity offered the rater for adequate observation."[1]

---

[1]William A. Mehrens and Irvin J. Lehmann, *Measurement and Evaluation in Education and Psychology*, Holt Rhenehart and Winston, New York, p. 380.

1

The studies of Ford,[2] Gross,[3] and Crow and Hammond,[4] have shown that individual rater variables, independent of training and experience, accounted for significant amounts of rating error. Thus these studies lend support to the notion that rater personality is an important factor in rating error. These observations, however, were only artifacts for the researchers, whose attention was focused elsewhere. The relationship between personality type and rating error has not been directly investigated, but these earlier studies provide a basis from which to start.

The value in identifying the relationship between personality and rating error is three-fold. First, it provides empirical evidence regarding the assumption concerning personality and its implicit relationship to rating error. Second, it might enable researchers to limit rater bias which would be an artifact to identifiable personality traits, which, if they were dominant in the sample, could produce high interrater reliability but poor validity. The third value is that it would test the predictive validity of the personality constructs and

---

[2]Adelbert Ford, "Neutralizing Inequalities in Rating," Personnel Journal, 1931, Vol. 9, pp. 466-469.

[3]C.F. Gross, "Intrajudge Consistency in Ratings of Heterogeneous Persons," Journal of Abnormal Psychology, 1961, Vol. 62, pp. 605-620.

[4]W.J. Crow and H.R. Hammond, "The Generality of Accuracy and Response Sets in Interpersonal Perception," Journal of Abnormal and Social Psychology, 1957, Vol. 54, pp. 384-369.

measures involved. Positive results would lend further support to the use of personality considerations when training and supervising counselors, managers, and other personnel involved in making evaluations.

Purpose of the Study

The purpose of this study is to examine the relation-ship between personality characteristics of raters and the type of rating errors they tend to make. Fifty-four raters were categorized according to personality type defined by the Myers-Briggs Type Indicator,[5] a personality test based on the theories of Carl Jung. Predictions were made about the kind of rating error which might be expected from certain different personality types. After being given the MBTI, the raters were asked to make a series of ratings of taped interview interactions and speeches. These ratings were analyzed and compared with the personality type of each rater's MBTI to determine whether the nature and degree of the rating errors varied according to the personality types as predicted.

Research Hypotheses

Five research hypotheses were tested. Each postulated a relationship between personality type and the nature of rating errors expected. The six hypotheses are as follows:

---

[5]Isabel Briggs Myers, The Myers-Briggs Type Indicator Manual, The Educational Testing Service, Princeton, N.J., 1962.

1. Ratings made by the Sensing/Judging personality
    type will be more Severe than those made by
    Intuitive/Feeling personality types.

2. The Range Restriction error of Perceiving
    personality types will be greater than those found
    with Judging types.

3. Ratings made by Feeling personality types will have
    more Leniency than those of Thinking types.

4. The Introvert's ratings will have less Reliability
    than will the Extravert's.

5. There will be more Interrater Reliability within
    personality type than in the sample as a whole.

Theory

The concepts underlying this research have their roots
in three areas: the rating error theory of the applied
psychologist, the response set theorist's work with
personality, and the personality theory of C.G. Jung as
operationalized in the Myers Briggs Type Indicator (MBTI).

Rating Error Constructs

The applied psychologists, in their work with the
development of criteria for rating the quality of rating
scales, have focused on four primary categories of rating
quality. These are Halo error, Leniency error, Range
Restriction/Central Tendency, and Interrater Reliability.

The error terms commonly used are reviewed and analyzed
in a comprehensive work by Saal, Downey, and Lahey[6] where

---

[6]Frank E. Saal, Ronald G. Downey, and Mary Anne
Lahey, "Rating the Ratings, Assessing the Quality of Rating
Data", Psychological Bulletin, 1980, Vol. 88, pp. 413-428.

the definitions of previous researchers are reviewed.

Halo error is the "tendency to attend to a global impression of each ratee rather than to carefully distinguish among levels of different performance dimensions...a rater's inability or unwillingness to distinguish among the dimensions of a given ratee's job behavior."[7] The Leniency/Severity errors are defined as ratings which are consistently too high or too low in relation to the mid-point of the scale or in relation to some established standard. Range restriction refers to raters who use only a narrow part of the rating scale, thus reducing the extent to which obtained ratings can discriminate among different ratees' performance level. Interrater reliability is the fourth type measure of rating quality, and is probably the most widely referred to in the use of rating scales. Interrater reliability is defined here as the "extent to which two or more raters independently provide similar ratings on given aspects of the individual's behavior...." Reliability is generally accepted as a form of consensual or convergent validity.[8] The conceptual definition of rating errors can be seen in Table 1.1.

The four types of rating error are the more frequently

---

[7]Ibid., p. 415.

[8]Ibid., p. 419.

Table 1.1

Definitions of Measures of Rating Quality

_____

Halo Error	"Tendency to attend to a global impression of each ratee rather than to carefully distinguish among levels of different performance dimensions."[9]
Leniency/Severity Error	Ratings are given by a rater which are consistently too high or too low in relation to the midpoint of a scale or in relation to some established standard.[10]
Range Restriction	"The extent to which obtained ratings discriminate among different ratees in terms of their respective performance levels."[11]
Interrater Reliability	"extent to which two or more raters independently provide similar ratings on given aspects of the individual's behavior..."[12]

_____

[9]Saal et. al., p. 415.

[10]Ibid., p. 417.

[11]Ibid., p. 417.

[12]Ibid., p. 419.

used concepts in applied psychology. In a related field of study, response set theory, different constructs are used yet the research done by response set theorists has implications for rating error research.

Response Set Theory

Response set theorists differ from applied psychologists in that much of their work has been done with objective tests rather than with rating scales. To them the response sets were seen as contaminating variables affecting the quality of their tests, much as rating error was seen by the applied psychologists. "In recent years, there has been considerable interest in treating the response set component of test scores, not as error variance, but as an expression of a personal stylistic variable."[13]

Efforts to understand the impact of personality on response sets led to studies which found correlations between extreme response sets and such personality traits as concreteness, rigidity, authoritarian personality, and intolerance of ambiguity. However, the results of such studies were not always consistent. Some studies found no significant correlations between response set and personality traits and others found results which

---

[13]Richard R. Schuz and Robert J. Foster, "A Factor Analytic Study of Acquiescence and Extreme Response Set," Educational and Psychological Measurement, Vol. XXIII NO. 3, 1963 p. 435.

occasionally contradicted earlier studies. The mixed results pointed to a weakness in the response set concept which becomes apparent when seen in conjuntion with rating error theory.

The extreme response set construct is represented by two constructs in rating error theory, leniency and severity. Each of these two terms has been shown to characterize opposite rater tendencies. Thus, research studying the relationship of personality to extreme response set would actually measure only traits common to both types of raters or traits common to the type of rater most predominant in the sample. This finding could explain the mixed results obtained in past research. Research on the relationship of extreme response set to personality does indicate that rating error may be related to personality. These indications were used to form the hypotheses of this study, along with the theory of Carl Jung.

Jung's Personality Theory

In this section the personality theory of Carl Jung will be outlined and the operationalization of his constructs in the Myer Briggs Type Indicator (MBTI) will be presented. The management styles which were derived from the MBTI will be discussed and related to rating error theory.

9

In his work <u>Psychological Types</u>[14] Jung reviewed and documented attempts since ancient times to characterize the typical differences between people. From his study of these past systems and from his own clinical experience, Jung developed his theory of psychological types. His typology is related to the task of rating error theory in a fundamental fashion. Jung states that a person's psychological type determines and limits his judgment throughout life.

Jung's primary concept of type was that each person has a preference for one of two attitudes toward the world, Introversion or Extraversion. Jung also posited four psychological functions. These consist of two perceiving functions, Sensation and Intuition, and two judging functions, Thinking and Feeling. According to Jung, one of the psychological functions will become the dominant force in shaping a person's psychological processes as well as his adaptation to the world.

Extraversion/Introversion

Jung sees the Extravert as the person whose life focuses around the external conditions in life. "When orientation by object predominates in such a way that decisions and actions are determined not by subjective views but by objective conditions, we speak of an extraverted

---

[14]C.G. Jung, <u>Psychological Types</u>, Princeton University Press, Princeton, N.J., 1971.

attitude. When this is habitual, we speak of an extraverted type. If a man thinks, feels and acts and actually lives in a way that is directly correlated with the objective conditions and their demands, he is extraverted."[15] The extraverted type then is more comfortable with the environment and usually more at ease with people and things.

Jung conceptualized the Introvert as differing from the Extravert in that instead of orienting himself to objective factors in the world he orients himself to subjective factors within his own disposition. In responding to external events the Introvert tends to rely on a subjective response rather than on a direct response to the event itself. Under stress, the Introvert tends to draw into himself rather than to move towards people as the Extravert would tend to do. Where the Extravert has the gift of action the Introvert has the gift of conceptualization and inner illumination.

It was therefore hypothesized that the Extravert would then likely be more in tune with environmental demands made by rating scales and would be less likely to make the subjective judgments the Introvert would make. This would translate into less rater reliability for the Introvert than for the Extravert.

In addition to the attitudes of Extraversion-Introversion, Jung postulated four psychological functions. The

---

[15]Jung, Op. Cit., p. 333.

four functions consisted of two perceiving functions,
Sensation and Intuition, and two evaluative or judging
functions, Thinking and Feeling.

Perception: Sensing/Intuiting

The process of perception referred to as Sensation
involves direct perception of the concrete physical
properties and details of the environment. The focus is on
practical facts, known qualities, and actualities. The
Sensing type person is known for precise work and attention
to details and routine. The Sensing type is usually
impatient with complexity and abstraction, being a steady
and realistic worker who enjoys using skills which have been
developed. The Intuitive process is an indirect rather than
a direct mode of perception. The person in whom Intuition
is the primary mode of perception looks at the relationship
between the object being perceived and other objects,
mediating perceptions in an unconscious way. So, rather
than looking at the individual tree, as the sensing type
would, the Intuitive would tend to see the tree as part of a
forest, looking at the bigger picture rather than at
details. The Intuitive generally enjoys learning new skills
more than actually applying old ones over a long period of
time and tends to see things from a global rather than a
specific perspective.

In relation to rating scales, the Intuitive could be
expected to differentiate between dimensions since the

strength is in looking at relationships on a theoretical level, where the sensing type may get lost in the details of the ratings and not make good dimensional differentiations. This would be particularly so where there were not clear behavioral definitions of each dimension.

Evaluative Processes: Thinking/Feeling

Thinking, according to Jung, is the psychological function which connects and orders ideas and thoughts. Persons in whom the thinking mode of evaluation is predominant utilize a logical process in objective, impersonal analysis to make judgments on the contexts of ideation. The thinking type tends to be critical of himself and others on the basis of their intellectual ideas, tending not to be aware of the affective components of people's perceptions.

Feeling is the psychological function which imparts a value rather than an objective judgment to the things a person perceives. Thus: "feeling is a kind of judgment, differing from intellectual judgment in that its aim is not to establish conceptual relations but to set up a subjective criterion of acceptance or rejection."[16] The focus for the feeling type then is on making judgments according to cultural and personal experiences. Feeling types operate best in activities involving human relationships and in activities which conform to their central values and

---

[16]Jung, Op. Cit., p. 434.

beliefs.

The primary characteristic relevant to a feeling type's activity as a rater is the sensitivity to the feelings and impulses of others and the value of harmony with others. One could expect that this tendency would make them more lenient as raters, in contrast with the thinking type, whose inclination to be critical of self and others might lead to severity errors on rating scales.

Myers Briggs Type Indicator

The MBTI translates Jung's concepts into four bipolar dimensions. The first is the attitude dimension, Introversion/Extraversion, the second is the perceiving dimension, with the functions of Sensation and Intuition at opposite poles, and third is the dimension judging process, with Thinking and Feeling at the poles. The final dimension of the MBTI was created to determine the preferred Extraverted psychological process, that of Perceiving or Judging. The result is a scale with Judging on one pole and Perceiving on the other. These four dimensions are presented in Table 1.2.

Judging/Perceiving

The dominant psychological process used in adaption to the environment determines the style with which the person adapts to the world. If the dominant process is the Judging one, the person will find decision-making easy. Because of

Table 1.2

Four Dimensions of the Myers-Briggs Type Indicator

(E) Extraversion _____	Introversion (I)
(S) Sensation _____	Intuition (N)
(T) Thinking _____	Feeling (F)
(J) Judging _____	Perceiving (P)

this preference for making decisions, the person's life will
be ordered and planned. This creates a life-style which is
regulated and controlled, and opinions which are readily
made and reluctantly changed.

On the other hand, a person whose dominant process is
one of the perceiving functions will find decisions are hard
to make because they always feel the need of more
information. They will have a life-style which emphasizes
more spontaneity and adaptability, and they will be
reluctant to judge themselves or others.

The person whose dominant function is a Judging one
could be expected to make ratings which give a clear
preference at one extreme or the other. Thus we would
expect them to be lower on range restriction error. The
person who prefers the Perceiving process would be expected
to make considerable range restriction error since the
perceiving-dominant person would be reluctant to judge
themselves or others.

Psychological Types

The four functions of the MBTI have been studied
extensively during the past thirty years and considerable
work has been compiled concerning their reliability and
validity. The different combinations of the four dimensions
form sixteen personality types. Table 1.3 shows the sixteen
types generated from the eight personality preferences.
Each preference is indicated by an initial representing the

Table 1.3

Table of Sixteen Personality Types of the Myers-Briggs Type Indicator

ISTJ	ISFJ	INFJ	INTJ
ISTP	ISFP	INFP	INTP
ESTP	ESFP	ENFP	ENTP
ESTJ	ESFJ	INFJ	ENTJ

(E) Extraversion		(I) Introversion	
(S) Sensation		(N) Intuition	
(T) Thinking		(F) Feeling	
(J) Judging		(P) Perceiving	

direction scored, thus a person preferring Extraversion (E), Sensing (S), Thinking (T), and Judging (J) would be referred to as ESTJ.

The sixteen personality types have been combined into four managerial types by Kiersey and Bates. Two of these styles are used in this study to predict rater error because they describe characteristics of managers which relate to how they evaluate and interact with personnel. These two styles are termed the Sensing/Judging style and the Intuitive/Feeling style. These are described as follows by Kiersey and Bates:[17]

The Sensing/Judging individual is described as a Traditional/Judicial manager. Persons with this style are seen as deciding things quickly and firmly. They have a tendency to see people as good or bad and they tend to emphasize the negative while taking the positive for granted. A personality style such as the Sensing/Judging would be expected to be most prone to severity errors.

The Intuitive/Feeling manager is known as the catalyst. Persons with this style are known for their sensitivity to staff morale, and for their ability to bring out the positive in people. Their weakness is tending to see individuals' personal needs above organizational needs. The

---

[17]David Keirsey, and Marilyn Bates, Please Understand Me, Prometheus Memesis Books, Del Mar, Ca., 1978, ch. V.

Intuitive/Feeling style rater should be prone to making more
leniency errors.

Overview

The research studies and theoretical works which
explore the relationship between personality and rating
error will be reviewed in Chapter II. In Chapter III the
design of the study is described, the test instruments are
presented, and the method of analysis outlined. The results
of the analysis are described in Chapter IV, and in Chapter
V the study is summarized, the conclusions are drawn and
directions for future research suggested.

CHAPTER II

LITERATURE REVIEW

The literature relevant to this research is drawn from three areas: applied psychology studies of rating scales and the nature of rating errors; response-set theory focusing on the relationship between response style and personality factors; and literature regarding Jung's theories and their operationalization in the Myers Briggs Type Indicator.

Literature on Rating

Rating Scales

A number of different rating scales are described in the literature: numerical, graphic, standard, cumulative points, forced choice,[1] comparative, paired comparison,[2] and the Behavioral Expectation Scale.[3] Those most used in applied psychology research are the numerical rating scale and some form of the graphic scale.

In the numerical scale "a sequence of defined numbers

---

[1]J.P. Guilford, Psychometric Methods, McGraw-Hill, New York, 1954, p. 263.

[2]W.A. Mehrens, and I. Lehmann, Measurement and Evaluation in Education and Psychology, Holt, Rinehart and Winston, New York, 1978, p. 355.

[3]John A. Bernardin, and P.C. Smith, "A Clarification of Some Issues Regarding the Development and Use of Behaviorally Anchored Rating Scales," Journal of Applied Psychology, 1981, Vol. 66, No. 4, p. 458.

is supplied to the observer."[4] Here the rater must select

a numerical value which represents his rating:

How would you rate the applicant's composure?

1   2   3   4   5   1 = very good
                             2 = good
                             3 = average
                             4 = poor
                             5 = very poor

The graphic rating scale consists of a continuum which

may or may not contain numbers. Even if it does, the rater

is not forced to select a number but may place the rating

anywhere on the scale:

calm, self	very nervous
assured	uncertain

Rating Error

Since the early use of rating scales, researchers have

observed that certain rater response patterns reduced the

quality and meaning of the rating results. Edward Thorndike

addressed this in his 1929 article "A Constant Error in

Psychological Ratings."[5] In this article Thorndike

described halo error, still considered one of the common

forms of rater error today. He observed that certain raters

were "unable to analyze out these different aspects of the

person's nature and achievement and rate each in indepen-

dence of each other....Their ratings were apparently

---

[4]Guilford p. 263.

[5]Edward Thorndike, "A Constant Error in Psychological
Ratings," Journal of Applied Psychology, 1929, Vol. 4, pp.
25-29.

affected by a marked tendency to think of the person in general as rather good or rather inferior and to color the judgments of the qualities by this general feeling."[6]

Other types of rater errors were first described by Kingsbury in 1922 when he discussed high and low raters, and the rater's "fear" of making distinctions.[7] This work was the first conceptualization of the concepts of Leniency/Severity and Central Tendency Error. They were not labeled as such until Kneeland's work[8] which addressed the tendency of raters to "rate well above the midpoint of the scales used"[9] and defined this as leniency.

In Ford's article[10], 1930, the term "severe" was first used. He analyzed ratings of factory foremen, noting that some of the foremen used only the high end of the scale while others used only the lower end. He labelled those who gave only high ratings as lenient and stated that they "may give too many men the benefit of the doubt." Ford labelled those who always rated low as severe and said of them that they have "possibly an unreasonably high standard of

---

[6]Thorndike, p. 25.

[7]F.A. Kingsbury, "Analyzing Ratings and Training Raters," Journal of Personnel Research, 1922, I, pp. 377-383.

[8]Natalie Kneeland, "That Lenient Tendency in Rating," Personnel Journal, 1929, pp. 356-366.

[9]Kneeland, p. 356.

[10]Adelbert Ford, "Neutralizing Inequalities in Rating," The Personnel Journal, 1930, Vol. IX, No. 6, pp. 466-489.

performance." He observed another group of men who rated "good men very high and poor men low"[11] and these he saw as the most effective raters. Ford also noted that in the Lenient and Severe rater there was a range restriction (failure to use the full distribution of the scale).

In addition Ford attempted to reduce the error in ratings. He noted, "We found evidence of wide differences in severity standards even where the greatest patience had been exercised in giving the foreman directions for scoring."[12] In fact, this error was so resistant to training and was so stable that he developed instead a system for correcting the error by designing a "correlation factor" which could be developed for each foreman and then applied to his ratings so that they would have more universal meaning. Ford's clear delineation of these rater tendencies forms an early basis for the idea that personality variables have a significant relationship to the type of rating error.

The first complete and systematic analysis of rating errors appeared in Guilford's book in 1954.[13] Here he describes the best-known rating errors as error of leniency and negative leniency or "hard rater error," error of central tendency, and halo effect. These error types are given operational definitions in this work. The less common

---

[11]Ford, p. 466.

[12]Ford, p. 467.

[13]Guilford, op. cit.

error types, logical error, contrast error, and proximity error are grouped into what Guilford called a residual error category. Guilford was very thorough in his explication of the statistical methods used to determine rating errors and this material will be reviewed in the analysis section.

Recent work with rating errors in applied psychology has been well summarized and elaborated in a work by Saal, Downey, and Lahey.[14] This work not only reviews the literature on rating errors, it compiles and summarizes current conceptual and operational definitions of the primary rating errors and offers evidence as to the soundness of those definitions. Their work was central in developing both the conceptual and operational definitions used in this study and will be elaborated on further in the appropriate sections.

While the typology of rating errors has become more specific in the field of applied psychology in recent years it is necessary to turn to the parallel field of response set theory to find research on the relationship between personality type and response styles.

Response Set Theory

Response set theory differs from rating error research in that the primary focus is on response styles as

_____

[14]F. Saal, R. Downey, and M.A. Lahey, "Rating the Ratings; Assessing the Quality of Rating Data," Psychological Bulletin, 1980, vol. 88, pp. 413-428.

"consistent patterns of responding to objective test items."[15] These response sets were usually seen as error variance which needed to be eliminated as much as possible. Berg[16] outlined five elements which determined response: chance, stimulus variables, response alternatives available, fractional antedating responses and subject variables. The subject variables category includes personality characteristics and is the area of research in response set theory which will be focused on in this study. This concern with subject variables is what led some response set theorists to begin to interpret the response style not merely as error but as a potential indicator of personality characteristics.[17] Hamilton[18] in his literature review summarized response styles as falling into four categories: acquiescence, deviation, social desirability, and extreme response set. It is the extreme response set studies which will be analyzed in this study since this concept closely parallels the rating error categories of leniency and severity.

Hamilton demonstrated that the extreme response style is a reliable response set which exists over time and across

[15]David Hamilton, "Personality Attributes Associated with Extreme Response Style," Psychological Bulletin, 1968, Vol. 69, p. 192.

[16]I.A. Berg, (Ed.,) Response Set in Personality Assessment, Aldine Publications, Chicago, 1966.

[17]Hamilton, p. 192

[18]Hamilton, op. cit.

tests. In addition he pointed to a number of studies which indicated extreme response set to be related to a number of personality attributes. These attributes were concreteness, abstractness, rigidity/flexibility, and intolerance of ambiguity. White and Harvey[19] carried out a correlational study between the concreteness-abstractness dimension and extreme response set and concrete modes of conceptual functioning as described by Harvey, Hunt, and Schroder.[20]

Shutz and Foster[21] designed a study to investigate the functional structure of several test response set measures. Analyzing the extreme response set of 150 college students, they found loading on Authoritarian and Inflexibility factors, supporting the contention that authoritarian personalities tend toward extreme response sets. In another study Brim and Hoff[22] obtained significant correlations between extreme response set and the desire for certainty or intolerance of ambiguity. Further support was lent to this contention

---

[19]B.J. White, and O.J. Harvey, "Effects of Personality and Stand on Judgment and Production of Statements about a Central Issue," _Journal of Experimental and Social Psychology_, 1965, I, pp. 334-347.

[20]O.J. Harvey, D.E. Hunt, and H.M. Schroeder, _Conceptual Systems and Personality Organization_, Wiley, New York, 1961.

[21]R.E. Shuts and R.J. Foster, "A Factor-analytic Study of Acquieseent and Extreme Response Set," _Educational and Psychological Measurement_, 1963, 23, 435-447.

[22]O. Brim and D. Hoff, "Individual and Situational Differences in the Desire for Certainty," _Journal of Abnormal and Social Psychology_, 1957, 54, pp. 225-228.

in a review of Cattell's studies by Damarin and Messick,[23] which found several factors associated with extreme response sets which could be interpreted as a need for certainty. These constructs closely parallel the Judging-Perceiving dimensions of the MBTI and were used in this study to predict the nature of rating error.

The results of studies in this area have not been uniform, however. A number of studies have failed to find correlations between extreme response set and personality. Borgatta and Glass[24] cited a number of studies of the relationship between extreme response set and Cattell's 16 PF. No significant relationships were found within college student samples. In a mental patient sample several relationships did occur. With a male sample of 17 there was a significant relationship between extreme response set and shrewdness, confident adequacy and phlegmatic/composed. Within the female sample of 10 there was a significant relationship to realistic/tough and radicalism. In a population of ten female prisoners there was a correlation with control, exacting, will power.

Borgatta and Glass[25] also examined studies of the

---

[23]F. Damarin and S. Messick, "Response Styles as Personality Variables," Research Bulletin # RB-65-10, Princeton, N.J., E.T.S., 1965.

[24]E.F. Borgatta and D.C. Glass, "Personality Concomitants of Extreme Response Set," The Journal of Social Psychology, 1961, 55, pp. 213-221.

[25]Ibid.

correlation between response sets and the Edwards Personal
Preference Scale in college students. For the 84 female
students there was a significant negative relationship
between extreme response set and exhibition score on the
Edwards and a significant positive relationship to the
deference score. For the 183 college males the only
significant relationship was a negative relationship to the
change score. In the study as a whole there was no
consistent relationship between the personality variables
measured by the Edwards and extreme response set. It should
be noted, however, that the Edwards and Cattell's 16 PF do
not measure characteristics which have shown the strongest
relationship to extreme response set.

A factor analytic study done by Zuckerman and
Norton[26] found results which appear to contradict the
result of extreme response set by Foster mentioned above.
In this study the extreme response set was correlated with a
non-authoritarian attitude. This suggests that the division
of extreme responses into severe and lenient by the rating
error theorists may lead to a more consistent correlation
with personality types than merely using the general term
"extreme response set."

If indeed extreme response set was not a single pattern
but a combination of two patterns the results of studies

---

[26]M. Zuckerman, J. Norton, and D.S. Sprague,
"Acquiescence and Extreme Sets and Their Role in Tests of
Authoritarianism and Parental Attitudes," Psychiatric
Research Reports, 1958, I, pp. 28-40.

would vary according to the dominant feature of the response set. For instance, if the extreme responses were all in the severity direction one might get a correlation with the authoritarian personality.

Research from the field of rating error indicates that leniency and severity ratings are not usually character- istics of the same person. This being so, it would appear that while important directions have been pointed out by extreme response set research, the refinement of rating error theory should yield even more accurate predictions of the relationship between rating error and the personality of the rater.

When generalizing the response set research to rating error constructs it should be mentioned that response set theory is based largely on responses made to self description questions. Ratings are generally of someone else's performance. The difference between how a person rates himself and how they rate others would limit the generalizability between response set research and rating error research. For the purposes of this study, however, the response set literature has been used as a source of trends since there has been so much more research correlating personality with response set then with rating error. For this use the difficulties in generalizability are not a serious problem.

Literature on Personality Type

The Myers-Briggs Type Indicator

A number of authors have assessed the MBTI's corres-
pondence to Jungian theories. Carlyn's analysis of studies
done by Stricker and Ross found that the Extraversion-
Introversion (E-I), Sensing-Intuition (S-N), and Thinking-
Feeling (T-F) scales were all "generally consistent with the
content of Jung's typological theory."[27]

Other content validity was shown in a study
by Bradway[28] which compared the self classification of
Jungian analysts to their results on the MBTI. The
comparison found 100% agreement on the E-I classification,
68% agreement on the S-N dimension, and 61% agreement on the
T-F classification. These levels of agreement were similar
to another Bradway study,[29] where MBTI classifications
were compared to the Gray-Wheelwright, also an indicator
designed to measure Jungian type. Here it was found that
there was 96% agreement on E-I, 75% on S-N, and 72% on the
S-N when Jungian analysts were studied. Another study cited
by Carlyn as supporting the context validity was done

---

[27]Marcia Carlyn, "An Assessment of the Myers-Briggs
Type Indicator," Journal of Personality Assessment, 1977,
Vol. 41, n. 468.

[28]K. Bradway, "Jung's Psychological Types:
Classification by Test Versus Classification by Self,"
Journal of Analytical Psychology, 1964, vol. 9, p. 130.

[29]K. Bradway, p. 34.

by Stricker and Ross,[30] involving a comparison of continuous scores between the Gray-Wheelright and the MBTI. The results showed a correlation of .79 between the two E-I scales, .58 between the S-N scales and .60 between the T and F scales. All of the correlations were significant at the .01 level. The MBTI has been used in a number of studies as a predictive instrument. Goldschmid[31] found it to have a moderate ability to predict the choice of major by college undergraduates. Other studies reported by Carlyn indicated that the MBTI has some ability to predict grade pointaverage and dropout rate, but that this predictability was not consistent. While some predictive studies have been done using the MBTI, the literature here is not as extensive as that of the construct validity studies.

The construct validity literature on the MBTI is that which gives the basis for the predictions made in this study. There have been considerable correlational studies done with the MBTI, many of which have been summarized by Carlyn.[32] Correlations with the E-I scales have shown the extravert to be "talkative, gregarious, and impulsive, with underlying needs for dominance, exhibition, and

---

[30]L. Stricker, J., and J. Ross, "Some Correlates of Jungian Personality Inventory," Psycholgical Reports, 1964, 14, pp. 623-643.

[31]M.L. Goldschmid, "Prediction of College Majors by Personality Type, "Journal of Counseling Psychology, 1967, Vol. 14, pp. 302-308.

[32]Carlyn, op. cit.

affiliation."[33] They tend to prefer active careers where they interact with others. The introverts were found to want to reflect before acting and preferred working alone. On aptitude tests they show strengths in abstract reasoning, reading abilities, and aesthetic values.

Sensing types were shown in Carlyn's literature review to have interests in that which is solid and real. They tend to work consistently and have respect for authority. They have a factual orientation and a strong need for order. The Intuitive types, on the other hand, have a high tolerance of complexity and they prefer open-ended instruction. They have a strong need for autonomy and change. The Intuitive type tends to be rated high in imagination by faculty.

The studies summarized by Carlyn further showed the Thinking types to be objective, analytical, and logical in making decisions. They have a strong need for order, autonomy, dominance, achievement and endurance. The Feeling types, on the other hand, have been shown through correlative studies to be extremely interested in human values and interpersonal relationships. They have strong needs for affiliation and further nurturance, are generally seen as "pleasant" and have more free-floating anxiety than Thinking types.

---

[33] Carlyn, p. 469.

Judging types cited in the Carlyn article were shown to be responsible, steady, industrious workers. They have a strong need for order and like to have things decided and settled. They have a high capacity for endurance and tend to prefer vocations requiring administrative skills particularly business careers. The Perceiving types were found to be spontaneous, flexible, and open-minded. they tended to score high on measures indicating impulsiveness and showed a strong need for autonomy. The Perceiving type did better on tests of abstract reasoning and scholastic aptitude but tended to get lower grades in school. The research showed that perceiving types enjoyed change and had a high tolerance for complexity.

Combination Types: Managerial Styles

Carlyn in this review also noted that combination types have been shown to be valid constructs. The major research cited showed type-combinations predominating in various fields. The ST type predominates in business and administration; the SF type sales and professions; the NF were reported to outnumber other types in fields involving counseling and writing; and the NT tended to go into science and research. More recent work on type combinations has been done by Kiersey and Bates.

The work done by Keirsey and Bates which is of particular interest for the purpose of this study is their

work with managerial styles.[34] They conceive of
temperaments resulting in the four managerial styles referred
to in Chapter 1: the Sensing/Judging SJ manager, the
Intuitive/Feeling NF manager, the Sensing/Perceiving SP
manager, and the Intuitive/Thinking NT manager. They see each
of the managerial types as having particular strengths and
weaknesses.

The SJ manager according to Kiersey and Bates is
decisive, enjoys the decision-making process, and is a
persevering and patient worker. According to their theory
the SJ types seldom make error of fact and they tend to be
outstanding at precision work. The SJ manager likes to get
things cleared, settled, and wrapped up. They are people
who know, respect, and follow rules. The weaknesses that
come with this style are that the SJ manager may decide
issues too quickly, or become impatient with delays and
complications. The SJ also has a tendency to believe that
some people are good and some bad, and that the latter
should be punished. The SJ manager tends to respond to
negative elements as they become tired and may become
blaming or denigrating. This last attribute of the SJ type
is most directly related to the process of ratings: the SJ
manager may rate people low. This particular style
contrasts most with the Intuitive/Feeling NF style of
management.

---

[34]David Kiersey and Marilyn Bates, <u>Please Understand
Me</u>, Prometheus Nemesis Books, Del Mar, CA., 1978.

The NF managers tend to see people's strengths. They are comfortable with unstructured meetings and quite sensitive to the organizational climate. The NFs easily forget negative disagreeable events of the past and look toward the future from a somewhat romantic position. The NF managers when at their best are very skilled at turning liability into asset. A weakness of the NF managers which may affect how they make ratings is a tendency to avoid unpleasantness. This, combined with the tendency to see people's strength would make them vulnerable to making leniency errors.

The other managerial types, Sensing/Perceiving SP and the Intuitive/Thinking NT, have styles which are not as easily translated into rating error constructs. The SP managers have the strengths of being very practical and concrete in problem solving. They can observe a system and see where it breaks down. They are adaptable, create change easily, and have acute powers of observation. If this theory is true, this type should be a most accurate rater and make less error than the other types. The NT manager has the strength of being a visionary. They have the inner workings of systems in both long and short-term perspective. Their weaknesses are that they have vision but would rather that someone else carry out the construction and execution. The NTs tend to be unaware of others' feelings and may be seen as cold and distant, but neither their strengths nor weaknesses appear to relate directly to the type of rating

errors this type would make.

The literature on managerial styles from Kiersey and Bates provides a clear indication of the types of errors which can be expected from the SJ type and the NF type. Hypothesis One is based on their premises, and positive results on this hypothesis should not only lend support to the notion that personality traits can be used to predict the nature of rating error, but also support the predictive validity of their particular use of the MBTI manager styles.

Summary

In this literature review the history and current trends in rating error theory, the contributions of response set theory, and the validity of the MBTI were discussed along with other literature which might indicate the nature of the rating error different personality types might make. Rating error research began with the early works of Thorndike, who studied Halo error, and progressed to the current status where a range of rating errors are identified. The operational definitions of these errors are diverse. The major rating errors discussed were Leniency/Severity error, which is universally understood to mean tendency to rate high or low. Halo error is understood to mean carrying over a bias for a given rater across the traits being rated. Range restriction error is the failure to use the full range of the scale.

The relationship of response set theory to rating error constructs was explored in the context of error variance in test responses which could be attributed to personality characteristics. The error variance was found to be similar to Leniency and Severity rating error. With this in mind the literature relating extreme response set to personality characteristics was explored as a source for predictions concerning the relationship of rating errors to personality.

The final section covered the literature validating the scales of the MBTI and looked at literature which led to making the predictions found in the research hypothesis.

CHAPTER III

DESIGN OF THE STUDY

In this chapter the sample, the measures, and the design will be described. The hypotheses and method of analysis will be presented.

Sample

Fifty-six students from two undergraduate classes composed the sample of raters for this study. The students were asked as a class if they would volunteer to participate in the research in exchange for interpretations of their MBTI personality profiles. The first group, consisting primarily of college sophomores, was an introductory Sociology class at Western Michigan University. Twenty-nine students, 17 females and 12 males, participated from this class of forty. The second class consisted of juniors in the Nutritional Science program at Michigan State University. Twenty-seven students, all of them female, participated from this class of thirty.

Certain personality characteristics in the Myers Briggs Type Indicator are highly correlated with gender differences, and this meant that the study sample, being predominantly female, reflected a higher proportion of certain traits. These traits will be described specifically later in the chapter.

Measures

Two types of test instruments were used in this study. The Myers-Briggs Type Indicator[1] was used to assess the personality characteristics of the raters in the sample, and several rating scales were used by the subjects in their tasks as raters. These rating scales were as follows: two interpersonal process scales developed by Truax[2]; a counselor effectiveness scale developed by Ivey[3]; and a rating scale used by judges in speaking contests to rate speeches.[4]

The Myers-Briggs Type Indicator

The MBTI is designed so that it measures four bipolar dimensions stemming from Jungian personality typology:

(E) Extraversion......Introversion (I)
(S) Sensation...........Intuition (N)
(T) Thinking..............Feeling (F)
(J) Judgment............Preception (P)

---

[1]Isabel Myers, MBTI Manual, Consulting Psychologists Press, Princton, N.J., 1962.

[2]Charles B. Truax, "A Scale for the Rating of Accurate Empathy, and "A Tentative Scale for the Rating of Unconditional Positive Regard," in Rogers, Gendlin, Kiesler & Truax (Eds) The Therapeutic Relationship and Its Impact, Madison, Wisc., 1967, pp. 555-579.

[3]A.E. Ivey, Microcounseling: Innovations in Interviewing Training, Charles C. Thomas, Springfield, Il., 1971, p. 183.

[4]Waldo W. Braden, (Ed) Speech Methods and Resources, Harper and Row, New York, 1971, p. 126.

Forced-choice items are used to indicate a preference for one pole of each dimension. Each question has one item which indicates a preference for one pole and one item for its opposite. Some items have been weighted more heavily in an attempt to offset social desirability.[5] The highest score on each dimension represents the type preference. The scoring manual provides a procedure for breaking ties.

In this study both Form F and Form G are used. Form F is the original form and consists of 166 items. Form G has been developed more recently and consists of 126 items. Studies have shown Form G to be equivalent to Form F, and that the two forms may be used interchangeably.[6]

A preference on each dimension yields a possible sixteen different personality types. These types were discussed in more depth in the theory section of Chapter I and Chapter II. In the sample population the distributions were evenly divided on the primary dimensions with the exception of the Thinking-Feeling dimension. There the sample had 27 percent Thinking and 73 percent Feeling. This distribution is similar to that found in the female population at large and is mirrored in our population sample which is predominantly female. The distribution in the sample of raters can be seen in Table 3.1.

---

[5]Isabel Myers, p. 86.

[6]Isabel Meyers, <u>MBTI Form G Manual</u>, p. 4.

Table 3.1

===

Distribution of MBTI Types for the Present Sample

(E) Extraversion   Introversion   (I)

          N = 29   N = 27
          % = 52   % = 48

(S) Sensing   Intuition   (N)

          N = 31   N = 25
          % = 55   % = 45

(T) Thinking   Feeling   (F)

          N = 15   N = 41
          % = 27   % = 73

(J) Judging   Perceiving   (P)

          N = 34   N = 22
          % = 60   % = 40

Structural Qualities of the MBTI   A considerable amount of testing has been done on the independence and reliability of the scales of the MBTI. In a comprehensive assessment of the MBTI, Carlyn found that the three type categories directly related to Jung's theory - Extraversion-Introversion, Sensing-Intuition, and Thinking-Feeling - were all relatively independent of each other. The Judging-Perceiving dimension was found to be consistently correlated to the Sensing-Intuition scale and occasionally correlated to several of the other dimensions.[7]

Two aspects of reliability have been investigated: internal consistency and stability of type of category. In her assessment of the MBTI, Carlyn described the two primary methods of measuring internal consistency with the MBTI. Phi Coefficient estimates are used with the Spearman-Brown prophecy formula. This estimate tends to underestimate the reliability, while the tetrachoric correlation coefficient together with the Spearman-Brown prophecy formula tends to give an inflated estimate of the reliability. Carlyn summarized the reliability estimates as follows: The low estimates for the Extraversion-Introversion scale range from .55 to .65 and the high estimates from .70 to .81; for Sensing-Intuition the lower estimates were .64 to .73 and the high from .82 to .92; for Thinking-Feeling the scores

---

[7]Marcia Carlyn, "An Assessment of the Myers-Briggs Type Indicator," Journal of Personality Assessment, 1977, Vol. 41, 5, p. 462.

range from .43 to .75 on the low side and .66 to .90 on the high side; and for the Judging-Perceiving scale the lows were .58 to .84 and the high estimates .76 to .84. Although there is considerable range in the estimated reliabilities they appear to be satisfactory.[8]

The split-half reliability of the MBTI type categories for the present sample of raters was found using the more conservative Phi Coefficient estimates. The reliabilities are displayed in Table 3.2. For Sample A using the MBTI Form G, the reliability was higher than that found in Sample B. The difference was especially great on the E-I scale and the T-F scale. In Sample A using Form G, the E-I reliability was .79 while for Sample B using Form F the reliability was .59. On the T-F scale Sample A with Form G the reliability was .90 while for Sample B using Form F the reliability was .47. These results confirm the improvement in reliability which some had predicted for Form G. The reliability on the other scales is good considering the conservative nature of the statistics used. The E-I and the T-F reliabilities on the Sample B could, however, weaken the study with unclear distinctions between personality types.

Studies of the stability of type category on test-retest studies were also summarized by Carlyn. The four studies which were summarized found that the proportion of agreement between the first testing and the retesting was

---

[8]Carlyn, p. 465.

Table 3.2

Reliability[*] of MBTI Type Categories

Sample	MBTI Type Category			
	E-I	S-N	T-F	J-P
Sample A 17 females, 12 males MBTI Form G	.79	.72	.90	.88
Sample B 27 females MBTI Form F	.59	.86	.47	.80

[*]Calculated using Phi Coefficients and applying

Spearman-Brown prophecy formula.

greater than chance. The majority of the subjects showed shifts on no more than one of the four dimensions. In three of the studies the stability of each scale was studied separately. All of these studies produced test-retest results which were reasonably stable.[9]

Rating Scales

Three of the rating scales used in the study, Scales 1, 2, and 4, were numerical and one was graphic, scale 3. The two Interpersonal Process scales of Truax, "A Scale for the Rating of Accurate Empathy,"[10] and "Tentative Scale for The Rating of Unconditional Positive Regard,"[11] are single-dimension numerical scales with well-defined rating levels. Ivey's Counselor Effectiveness Scale,[12] a graphic scale, has 25 dimensions, 15 of which were used in this study. These dimensions are defined by a key word describing the extremes on each end of a line with seven blank spaces between the extremes. The fourth scale, that used for the evaluation of speeches,[13] has seven dimensions, with each dimension briefly described and given a rating scale of one through seven, with one being poor and seven designated as

---

[9]Carlyn, p. 467.

[10]T-uax, p. 555.

[11]T-uax, p. 569.

[12]Ivey, p. 183.

[13]Braden, p. 126.

excellent. Copies of the scales used are found in Appendix A.

The two scales of Truax were designed to assess brief interactions between a client and a counselor. They have been used in assessing interactions as brief as two-counselor and one-client-statement episodes, to interactions lasting up to four minutes.

Forms of the two interpersonal process scales have been used in many studies. The reliability of these scales based on correlations between raters' ratings has been moderately good, according to Truax and Carkhuff. The Accurate Empathy scale showed a higher reliability than the Unconditional Positive Regard scale, ranging from a high of .95 to a low of .43. The median from twenty-five studies was better than .80. The reliability for the Unconditional Positive Regard scale ranged from a high of .95 to a low of .25 with a median of .60. The range is great for both of these scales and probably reflects the differences in the type of rater and degree of training.[4]

The Truax Accurate Empathy scale is a numerical scale with nine levels of empathy. The lowest level is "an almost complete lack of empathy" and the scale continues to "a level where the therapist unerringly responds to the client's full range of feeling and recognizes each emotional

---

[14] C.B. Truax and R.R. Carkhuff, Toward Effective Counseling and Psychotherapy, Chicago, Aldine Press, 1967.

nuance and deeply hidden feeling."[15] The Truax

Unconditional Positive Regard scale is a nominal scale with

five levels. This scale has a continuum, "beginning with an

almost complete lack of Unconditional Positive Regard and

continuing to a level where the therapist unerringly

communicates to the client a deep and genuine caring for him

as a person with human potentialities, uncontaminated by

evaluation of his thoughts and behaviors."[16]

The reliability scores for our sample were opposite the

trends in the studies cited by Truax and Carkhuff. The

Unconditional Positive Regard scale had a reliability of .95

while the Accurate Empathy scale's reliability was .43.

These reliability scores were determined through the

intraclass correlation method, and are displayed with the

reliability of the other two scales in Table 3.3.

Both the Truax scales were transformed into seven-

level scales for this study. Audio-taped dialogues of those

written by Truax were played to the subjects, thus assuring

that the "correct levels" of interpersonal functioning of

the counselors were as Truax defined them. The vignettes

used to represent the different levels of counselor response

can be seen in Appendix B.

---

[15]C.R. Rogers, E.T. Gendlin, D. Kiesler, and C.B
Truax, The Therapeutic Relationship and Its Impact: A study
of Psychotherapy with Schizophrenics. Madison: University
of Wisconsin Press, 1966, p. 569.

[16]Rogers et al., p. 555.

Table 3.3

Reliability of Rating Scales

Scale #	What was Measured	Reliability*
1	Accurate Empathy	.43
2	Unconditional Positive Regard	.95
3	Counselor Effectiveness	.58
4	Speaker Evaluation	.79

* The reliability was calculated by the intraclass correlation method.

48

Ivey's Rating Scale of Counselor Effectiveness[17] was created to measure both counselor effectiveness and client attitude. This instrument has been shown to be a reliable and valid instrument even when used by inexperienced raters. In a parallel form reliability study, Ivey found a coefficiency of equivalence of .975.[18] The rating scale consists of twenty-five items placed in a semantic differential format describing counselor qualities. There is a clear valence to each item since the scale is designed to differentiate between "good" and "bad" counselors. The extreme "good" rating was designated a seven and the extreme "bad" a one on the scale. Five intermediate levels were provided.

Unlike the interpersonal scales, there is no specific process being rated which would have a correct level of response. Instead, the scale was used to rate the global impressions gained by the raters as they listened to the counselors' respond to the clients in the taped vignettes used for the first two rating scales. In this study the scale has been modified from 25 to 15 items. This may have lowered the reliability of the instrument, but that actually strengthened the research design, since a range of reliability was desirable.

The reliability of Ivey's Counselor Effectiveness scale

---

[17]Ivey, p. 183.

[18]Ivey, op. cit.

as used in this study was .58, which was better than that of the Accurate Empathy scale but lower than the reliability of the other two scales. That the reliability was no higher than this is not surprising since the ratings were made on relatively little data and with little explanation of the meaning of the traits measured. This was done intentionally to generate a measure with lower reliability so that the impact of a poor rating scale on the rating error and personality interaction could be observed.

The scale for the rating of speakers was taken from a college level textbook in debate and public speaking[19] and is representative of many scales developed to guide the evaluation of speeches. It is a standard numerical scale, with ratings of "one" indicating a poor performance on a particular dimension and "seven" indicating an excellent performance. The activity of rating a speech was added to the study design in order to control for the effects on rater judgments which might result from the interactional nature of the material being judged in counselor-client vignettes. The final scale developed to rate speakers was found to have a .79 reliability, thus providing a reliable scale which was not rating counselors or the counseling process.

---

[19]Braden, p. 126.

Design of the Study

The major premise of this study was that the nature of rating errors can be predicted by assessing the personality characteristics of the rater. Prediction of rating error was based on personality characteristics measured by the Myers Briggs Type Indicator. The MBTI was administered to each of two undergraduate classes; then the classes heard taped counselor-client vignettes and three speeches. Three counselors were rated on three different scales and seventeen different dimensions. Three speakers were rated on seven dimensions. A detailed outline of how the scales and the taped vignettes were presented is found in Appendix C.

The ANOVA method of testing rating error was presented by Saal et. al. as a method best used when a complete design is possible.[20] This type of analysis allows the comparison of discrete components of the variance found in ratings. The design of this study makes possible the use of ANOVA statistical procedures by providing a sample where all raters observed all ratees, on all dimensions. This allows more powerful analysis of rater error.

Operational Definitions

In this study the ANOVA method was used whenever possible, but some types of error were best measured by more

---

[20]Frank Saal, Ronald Downey, Mary Anne Lahey, "Rating the Ratings: Assessing the Quality of Rating Data," Psychological Bulletin, 1980, Vol. 88, p. 424.

traditional means. The rating error terms are defined as follows:

Leniency/Severity

Leniency and Severity error was defined as the relationship of mean ratings to each other. The higher ratings were considered Lenient and the lower scores considered Severe.[21]

Range Restriction

A comparison of ratee main effect is the basis for Range Restriction calculation. The absence of ratee main effect is considered Range Restriction.[22]

Interrater Reliability

Two methods of measuring Interrater Reliability were used. First, intraclass correlations were used to measure reliability when sample units were small enough for the two-way analysis of variance procedure to be carried out. The second method was used when comparing larger groups of raters. Here correlations were calculated between pairs of raters rating the same individual on the same dimension. These correlations are summed through the use of Z transformations, and larger correlations indicate greater reliability.[23]

Rating Accuracy

It was possible to define accuracy for scales one and two. Rating accuracy was defined as the mean distance between the rater's rating and the predetermined correct rating.[24]

---

[21]Saal, Downey and Lahey, p. 417.

[22]Ibid., p. 418.

[23]Ibid., p. 422.

[24]Ibid., p. 417.

Methods of Analysis

Each hypothesis, with its method of analysis is presented below. The reasoning for the alternate hypotheses is also given.

Hypothesis One:

Null hypothesis: There is no difference between the mean ratings of Sensing/Judging and Intuitive/Feeling types.

Alternate hypothesis: Sensing/Judging types have lower (more severe) ratings than Intuitive/Feeling types who have higher (more lenient) ratings.

The alternate hypotheses was created from the theory of managerial styles developed by Kiersey and Bates where they postulate that the Sensing/Judging type will be more critical in their style of management and tend to see the negatives more than the positives. This should result in more severe ratings by the Sensing/Judging types. The intuitive/Feeling types on the other hand tend to be more aware of the employees feelings and this should if anything make their ratings more lenient. The method for determining the difference between the means was a one-way ANOVA.

Hypothesis Two:

Null hypothesis: There is no difference in the frequency of ratee main effect between Perceiving types and Judging types.

Alternate hypothesis: The frequency of ratee main effect for Perceiving types is less than the frequency for Judging types, indicating more Range Restriction in the Perceiving type.

The alternate hypotheses was developed from the MBTI descriptions of personality type which suggests that the Judging type will have clear opinions about the events they encounter and that they readily make desisions. The Perceiving type on the other hand is described as reluctant to make decisions and prefers to withhold judgement. These factors should result in the Judging type being less prone to Range Restriction error than the Perceiving type. Range Restriction was determined by assessing the frequency of ratee main effect. Ratee main effect is considered a measure of Range Restriction.

Ratee main effect was determined for both personality groups, Judging and Perceiving. Once the significance of ratee main effect for each group was determined the frequencies were compared between groups. The ratee main effects were determined according to the formula:

$$\frac{MS\ (Ratees)}{MS\ (raters\ x\ Ratees).}[25]$$

A two-way ANOVA of Rater X Ratee produced the mean squares used. The two-way ANOVA was done for each Rater group on each of the four rating scales. The design is presented in table 3.4.

Some data was lost in using this design because the computer could not handle more than twelve raters and six ratees at one time. In order to minimize the data loss

---

[25]Ibid., p. 422.

Table 3.4

Design of the Two-Way ANOVA for Rater by Ratee Interaction

Rater (of a Given Personality Type)

	1	2	3	4	5	6	7	8	9	10	11	12
Ratee A	*											
Ratee B												
Ratee C												

* Ratings from the scale being analyzed fill each cell.

and improve the research design, the rater groups were divided into Sample A and Sample B. The groups were further reduced by limiting those raters whose MBTI scores were less clearly differentiated on a given type dimension. This was done by removing those raters whose scores were less than 9 on a given type dimension. This score of 9 is commonly accepted as the level at which the score obtains greater type stability. There is, however, no published research to support validity of this common practice. For the purpose of this study it was a convenient method of reducing the sample size while increasing the probable reliability of the type categories. The final design yielded eight tests for significance of ratee main effect for each group of raters. Difference in the frequencies of ratee main effect between the Perceiving type and the Judging type was tested with a Chi Square statistic.

Hypothesis Three:

> Null Hypothesis: The mean ratings for Feeling types
> the same as the mean of the ratings made by
> Thinking types.

> Alternate Hypothesis: The mean ratings made by Feeling
> types is higher (more lenient) than the mean of
> the ratings made by Thinking types.

The alternate hypothesis was developed from MBTI descritptions of the Thinking and Feeling types which suggests that the Thinking type base their judgments on logic and systematic evaluation. The Feeling types base their judgments on other subjective value systems and are

often influenced by the impact of their judgment on the person being judged. This could give the Feeling type the tendency to rate lenient while the Thinking type should not be suceptable to that error.

The testing for differences between the mean ratings of Thinking types and Feeling Types is analyzed with a one-way ANOVA as was done with Hypothesis 1.

Hypothesis Four:

> Null hypothesis: The interrater reliability, as measured by intraclass correlation, is the same for the Introverted type as it is for the Extraverted type.

> Alternate hypothesis: The interrater reliability, as measured by intraclass correlation, is greater for the Extraverted type than for the Introverted type.

The alternate hypothesis is developed from the theory of Jung which suggests that the Introvert's basic stance toward the world is more subjective, whereas the stance of the Extravert is primarily objective. This means that the Introvert responds more to interanl stimuli than to external stimuli. The Extravert on the other hand is more responsive to the external events. This would result in ratings for the Introvert which were more variable between raters since the Introverted rater would be less responsive to the external event, and more responsive to their own subjective experience. The Extravert on the other hand should have greater Interrater Reliability since they are theoretically more responsive to the environmental cues, the ratee, than

they are to their own subjective experience.

The testing of this hypothesis involves the calculation of intraclass correlations for each rater group. The intraclass correlations were calculated with the following formula:[26]

$$\frac{MS\ Ratees - MS\ RATERS}{MS\ Ratee}$$

The reliability was calculated for each rater group on each of the scales using Sample A and Sample B separately as was done in Hypothesis two. The differences between reliability scores were calculated with the standard formula found in Blalock as follows:[27]

$$q_{z_1} = z_2 = \frac{1}{N-3} + \frac{1}{N-3} \qquad z = \frac{(z-z)-0}{q_{z_1}-z_2}$$

Hypothesis Five:

Null hypothesis: The Interrater Reliability is the same within type groups of raters as it is for the whole population of raters.

Alternate hypothesis: Interrater Reliability is greater within type groups of raters than it is for the population of raters as a whole.

The alternte hypothesis was developed from the notion tht the ratings of similar groups should be more highly correlated with each other than groups of divergent natures. To test this hypothesis, correlations of all possible

[26]Saal et. al., pp. 422.

[27]Hubert M. Blalock, Social Statistics, McGraw-Hill, Inc., 1972, p. 406.

combinations of raters were calculated. The average correlation based on Fisher's r to Z transformation[28] was calculated for both the sample as a whole and for each of the personality types. Tests for the difference between correlations were calculated using the same formula used in Hypothesis Four.

Summary

In chapter three the sample was presented, the measures used were discussed, the design was outlined and the testable hypotheses presented with their methods of analysis.

The sample of fifty-six undergraduates rated was presented their distribution on the MBTI. The predominant characteristic of the distribution was fairly equal except for the Thinking/Feeling scale which was divided 27% Thinking and 73% Feeling. This was attributed to the fact that the sample was predominantly female and that the distribution found is similar to the distribution found in the female population at large.

The MBTI was shown to be a personality test with moderately good reliability with the individual scale reliabilities ranging for the most part from .65 to .85 in a wide range of studies. The rating scales used were shown to have adequate reliability in past studies, and the

---

[28]Saal, et. al. pp. 422.

modification for this research was discussed.

The design of the research was presented with its unique characteristic of allowing a large number of raters to rate the same ratees on the same dimensions. This allows both for the comparison of large groups of raters and the use of ANOVA procedures when assessing for rating error. The operational definitions of the rating error terms were presented as were the testable hypotheses. The methods of analysis included one ANOVA to compare means, two-way ANOVA to test for ratee main effect and to be used in intraclass correlations, as well as person-product correlations used with Fisher's r to Z transformation.

CHAPTER IV

PRESENTATION OF FINDINGS

In this chapter the results of the analysis are presented. The findings of the original Five hypotheses and two additional hypotheses stemming from those findings are reported.

Hypothesis One

Null hypothesis: There is no difference between the mean ratings of Sensing/Judging types and those of Intuitive/Feeling types.

Alternative hypothesis: Sensing/Judging types have more Severe ratings than those made by Intuitive/Feeling types.

According to the personality theory of Carl Jung as interpreted by Kiersey and Bates, Sensing/Judging (S/J) managerial types would be expected to be very critical in their style, while Intuitive/Feeling (N/F) types would have difficulty being critical when they need to be. These tendencies should result in the S/J's ratings being lower (more Severe) than the N/F's. A one way analysis of variance was used to test for differences between the two groups.

The null hypothesis was rejected and the alternate accepted. While significant differences were found in the predicted direction, less than one percent of the variance is accounted for by the difference in the mean ratings. Table 4.1 shows the results of the analysis.

Table 4.1

Comparison of Mean Ratings of Sensing/Judging and Intuitive/Feeling Types.

Source of variation:	df	MS	F	Probability
Between groups	1	21.425	8.231	.004*
Within groups	3016	2.603		

* Rejected null at .05 level.

Hypothesis Two

Null Hypothesis: There is no difference in the frequency of ratee main effect between Perceiving types and Judging types.

Alternate Hypothesis: Perceiving types have less frequent ratee main effect than Judging type.

The analysis of the frequency in which ratee main effect was found in the Judging and Perceiving types yielded no significant differences using a chi square analysis. While the differences were not significant there was a pattern in the direction opposite to that predicted. The pattern was especially apparent when all personality types were compared. This indication was used to develop the exploratory Hypothesis Seven. The distribution of the ratee main effect for all testable type categories can be seen in Table 4.2.

Hypothesis Three

Null hypothesis: There is no difference between the mean ratings of Feeling types and the mean rating of Thinking types.

Alternative hypothesis: The mean ratings of Feeling types is higher than the mean ratings of Thinking types.

A one way analysis of variance comparing the main ratings of the two groups failed to show significant differences. The results of the analysis can be seen in Table 4.3.

Table 4.2

Comparison of Ratee Main Effects for MBTI Personality Types

Extraverted Raters					Introverted Raters				
sample	scale	df	F		sample	scale	df	F	
A	1	(2,22)	2.09		A	1	(2,22)	.2	
A	2	(2,22)	35.98	*	A	2	(2,22)	36.18	*
A	3	(2,22)	6.28	*	A	3	(2,22)	1.67	
A	4	(2,22)	7.55	*	A	4	(2,22)	3.84	*
B	1	(2,16)	8.9	*	B	1	(2,16)	6.8	*
B	2	(2,16)	13.2	*	B	2	(2,16)	13.9	*
B	3	(2,16)	3.51		B	3	(2,16)	1.83	
B	4	(2,16)	3.79	*	B	4	(2,16)	7.76	*

Freq. Ratee Main Effect = 6   Freq. Ratee Main Effect = 5

Sensing Raters					Intuitive Raters				
sample	scale	df	F		sample	scale	df	F	
A	1	(2,22)	.94		A	1	(2,14)	.5	
A	2	(2,22)	23.5	*	A	2	(2,14)	40.3	*
A	3	(2,22)	4.94	*	A	3	(2,14)	6.24	*
A	4	(2,22)	4.77	*	A	4	(2,14)	2.3	
B	1	(2,26)	4.5	*	B	1	(2,20)	5.49	*
B	2	(2,26)	17.5	*	B	2	(2,20)	41.8	*
B	3	(2,26)	2.7		B	3	(2,20)	2.55	
B	4	(2,26)	14.3	*	B	4	(2,20)	6.54	*

Freq. Ratee Main Effect = 6   Freq. Ratee Main Effect = 5

Judging Raters					Perceiving Raters				
sample	scale	df	F		sample	scale	df	F	
A	1	(2,24)	1.36		A	1	(2,22)	1.624	
A	2	(2,24)	25.99	*	A	2	(2,22)	54.46	*
A	3	(2,24)	2.9		A	3	(2,22)	6.5	*
A	4	(2,24)	2.29		A	4	(2,22)	6.67	*
B	1	(2,14)	3.21		B	1	(2,10)	.599	
B	2	(2,14)	18.1	*	B	2	(2,10)	13.77	*
B	3	(2,14)	2.26		B	3	(2,10)	.81	
B	4	(2,14)	11.4	*	B	4	(2,10)	4.58	*

Freq. Ratee Main Effect = 3   Freq. Ratee Main Effect = 5

* Statistical significance at .05 level.

Table 4.3

Comparison of Mean Ratings of Thinking and Feeling Types.

Source of Variance	df	MS	F	probability
Between Groups	1	.720	.278	.598[*]
Within Groups	4020	2.591		

[*] Failed to reject null at .05.

Hypothesis Four

Null Hypothesis: The Interrater Reliability as
measured by intraclass correlation is the same for
the Introvert as it is for the Extraverted type.

Alternate Hypothesis: The Interrater Reliability as
measured by intraclass correlation will be greater
for the Extraverted type than the Introverted
type.

The analysis showed no significant difference between
the Interrater Reliability of Introverts and Extraverts.
The Extraverts Interrater Reliability was .80 and the
Introverts .62, across all the scales; see Table 4.4. Given
that there was some difference, further analysis was done on
each scale to determine if the Reliability of the scales
would affect the Interrater Reliability of the raters. The
analysis showed that the differences between the Extraverts'
Interrater Reliability and the Introverts' Interrater
Reliability was negligible on the scales which were very
reliable but the difference were considerable on scales with
low reliability. On Scale 2, the most reliable scale, the
Interrater Reliability scores were identical at .95 and yet
on the two lowest Reliability scales, 1 and 3, the
Interrater Reliability for the Extraverts was .70 and .77
compared to .42 and .29 for the Introvert. The Interrater
Reliability of all the personality types can be seen for
each scale in Table 4.5.

Hypothesis Five

Null Hypothesis: The Interrater Reliability will be
the same within rater type groups as it is for the
whole population of raters.

Table 4.4

Reliability[1] for Six MBTI Types Across all Scales

MBTI Type	Reliability	MBTI Type	Reliability
Extraversion	.80	Introversion	.62
Sensing	.71	Intuition	.81
Judging	.70	Perceiving	.59

[1] Reliability derived from intraclass correlations.

Table 4.5

Reliability[1] for Six MBTI Types for each Scale

MBTI Type	Reliability	MBTI Type	Reliability
Scale 1: Accurate Empathy			
Extraversion	.70	Introversion	.42
Sensing	.47	Intuition	.19
Judging	.38	Perceiving	.41
Scale 2: Unconditional Positive Regard			
Extraversion	.95	Introversion	.95
Sensing	.95	Intuition	.98
Judging	.95	Perceiving	.95
Scale 3: Counselor Effectiveness			
Extraversion	.77	Introversion	.29
Sensing	.60	Intuition	.72
Judging	.71	Perceiving	.42
Scale 4: Speaker Evaluation			
Extraversion	.80	Introversion	.84
Sensing	.81	Intuition	.70
Judging	.78	Perceiving	.81

[1] Reliability derived from intraclass correlation.

Alternate Hypothesis: Interrater reliability is
greater within rater type groups than it is for
the whole population of raters.

Comparison of the correlations of raters within type

group with the correlations of raters in the whole rater

population showed no significant differences in the level of

correlation. The results yielded correlations both above

and below that of the population of all raters. The results

can be seen in Table 4.6.

Exploratory Findings

Two additional hypotheses were developed from the results

of the original five hypotheses. The exploratory hypotheses

were developed to follow up trends observed in the original

analysis by studying the accuracy of the ratings. The accuracy

was determined by comparing the rating of the rater with the

predetermined "correct" rating. "Correct" ratings were

available for Scales 1 and 2, where Truax's examples of

different performance levels were used to create the vignettes

which were rated. Comparisons were made between the

personality style of the raters and the mean variance from the

"correct" rating.

Hypothesis Six

Null Hypothesis: There will be no difference in
accuracy of the ratings of the Sensing/Judging
type and those of the Intuitive/Feeling type.

The analysis showed that the ratings of the

Sensing/Judging type were significantly less accurate on

Table 4.6

Correlation Between Raters' Ratings Within MBTI Type and for the Sample as a Whole

MBTI Type	Correlation	MBTI Type	Correlation
Extraversion	.27	Introversion	.23
Sensing	.24	Intuition	.26
Thinking	.22	Feeling	.26
Judging	.23	Perceiving	.29

Overall Correlation Between Raters' Ratings was .25.

Scale Two rating Unconditional Positive Regard, but not

significantly less accurate than the Intuitive/Feeling types on

the Accurate Empathy Scale. Interestingly, Scale Two is the

more reliable of the two scales and yet that is where the

differences in accuracy occurred. The results of the analysis

can be seen in Table 4.7.

Hypothesis Seven

> Null Hypothesis: There is no significant difference
> between the accuracy of ratings of the Judging
> type and the Perceiving type.

> Alternate Hypothesis: The ratings of the Judging type
> are less accurate than the ratings of the
> Perceiving type.

The alternate hypothesis was developed from the trends observed

in Hypothesis Two where the Judging types appeared to be less

reliable raters than the other types. The results indicated

that on Scale Two again the ratings of the Judging type were

significantly less accurate than those of the Perceiving type,

and that on Scale One there was no difference in the accuracy

of the two type's ratings. This finding substantiates the

indication in Hypothesis One that the Judging type's ratings

appeared less reliable than the Perceiving type's. These

findings are displayed in Table 4.8.

Summary

A relationship of personality to rating error was found in

three of seven hypotheses tested. It was found that the

ratings of Sensing/Judging types were significantly more Severe

Table 4.7

Comparison of the Mean Variance From the Correct Rating
for Sensing/Judging vs. Intuitive/Feeling Types.

Scale	Source of Variance	df	MS	F
1	Between Group	1	5.2	.49
1	Within Group	40	10.45	
2	Between Group	1	24.68	5.06*
2	Within Group	40	4.87	

* Significant at .05 level.

Table 4.8

Comparison of the Mean Variance From the Correct Rating
for Judging vs. Perceiving Types.

Scale	Source of Variance	df	MS	F
1	Between Group	1	17.75	3.63*
1	Within Group	54	4.88	
2	Between Group	1	.85	.083
2	Within Group	54	10.14	

* Significant at .05 level.

than the ratings of Intuitive types across all of the scales used in the study.

Ratings of the Sensing/Judging types were significantly less accurate than those of the Intuitive/Feeling types when rating Unconditional Positive Regard; there was, however, no difference when rating Accurate Empathy. It was also found that the Judging type rater was less accurate than the Perceiving type when rating Unconditional Positive Regard, and again there was no difference in the accuracy when rating Accurate Empathy.

No significant relationship was found between the Range Restriction error of Judging vs. Perceiving types; nor was any difference found in Severity/Lenience error between Thinking and Feeling types. The data also failed to find any statistically significant relationship between Extraversion and Introversion and rating error, though some consistent patterns did emerge. There was no significant difference in the reliability of ratings within type group vs. the reliability of ratings in the sample as a whole.

CHAPTER V

SUMMARY AND CONCLUSIONS

The question addressed in this study was what influence does the personality of raters have on the ratings they make. Types of rating error were explored with the goal of finding those errors which appeared to have the strongest theoretical and empirical link to the personality constructs of C.G. Jung, as operationalized in the Myers-Briggs Type Indicator (MBTI).[1]

Summary of the Study

The review of literature on rating error yielded a consensus on the primary measures of rating quality: Leniency/Severity error, Halo error, Range restriction, and Interrater Reliability. Since little research had been done relating rating error to personality type, parallel literature was searched for empirical indications of the relationship of rating patterns to personality. Response set research yielded indications of the relationship between personality traits and extreme response set. Research in this area, combined with the theories of C.G. Jung, particularly the concept on which the MBTI is based, resulted in five hypotheses. These hypotheses predicted the existence of relationships between rating error and

---

[1] Isabel Briggs Myers, _The Meyers Briggs Type Indicator Manuel_, Consulting Psychologists Press, 1962.

73

personality, and the nature of those relationships.

In the study each participant was presented with four different rating tasks. The tasks involved the use of selected rating scales. Two of the scales were developed by Truax, measuring Accurate Empathy and Unconditional Positive Regard.[2] Another scale developed by Ivey measured counselor effectiveness,[3] and the fourth was a scale designed especially for this study to measure the effectiveness of public speakers. All of the scales were modified to a 1 to 7 Likert format. The raters used the scales to rate audio-taped vignettes of counselor-client interaction and three speeches designed for use with the scale. All of the raters rated all of the taped interactions or speeches. Having all the raters rate all the segments on all the dimensions, (a total of 72 ratings per rater), while using such a large number of raters, allowed for the use of a wide range of statistical procedures to analyze the rating errors of the different personality groups.

A variety of methods were chosen as measures of rating error. For the purposes of this study the rating errors were operationally defined as follows:

---

[2]Charles B. Truax, Op. Cit.

[3]G.E. Ivey, Op. Cit.

[4]Saal et. al., p. 417.

1. <u>Leniency/Severity</u> error was defined as the relationship of mean ratings to each other. The higher ratings were considered Lenient and the lower scores considered Severe.[4]

2. <u>Range Restriction</u> error was defined as the absence of ratee main effect.[5]

3. <u>Interrater Reliability</u> was calculated in two ways. First, intraclass correlations were used to measure reliability when units of comparison were small enough to permit the use of ANOVA. The second method was used when large group comparisons were needed. In this case correlations were calculated between pairs of raters rating the same individual on the dame dimension. These correlations were summed through the use of Z transformation and larger correlations were assumed to represent greater reliability.[6]

4. The <u>rating error</u> for scales 1 and 2 was determined by the mean difference between the rating given by a rater and the predetermined correct rating.[7]

The Myers Briggs Type Indicator (MBTI) measures four dimensions of personality: Extraversion-Introversion, Sensing-Intuition, Thinking-Feeling, and Judging-Perceiving. The hypotheses developed for this study were based upon Jungian theory, on which the MBTI is based as well as upon the recent work of Keirsey and Bates[8] which related the MBTI profiles to management style. When these theories were combined with the rating error constructs, the following hypotheses were generated:

1. Ratings made by Sensing/Judging types will be more Severe than those made by Intuitive/Feeling types.

---

[5]<u>Ibid.</u>, p. 422.

[6]<u>Ibid.</u>, p. 422.

[7]<u>Ibid.</u>, p. 417.

[8]David Keirsey and Marilyn Bates, Op. Cit.

2. The Range Restriction error of Perceiving types will be greater than that of the Judging types.

3. Ratings made by Feeling types will have more Leniency than those of Thinking types.

4. The Introvert's ratings will have less Reliability than will the Extravert's.

5. There will be more Interrater Reliability within personality type than in the sample as a whole.

These hypotheses were tested using 56 raters from undergraduate classes in sociology and nutritional science. The sample was predominately female and their personality types as measured by MBTI were fairly evenly distributed with the exception of the Thinking-Feeling dimension. Because of the largely female sample there was a preponderance of Feeling types, the same as there is in the female population as a whole.

The analysis yielded a number of significant relationships between personality type and rating errors. These relationships were found primarily with those personality characteristics which clearly have an impact on the evaluative process and when evaluating measurements of rating error which allowed the use of powerful statistical procedures.

Discussion of the Findings

Hypothesis One

Hypothesis One was supported by the analysis showing a significant difference between the Sensing/Judging type's ratings and those of the Intuitive/Feeling type. These

differences were in the predicted direction with the Sensing/Judging type's ratings being more Severe and the Intuitive/Feeling type's being more Lenient. While these differences were not great in magnitude, they were consistent across all scales, thus yielding a statistically significant result. This result supports the notion that the nature of rating errors can be predicted according to personality type. It begins to define the nature of that relationship, and it gives support to the notion of management styles of Kiersey and Bates which indicates that a contrast between Sensing/Judging and Intuitive/Feeling types would occur.

Hypothesis Two

The analysis showed that the Range Restriction error between Perceiving personality types and Judging types did not differ significantly. The prediction of this hypothesis that the Judging types who were characterized as making quick decisions and having strong opinions would make less Range Restriction error in comparison to the Perceiving types who are characterized as being hesitant to make decisions did not hold true in this sample.

Hypothesis Three

Hypothesis Three predicted that the Thinking types would make lower ratings than the Feeling types. The analysis showed that such was not the case. The lack of difference between the two populations could be attributed

to two factors: one, that there was a strong imbalance between the number of Thinking and Feeling types (15 to 41), and, two, that the variable reliability was .90 in one sample and .47 in the other. Despite these difficulties the number of ratings made was so large that differences between the two types had a high chance of being identified if they did, in fact, exist.

Hypothesis Four

Hypothesis Four predicted that the reliability of the ratings made by the Extravert would be greater than those of the Introvert. The hypothesis was based on the concept that the Introvert is more subjectively oriented and therefore would make less accurate observations of the world than would the Extravert, whose orientation is more toward the external world. The results showed no statistically significant difference between the reliability of the Extravert and the Introvert on all the scales taken together. When the results were compared by individual scales, there were still no statistical differences, yet the spread between the scores of the two types showed a pattern which could indicate direction for future research. The two personality types had similar reliability scores on the scales which had high reliability, but on the scales which had lower reliability, the scores were quite divergent.

Hypothesis Five

Hypothesis Five predicted that the reliability within personality groups would be greater than the reliability of the population of raters as a whole. The results of this analysis did not show any differences between the ratings of given type groups and the combined reliability of all the raters. There was some variation of the reliability between the different types, but none that was significant. The direction of the differences supported the other measures of rating error so it is possible that, if a design could be developed to increase the power of the study, some difference might be found in comparisons of reliability on this level. However, other directions for research appear more promising on the basis of this study.

Exploratory Hypotheses

The two exploratory hypotheses were used with scales one and two which had predetermined correct ratings thus allowing for a comparison of rater accuracy.

The first comparison made was between the Sensing/Judging types and the Intuitive/Feeling types. The analysis showed no difference in accuracy on Scale 1, but rating Accurate Empathy showed the Sensing/Judging types to be significantly less accurate in their ratings than the Intuitive/Feeling types on Scale 2 which rates Unconditional Positive Regard. The result is interesting for several reasons; first the

Unconditional Positive Regard scale was the most reliable in the study; thus the differences in ratings between different personality types is not necessarily most likely to occur when the scale has low reliability though it is logical to assume that it might. Other variables may have more impact. In this case it is possible to make the conjecture that the difference lies in the interaction between the rater's personality and the nature of the rating task. It does seem likely that the Sensing/Judging types who are described as viewing people as either good or bad[9] would have difficulty accurately assessing Unconditional Positive Regard.

The second exploratory hypothesis again used the first two scales to check a pattern observed in Hypothesis Two which was not statistically significant. The pattern was that the Judging type personality appeared to have less ratee main effect. It was hypothesized that if the Judging type had low ratee main effect, which is seen as an indicator of Range Restriction and poor Reliability, it would show up in the lack of accuracy in their ratings on Scale One and Two. The Judging type was no less accurate on Scale One, but was significantly less accurate than the Perceiving type on Scale Two. The finding supported the indications of lower Reliability, and suggests that pursuing

---

[9]David Keirsey and Marilyn Bates, Op. Cit. pp. 142.

research in this direction would be productive.

If the result of the Judging type being a less reliable rater were substantiated in further research, the question of theoretical prediction would need to be studied. It was thought initially that the Judging type would be more accurate than the Preceiving types because their readiness to make judgments and strong opinions would keep them from making range restriction errors. It is also possible that such a strength of opinion could work in the opposite direction by reducing their responsiveness to differences in ratees. It would be possible to test such a hypothesis by using a design which contained a large number of predetermined correct ratings on a variety of rating tasks.

Limitations of the Current Study

One limitation of this study is that the sample was largely undergraduate females. The generalizability is therefore limited to undergraduate social science majors who are female. Another limitation is that while significant differences were found in the ratings of several personality types, the proportion of the variance accounted for was small. This finding suggests that while there is support for the theoretical links between personality and the nature and degree of rating error, an insufficient proportion of the variance is accounted for, so that a practical tool for selection of raters has not been established. It is

possible, however, that future research could account for sufficient amounts of the variance for personality assessment to become a tool in decisions regarding rater selection.

Recommendations for Further Research

The aspects of this study which relate to future research are sample composition; design; personality type considerations; and rating scales and rating tasks.

Sample Composition

The Sample of this study was undergraduate students, predominantly female, who rated counselors and public speakers. Future research is needed with different populations. It appears particularly important to repeat this research with a predominantly male sample, and a sample whose profession is closer to the one to which one wishes to generalize , i.e., managers, supervisors, teachers or other people who are in positions where they are called upon to rate the performance of others.

Design Considerations

Design difficulties resulted when rating error was measured using methods involving a two-way analysis of variance. The methods used to measure Range Restriction produced an assessment of the quality of the raters' rating. It assessed ratee main effect as a ratio with the variance attributable to the rater-ratee interaction. While

rater-ratee interaction is a meaningful measure of the quality of the ratings, it is difficult to develop powerful methods of comparing groups on these dimensions. The sample in this study was larger than most studies of raters and the ANOVA statistic became cumbersome with the two-way interactions needed. It is suggested that if researchers in the future continue to use this measure of rating error they build into their design a larger number of small units of analysis, and that they consider the limitations of their computers when designing the study.

Personality Type Considerations

Certain MBTI personality types seem to be likely topics for future research. In addition, some of the broader implications of personality interaction with rating presses also should be considered in designing future research.

The finding that Sensing/Judging types rated consistently lower than Intuitive/Feeling types emphasized two important considerations for future research. First, personality characteristics which are clearly related to the process of making judgments are more likely to be predictive of rating error. Secondly, a typology which uses combinations of two MBTI factors may be more useful than categories which use only one factor. This is further supported by the results of data used to test Hypothesis Six and Seven where, although the S/J and J were both significantly in error, the result is more marked for the

S/J than for the J alone. Thus it appears that further research should use combination MBTI types which directly relate to the evaluating and judging process.

Several implications for the use of MBTI categories come from the results of the prediction that the Judging type would make less range restriction error than the Perceiving type, which turned out to be false. It appears that this result is because of the lack of reliability and accuracy of the Judging types' ratings. It seems that the hesitancy of Perceiving types to make judgments does not restrict the range of their responses. What emerges as an area for future research is the possibility that the Judging type may be consistently less reliable and less accurate rater; there is sufficient indication of this tendency to merit further exploration.

Other implications for further research come out of the data on the Thinking/Feeling factors. The study's findings suggest that given the difficulties in the reliability of this scale, and the difficulties with male-female distribution in the population, the Thinking/Feeling scale by itself is not the best area on which to focus research effort in the future, especially in the area of Severity/Leniency error. The result relating to Hypothesis One showed that Feeling preference in conjunction with Intuition is a good predictor when compared to the Sensing/Judging combination, and it is in combination with other personality dimensions that the Thinking/Feeling

dimension is most likely to be useful in further research into the relationship of personality to rating error.

The personality dimensions of Introversion/ Extraversion is another area for future research. A response pattern emerged in this study which indicated the Extraverts may be more resilient to poorly constructed scales than the Introvert. Further substantiation of the Extravert's resiliency could have important implications. It would be useful not only in selection of reliable raters for low reliability rating tasks but also for high ambiguity situations such as hiring or student selection processes when the criteria are not clearly spelled out.

Rating Scales and Rating Tasks

The results of this study have shown that it is important to consider the Interrater Reliability of the rating scales when studying the relationship of personality to rating error. It appears that in some instances the effect of personality on rating error is increased by low Reliability, as was the case for the Interrater Reliability of Extraverts' and Introverts' ratings. In other cases, the high Reliability of the scale may have facilitated finding differences in rating error. This is possibly the case with reactions to Scale Two, which rated Unconditional Positive Regard. This scale had the highest Reliability of all the scales and it was the one on which personality differences most affected the accuracy of ratings. In designing future

research, depending on the objectives of the study, it may be important to have high Reliability in some cases and low Reliability in others.

In addition to considering Reliability of scale, it is important to consider the nature of the task which is being rated. The Sensing/Judging lack of accuracy in rating Unconditional Positive Regard implies that certain personality types may have difficulty recognizing certain interaction patterns. Either this should be taken into account in designing a study or it could be the focus of a study itself.

Besides the implications described above, further study might have implications for clinical supervisors of a client-centered orientation. It could have immediate relevance because the ability to accurately judge another counselor's skill in giving unconditional positive regard is an important part of selecting, training, and evaluating a counselor's performance. Future studies exploring the relationship of personality to a supervisor's proficiency might focus on rating accuracy on a variety of scales developed to measure counselor effectiveness. There is a wealth of research studying the therapeutic process itself, but little studying a person's ability to assess that process. The present study produces results which indicate that the ability to assess therapeutic conditions may be related to personality.

Conclusion

This study directly analyzed the relationship of personality to rating error. The results show that the relationship of some MBTI types to certain rating errors can be predicted. The relationships found provide a basis for replication and a focus for further research. A number of predicted relationships were not found; however there were indications as to where future research might find significant results. There was sufficient evidence to indicate that future research could be valuable.

APPENDIX A

RATING SCALES USED IN THE RESEARCH

APPENDIX A

A Scale for the Rating of Accurate Empathy

Note: You will make only one mark on this sheet for each counselor.

Counselor A	Counselor B	Counselor C		
1	1	1	Level 1	Therapist seems completely unaware of even the most conspicuous of the client's feelings. His responses are not appropriate to the mood and content of the client's statements and there is no determinable quality of empathy, hence, no accuracy whatsoever.
2	2	2	Level 2	
3	3	3	Level 3	
4	4	4	Level 4	Therapist accurately responds to all of the client's more readily discernible feelings. He shows awareness of many feelings and experiences which are not so evident, too, but in these he tends to be somewhat inaccurate in his understanding
5	5	5	Level 5	
6	6	6	Level 6	
7	7	7	Level 7	Therapist unerringly responds to the client's full range of feeling in their exact intensity. Without hesitation he recognizes each emotional nuance and communicates an understanding of every deepest feeling.

Scale for the Rating of Unconditional Positive Regard

Note: You will make only one mark on this sheet for each counselor.

Counselor A	Counselor B	Counselor C		
1	1	1	Level 1	The therapist is actively offering advice or giving clear negative regard. He may be telling the client what would be "best" for him or may be in other ways actively either approving or disapproving of his behavior.
2	2	2	Level 2	
3	3	3	Level 3	
4	4	4	Level 4	The therapist indicates a positive caring for the client but is a semi-possessive caring in the sense that he communicates to the client that what the client does, or does not do, matters to him.
5	5	5	Level 5	
6	6	6	Level 6	
7	7	7	Level 7	The therapist communicates positive regard without restriction. There is a deep respect for the client's worth as a person and his rights as a free individual

Rating Scale of Counselor Effectiveness

Counselor A

```
      sensitive ___:___:___:___:___:___:___insensitive

        skilled ___:___:___:___:___:___:___unskilled

        nervous ___:___:___:___:___:___:___calm

      confident ___:___:___:___:___:___:___hesitant

      attentive ___:___:___:___:___:___:___unattentive

         gloomy ___:___:___:___:___:___:___cheerful

     intellient ___:___:___:___:___:___:___unintelligent

  irresponsible ___:___:___:___:___:___:___responsible

        sincere ___:___:___:___:___:___:___insincere

       apathetc ___:___:___:___:___:___:___enthusiastic

          tense ___:___:___:___:___:___:___relaxed

       sociable ___:___:___:___:___:___:___unsociable

        shallow ___:___:___:___:___:___:___deep

       careless ___:___:___:___:___:___:___careful

         polite ___:___:___:___:___:___:___rude
```

Scale for the Evaluation of Speeches

Speaker A

	poor					excellent	
Suitability of Subject: is the subject timely and Worthwhile?	1	2	3	4	5	6	7
Thought Content: Does it have depth? Is the approach fresh and challenging?	1	2	3	4	5	6	7
Organization: Is the introduction adequate? Are points apparent? Are transitions clear? Is the conclusion adequate?	1	2	3	4	5	6	7
Development of Ideas: Is there adequate use of repetition, example and illustration etc.?	1	2	3	4	5	6	7
Use of Language: Does the wording have simplicity, accurateness, vividness and forcefulness?	1	2	3	4	5	6	7
Voice and Diction: Is the voice pleasant and appealing? Is there adequate pronunciation and enunciation?	1	2	3	4	5	6	7
Communication: Does the speaker make contact with the audience? Is he sincere, direct and persuasive?	1	2	3	4	5	6	7

APPENDIX B

VIGNETTES USED TO ESTABLISH CORRECT RATING FOR ACCURATE EMPATHY AND UNCONDITITIONAL POSITIVE REGARD

Appendix B

Vignettes Used to Establish Correct Rating for Accurate
Empathy and Unconditional Positive Regard.

The vignettes used with Scales One and Two are
presented here. They are examples of interactions developed
by Truax to depict the various levels of Accurate Empathy
and Unconditional Positive Regard. Level One is the lowest
rating, level Seven the highest.

Accurate Empathy

Level One:

"C: I wonder if it's my educational background or if it's
     me.
T: M-hm.
C: You know what I mean.
T: Yeah.
C: [pause] I guess if I could just solve that I'd know just
    about where to hit, huh?
T: M-hm, m-hm. Now that you know, a way, if you knew for
    sure, that your lack, if that's what it is -- I can't be
    sure of that yet [C: No] is really so, that it, it might
    even feel as though it's something that you just
    couldn't receive, that it, if, that would be it?
C: Well -- I -- I didn't, uh, I don't quite follow you --
    clearly.
T: Well [pause], I guess, I was, I was thinking that --
    that you perhaps thought that, that if you could be sure
    that, the, uh, that there were tools that, that you
    didn't have, that, perhaps that could mean that these --
    uh -- tools that you had lacked -- way back there in,
    um, high school [C: Yeah] and perhaps just couldn't
    perceive now and, ah --
C: Eh, yes, or I might put it this way, um -- [pause] If I
    knew that it was, um, let's just take it this way -- if
    I knew that it was my educational background, there
    would be a possibility of going back [T: Oh, so I missed
    that now, I mean now, and, uh] and really getting myself
    equipped.

T: I see, I was -- uh -- I thought you were saying in some ways that um, um, you thought that if, if that was so, you were just kind of doomed.

C: No, I mean --

T: I see -- [interrupts]

C: Uh, not doomed. Well, let's take it this way, um, as I said, if, uh, it's my educational background, then I could go back and, catch myself up [T: I see --] and come up --

T: Um."[1]

Level Four

"C: I gave her her opportunity . . .

T: Mhm.

C: . . . and she kicked it over. [heatedly]

T: Mhm -- first time you ever gave her that chance, and -- she didn't take it? [inquiring gently]

C: No! She came back and stayed less than two weeks -- a little more than a week -- and went right straight back to it. [shrilly] So that within itself is indicative that she didn't want it. [excitedly] [T answers "Mhm" after each sentence.]

T: Mhm, mhm -- it feels like it's sort of thrown -- right up in your face. [gently]

C: Yah -- and now I would really be -- crawling. . .

T: Mhm.

C: . . . if I didn't demand some kind of assurances -- that, that things was over with. [firmly]

T: Mhm, mhm, it would be -- pretty stupid to -- put yourself in that -- same position wher it could be sort of -- done to you all over again. [warmly]

C: Well, it could be -- yes! I would be very stupid! [shrilly]

T: Mhm.

C: . . . because if it's not him -- it might be someone else. [emphatically]"[2]

Level Seven

T: ...I s'pose, one of the things you were saying there was, "I may seem pretty hard on the outside to other people but I do have feelings."

C: Yeah, I've got feelings. But most of'em I don't let 'em off.

T: M-hm. Kinda hide them. [C, faintly: Yeah.] [long pause]

---

[1]Truax, op. cit., p. 557.

[2]Ibid., p. 562.

C: I guess the only reason that I try to hide 'em, is, seein' that I'm small, I guess I got to be a tough guy or somethin'.

T: M-hm

C: That's the way I, think, I think people might think about me.

T: Mm. "little afraid to show my feelings. They might think I was weak, 'n' take advantage of me or something. They might hurt me if they -- knew I could be hurt."

C: I think they'd try anyway.

T: "If they really knew I had feelings, they, they really might try and hurt me." [long pause]

C: I guess I don't want'em to know that I got'em.

T: Mm.

C: 'Cause then they couldn't if they wanted to.

T: "So I'd be safe if I, if I seem like a, as though I was real hard on the outside. If they thought I was real hard, I'd be safe."[3]

Unconditional Positive Regard

Level One:

"C: ....and I don't, I don't know what sort of a job will be offered me, but -- eh ---

T: It might not be the best in the world.

C: I'm sure it won't. [T: And uh.] But --

T: But if you can make up your mind to stomach some of the unpleasantness of things [C: M-hm] you have to go through -- you'll get through it. [C: Yeah, I know I will.] and , ah, you'll get out of here.

C: I certainly, uh, I just, I just know that I have to do it, so I'm going to do it but -- it's awfully easy for me, to -- [sighs] well, more than pull in my shell, I-I just hibernate. I just, uh -- well, just don't do a darn -- thing.

T: It's your own fault. [severely]

C: Sure it is. I know it is [pause] But it seems like whenever I -- here -- here's the thing. Whenever I get to the stage where I'm making active plans for myself, then they say I'm high. An'

T: In other words they criticize you that --

C: Yeah.

---

[3]Ibid., p. 569.

T: So tender little lady is gonna really crawl into her shell. [C: Well, I'll say "okay."] "If they're gonna throw, if they're gonna shoot arrows at me, I'll just crawl behind my shield and I won't come out of it." [forcefully]
C: That's right. [sadly]
T: And that's worse. [quickly]"[4]

Level Four:

"C: It's gettin' so I can't even -- can't even sleep at night anymore -- roll and toss all, toss all night long --
T: Pretty upset?
C: Oh, well, just lay there and think of everything -- and some of the guys that come in after I did. There, there's some of them guys what of gone home, 'n' I'm still in here.
T: It's sort of up to you when you, as to when you go.
C: You can't do anything?
T: Well, I said, I sort of feel you have been -- ah -- you've been holding down that job -- you still work in the kitchen, don't ya?
C: Yeah -- [mumbled]
T: O.K., but you -- you been holding that job, and you have your card, well, O.K. You fouled up somewhere, but you'll have your card again. And, well, you, in a sense showed the staff that you can handle these things, without getting into difficulties, you are on your way home.
C: That doggone kitchen detail, detail -- seven cents a day -- just ta scribble bunch of junk. [mumbled]
T: Well, you're sure as hell not gonna get rich on it. -- What about this trouble, talking about money -- what about this trouble you were raising the last time? About borrowing some money from this gal, have you come to any decision on that?
C: Well [pause] I'd rather not say, I ain't gonna say nothin' as long as that tape recorder's on.
T: Want me to turn it off for a while" -- It's a part of the project. That's why I sort of feel it's your responsibility to -- to record these things."[5]

---

[4] Ibid., p. 571.

[5] Ibid., p. 575.

Level Seven:

"T: And I can sort of sense -- and when you want to, when you feel like it, I'd be glad if you shared some of those --

C: What? [abruptly]

T: I said, when you want to, and when you feel like it, I'd be glad if you shared some of those feelings with me -- [Client, breaking in and speaking with Therapist: Why, why -- whoa, whoa, whoa --] I'd like to just sort of see 'm --

C: Why, you gettin' rich off this silent character or somep'n or what? [raucous laughing sound] Ten, fifteen, twenty dollars an hour? [loudly] Then he just sits here -- an' that's it, huh? Oh, I know -- [mumbling]

T: I'd say that's -- that's a good point -- what ya mean -- [softly]

C: Oh, I don't know -- [pause]

T: Well, that -- uh, makes me say something stupid -- uh [laughs] -- I sometimes get paid fifteen, twenty dollars an hour, but that, I'm not getting paid --

C: [interrupting loudly, overtalking Therapist] Why, the state's paying ya that now, ain't they?

T: Not for you, no. I thought you might think that.

C: Who is, then? [insistently]

T: No, I get a salary from the University for doing research. [calmly]

C: Oh -- research! [incredulously]

T: M-hm -- [pause]

C: I think that's just a -- roundabout way to put it -- th-that's what, that's what I think.

T: Well, let's put it this way: I get it, but -- I get exactly the same salary whether -- I see you or not [gently]

C: Oh, there, there probably is a -- there probably is a -- that type doctors there, but -- uh, but I wouldn't call it research! [scornfully] -- I, I, I, I, I, I, I don't know, I don' know, I don' care -- I don' -- I -- [ending in angry confusion]

T: [speaking with conviction] Well, I'd like to know you -- that, that's not research."[6]

---

[6]Ibid., p. 579.

APPENDIX C

RESEARCH PROCEDURES

APPENDIX C

Research Procedures

Step 1:   Handing out research package:

Hello, I'm Tom Holmes, I appreciate your
willingness to participate in this research. I
think you will find the study interesting and the
feedback after the study useful. I am going to
handout the forms you will use in this research.
If you chose not to participate, please let me
know as I am handing out the material. Do not
open the envelope until I instruct you to do so.

Step 2:   Introduction to the overall experiment:

Please open the envelope and remove the stapled
booklet. The directions for the study are on the
top page. Please read them to yourself as I read
them aloud. (read directions)

Step 3:   Orientation to the accurate empathy scale:

Now turn to page 1 titled "A Scale for the rating
of Accurate Empathy". You will notice that there
are seven possible ratings on this scale. Next to
rating levels 1, 4, and 7 are descriptions of
those rating levels of accurate empathy. Read
these to yourself as I read them to you, starting
with a level 1 response. (read levels 1,4, and 7)

I will be playing tape recordings of counselors
working with clients, you will make your ratings
of the counselors responses according to the
descriptions on the scale. Do the counselor
responses represent level 1, 2, 3, 4, 5, 6, or 7
on the accurate empathy scale.

You will notice that on the left hand side of the
scale there are three columns labeled: Counselor
A, Counselor B, and Counselor C. In each column
are numbers corresponding to the level of accurate
empathy. If in your judgment Counselor A showed a
very high degree of accurate empathy then you
should circle 7 under Counselor A. If you feel
that Counselor B exhibited an all most complete
lack of accurate empathy then you should circle a
1 under Counselor b. If you believe the
counselors response to be somewhere in between you
should circle the number which you feel best

describes your opinion of the counselors performance. You will circle only one number for each counselor.

Step 4: Introduction to the counselor/client tapes:

The tapes you are about to hear are reenactments of actual counselor/client interactions. Tape recordings such as these are often used to assess students who are being trained as therapists. The dialogues you will hear are the result of a past project at another university. The clients responses have been edited and some of the clients were in a residential treatment setting at the time of the counseling sessions.

I will play a short sample of a counselors work with a client. Listen carefully to the counselors responses. It is the counselors responses which you are rating, not the client. Make your judgment as quickly as possible. When you have decided which level of accurate empathy you feel best describes the counselors performance, indicate that by circling the corresponding number in that counselors column. Remember 1 is the lowest level, 7 the highest.

Step 5: Playing the tapes:

I am now going to play the tape for Counselor A. Rate his level of accurate empathy. The first person to speak on this recording is the therapist. (play tape) 0 - 33 Rate the counselors responses not the client.

Now please record your rating of Counselor A's level of accurate empathy in the proper column. We will now repeat the process for Counselor B. In this recording the client begins. (play tape of Counselor B) 33 - 55.

Record your rating for Counselor B's level of accurate empathy....Now here is the recording of Counselor C. The client begins this recording. (play tape) 55 - 98.

Please rate Counselor C's level of accurate empathy....That concludes the accurate empathy ratings.

Step 6:   Orientation to the Unconditional Positive Regard
scale:

> The next counselor characteristic you will rate is
> the level of unconditional positive regard.
> Please turn to page two where you will find the
> scale for rating this dimension. When the
> therapist is communicating a low level of
> unconditional positive regard he appears as
> described by the narrative next to level one on
> the scale. (read level 1)....A description of a
> mid-range response is found next to level 4.
> (read level 4).....The highest level of
> unconditional positive regard is described for
> level seven. (read level 7)...

> Your ratings are to be recorded in the same way
> they were on the last scale. You will make one
> rating for each counselor in the proper column for
> that counselor.

Step 7:   Presentation of the unconditional positive regard
tapes:

> The same three counselors A, B, and C will be
> presented again in the same order. Remember you
> are rating the counselors responses not the
> client.

> Here is the recording of Counselor A, please judge
> his level of unconditional positive regard. The
> client will begin speaking first. (play tape) 95
> - 122.

> Now rate Counselor A's level of unconditional
> positive regard from one to 7 and record you
> ratings in the appropriate column. Here is the
> recording of Counselor B. In this dialogue the
> client again begins. (play tape) 122 - 154.

> Rate Counselor B. Now here is Counselor C. The
> first voice you will hear is Counselor C. (play
> tape) 155 - 186.

> Mark your ratings for Counselor C.

> That concludes the unconditional positive regard
> rating.

Step 8:   Introduction to the counselor effectiveness scale:

> I would now like you to think of the counselors in a more general sense. I want you to rate each counselor on a number of characteristics.
>
> Please turn to the next page. Here you find a list of counselor characteristics. Read down the list with me. (read list).
>
> You will be rating each counselor on these characteristics. You are to place a checkmark at the point on the line which corresponds to your opinion as to how a counselor rates each characteristic.
>
> For example: if you feel counselor A is very sensitive you would place a check right next to sensitive. If on the other hand you felt he wa insensitive you would place a mark next to insensitive, and of course if felt he was somewhere inbetween you would place a mark at the point which most accurately described him in your mind. You will do the same for each characteristic, making one check on each line.
>
> There are three copies on this scale, one for each counselor. On the top of each page is an indication as to which counselor the scale is for.

Step 9:   Rating the counselors:

> I will replay several samples of each counselors responses in order to remind you of each counselor. While listening to the tapes please put a checkmark indicating your assessment of that counselor on each characteristic. Record your first impression.
>
> Begin by rating counselor A. The scale you are using should have counselor A at the top. I will play the excerpts from counselor A. Please mark your scales for the characteristics listed. (play Counselor A tape) 188 -209. When you are finished rating please look up.
>
> Now turn to the next page and find the scale for Counselor B. Please rate Counselor B as I play the tape of several of his responses. (play tape). 210 - 224.

Turn to the next page and find the scale for counselor C. Rate counselor C as I play several of his responses. (play tape) Look up when you are finished. 225 - 244.

That concludes the counselor ratings.

Step 10: Introduction to the evaluation of speakers:

In this section you will be rating the performance of individuals as they present short speeches.

The speakers will be rated on seven dimensions which can be found on the rating scale for the evaluation of speeches. This can be found on page 6.

Step 11: Presentation of the scale for the evaluation of
speeches:

The seven dimensions to be rated are listed here with there explanation. Please read with me as I go over the seven dimensions. (read dimensions)

The ratings are again on a seven point scale. Seven is the highest rating and one is the lowest rating. After listening to the speaker you will make a judgment as to their level of performance on the various dimensions. For example: if you feel speaker A chose a subject which was very timely and worthwhile, then you would rate him at 7 on the dimension. If you felt that the organization of his speech was very poor you would circle 1 next to that dimension. You will circle one number representing ;your rating for each dimension shown on the scale. There is a separate rating sheet for each speaker. It is noted at the top of each sheet which speaker the scale is for.

Step 12: Rating of the Speakers:

The page number should be 6 and the designation at the top of the page should indicate speaker A. I will now play the recording of speaker A. (play tape) 333 - 355. Now please rate speaker A on the dimensions listed on the scale. Remember 7 is the highest and 1 the lowest.

Now turn to page 7. There you should find the rating scale for speaker B. Here is the tape of speaker B. (play tape). Now rate speaker B. 355 - 375.

Turn to page 8 where you will find the rating scale for speaker C. Here is speaker C. (play tape) Now please rate speaker C on the rating scale. 375 - 403.

This concludes the research section. Thank you very much for your assistance.

BIBLIOGRAPHY

BIBLIOGRAPHY

Arthur, A.Z. "Response Bias in Semantic Differential.
"British Journal of Sociology and Clinical
Psychology, 1966, Vol. 5, pp. 103-107.

Barrett, Gerald, Phillips, James, Alexander, Ralph.
"Concurrent and Predictive Validity Designs: A Critical
Reanalysis." Journal of Applied Psychology, 1967,
Vol. 51, No. 2, pp. 1-6.

Berg, I.A. (Ed.). Response Set in Personality Assessment.
Chicago: Aldine, 1966.

Bernardin, John. "A Recomparison of Behavioral Expectation
Scales to Sumated Scales." Journal of Applied
Psychology, 1976, Vol. 61, No. 5, pp. 564-570.

Bernardin, H. John and Smith, Patricia Cain. "A Clarificaton
of Some Issues Regrading the Development and Use of
Behaviorally Anchored Rating Scales." Journal of
Applied Psychology, 1981, Vol. 66, No. 4, pp.
458-463.

Bernardin, H. John. "Effects of Rater Training on Leniency
and Halo Errors in Student Ratings of Instructors."
Journal of Applied Psychology, 1978, Vol. 63, No. 3,
pp. 301-308.

Blalock, Hubert, M. Social Statistics, McGraw-Hill, Inc.
1972.

Borgatta, Edgar F., and Glass, David C. "Personality
Concomitants of Extreme Response Set." The Journal Of
Social Psychology, 1961, 55, pp. 213-221.

Borman, Walter C. "Consistency of Rating Accuracy and Rating
Errors in the Judgement of Human Performance."
Organizational Behavior and Human Performance, 1977,
Vol. 20, pp. 238-252.

Borman, Walter C. "Effects of Instructions to avoid Halo
Error on Reliability and Validity of Performance
Evaluation Ratings." Journal of Applied Psychology,
1975, Vol. 60, No. 5, pp. 556-560.

Borman, Walter C. and Dunette, Marvin, D. "Behavior-Based
Versus Trait-Oriented Performance Ratings: an Empirical
Study." Journal of Applied Psychology, 1975, Vol. 60,
No. 5, pp. 561-565.

Braden, Waldo, W. (Editor) Speech Methods and Resources, Harper & Row, N.Y.

Bradway, K. "Jung's Psychological Types: Classification by Test versus Classification by Self." Journal of Analytical Psychology, 1964, 9, pp. 129-135.

Brim, Orville and Hoff, David B. "Individual and Situtional Differences in the Desire for Certainty." Journal of Abnormal and Social Psychology, 1957, 54, pp. 225-228.

Broen, William E., Jr., and Wirt, Robert D. "Varieties of Response Sets." Journal of Counseling Psychology, 1958, Vol. 22, No. e, pp. 237-240.

Bucker, Donald N. "The Predictability of Ratings as a Function of Interrater Agreement." Journal of Applied Psychology, 1959, Vol. 43, No. 1, pp. 60-64.

Burnaska, Robert F. and Hollmann, Thomas D. "An Empirical Comparison of the Relative Effects of Rater Response Biases of Three Rating Scale Formats" Journal of Applied Psychology, 1974, Vol. 79, No. 3, pp. 307-312.

Carlyn, Marcia. "An Assessment of the Myers-Briggs Type Indicator." Journal of Personality Assessment, 1977, Vol. 41, pp. 461-473.

Cascio, Mayne F. and Valenzi, Enzo R. "Behaviorally Anchored Rating Scales: Effects of Education and Job Experience of Raters and Ratees". Journal of Appleid Psychology, 1977, Vol. 62, No. 3, pp. 378-382.

Couch, Arthur and Keniston, Denneth. "Yeasayers and Naysayer: Agreeing Response Set as a Personality Variable." Journal of Abnormal and Social Psychology, 1960,0Vol. 60, No. 2, pp. 151-173.

Damarin, F., and Messick, S. "Response Styles as Personality Variables: A Theoretical Integration of Multivariate Research." (Research Belletin No. RB-65-10), Princeton, N.J.: Educational Testing Service, 1965.

De Coths, Thomas A. "An Analysis of the External Validity and Applied Relevance of Three Rating Formats." Organizational Behavior and Human Performance, 1977, Vol. 19, pp. 247-266.

Di Teverio, John Kesley. "The Strength of Sensing-Intuition Preference on the Myers-Briggs Type Indicator as Related to Empathetic Discrimination of Overt or Covert Feeling Messages of Others." Unpublished Doctoral dissertation, Michigan State University, 1976.

Doonan, Robert Joseph. "An Analysis of Rating Methodologies of Empathy, Warmth, and Genuineness. "Doctoral Dissertation, Auburn University, 1978, Dissertation Abstracts International, pp. 2978-B 2979-B.

Ford, Alexlbert. "Neutralizing Inequalities in Rating." The Personnel Journal, 1930, Vol. IX, No. 6, pp. 466-489.

Freeberg, Norman E. "Relevance of Rater-Ratee Acquaintance in the Validity and Reliability of Ratings. "Journal of Applied Psychology, 1969, Vol. 53, No. 6, pp. 518-524.

Goldshmidt, M.L. "Prediction of College Major by Personality Type." Journal of Counseling Psychology, 1967, 14, pp. 302-308.

Greenwood, John M. and McNamara, Walter J. "Interrater Reliability in Situational Tests." Journal of Applied Psychology, 1967, Vol. 51, No. 2, pp. 101-106.

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

Hamilton, David. "Personality Attributes Associated with Extreme Response Style." Psychological Bulletin, 1968, Vol. 69, No. 3, pp. 192-203.

Harvey, O.J., Hunt, D.E., and Shcroeder, H.M. Conceptual Systems and Personality Organization, New York: Wiley, 1961.

Ivey, A.E. Microcounseling: Innovations in Interviewing Training", Springfield, Ill.: Charles C. Thomas, 1971.

Jung, C.G. Psychological Types. Rev. C.G. Hull. Translated by H.G. Baynes, Princeton University Press, Princeton, N.J., 1971.

Keirsey, David, and Bates, Marilyn. Please Understand Me, Prometheus Nemesis Books, Del Mar, California, 1978.

Kilmoski, Richard J. and London, Manuel. "Role of Rater in Performance Appraisal." Journal of Applied Psychology, 1974, Vol. 59, No. 4, pp. 445-451.

Kingsburg, F.A. "Analyzing Ratings and Training Raters." Journal of Personal Research, 1922, 1, pp. 377-383.

Kneeland, Natalie. "That Lenient Tendency in Rating." Personnel Journal, 1929, 7, pp. 356-366.

Lahey, Mary Anne and Saal, Frank E. "Evidence Incompatible with a Cognitive Theory of Rating Behavior." Journal of Applied Psychology, 1981, Vol. 66, No. 6, pp. 706-715.

Lawlis, G. Frank and Lu, Elba. "Judgment of Counseling Process: Reliability, Agreement, and Error." Psychological Bulletin, 1972, Vol. 78, No. 1, pp. 17-20.

Lee, Raymond, and Malone, Michael, Greco, Susan. "Multitrait- Multimethod- Multirater Analysis of Performance Ratings for Law Enforcement Personnel." Journal of Applied Psychology, 1981, Vol. 66, No. 5, pp. 625-632.

McGee, Richard. "Response Style and Personality Variable: By What Criterion?" Psychological Bulletin, 1962, Vol. 59, No. 4, pp. 284-295.

McGee, Richard. "The Relationship Between Response Style and Personality Variables." Journal of Abnormal and Social Psychology, 1962, Vol. 5, No. 5, pp. 347-357.

Mehrens, William A. and Lehmann, Irvin. Measurement and Evaluation in Education and Psychology, Holt, Rinehart and Winston, New York, 1978.

Myers, Isabel Briggs. The Myers-Briggs Type Indicator Manual, Educational Testing Service, Princeton, N.J., 1962.

Newcomb, Theodore. "An Experiment Designed to Test the Validity of a Rating Technique." Journal of Educational Psychology, 1931, Vol. 22, pp. 279-288.

Rogers, C.R., Gendlin, E.T., Kiesler, D. and Truax, C.B. The Therapeutic Relationship and its Impact: A Study of Psychotherapy with Schiophrenics." Madison: University of Wisconson Press, 1966.