This is to certify that the

thesis entitled

THE DEVELOPMENT AND VALIDATION OF THE AFFECT RECOGNITION AND RESPONSE SCALE, A MEASURE OF EMPATHIC ABILITY

presented by

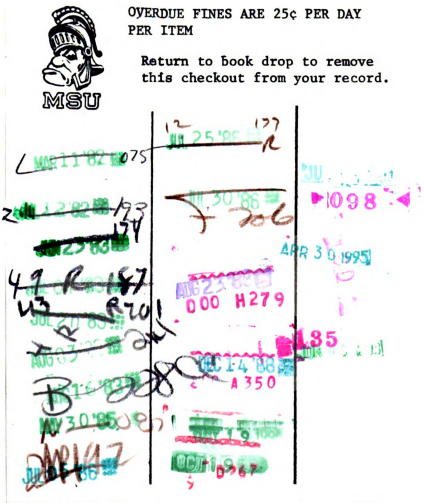
Margaret Ann Parsons

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling, Personnel Services and Educational Psychology

Major professor

Date 5/7/79


MSU
OVERDUE FINES ARE 25¢ PER DAY
PER ITEM
Return to book drop to remove
this checkout from your record.
APR 30 1995
000 H279
A350

THE DEVELOPMENT AND VALIDATION OF THE AFFECT RECOGNITION AND RESPONSE SCALE, A MEASURE OF EMPATHIC ABILITY

By

Margaret Ann Parsons

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services
and Educational Psychology

1979

ABSTRACT

THE DEVELOPMENT AND VALIDATION OF THE AFFECT RECOGNITION AND RESPONSE SCALE, A MEASURE OF EMPATHIC ABILITY

By

Margaret Ann Parsons

Empathy, the ability to recognize what another person is feeling and respond appropriately to those feelings, is a necessary part of the counseling relationship. Many attempts have been made to develop an instrument to adequately measure this ability. To date, however, there is no instrument available which has been adequately validated for use in graduate admissions programs in psychology. The purpose of this study was to develop such an instrument and provide initial validation data for it.

The Affect Recognition and Response Scale combines two areas of theory and research which have been previously studied separately: the ability to recognize emotions through facial expressions of affect and the measurement of the ability to respond empathically to another person. The first subtest of the Affect Recognition and Response Scale uses slides of facial affect to measure the ability to recognize emotions. The second and third subtests use color videotape and written stimulus situations with multiple-choice answers to measure the ability to respond empathically.

Sixty-five subjects were selected from each of two populations, one population of students in professions which regularly require the use of empathic skills and a population of students in professions which do not require the regular use of empathic skills. Subjects in Group I were graduate students in clinical and counseling psychology at Michigan State University and Central Michigan University. Subjects in Group II were graduate students in engineering, mathematics and the physical sciences at Michigan State University.

A multiple measures design with two crossed factors of group and sex was used. In addition, supervisors' ratings of affective skills were obtained for subjects in Group I who were currently engaged in clinical work. The scale was administered to subjects in small groups, demographic data was obtained on a Biographical Data Sheet, and supervisors' ratings were obtained for subjects in Group I.

Reliability for the total scale was estimated to be .853. The average percent of agreement among expert judges for all items on the scale was .94. Factor analysis results indicated one main factor for the scale corresponding to the ability to respond empathically, regardless of stimulus situation format. The factor analysis structure did not correspond to the subtest structure of the scale, nor was there a secondary factor structure corresponding to the emotion categories for the slides of facial affect used in the first subtest.

The major results of the study were:

1. Graduate students in clinical and counseling psychology scored significantly higher than graduate students in engineering, mathematics and the physical sciences on the second and third subtests. There were no significant differences between groups on the first subtest (slides of facial affect.

2. There was no relationship between subtest scores and supervisors' ratings of affective skills.

3. There were no significant differences between men and women on any of the subtests.

4. There was a slight but significant positive relationship between scores on each of the subtests and graduate grade-point average.

The scale was found to be within acceptable limits for reliability and some initial positive validation data was obtained. While the recognition of emotions may be a necessary condition for the ability to respond adequately, this ability seems to be widespread. Further study of how the expression of emotions is modified by display rules and how groups differ in the ability to recognize such modified expressions is needed. Further validation of the Affect Recognition and Response Scale is indicated, using different criterion groups and different behavioral measures of empathic ability.

DEDICATION

To Bill, John, and Mike

ACKNOWLEDGMENTS

Many people have provided help, support and understanding during the completion of this work: my family, especially, Chuck Bassos, Bill Hinds, Judith Krupka, Bill Farquhar, Don Grummon, Diane Singleton, Margaret Zerba, Martha Anderson and Doug Miller.

TABLE OF CONTENTS

Page

APPENDICES

LIST OF TABLES

CHAPTER I

THE PROBLEM

The ability to be able to recognize what another person is feeling and respond appropriately to those feelings may be considered an essential part of the counseling relationship. Rogers (1951, 1957) has labeled this ability empathy, and his work has been expanded on by Truax and Carkhuff (1967), who have sought to operationalize definitions of empathy and study the component abilities included in this skill. Carkhuff (1969) defines empathy as the ability to determine what a client is feeling and to communicate this understanding to the client. He has labeled these two components the ability to discriminate and the ability to communicate.

A related field of research has developed which has sought to operationalize and measure the recognition of affect through the study of facial expressions of emotion. Izard (1971) has studied emotion recognition and labeling across cultures using a set of slides showing different affect categories which he has developed. Ekman and his associates (1972) have also developed slides to study the expression of facial affect and the rules which appear to govern the display of emotion and the recognition of facial expressions.

While it seems that the ability to respond empathically to another person rests on the prior ability to recognize that person's

feelings and that one of the major cues to another's feelings is his/her facial expression, little work has been done to integrate these two fields of study. One of the intentions of this study will be to integrate the related research fields of empathy and the study of facial expressions of emotion.

Need

There have been periodic attempts in the literature to point out the necessity for making selection decisions in graduate programs in psychology based on criteria which include noncognitive dimensions, particularly empathic ability (Sub-Committee on Counselor Trainee Selection, Division of Counseling Psychology, American Psychological Association, 1954; Santavicca, 1959; Stripling and Lester, 1963; Patterson, 1962; Carkhuff, 1969d; Hurst, 1973; Jones, 1974; Hurst and Shatkin, 1974). The need for an instrument to assess empathic skills is also evident in related areas. In the field of medical education, for example, increasing emphasis is being placed on the importance of the doctor-patient relationship (Schofield, 1966; Turner, et al., 1974) and the desirability of being able to screen applicants to medical school on non-cognitive criteria. Chief among these criteria is empathic ability. As yet no adequate measure has been found for this purpose.

Admissions decisions, both in graduate psychology programs and in medical schools, continue to be made largely on the basis of cognitive variables (Rawls, Rawls, and Harrison, 1969; Hurst, 1973). Interviews have been used to assess noncognitive areas, but are generally

unreliable and costly in terms of faculty and applicant time (Sax, 1968; Schwab, 1969; Austin, 1972). Several studies have shown that the traditional selection measures, i.e., undergraduate grade-point average, Graduate Record Examination scores, and letters of recommendation, although predictive of academic success, are unrelated to ratings of empathy (Bergin and Solomon, 1963; McGreevy, 1967; Wiggins and Blackburn, 1969).

There are uses for a valid and reliable measure of empathic ability in other areas. Many paraprofessional training programs, such as volunteer crisis centers, could use such an instrument for screening trainees or as a stimulus for training and discussion. A variety of instruments are currently available for the measurement of empathy skills. None of these measures, however, has been widely validated or used for selection purposes. As Hurst and Shatkin point out:

> . . . direct new measures must be introduced and validated for admissions purposes if admissions procedures are to be defended on just about any level.[1]

Purpose

The purpose of this study is to provide initial validity data for the Affect Recognition and Response Scale. This scale has been developed for use as an admissions screening device and consists of three subtests. Each subtest uses a different stimulus situation. Subtest 1 has a set of slides of facial affect, Subtest 2 has a color

---

[1]Hurst, Michael, and Shatkin, Stephen, "Relationship Between Standardized Admissions Variables and Certain Interpersonal Skills," Counselor Education and Supervision (September 1974), p. 32.

videotape, and Subtest 3 uses written stimulus situations. Two skills are measured: the ability to recognize affect and the ability to respond empathically.

The study compared responses of two groups of graduate students, one group from a population of students in professions which regularly require the use of empathic skills, and one group from a population of students in professions which do not require regular use of empathic skills. In addition, supervisor's ratings of affective skills were obtained for students in the first group who were currently involved in direct clinical service, and these ratings were compared with their scores on the Affect Recognition and Response Scale.

Theory

Two main theoretical discussions will be presented here. The first is a theory of emotion, based largely on the work of Ekman, Izard, Tomkins, and Plutchik. Ekman's approach will be emphasized since it integrates much of the work of the other theorists. Secondly, a theoretical explanation of empathy as the recognition and response to emotion and the measurement of empathy will be presented.

Theories of Emotion

Several different theories of emotion have been developed in an attempt to explain affective experience. Physiological explanations for emotion focus on observed bodily states and changes. From this viewpoint emotions are described as changes in glandular secretions, neural activity, and movements of the musculature, particularly

the facial musculature. Varying degrees of emotional intensity may be measured by observing physiological changes, but discrete emotions are not postulated.

Schlosberg (1954) added the concept of subjective experiencing of feelings in his theory of emotions but postulated only three different dimensions of emotion. These three dimensions, pleasant-unpleasant, attention-rejection, and activity-rest, were the only divisions of emotional experience which he made. Plutchik (1962) enlarged on Schlosberg's notions of emotional dimensions to develop discrete categories of emotions. Izard, Tomkins, and Ekman have all developed theories based on the concept of discrete emotion categories, although the categories used have varied from theorist to theorist.

Definitions of Emotion

Definitions of emotion generally focus on one or more of the following areas: neurophysiological changes; patterns of muscle activity, including movements of the facial musculature; subjective experience of feelings; eliciting stimuli; verbal responses; and behavioral or interactive consequences of the experience of emotions. Theorists differ chiefly according to which of these aspects they include and which they do not. Schlosberg, for example, focuses on the subjective experience as the chief component of emotion.

Hebb (1946), in work with primates, defines emotion as neurophysiological states which are inferred from behavior and which have interactive and behavioral consequences.

Plutchik, while acknowledging the subjective experiencing of emotions, considers this to be a sufficient component, but not a necessary one, similar to the psychoanalytic view that emotional experience may be repressed and thus not experienced subjectively, even though the emotion may be obvious to an observer. He bases his definition on the components of neurophysiological changes, patterned muscle activity, eliciting stimuli, and behavioral consequences.

Tomkins focuses his definition of emotion on muscle and glandular responses, particularly the patterns of facial muscle changes associated with different emotions, and also includes the idea of feedback, which may modify or change the emotion or the subjective experience of the emotion.

Ekman defines emotion rather loosely, including the concepts of physiological responses, motor responses (including facial muscle patterning), verbal responses, and interactive consequences of certain behavior. Izard incorporated all of the previously mentioned components of emotion in his definition, although emphasizing more than other theorists the subjective experience and the importance of the facial muscle responses. His definition is perhaps the most concise and inclusive of the theorists:

> When neurochemical activity, via innate programs, produces patterned facial and bodily activities, and the feedback from these activities is transformed into conscious form, the result is a discrete fundamental emotion which is both a motivating and a meaningful cue-producing experience.[2]

---

[2]Izard, Carroll, The Face of Emotion (New York: Appleton-Century-Crofts, 1971), p. 185.

For the purpose of this study, emotion is defined as the feeling state of a person, whether or not subjectively experienced, which is accompanied by specific neurophysiological and muscular responses.

Nature of Emotion

Much of what is known about the nature of emotions and the experience of emotions is contained in the various definitions which have been presented. In common usage, emotion most often refers to the subjective feeling state of a person, such as a feeling of anger, happiness, or sadness. This subjective experience, however, may or may not be present. Clinical experience, as well as personality theory, supports the assumption that a person may be experiencing an emotion while not consciously aware of it or willing to acknowledge it. In addition, feedback from the emotion, such as awareness of physiological changes, or awareness and analysis of the emotion itself, will often change the subjective experience of that emotion.

Specific neurophysiological changes take place in the body when an emotion is experienced, and these changes vary with the emotion. The hypothalamus seems to be a particular site for electro-chemical activity during emotions and, in fact, injections of certain chemicals or electrical stimulation of specific brain areas will result in subjective emotional experiences.

Duchenne was the first to extensively map the changes in facial musculature associated with different emotional responses. Izard and Ekman have continued this work with Izard focusing on the underlying muscle structures of the face which change with the

different emotion categories and Ekman emphasizing the external changes in the appearance of the face with the different emotions.

An important aspect of emotion is that of behavior. Plutchik defines his emotion categories in terms of disposition to various behaviors; Darwin saw the different emotion expressions developing from different behaviors involved in the experience of these emotions, such as the infant's smile developing from the sucking movements of the mouth during feeding. A simple division of emotion-related behavior is the familiar "fight or flight" response, with fighting associated with the emotion of anger and flight with the emotion of fear. Other emotions show similar characteristic patterns of behavior. However, as one moves up the phylogenetic scale the behaviors associated with different emotions become increasingly complex. As will be discussed in the section on display rules, man has learned many ways of disguising or inhibiting behavior which would ordinarily accompany the experience of a given emotion.

Contemporary theorists generally divide emotions into a number of discrete categories, although the exact number of categories and the labels for distinct emotions vary from theorist to theorist. Ekman has chosen a set of discrete emotion categories based on previous research, with each category chosen having been found by more than one researcher to be recognizable in a literate culture. These six categories were used as the basis for the recognition of facial affect in this study. The six categories are: anger, pleasure, fear, distress, surprise, and disgust.

Facial Expressions of Emotion

The face is the primary site for the expression of emotions in humans, partly because of the flexibility and responsiveness of the facial muscles, partly because of the large part played by face-to-face contact in human communication. Current debate centers on whether facial expression, and in fact, emotional experience, is universal and cross-cultural or varies from culture to culture. Many of the studies done in this area have looked at the ability of subjects in different cultures to agree on the facial expression of different emotions. Ekman's position is that the experience and expression of emotion is both universal and culture-bound. The initial emotional experience and the facial expression of the primary emotions are generally the same across cultures, while the subsequent behavior varies widely from culture to culture.

The initial emotional response and expression is an innate, reflexive behavior, apparent from birth. Each member of a given society, however, learns from childhood a set of rules, which Ekman calls display rules, which govern emotional experience and expression. By the time a person reaches adulthood his/her emotional expression is to a large extent governed by learned display rules, even though the initial emotional experience is innate. These display rules are used to intensify, deintensify, neutralize, or mask the expression of an emotion. Thus, in a culture where aggressiveness is frowned upon, the display rule for anger may serve to mask its expression with a smile. In addition to culturally differing display rules, cultures may differ in the eliciting circumstances for different emotions, the

behavioral consequences of the emotion, and attitudes about certain emotions.

Recognition of Emotions

Recognition of emotions is an important component of empathic ability. In order for a person to respond accurately and helpfully to the emotion of another, she/he must first be able to recognize and accurately label that emotion. Much of the work that has been done in the area of recognition of emotion from nonverbal clues has used facial expressions of emotion for stimulus cues. Two main problems remain in this area, that of generalizability of results and conclusions, and that of how to develop valid stimulus situations. Generalizability covers several areas. Are results consistent across judges, across subjects, and across situations? This study tests the hypothesis that the ability to accurately recognize facial expressions of emotion is not consistent across judges but rather is a skill which varies from person to person. Previous research on the judgment of facial expression has often assumed that the ability to accurately label facial expression of emotion is consistent across the general population, at least within a given culture.

Generality across subjects assumes that there is little variation in the ability to portray facial expressions of emotion. While research indicates some generality for primary or "pure" expressions of emotion, the spontaneous expression of emotion is governed largely by culturally learned display rules, as discussed previously. Research using posed expressions suggests that there is in fact variation in the ability to accurately portray facial expressions.

The third aspect of generality is generality across situation, e.g., is the expression of fear the same when a person is alone or with others, with friends or strangers? Again, culturally learned display rules seem to restrict the generality of expression across situations, such that the expression of an emotion, for example, may be culturally acceptable when one is with family members but will be neutralized or masked when one is in public.

The problem of posed versus spontaneous expressions of emotion is related to difficulties of generality. While posed expressions appear to be "artificial," research findings indicate that they are more easily agreed upon by judges than spontaneous expressions. Aside from methodological difficulties inherent in obtaining spontaneous expressions of emotion, there seems to be little difference in the actual expression of posed and spontaneous emotions (Coleman, 1949). In addition, as Plutchik and Izard have pointed out, primary emotions occur infrequently in spontaneous situations. Learned behavior quickly overrides the pure expression of emotion, resulting in affect blends, substitutions, and masking.

Definitions of Empathy

Empathy is defined in this study as the ability to recognize another's emotion and respond to that emotion. This definition rests on the theoretical foundation of work begun by Rogers and continued by Truax and Carkhuff. Rogers defined empathy as ". . . a state of perceiving the internal frame of reference of another with accuracy, and with the emotional components and meanings which pertain thereto,

as if one were the other person, but without ever losing the 'as if' condition."[3] Carkhuff further refined this definition to include two separate components, which he labeled the ability to discriminate and the ability to communicate. He considered both to be necessary for effective empathic ability, unlike Rogers, who put more emphasis on the ability to perceive emotion.

Tagiuri, in his review of the literature (1965), defines empathy as the ability to accurately perceive or judge others, and provides one of the few discussions of the link between the work done in judging facial affect and the work in the area of empathy measurement, although confining himself to studies which defined empathy as predictive ability. In his discussion, however, Tagiuri does go further to enlarge the definition of empathy to include several independent abilities, including the ability to discriminate distinct emotions, as in the studies of judging of facial affect.

Measurement of Empathy

The development of the Affect Recognition and Response Scale is based on the analog model of selection proposed by Carkhuff: "The best index of a future criterion is a previous index of that criterion."[4] Much of the previous research on the measurement of empathy has attempted to measure empathy indirectly, through the use of personality inventories, standardized tests, or measures of

---

[3]Rogers, Carl, On Becoming a Person (Boston: Houghton Mifflin Co., 1961), p. 284.

[4]Carkhuff, Robert, Helping and Human Relations, Vol. 1 (New York: Holt, Rinehart and Winston, Inc., 1969), p. 85.

related characteristics. Those instruments which have shown the most promise have been those most directly related to the actual use of the ability to be measured, i.e., some form of stimulus situation to which the subject can make an empathic response.

Just as empathic ability involves more than one dimension, so too does the stimulus to which subjects respond. Thus, the ability to recognize emotions will vary according to whether the stimulus cue involves the face or body, one or more than one person, or nonverbal as well as verbal cues, as in the use of a videotape or written stimulus situation.

General Hypotheses

The following general hypotheses are tested in this study:

1. Subjects in graduate programs which require the use of one-to-one interpersonal skills will have higher scores on a test of empathic ability than subjects in graduate programs which do not require the use of these skills.

2. Subjects with higher supervisor's ratings of affective skills in clinical settings will score higher on a measure of empathic ability.

3. There will be no difference between men and women in scores on a measure of empathic ability.

Definitions of Terms

Empathy. The ability to determine what another person is feeling and to communicate this understanding to the other person.

Affect. The feeling or emotional state of a person at a given time.

Facial Affect. The nonverbal communication of an emotional state or feeling through facial expression, involving patterned movements of the facial musculature.

Nonverbal Communication. The expression of affect without the use of words, primarily through facial expression, posture, gestures, and voice qualities such as pitch and volume.

Emotion. The feeling state of a person, whether or not subjectively experienced, which is accompanied by specific neurophysiological and muscular responses.

Stimulus Situation. A test item which presents a person or persons expressing an emotion, to which the subject can make a response. Stimulus situations used in the present study in the Affect Recognition and Response Scale are slides of facial affect, videotape vignettes, and written vignettes.

Criterion. A direct and independent measure of the variable to be tested, in this study, the variable of empathic ability.

Criterion Groups. Groups selected because of differences on some criterion measure.

Delimitations

The samples used in this study are restricted to volunteer subjects, rather than being randomly drawn from their respective populations. The samples were generally nonminority (only three minority subjects in Group I and none in Group II) and contained a higher percentage of males than females. Although the percentage of female and minority subjects in the sample is probably representative

of their numbers in the populations from which they were drawn, caution should be used in applying the results of this study to future admissions programs which may include larger numbers of both female and minority applicants.

The use of criterion groups assumes that a group of subjects identified as having high empathy skills will score higher on a measure of empathy than a group of subjects identified as low in empathy skills, if the scale is indeed a valid measure of empathic ability (Hambleton and Novick, 1973). Various attempts have been made to identify high and low empathy criterion groups, but this remains a major difficulty in empathy measurement. The assumption made in this study is that advanced graduate students in counseling and clinical psychology, the majority of whom have completed a supervised practicum, have more highly developed levels of empathic ability than do graduate students in fields not requiring training or practice in the use of empathic abilities.

In Chapter V, the implications of these limitations in interpreting the results of this study will be discussed more fully.

Overview

In Chapter II the literature relevant to facial affect and emotion and the measurement of empathic ability will be reviewed. In Chapter III the design and analysis of the study will be presented, including a description of the samples and methodology and a description of the analyses used. The development of the Affect Recognition and Response Scale will also be described in Chapter III, along with

reliability and factor analysis data. In Chapter IV the analysis of the results will be given. Chapter V will contain the summary and conclusions, as well as a discussion of the implications for future research.

CHAPTER II

RELATED RESEARCH

This chapter will present an overview of research in several areas. First, some general findings on recognition of emotion will be given. A more extensive summary of research findings specifically dealing with the area of recognition and labeling of the facial expression on affect will then be presented. A section on definitions of empathy and research in the measurement of empathic ability will then be given. Finally implications of the related research for the present study will be discussed.

Recognition of Emotion: Nonverbal Behavior

Many authors (Barbara, 1956; Berger, 1958; Dittman and Wynne, 1961; Ekman, 1965) have pointed out the importance of the study of nonverbal behavior, particularly in a psychotherapy setting. Nonverbal behavior involves any aspect of communication other than verbal content, such as body movements, facial expression, voice rate, pitch, and length of speech pauses. Researchers have studied the usefulness of attending to nonverbal communication during interactions and the degree to which judges can agree on different aspects of nonverbal communication.

Davitz and Davitz (1959a, 1959b) in two separate studies looked at the ability of judges to accurately identify feelings

expressed in content-free speech. They found a significant negative correlation between similarity between feelings and the accuracy with which those feelings were discriminated. In addition, some feelings were more frequently identified correctly, with anger being most frequently identified and pride least frequently identified correctly. They concluded that feelings which are subjectively experienced as similar will be more difficult to differentiate than feelings experienced as disparate.

Dittman and Wynne (1961), however, in their study of voice characteristics such as stress and pitch, while finding that such characteristics could be reliably coded by different judges, were unable to find consistent patterns to correspond with the different emotions expressed, using excerpts from a therapy interview and from a recorded radio program. Dittman (1962) also studied the relationship between body movements and moods (emotions) in therapy interviews and was able to find a relationship between moods and frequency of body movements, although he concluded that these patterns are unique to each individual.

Starkweather (1956) reviewed several studies of vocal cues in nonverbal communication and concluded that vocal cues are useful in indicating the presence of strong emotional states. Eldred and Price (1958) studied vocalization patterns in psychotherapy interviews and were able to find high agreement among judges on different patterns of vocal cues which correspond to different emotions of the client in the interview.

Ekman has done several studies focusing on the nonverbal communication value of body movements and body posture, in addition to his work in the area of recognition of facial affect. In one series of four experiments (Ekman, 1964), judges were asked to match verbal excerpts from interviews with photographs of body position taken during the interviews. Results showed that the judges could correctly match verbal with nonverbal behaviors significantly better than chance. In a second series of experiments using stressed and unstressed interviews for stimulus situations (Ekman, 1965), Ekman used both videotapes and still photos as cues for judges. Ekman concluded that whole body stimulus photos were better than either face-only photos or body-only photos since different parts of the body transmit different types of nonverbal information. He found no difference, however, in accuracy between still photos and motion pictures. Judges were initially allowed to see only one photo from an interview to eliminate situational and contextual cues, but Ekman found that adding more photos from the interview did not improve accuracy. However, he did find that judges could not accurately distinguish between stressed and unstressed conditions when shown the subject only but could make the distinction when shown a photo of both the subject and the interviewer. He suggests that one reason for this may be that the subject tries to conceal his emotions, an idea he later expanded into his concept of display rules.

Recognition of Emotion: Facial Expression of Affect

Development of Stimulus Photographs

Many different sets of stimulus photographs of facial affect have been developed over the years and used in research in the recognition of facial expressions of emotion. One of the earliest sets was that developed by Ruckmick (1921) using a female drama student who practiced and posed various expressions. While the quality of the pictures is generally good, they are limited in the number of emotion categories represented and the fact that only one stimulus subject was used.

Frois-Wittman (1930) also developed a set of photographs, using himself as the model, and posed various facial expressions. He attempted to eliminate situational cues, hand gestures, or other distractions from the expression portrayed. In addition, he developed a set of drawings of different facial expression. While his photos were an improvement over available pictures, they still represented only a limited number of emotion categories and used only one stimulus subject.

Coleman (1949) reviewed the literature on the facial expression of affect and concluded that posed pictures were too artificial to be useful. He developed a set of stimulus photographs using various stimulus conditions to elicit spontaneous expressions of emotion. The difficulty in this method can best be understood by describing the elicitors which Coleman used: subjects were given a sudden very loud blast on an electric horn, received a severe

electric shock, and were required to gradually crush a snail through the use of both index fingers. Coleman includes a section on the extremely negative reactions of subjects used to make these photos. The actual value of such spontaneously obtained photos will be discussed further in the next section on recognition labeling experiments. Coleman also asked each subject to act the emotions they originally experienced spontaneously and so obtained a second set of stimulus photos.

It has been difficult to obtain spontaneous examples of certain emotions expressed with sufficient intensity because the experiencing of these emotions is naturally defended against, such as shame or fear. Based on previously learned display rules, primary emotions occur only infrequently in spontaneous expression. Affect blends are much more common, or masking of the expression of emotions. Inman (1976) in his study of facial expression using slow motion and normal speed videotape stimulus situations, found that raters recorded a greater number of emotions for the slow motion tape than for the normal speed tape. Ekman theorized that the initial expression of a primary emotion is often displayed for a few micro-seconds, but is quickly masked, thus making it difficult to obtain photographs of expressions of strong primary emotions.

More recently, both Izard (1971) and Ekman (1976) have developed sets of slides of facial expressions of affect for use in research. These slides have used trained actors to pose various expressions of emotions. Both sets have the advantage of including a variety of stimulus subjects, both male and female, and a complete

representation for the categories of emotion being studied. Izard's slides appear somewhat dated today, and one recent study (Zerba, 1977) showed low homogeneity coefficients for items within each emotion category for his set of slides.

The slides developed by Ekman (1976) were made using trained subjects who posed a variety of emotions. The technical quality is an improvement over other available pictures, and Ekman reports a high percentage of agreement among college students used as judges for each slide available in his final set.

Recognition-Labeling Experiments

Research in the area of recognition and labeling of emotions has gone on for many years. Early studies in the area of recognition and labeling of facial expression were often hindered by the lack of adequate stimulus materials. Nevertheless, some answers to questions about facial expression were provided. These questions were whether judges can accurately identify expressions of emotion, whether some emotions are easier to identify than others, whether the identification of emotion is dependent on situational or interactional cues, whether some persons are more accurate judges of emotions than others, and whether it is easier to judge spontaneous or posed expressions accurately.

In one of the early studies of labeling of facial expressions, Ruckmick (1921) asked observers to label the emotion expressed in each of his series of photographs of a young woman. Although he performed no statistical analysis of the results, he did find some

agreement and accuracy of judgment. Primary emotions were judged more accurately, and with better agreement, than secondary emotions. Accuracy was made more difficult by the fact that judges were asked to label thirty-five separate expressions, each supposedly different. The intended distinctions between such categories as resentment and sulkiness or haughtiness and defiance are difficult to distinguish in theory much less as distinct facial expressions.

Frois-Wittman (1930) conducted a series of experiments in recognition-labeling using pictures of himself which he had developed and a set of pictures of facial expression. Judges in the experiments were college students. Each judge was given a list of forty-three terms compiled from previous researchers and asked to label the expressions presented in the photos and drawings. The median for agreement on the pictures was 37.5 percent, not low considering the possibility of forty-three different labels, each considered as a separate category. In general, Frois-Wittman found a wide scatter for labels, with one or more modal frequencies. Pictures with more than one modal label showed a logical relationship between the modes, e.g., anger and hate being the two modes for a given picture. Frois-Wittman also studied patterns of muscular involvement in each expression and found distinct patterns of muscle involvement for each expression which had appeared as a modal frequency for at least one picture. Frequent disagreements were found between judgments on the whole face and judgments of separate features and Frois-Wittman concluded that the meaning of a given pattern of muscle involvement, e.g., raising of the eyebrows, differed, depending on the rest of the

facial muscle pattern. Given the limitations of the stimulus pictures with which he worked and the large number of classifications, often representing fine distinctions of meaning, Frois-Wittman was able to show a significant agreement in the ability of observers to recognize an emotion expressed in a stimulus picture.

In a follow-up study using the Frois-Wittman pictures, Hulin and Katz (1935) used seventy-two pictures and asked judges to sort the pictures into groups according to whether the pictures showed approximately the same facial expression. Results showed a wide scatter in agreement among judges, with some cases of high percentage of agreement. Unfortunately, Hulin and Katz did not report which emotion categories were chosen by observers or ask them to label the groupings they chose. Results are reported as the percent agreement of observers on the similarity between any two pictures, making it impossible to assess the agreement in labeling any one picture.

Coleman (1949), in addition to reviewing the literature on studies of facial expression of emotion, used both spontaneous and posed stimulus pictures. The situations used to elicit the spontaneous emotions have been previously described. Coleman used undergraduate students in psychology as judges and asked them to match the expression shown on the stimulus photo with the list of situations. He concluded that laughter was the most easily identified emotion and also found that the acted situations were equal to or better than the natural expressions in ease of accurate identification. There were no differences between male and female judges in accuracy of identification. In his review of other studies, however, Coleman

cites both studies with no sex differences and studies in which females were more accurate than males. Differences in methodology reduce the comparability of these studies, however. Full-face photos were judged more accurately than either mouth-region or eye-region photos. Coleman's findings add support for the notion of discrete, identifiable emotion categories.

In 1965 Tagiuri reviewed studies in the area of recognition of emotion, including recognition of facial expressions of emotion. He discusses five problems that have not only hindered research in this area, but make comparison of results across studies difficult. These problems remain relevant to current research in this area. The first problem which Tagiuri discusses is the variability in stimulus situations presented, i.e., still photos, motion pictures, drawings. At the time of his review there was no accurate standardized set of pictures of facial expression available, and each researcher generally developed his own set of stimulus pictures, none of which was comparable to any other. Secondly, there is a difference in the task, either recognition or labeling of emotion categories. As Tagiuri points out, the task of labeling an expression without preselected categories is not the same task as that of selecting one label from a set which has been preselected by the researcher.

A related problem, and one which was particularly apparent in early research on recognition is the variability in emotion categories of labels used. As previously mentioned, these range from Schlosberg's three dimensions to Frois-Wittman's list of forty-three different terms for emotional expressions. Studies have also varied

in the contextual or situational cues provided for the judges. Tagiuri concludes that the more situational information available to a judge, the more accurate will be his/her judgment. The final problem which he mentions is that of the sampling of emotional expressions. Ruckmick and Frois-Wittman, for example, both used only a single subject in posing their sets of photographs. Thus, these photos are open to possible distortion based on idiosyncracies of the subjects used. Izard and Ekman were the first to develop sets of stimulus slides using a number of different subjects of both sexes. Ekman also includes more of an age range in the subjects which he used.

Tagiuri presents several important conclusions in his review. He finds no consistent sex differences in recognition of affect, but does conclude that there is some relationship between the ability to judge emotion and level of intelligence. His review of studies of subjects who are blind, and thus have not learned emotional expressions or cues from others, shows that there are some innate patterns of expression of emotion. He also concludes that some expressions are more easily discriminated than others and that the expression of a specific emotion varies with the sequence of emotions and the situation, as in the masking of an emotion to fit what is considered appropriate in a situation. Tagiuri cites studies to support both the cross-cultural and the universal positions of emotional expression and suggests that the universal similarities may reflect innate aspects of expression and recognition.

Tomkins and McCarter (1964) used a set of sixty-nine posed photographs and a sample of twenty-four urban firemen in a

recognition experiment. Each subject was asked to identify the photos according to a set of nine emotion categories, including neutrality. An average correlation of .86 between judgments and the affect which was intended was found. Tomkins and McCarter also found some systematic confusion between emotion categories and some individual idiosyncracies for individual subjects. Affects most likely to be confused were those most similar to each other, e.g., distress and shame, interest and neutrality. Tomkins and McCarter theorize that affects which are triggered by similar situations are most likely to be confused with each other. Individual idiosyncracies may develop because of learning, as when a child is taught that the display rule for anger is to mask its expression with a smile, or because of an individual tendency to continuously experience only one affect or expect only one affect from others, as when a person who is continuously hostile is not able to accurately express enjoyment or recognize its expression in others.

Ekman and Friesen (1971) conducted an experiment to measure the universality of facial recognition with a preliterate tribe from New Guinea. Each subject was told a story focused on one emotion and then asked to pick one of three pictures of facial expression which matched the story. The percent of subjects choosing the correct picture was generally better than 75 percent, with the exception of the fear and surprise categories, which were not accurately discriminated. In a subsequent experiment, American college students were shown videotaped facial expressions of the

New Guinea natives and were able to accurately identify the emotions being portrayed.

Izard (1971) conducted a series of experiments in emotion labeling and recognition using a sample of American and foreign college students. His stimulus photos were a set of slides of facial affect which he had developed. In the Emotion Recognition Experiment each subject was asked to choose an emotion from a list of eight emotions provided. The average agreement for each slide was 78 percent, with a high degree of agreement across cultures (American, European, Oriental, and African). The same subjects were also asked to provide their own labels for the slides, before participating in the Emotion Recognition Experiment. The average agreement for labeling was 56 percent for females and 50 percent for males, with lesser degree of similarity across cultures than for the recognition task.

Definitions of Empathy

There are many definitions of empathy which have been proposed over the years. The first major definition was that of Dymond (1949), who defined empathy as the ability of the subject to accurately predict another's feelings, attitudes, or opinions. This definition was used both by Dymond and by Kerr and Speroff (1954) as the basis for tests of empathic ability. Smith (1966) used predictive empathy, which he called sensitivity to people, as the basis for both a training program and a measure of ability.

Cohen (1973) and Feshbach and Feshbach (1969) defined empathy as the ability to vicariously experience the emotions of another person and used slides as the stimulus situations for testing this concept with small children. Stotland and Dunn (1963) carried this idea one step further and defined the vicarious experiencing of emotion in physiological terms, measuring empathic ability by checking physiological changes in their subjects concomitant with the actual experiencing of the same emotion as another person. Chapin (1942) considered empathy to be the equivalent of social insight and devised a test to measure a person's knowledge of social skills in a variety of situations.

A major influence on research in this area has been the definition of empathy given by Rogers, which was discussed in Chapter I. For Rogers, empathic understanding was one of the three necessary and sufficient conditions for change in therapy (1957).

Another aspect of empathy which has been studied particularly by Izard (1971) and Ekman and Friesen (1975) is the nonverbal expression of affect, particularly facial expression and body posture as expressions of affect. The ability to correctly label the nonverbal cues to another's feelings is an important expansion of the definition of empathy.

Measures of Empathy

Although a large number of instruments in various forms have been developed, the difficulty of operationalizing the concepts, the multifaceted nature of the construct of empathy, the process-content

distinction in interviews, and problems in identifying adequate criterion groups continue to pose difficulties for research. A number of authors have specifically addressed these problems (Wolf and Murray, 1936; Taft, 1955; Strunk, 1957; Carkhuff, 1969c; Gormally and Hill, 1974), but they have yet to be satisfactorily resolved.

Astin (1957) used two different measures of empathy, based on different definitions, and found that one measure discriminated between counselors and noncounselors, and the other did not. Similarly, Hayden (1955) used a measure of predictive empathy and ratings of group members' empathy by group leaders. His results were not significant, and he concluded that predictive ability is not the best definition of empathy. Hastorf and Bender (1952) discuss the confounding effects of projection and perceived similarity on predictive empathy measurement.

Truax and Carkhuff (1967) have been careful to differentiate between the ability to discriminate affect and the ability to communicate empathically, as have other investigators (Chandler, 1970; Jarrett, et al., 1972; and Jones, 1974). Both Chandler and Heilman (1972) factor-analyzed the data from several different empathy measures and concluded that empathy is a many-faceted concept, rather than a single construct.

The written test has had the most extensive use among instruments measuring empathic ability. Within that format there is considerable variation in the construction of the tests. Early measures of predictive ability (Dymond, 1949; Kerr and Speroff, 1954)

involved rating, in written form, how others would respond to a given test or situation. Dymond required the subject to predict others' self-ratings for six personality traits, whereas Kerr required that the subject predict how people in general would respond to music, magazine selections and interpersonal situations.

Attempts have also been made to measure the process and relationship aspect of interactions (Barrett-Lennard, 1962; Linden, et al., 1965; Dilley and Tierney, 1969). Often these measures have been used concurrently with other measures of therapist empathy (Truax, 1966; McWhirter, 1972; Kurtz and Grummon, 1972), but the results have generally shown little correlation between the empathy measures and the client's perceptions either of the therapy relationship or of the therapist's empathic ability.

An assortment of other empathy measures are based on the semantic differential (Bellucci, 1971), word association techniques (Kandler and Hyde, 1953), physiological measures (Stotland and Dunn, 1963), or developed from other tests such as the Minnesota Multiphasic Personality Inventory (MMPI) (Hogan, 1969; Hurst, 1973).

A variation of the written instrument has been utilized which requires that the subject respond in writing to a given stimulus (Astin, 1957; O'Hern, 1962; Carkhuff, 1969c). Sidman (1968) developed a test based on responses to questions about short stories. In addition, Carkhuff designed his Index of Discrimination and Index of Communication to be used either in a written form or with both the stimulus situations and the subject's responses

recorded on tape and was able to demonstrate that the two forms were equivalent.

Instruments have also been designed which utilize situations with a number of possible responses provided in a multiple-choice format. This has the advantage that the responses can be more readily rated (Chapin, 1942; Porter, 1950; Kerr and Speroff, 1954; Ashby, et al., 1957; Craig, 1959). Still another variation is the Interaction Maze described by Gazda (1974). In the Interaction Maze, the subject is presented with a series of stimulus situations centered on one problem. The subject moves back and forth through the series of responses, depending on which response he chooses. Thus the instrument more closely resembles a real-life interaction where an interviewer will elicit more information with a facilitative response or stop communication with a judgmental response. Bernstein and his associates (Bernstein, et al., 1954; Rasche, et al., 1973) have developed and validated an objectively scored instrument which is used with medical students in a doctor-patient relationship course.

Most other measures of empathic ability focus on the ability of the subject to perceive and discriminate affect and to communicate this perception to the client. There are several variations, but the general format involves a presentation of several stimulus situations (tape recordings, film) to which the subject responds. The responses are then rated using various scales. The most widely used rating scale is the Accurate Empathy Scale developed by Truax (Truax and Carkhuff, 1967; Walker, 1969; Spadone, 1974), with variations developed by Carkhuff (1969c), Smith (1971), and others

(Chandler, 1970; Mickelson and Stevic, 1971; Gazda, 1974). Adler and Enelow (1966), Passons and Olson (1969), and Guerney, et al. (1968) have also developed scales for rating responses.

Carkhuff's rating scales have been criticized by a number of authors (Chinsky and Rappaport, 1970; Rappaport and Chinsky, 1972; Gormally and Hill, 1974; Horwitz, 1977; Thoresen, 1977). The rating scales require training judges and are not usable for large-scale testing. In addition, obtaining adequate inter-judge reliability has been a persistent problem. Many of Carkhuff's findings are biased by his use of the same rating scale for both pre- and post-training measures and for the actual training sessions. Thus, his subjects were measured in empathic ability using the Accurate Empathy Scale, before training, trained to give correct responses to the same scale, and then measured with the scale after training.

Demographic Variables

Demographic variables have been studied extensively to determine their effect on empathic ability, but results either do not reach significance, or are contradictory. Taft (1955) reviewed numerous studies involving the correlates of the ability to judge others (his definition of empathy). His variables include age, sex, family background and sibling rank, and intelligence and perception, but there were no consistent results for any of these variables. Some investigators report differences in results by sex (Cantrell, 1967; Johnson, et al., 1967; Sidman, 1968; Feshbach and Feshbach, 1969; Huber, 1972; Cohen, 1973; Veeser, 1974), while

others report no differences (Taft, 1955; Cohen and Struening, 1962; O'Hern, 1962; Blumstein, 1972). Where sex differences are found, females display the greater empathic ability.

The effects of birth order have also been found to be contradictory (Stotland and Walsh, 1963; Stotland and Dunn, 1963; Cantrell, 1967; Cohen, 1973). Prior training or experience also appears to have an inconsistent effect. Cohen and Struening (1962), Greenberg, et al. (1969), Huber (1972), and Veeser (1974) all report positive effects of training and experience in the development of empathic ability. Campbell (1962), however, found no differential effects due to experience and training. And Carkhuff (1969a; Carkhuff, et al., 1970) reports a decrease in empathic ability as the result of professional training.

Use of Standardized Tests

Various standardized tests have been administered in an attempt to predict empathic ability using personality variables. One series of such studies is based on the work of Whitehorn and Betz. In Whitehorn's original study (1960), psychiatrists were divided into two groups according to success rates with schizophrenic patients, and then administered the Strong Vocational Interest Blank (SVIB). Whitehorn found significantly different response patterns between the two groups, and he labeled these two groups the 'A' and 'B' therapists. Subsequent studies, however, have not upheld the clear-cut distinction of 'A' therapists as effective clinicians and 'B' therapists as ineffective (Boyd, 1970; Scott and Kemp, 1971).

A considerable amount of work has been done using the Minnesota Multiphasic Personality Inventory (MMPI) in an attempt to find correlations with empathic ability (Vesprani, 1969; Blumstein, 1972; Jones, 1974) or counselor effectiveness (Brams, 1961; Johnson, 1967; McGreevy, 1967). Results for studies using the MMPI show mixed results, with occasional significant correlations for some subscales. Hurst, in his review of the literature (1973), reports consistent negative correlations between the Depression and Psychasthenia Scales and empathy measures, but Brams (1961) and Jones (1967) did not find the same results. Their general conclusion was that the MMPI is not a useful measure for screening for empathic ability.

The Edwards Personal Preference Schedule (EPPS) has also been used extensively in research. Hogan (1969) reports a significant negative correlation between the Social Desirability subscale and scores on his empathy test. Morris (1971) reports that in the literature the results have been variable and sometimes contradictory. Results in general are similar to those found for the MMPI: some subscales of the EPPS correlate with the criterion measures used, but the significant subscales are not the same from study to study, and results are sometimes contradictory (Bergin and Solomon, 1963; Stefflre, et al., 1963; Lawton, 1965; Johnson, et al., 1967; Vesprani, 1969; Charles, 1973).

A third area of study has focused on the related concepts of authoritarianism, dogmatism, and openness, most often using the Rokeach Dogmatism Scale as the measure. Milliken and Paterson (1967)

and Stefflre, et al. (1962) found significant discrimination for subscales of the Dogmatism scale, but Passons and Olsen (1969) report no correlation between empathic sensitivity and scores on the Dogmatism scale. Allen (1967) measured openness by a special scoring of Rorschach protocols and concluded that openness was related to effectiveness of therapy.

A large number of studies have used other instruments. Except for the three already discussed, however, no measure has been used extensively, or with consistent findings. Instruments used include the Guilford-Martin Inventory (Halpern, 1954), Personal Orientation Inventory (Winborn and Rowe, 1972), Strong Vocational Interest Blank (Stefflre, et al., 1962); Berkeley Public Opinion Questionnaire (Brams, 1961); Myers-Briggs Type Indicator (Gough, 1960; Hogan, 1969; Boles, 1975); and the Omnibus Personality Inventory (Gruberg, 1969).

Criterion Groups

There have been differing approaches to the use of criterion groups for empathy research. Criterion groups have been designated on such bases as self-report, peer, or faculty ratings (Bandura, 1956; Stefflre, et al., 1962; Lawton, 1965; Allen, 1967), or by using various measures of empathy to divide subjects into high and low empathy groups for concurrent validation (Sidman, 1968; Dilley and Tierney, 1969; Feshbach and Feshbach, 1969; Blumstein, 1972).

Sandler (1972) compared female nonprofessional mental health workers (high empathy group) with a control group of adult women on

several measures, including the Hogan Empathy Scale and found that the experimental group scored significantly higher on the empathy measure. O'Hern (1962) developed the Sensitivity Scale to measure empathic ability using taped client problems as stimulus situations. The instrument was administered to counselor candidates and discriminated at a significant level between those judged most and least effective by staff ratings. It did not, however, discriminate between those judged most and least sensitive as counselors. Milliken and Paterson (1967) divided counselor candidates into two criterion groups ("good" and "bad" counselors) according to ratings by both supervisors and coached clients on their Counselor Effectiveness Scale.

Mickelson and Stivic (1971) divided counselors into facilitative or non-facilitative counselor groups according to rankings based on responses to taped stimulus situations. Their study tested the effectiveness of verbal reinforcement techniques in eliciting client information-seeking behavior, and results showed significant differences for the facilitative and non-facilitative counselors in the predicted direction. Carkhuff, Kratochvil, and Friel (1968) used first and fourth year clinical and nonclinical graduate students as criterion groups to study the effects of training on counselor effectiveness but did not find significant results. Campbell (1962) used experienced and inexperienced counselors to study counseling subrole behaviors but found few differences between the two groups.

Astin (1957) used counselors and non-counselors as criterion groups and administered a test of predictive empathy ability and a

situational test of empathy ability. Results showed that the situational test discriminated between counselors and non-counselors, and a predictive test did not. Allen (1967) found a correlation between psychological openness (defined as self-awareness and awareness of one's own feelings) and supervisor's ratings of practicum students. Similarly, Bandura (1956) found a relationship between therapist's anxiety and supervisor's rating of competence.

Veeser (1974) developed an instrument to measure sensitivity to both verbal and nonverbal emotional cues and found that psychology graduate students scored higher than engineering graduate students or undergraduate students on both the verbal and nonverbal measures.

Implications of Related Research

The review of related research suggests that while many instruments have been developed in attempts to measure empathic ability, there is currently no instrument available which has been adequately validated or which includes the aspect of recognition of nonverbal expressions of emotion through facial affect. There is support for the concept of discrete emotion categories and for the use of an analog model of measurement as the most likely to prove valid.

Posed stimulus situations have proven better than spontaneous expressions for accuracy of judgments, are more easily standardized, and may help to increase comparability of research results in the future.

Research findings on sex differences both on measures of empathic ability and in recognition and labeling of facial affect have been consistently inconclusive, showing either no differences or higher ability for women. With the gradual eliminating of sex role stereotypes and the greater acceptance of empathic behavior for men, any differences which may earlier have existed may be disappearing.

The use of criterion groups in previous research has focused on distinctions of training and experience or has used some measure of performance, such as supervisor's ratings, to designate high and low empathy groups. In this study the criterion groups are designated by both training and experience, and supervisor's ratings are used as an additional validity check.

CHAPTER III

DESIGN OF THE STUDY

The design of the study involved administering the Affect Recognition and Response Scale to subjects in the two designated groups and having each subject complete a Biographical Data Sheet which provided data on the demographic characteristics of the samples. Supervisor's ratings of affective skills were obtained for subjects in Group I (graduate students in majors requiring the use of one-to-one interpersonal skills) who were currently involved in clinical work. Tests of significance were applied to test scores and supervisor's rating scores to test the major hypotheses. A description of the sample, design, methodology, and analysis used is presented in this chapter. A description of the development of the Affect Recognition and Response Scale, including expert judges' data, reliability and item analysis data, and factor analysis results, is also included.

Sample

A sample of sixty-five subjects was obtained from each of two populations. The population sampled for Group I consisted of graduate students majoring in counseling and clinical psychology from Michigan State University and Central Michigan University. The majority of the subjects had completed at least one year of coursework and were

currently involved in clinical work, either in a practicum, internship, or job setting. Students at Central Michigan University were contacted through announcements in classrooms and the department office. Ten students agreed to participate as subjects and were tested in one session. Subjects at Michigan State University were individually contacted by telephone. Lists of doctoral graduate students were obtained from the departments of counseling and clinical psychology. An attempt was made to contact each student on the list because of the large size of the sample required relative to the total available population. Each student was asked to participate in a one and a half hour testing session for research purposes and told that the purpose of the study would be explained at the end of the testing session. Each subject was also paid for his/her participation. Fifty-five students from Michigan State University agreed to participate as subjects for the study.

The population sampled for Group II consisted of graduate students majoring in engineering, mathematics, and the physical sciences at Michigan State University. Lists of graduate students in the engineering and mathematics departments were obtained from the respective departments. All foreign students were eliminated from the lists to avoid the effect of cultural differences, particularly in the recognition of facial affect. Each of the remaining students was individually contacted by telephone and asked to participate in the study.

After contacting all students in these departments it was not possible to obtain enough subjects, so graduate students majoring in

the physical sciences were also included. Names of these students were obtained through the Michigan State University student directory. Since the directory gives information on a student's class rank, major, and home address, in addition to name and telephone number, it is possible to identify non-foreign graduate students in the desired majors. Again an attempt was made to contact almost all of the available students due to the large sample size required and a high refusal rate for participation.

Due to the relative lack of both women and minority graduate students in the population sampled for Group II and the lack of minority students in the population sampled for Group I, it was not possible to obtain equal numbers of female and male subjects for either group, nor was it possible to obtain sufficient minority subjects to test any hypotheses about differences in test scores due to race. Table 3.1 presents demographic data for each sample.

Development and Description of the Affect Recognition and Response Scale

The Affect Recognition and Response Scale is a revised form of the Empathy Skills Rating Scale (Krupka and Parsons, 1978), which was developed as a measure of empathic ability for use in medical school admissions screening under a grant from the National Fund for Medical Education. The Affect Recognition and Response Scale consists of three subtests using a set of slides of facial expressions of emotions and a series of written and color videotape vignettes (see Appendix A for a copy of the test packet, with sample items). The Empathy Skills Rating Scale consists of five subtests, including

TABLE 3.1.--Demographic Data for Group I and Group II.

Variable	Group I	Group II
Sex		
Male	35	51
Female	29	14
Non-response	1	0
Race		
Minority	3	0
Non-minority	62	65
Degree		
M.A./M.S.	28	18
Ph.D.	35	45
Non-response	2	2
Age		
Mean	28	27
Range	22 to 39	21 to 46
Number of Children		
Mean	.41	.31
Range	0 to 3	0 to 3
Undergraduate G.P.A.[a]		
Mean	3.29	3.36
S.D.[b]	.46	.37
Graduate G.P.A.		
Mean	3.80	3.64
S.D.	.16	.24

[a]Grade-point average.

[b]Standard deviation.

a set of postural line drawings, a series of color videotape vignettes, and a series of written stimulus situations. Subtests 2 and 3 of the Affect Recognition and Response Scale, including the color videotape vignettes, are taken directly from the Empathy Skills Rating Scale. Subtest 1 was added to this, using slides and emotion categories developed by Ekman and Friesen (1976).

Item Pool

An initial pool of six hundred written stimulus situations was generated by the test developers for the Empathy Skills Rating Scale. These items consisted of brief statements, usually no more than two or three sentences, covering a wide range of expressed affect and subject matter, such as hostility, enjoyment, depression, and fear; and sexuality, death, and racial issues. An attempt was made to develop stimulus situations which were brief, contained the expression of only one emotion, either overtly or covertly, and covered a wide range of topics, emotion categories, and levels of emotional intensity. A selection of stimulus subjects was also made so that they covered an age range from children to the elderly and included men, women, and minority as well as nonminority subjects.

Pilot Videotape

A black and white pilot videotape was produced using vignettes developed from selected stimulus situations taken from the original pool of items which had been generated. These pilot vignettes used trained role-players and included two people in an interaction in

each vignette. Each of these vignettes lasted between thirty and forty-five seconds.

Pilot Studies

An initial form of the Empathy Skills Rating Scale was developed, including the pilot videotape, and administered to two separate groups with a total of ten subjects, five male and five female. Each group took the scale, filled out an extensive debriefing questionnaire and participated in a debriefing session which was recorded and transcribed. Based on results from these pilot studies, the scale was revised and a new color videotape was produced.

This revised form of the scale was administered to a third pilot group of subjects, consisting of sixteen male and female undergraduate students in introductory psychology courses. Subjects in all three pilot groups were presented written and videotape stimulus situations and asked to write their own responses. These responses were used to develop multiple-choice answers for Subtests 2 and 3 of the scale.

Description of Subtest 1: Slides of Facial Affect

Several attempts were made to develop slides of facial affect which would be suitable for the scale. An initial set of slides was reproduced from works of art, but there was little agreement among pilot subjects on the emotion expressed in each slide. A second set of slides was then developed from the color videotape vignettes, using stop-action equipment for the videotape, but it was not possible to

cover a range of distinct emotion categories. A third set of slides was made from a black and white videotape[5] which portrayed a variety of subjects expressing emotions. Although this set of slides was of better technical quality, it still did not provide enough different clearly expressed poses for each emotion category.

The set of slides finally used for the scale was chosen from the Pictures of Facial Affect developed by Ekman and Friesen (1976). These are a set of 110 slides of facial expressions of emotion, using more than a dozen different persons who were trained to contract or relax different facial muscles associated with various facial expressions, so as to pose a specific facial expression for a given emotion category. The six emotion categories used were those which have generally been included by most theorists in the area of emotion: pleasure, distress, fear, anger, disgust, and surprise. Data on reliability and validity for the entire set of Pictures of Facial Affect are presented in a brochure which accompanies the set of slides (see Appendix B). All slides included in the set met a criterion of 70 percent or better agreement among observers. From this set of 110 slides, 36 (six from each emotion category) were originally selected for Subtest 1 of the scale. A Table of Emotions (see Appendix A) was developed for use with the slides. Each emotion category contains the main emotion description and a subset of synonyms denoting varying degrees of intensity for the main emotion category.

---

[5]Videotape courtesy of Bob Wilson, College of Education, Michigan State University.

These 36 slides were administered to a group of expert judges. Based on their response the "fear" category of emotions was dropped, since only two slides in this category met the criterion of 80 percent or better agreement among judges. Five slides which met this criterion were selected in each of the remaining categories for the final version of the scale, and the Table of Emotions was revised to eliminate the "fear" category. Thus, the final form of Subtest 1 contains five slides in each of five emotion categories, all of which met the criterion of 80 percent or better agreement among expert judges.

Development of Subtest 2:
Videotape Vignettes

Based on results of the pilot studies using a black and white test videotape, a new color videotape was developed. Trained actors were used to enact short (15-20 seconds) vignettes using a script developed from the original item pool. Twenty different vignettes with ten different actors ranging in age from seven to sixty-five were filmed. This videotape was then edited, based on technical quality and realism of the vignettes, to produce a final version of the videotape containing fifteen vignettes, each followed by one minute of blank tape for response time. The videotape vignettes, which were the stimulus situations for Subtest 2 of the scale, were administered to the expert judges with a set of multiple-choice response answers. Based on these expert judges' responses, four of the items were dropped from the subtest due to lack of agreement. Since it was not possible to edit these vignettes from the tape, they were administered to all subjects as part of the scale but were not scored or included in any of

the data analyses. Thus, the final version of Subtest 2 contains eleven videotape vignette stimulus situations with multiple-choice responses, all of which met the criterion of 80 percent or better agreement among the expert judges.

Development of Subtest 3:
Written Stimulus Situations

Fifteen written stimulus situations, taken from the original item pool, were included in the initial form of Subtest 3 which was administered to the expert judges. Based on their responses, one item was dropped from the scale. The final version of Subtest 3 contains fourteen items, all of which met the criterion of 80 percent or better agreement among the expert judges.

Expert Judges' Data

The Affect Recognition and Response Scale was individually administered to five expert judges, three female and two male. All had doctoral degrees in counseling psychology and were engaged in clinical work. Data from the expert judges' responses were used as the basis for the development of a scoring key for the scale. The form of the scale given to the judges included 36 slides, six in each of the six emotion categories (Subtest 1), 15 videotape vignettes (Subtest 2), and 15 written stimulus siutations (Subtest 3). Judges were asked to rank order the multiple-choice responses for Subtests 2 and 3 from one to four, with one being the least helpful response and four being the most helpful response. Data from the expert judges' response for Subtest 1 are shown in Table 3.2.

TABLE 3.2.--Number of Expert Judges Choosing Each Emotion Category for Slides of Facial Affect (Subtest 1).

Slide Number	Emotion Category					
	Anger	Pleasure	Distress	Disgust	Fear[a]	Surprise
1		5				
2			4		1	
3	5					
4						5
5			5			
6	1		4			
7		5				
8		5				
9						5
10				5		
11		5				
12	1			4		
13	5					
14			4	1		
15	5					
16			5			
17						5
18		1				4
19	5					
20				5		
21				5		
22	1			4		
23	4	1				
24		5				
25						5

[a]The "fear" category was dropped from the final form of Subtest 1.

A criterion of 80 percent agreement (four of the five expert judges) was used to retain slides for the scale. Thirty-one of the slides (86 percent) met this criterion. Only two of the six slides in the "fear" category met this criterion, however, so this category was dropped from the scale. Since only five of the six slides in the "anger" category met the criterion, one slide was randomly dropped from each of the remaining categories to equalize the number of slides in each category. Thus, the final form of the scale contains a total of 25 slides, five in each of five emotion categories (anger, pleasure, distress, disgust, and surprise).

Data from the expert judges' responses for Subtests 2 and 3 are shown in Table 3.3. Using the criterion of 80 percent agreement, it was not possible to assign a ranking from one to four for responses to each of these items. A decision was made to aggregate rankings of three and four to a single high ranking (correct response) and rankings of one and two to a single low ranking (incorrect response). Using the criterion of 80 percent agreement on this high-low ranking, it was possible to retain eleven items for Subtest 2 and fourteen items for Subtest 3. Thus, for each of these items subjects received one point for either of the two correct responses and zero points for either of the two incorrect responses. The average percent of agreement among the judges was calculated for each subtest and for the total scale. For the total scale this was 94 percent.

TABLE 3.3.--Number of Expert Judges Rating Each Response Choice Correct or Incorrect for Subtest 2 and Subtest 3.

Item Number and Rating		Response Choice			
		a	b	c	d
26	Correct	0	4	5	1
	Incorrect	5	1	0	4
27*	Correct	3	0	2	5
	Incorrect	2	5	3	0
28	Correct	0	5	5	0
	Incorrect	5	0	0	5
29	Correct	5	4	1	0
	Incorrect	0	1	4	5
30	Correct	0	0	5	5
	Incorrect	5	5	0	0
31	Correct	5	4	1	0
	Incorrect	0	1	4	5
32	Correct	0	0	5	5
	Incorrect	5	5	0	0
33	Correct	0	5	0	5
	Incorrect	5	0	5	0
34	Correct	5	5	0	0
	Incorrect	0	0	5	5
35*	Correct	3	2	5	0
	Incorrect	2	3	0	5
36*	Correct	5	0	3	2
	Incorrect	0	5	2	3
37	Correct	0	5	5	0
	Incorrect	5	0	0	5
38	Correct	0	0	5	5
	Incorrect	5	5	0	0
39*	Correct	5	2	0	3
	Incorrect	0	3	5	2

TABLE 3.3--continued.

Item Number and Rating		Response Choice			
		a	b	c	d
40	Correct	5	0	0	5
	Incorrect	0	5	5	0
41	Correct	5	0	5	0
	Incorrect	0	5	0	5
42	Correct	0	5	4	1
	Incorrect	5	0	1	4
43	Correct	0	0	5	5
	Incorrect	5	5	0	0
44	Correct	4	1	5	0
	Incorrect	1	4	0	5
45	Correct	5	0	0	5
	Incorrect	0	5	5	0
46	Correct	5	5	0	0
	Incorrect	0	0	5	5
47	Correct	0	0	5	5
	Incorrect	5	5	0	0
48	Correct	5	5	0	0
	Incorrect	0	0	5	5
49	Correct	0	5	0	5
	Incorrect	5	0	5	0
50	Correct	0	5	0	5
	Incorrect	5	0	5	0
51	Correct	4	5	0	1
	Incorrect	1	0	5	4
52	Correct	0	5	5	0
	Incorrect	5	0	0	5
53	Correct	5	0	1	4
	Incorrect	0	5	4	1
54	Correct	4	1	5	0
	Incorrect	1	4	0	5

*Item omitted in final form of scale.

Subtest Score Correlation Matrix

Pearson product-moment correlation coefficients were calculated for subtest scores for all subjects. The correlation matrix is presented in Table 3.4. The correlation matrix for individual item scores was also calculated and is given in Appendix C. For the subtest scores, all correlations are significant at the .001 level. The high correlation between Subtests 2 and 3 raises a question of the need for the two separate formats and whether these two subtests are in fact measuring different constructs. These questions will be discussed further in the section on factor analysis.

TABLE 3.4.--Subtest Score Correlation Matrix.

	Subtest 1 (Slides)	Subtest 2 (Videotape)	Subtest 3 (Written)
Subtest 1 (Slides)	1.000		
Subtest 2 (Videotape)	.264*	1.000	
Subtest 3 (Written)	.332*	.752*	1.000

*All correlations significant at the $p < .001$ level.

Reliability

Reliability for the Affect Recognition and Response Scale was calculated using the Kuder Richardson formula #20, which calculates an internal consistency coefficient using all possible split-half

combinations of items. A reliability coefficient was calculated for the scale as a whole and for each of the individual subtests. For the entire scale the reliability estimate was .853. Reliability estimates for each subtest were: Subtest 1, .416; Subtest 2, .799; Subtest 3, .804 (Table 3.5). Item-total reliability statistics were also calculated and are presented in Appendix D.

TABLE 3.5.--Reliability and Item Analysis Data for Subtest Scores and Total Scale Scores.

Characteristic	Subtest 1	Subtest 2	Subtest 3	Total Scale
Mean Item Difficulty	13.6	24.1	24.6	18.6
Mean Item Discrimination	13.6	51.6	60.3	31.4
Kuder-Richardson #20 Reliability Coefficient	.416	.799	.804	.853
Standard Error of Measurement	1.52	1.19	1.35	2.44

Reliability estimates for the scale as a whole and for Subtests 2 and 3 were within acceptable limits. The reliability estimate for Subtest 1, however, is quite low, particularly considering the number of items in the subtest. Zerba (1977) reported similarly low reliability for recognition and labeling tasks using slides and emotion categories developed by Izard. Indices of discrimination and difficulty for subtest scores indicate that items in Subtests 2 and 3 were more difficult than items in Subtest 1 and

discriminated more highly between subjects in the upper and lower scoring groups. Indices of difficulty and discrimination for individual items are presented in Appendix E.

Factor Analysis

In order to examine the underlying structure of the scale, a factor analysis using all items was done. A principal components factor analysis with no assumptions about expected structure was first performed, followed by a varimax rotation with no preset number of factors to be extracted. An eigen value of 1.00 or greater was used as the criterion for determining the number of factors extracted by varimax rotation. A total of ten factors emerged for the scale. The minimum value for factor loadings was set at an absolute value of .40; factor loadings for items in each of the ten factors are shown in Table 3.6. Item-factor correlations for all of the 54 items with each of the 10 factors are presented in Appendix F.

The factor structure which emerged from this initial analysis did not conform to the subtest structure of the test. Rather, there was one main factor with high factor loadings for most of the items in Subtests 2 and 3, a second factor with high loadings for slides 1 and 24 from the "pleasure" emotion category for Subtest 1, a third factor with high loadings for two items from Subtest 3, and additional factors with high loadings for one, two or three individual slides from Subtest 1. Thus, it appears that Subtests 2 and 3, rather than measuring different abilities or constructs, are essentially measuring the same thing. In addition, there seems to be

TABLE 3.6.--Varimax Rotated Factor Matrix for All Items with Minimum Eigen Value of 1.00.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
+1.00		S1		S10						
.90								S22		
.80										
.70	T29 T37 T40 W47 W52	S24	W50			S8				
.60	T28 T38 W45 W48 T30 W42 W49		W44		S18				S4	S23
.50	T31 W43 W46						S3 S21		S5	
.40	T32				S17 W51					
+ .30										
0										
- .30										
.40										
.50										
.60										
.70										
.80										
.90										
-1.00										

S = Slides (Subtest 1)
T = Tape (Subtest 2)
W = Written (Subtest 3)

no secondary structure conforming to emotion categories for the slides of facial affect (Subtest 1).

Since the initial factor analysis did not reflect the subtest structure of the scale, a second factor analysis was performed, presetting the number of factors to three. The correlations between items and factors, again set at an absolute value of .40, are presented in Table 3.7. This factor analysis yielded results similar to the initial one. Factor 1 components include most of the items from Subtests 2 and 3, Factor 2 components are slides 1 and 24 and the first item in Subtest 2 (videotape vignettes), and Factor 3 components include two slides from Subtest 1 and one item from Subtest 2. Again, the main factor appears to be a measure of ability to respond empathically, unrelated to the form of the stimulus situations presented.

A third factor analysis was performed using only items from Subtest 1 (slides of facial affect) to examine the relationships among these items and determine whether the factors correspond to the five emotion categories. For this analysis, the number of factors for varimax rotation was preset to five to correspond to the number of emotion categories. The correlations between items and factors, set at an absolute value of .40, are shown in Table 3.8. Factor 1 corresponded to the second factor found in the previous attempts, with high factor loadings for slides 1 and 24 from the "pleasure" category. The other slides from the "pleasure" category did not have high correlations with this factor, however (one of these slides, slide 7, was not included in any of the factor analyses because of

TABLE 3.7.--Varimax Rotated Factor Matrix for All Items with the Number of Factors Preset to Three.

	Factor 1	Factor 2	Factor 3
+1.00			
.90		S1	
.80	T40		
.70	T37	S24	
.60	T29 T38 W45 W48 W52		
.50	T31 W49 W47 T28 T33 W42 W43 W46		S10 S13
.40	T30	T32	T28
+ .30			
0			
- .30			
.40			
.50			
.60			
.70			
.80			
.90			
-1.00			

S = Slides (Subtest 1)

T = Tape (Subtest 2)

W = Written (Subtest 3)

TABLE 3.8.--Varimax Rotated Factor Matrix for Subtest 1 (Slides of Facial Affect) with the Number of Factors Preset to Five.

	F1	F2	F3	F4	F5
+1.00					
.90	P1				
.80					
.70	P24				
.60			A15		
.50		D10		S4 Di5	
.40		D22		D21	
+ .30					
0					
- .30					
.40					Di2 S17
.50					
.60					
.70					
.80					
.90					
-1.00					

P = Pleasure

D = Disgust

A = Anger

S = Surprise

Di = Distress

TABLE 3.8.--Varimax Rotated Factor Matrix for Subtest 1 (Slides of Facial Affect) with the Number of Factors Preset to Five.

	F1	F2	F3	F4	F5
+1.00					
.90	P1				
.80					
.70	P24				
.60			A15		
.50		D10		S4 Di5	
.40		D22		D21	
+ .30					
0					
- .30					
.40					Di2 S17
.50					
.60					
.70					
.80					
.90					
-1.00					

P = Pleasure

D = Disgust

A = Anger

S = Surprise

Di = Distress

zero variance). Overall, the factors which emerged on this analysis did not correspond to the emotion categories used with the slides.

The results of the three factor analyses suggest one main factor for the scale, corresponding to the measurement of the ability to respond empathically, regardless of the form of the stimulus situation. A second consistent factor corresponded to two of the slides expressing a facial affect of pleasure, but the other slides from this emotion category were not components of this factor.

Description of Other Instruments

The Biographical Data Sheet was developed to collect information on characteristics of the samples (see Appendix A for a copy of the Biographical Data Sheet). The Supervisor's Rating Scale (see Appendix A for a copy of the scale and cover letter) is based on rating scales used by the Department of Psychiatry, College of Human Medicine, at Michigan State University to rate medical students in psychiatry clerkships. The scale is an eight-point, behaviorally-oriented interval rating scale which measures affective skills. A high rating on the scale is defined as mastery of the ability to respond to a client's affect without avoidance and the ability to recognize and respond to both surface expressions of feelings and deeper levels of affect. A low rating indicates ineffective use of these behaviors; the therapist is unable to accurately label a client's feelings, avoids responding to the client's affect, and is unable to recognize or respond to even the obvious expressed surface feelings of the client.

Research Design

This study employed a multiple measures design with two crossed factors to test for group and sex main effects and a group by sex interaction effect. The dependent variables were the three subtest scores on the Affect Recognition and Response Scale. Group and sex were the independent variables. The research design is presented in diagram form in Table 3.9. A second design was used for the regression analysis with supervisor's ratings for subjects as the dependent variable and subtest scores as the independent variables (Table 3.10).

Apparatus

The Affect Recognition and Response Scale consists of a set of 25 slides of facial expressions of affect developed by Ekman and Friesen, a 3/4 inch color videotape cassette, developed by Krupka and Parsons, and a test booklet. A 35 mm. slide projector and screen are required, as well as a color videotape monitor and playback equipment. Although either portable or stationary videotape equipment may be used with the videotape cassette, a portable monitor and playback unit were used for all test administrations in this study. A stop watch was also used to time slide presentations for the first subtest so that each slide was presented for exactly thirty seconds. Number two lead pencils and mark-sense answer sheets were used to enable machine scoring of all test responses.

TABLE 3.9.--Design of the Study for Group and Sex Effects.

		T1	T2	T3
		I1 ... I25	I26 ... I40	I41 ... I54
P1	G1	S1 S35		
	G2	S36 S65		
P2	G1	S66 S116		
	G2	S117 S130		

P1 = Counseling and Clinical Psychology Majors (Group I)
P2 = Engineering/Mathematics/Science Majors (Group II)
G1 = Male
G2 = Female
S = Subject
T1 = Subtest 1 Score
T2 = Subtest 2 Score
T3 = Subtest 3 Score
I = Item Score

TABLE 3.10.--Research Design for Supervisor's Ratings Analysis.

	R1
P1	S1 S37

P1 = Group I

R1 = Supervisor's Ratings

S = Subject

Methodology

Administration of the Affect Recognition and Response Scale was carried out in small groups, ranging in size from two to fourteen subjects. Subjects were assigned to testing sessions according to their availability, except for subjects from Central Michigan University, who were all tested in one session at a previously designated time. All testing was carried out in well-lit, quiet rooms, with the same person administering the test for all subjects. Testing sessions lasted approximately one and a half hours. Due to limitations of subjects' available time and restrictions on the availability of videotape equipment, all testing sessions at Michigan State University were carried out on evenings and weekends.

At the beginning of each testing session, each subject was handed a test packet containing a consent form (see Appendix A for a copy of the consent form), a mark-sense answer sheet, number two lead pencil, test booklet, and Biographical Data Sheet. Subjects in Group I also received a Supervisory Rating Scale, cover letter and stamped addressed envelope to be handed to their clinical supervisor. Answer sheets, Biographical Data Sheets and Supervisory Rating Forms were all pre-coded with identifying numbers so that all materials for each subject could be identified and to insure confidentiality of all materials.

Subjects were asked to read and sign the consent forms, and the Affect Recognition and Response Scale was then administered. Subjects were asked to read the introduction and instructions for Subtest 1 and then given a few minutes to look over the emotion

categories to be used with the slides of facial affect. Each slide was then presented for thirty seconds. After completing the first subtest, subjects were asked to read the instructions for Subtest 2 and then shown the color videotape. The videotape consisted of fifteen short vignettes followed by one minute of blank tape to allow time for responding. Upon completion of the second subtest subjects were asked to read the instructions for Subtest 3, which consisted of fourteen written vignettes, and complete the responses to these items. After completing the third subtest, each subject was asked to fill out the Biographical Data Sheet, and at the end of the testing session questions regarding the study were answered.

Subjects in Group I were asked to give the Supervisory Rating Form, cover letter, and addressed envelope to their clinical supervisors, to be filled out and returned by mail. All forms were identified only by code number to protect confidentiality. Completed Supervisory Rating Forms were received for 37 of the 65 subjects in Group I.

Hypotheses

The following research hypotheses were developed to test the validity of the Affect Recognition and Response Scale:

I. Differences between criterion groups.

Null Hypothesis: There will be no differences between subjects in Group I and subjects in Group II in mean scores for each of the subtests of the Affect Recognition and Response Scale.

H_0: $M_I = M_{II}$

Alternative Hypothesis: Subjects in Group I will have higher mean scores than subjects in Group II for each of the subtests of the Affect Recognition and Response Scale.

H_A: $M_I > M_{II}$

II. Prediction of Supervisor's Ratings

Null Hypothesis: There will be no relationship between subtest scores on the Affect Recognition and Response Scale and ratings of affective skills for subjects in Group I.

Alternative Hypothesis: There will be a positive relationship between subtest scores on the Affect Recognition and Response Scale and ratings of affective skills for subjects in Group I.

III. Differences between sexes.

Null Hypothesis: There will be no differences between men and women in mean scores for each of the subtests of the Affect Recognition and Response Scale.

H_0: $M_M = M_W$

Alternative Hypothesis: Women will have higher mean scores than men for each subtest of the Affect Recognition and Response Scale.

H_A: $M_W > M_M$

Where,

M_I = Mean subtest scores for subjects in Group I.

M_{II} = Mean subtest scores for subjects in Group II.

M_M = Mean subtest scores for male subjects.

M_W = Mean subtest scores for female subjects.

Analysis

The main purpose of the study was to provide validation data for a measure of empathic ability. For this reason, several different aspects of scale validation were incorporated into the study. These included administering the test to a group of expert judges to determine agreement on correct item responses, performing factor analyses on the scale to examine the underlying structure, a criterion-based test of construct validity using subjects in groups which differed on the dimension of empathic ability, and a test of construct validity using an independent criterion measure (supervisor's ratings of affective skills). Data on expert judges' agreement and the factor analysis of the scale have been presented in a previous section. The following analysis was used to test the hypotheses related to criterion-based validation.

A two-way multivariate analysis of variance for multiple measures taken at one time was used to test the first and third hypotheses, with a significance level of .05 or less. For those hypotheses, univariate analyses of variance were then used to determine which subtest scores were significantly different for subjects in Group I and Group II. For the univariate analyses the alpha level was set at .017 by dividing the .05 alpha level equally among the three subtest scores. A regression analysis with a significance level of .05 or less was used to test the second hypothesis and determine the amount of variance accounted for by subtest scores.

Summary

Sixty-five subjects were obtained for each of the two criterion groups. Subjects in Group I were drawn from a population of graduate students majoring in counseling and clinical psychology at Michigan State University and Central Michigan University. Subjects in Group II were drawn from a population of graduate students majoring in engineering, mathematics, and the physical sciences at Michigan State University. All subjects were paid volunteers, and no attempt was made to randomize selection from either population. Because of skewed distributions in the populations sampled, it was not possible to obtain equal numbers of female and male subjects.

The Affect Recognition and Response Scale was developed and tested for construct validity. Supervisor's ratings were obtained for subjects in Group I who were currently involved in clinical work, using the Supervisory Rating Scale, an eight-point scale measuring affective ability which was developed for the study. Data on sample characteristics was also obtained using a Biographical Data Sheet, also developed for the study.

A multiple measures design with two crossed factors of group and sex was used to test the main hypotheses. Scores on the three subtests of the Affect Recognition and Response Scale were the dependent variables. Research methodology involved administering the Affect Recognition and Response Scale to subjects, obtaining responses on the Biographical Data Sheet, and obtaining supervisor's ratings for subjects in Group I.

Statistical hypotheses were formulated to test the differences between subjects in Group I and Group II, the differences between male and female subjects, and the relationship between subtest scores and supervisor's ratings of affective skills. A two-way multivariate analysis of variance for multiple measures was used to test for group and sex differences. Additional univariate analyses of variance were used to determine which subtest scores were significantly different for groups and for sexes.

The development of the Affect Recognition and Response Scale, a description of the scale, and data on expert judges' ratings, reliability, and factor analysis results were also presented. Expert judges' agreement for the scale was 94 percent. Overall scale reliability, using a measure of split-half reliability, was estimated at .853. Factor analysis results indicated one main factor corresponding to empathic responding, regardless of stimulus situation format.

The results of the hypotheses tests and an interpretation of these results will be presented in Chapter IV.

CHAPTER IV

ANALYSIS OF RESULTS

The statistical hypotheses, an analysis of the data, and a summary of the results of the hypothesis tests are presented. The first and third hypotheses were tested by a multivariate analysis of variance. When the results of the multivariate analysis were significant at the .05 level additional univariate analyses of variance were used to determine which subtest scores were significantly different. The second hypothesis was tested using stepwise multiple regression analysis.

Hypothesis I: Differences Between Criterion Groups

Null Hypothesis: There will be no differences between subjects in Group I and subjects in Group II in mean scores for each of the subtests of the Affect Recognition and Response Scale.

H_0: $M_I = M_{II}$

Alternative Hypothesis: Subjects in Group I will have higher mean scores than subjects in Group II for each of the subtests of the Affect Recognition and Response Scale.

H_A: $M_I > M_{II}$

Where,

M_I = Mean subtest scores for subjects in Group I.

M_{II} = Mean subtest scores for subjects in Group II.

Significant differences were found between groups for the three subtest scores ($F = 77.13$, $p < .00001$). Cell means, standard deviations, and approximate overall F value are shown in Table 4.1.

TABLE 4.1.--Summary Data for Multivariate Analysis of Group Effects.

Group and Subtest	Mean Score	Standard Deviation	95 Percent Confidence Interval
Group I			
Subtest 1	21.80	1.82	21.34 to 22.25
Subtest 2	10.02	1.12	9.74 to 10.30
Subtest 3	12.86	1.33	12.53 to 13.19
Group II			
Subtest 1	21.06	2.36	20.48 to 21.64
Subtest 2	6.53	1.88	6.07 to 6.99
Subtest 3	8.32	2.52	7.70 to 8.94

Approximate overall F Value = 77.133

Significance of F Value = .00001

Univariate analyses of variance were performed to examine group differences individually for each of the three subtest scores. For the univariate analyses the alpha level was set at .017 by dividing

the .05 overall alpha level equally among the three subtests. The results of the univariate analyses are presented in Table 4.2.

TABLE 4.2.--Univariate Analyses of Variance for Group Effects.

Variable	Hypothesis Mean Squares	F Ratio	Significance of F Ratio
Subtest 1	17.61	4.04	.04666
Subtest 2	394.70	161.86	.00001*
Subtest 3	670.07	165.10	.00001*

*Significant at the .017 level.

Differences in mean scores on Subtest 1 (slides of facial affect) were not significant at the .017 level. Thus, the null hypothesis of no differences between groups was not rejected for scores on Subtest 1. Differences in mean scores on Subtest 2 (videotape vignettes) and Subtest 3 (written stimulus situations) were significant at the .017 level. An examination of mean scores indicates that the differences were in the predicted direction. The null hypothesis of no difference in mean scores was therefore rejected in favor of the alternative hypothesis. Subjects in Group I did score higher than subjects in Group II on Subtest 2 and Subtest 3. The mean difference in scores for Subtest 2 was 3.49. The mean difference in scores for Subtest 3 was 4.54.

Hypothesis II: Prediction of Supervisors' Ratings

Null Hypothesis: There will be no relationship between subtest scores on the Affect Recognition and Response Scale and ratings of affective skills for subjects in Group I.

Alternative Hypothesis: There will be a positive relationship between subtest scores on the Affect Recognition and Response Scale and ratings of affective skills for subjects in Group I.

Hypothesis II was tested using a multiple regression equation with the three subtest scores as independent variables and supervisor's ratings of affective skills as the dependent variable. The results of the regression analysis were not significant at the .05 level for the three subtest scores entered together into the regression equation or for each subtest score entered independently in the stepwise regression analysis. The three subtest scores together accounted for only 2.0 percent of the variance in supervisor's ratings of affective skills. A summary of the multiple regression analysis results is presented in Table 4.3. The null hypothesis was not rejected in favor of the alternative hypothesis. There appears to be no relationship between subtest scores on the Affect Recognition and Response Scale and ratings of affective skills for subjects in Group I.

TABLE 4.3.--Stepwise Multiple Regression Summary Table for Prediction of Supervisors' Ratings from Subtest Scores.

Independent Variable	Beta	Standard Error of Beta	R^2	Overall F Value	Significance of F Value
Subtest 1	.108	.140	.017	.617	.437
Subtest 2	-.119	.378	.020	.347	.709
Subtest 3	-.027	.244	.020	.229	.876

Hypothesis III: Differences between Male and Female Subjects

Null Hypothesis: There will be no differences between male and female subjects in mean scores for each of the subtests of the Affect Recognition and Response Scale.

H_0: $M_M = M_W$

Alternative Hypothesis: Female subjects will have higher mean scores than male subjects for each subtest of the Affect Recognition and Response Scale.

H_A: $M_W > M_M$

Where:

M_M = Mean subtest scores for male subjects.

M_W = Mean subtest scores for female subjects.

Significant differences were not found between male and female subjects for the three subtest scores of the Affect Recognition and Response Scale. Therefore the null hypothesis of no differences between men and women for each of the subtest scores was not rejected. Since the results of the multivariate analysis were not significant at the .05 level, univariate analyses of variance were not performed for individual subtest scores. Cell means, standard deviations and approximate overall F value are given in Table 4.4.

TABLE 4.4.--Summary Data for Multivariate Analysis of Sex Effects.

Sex and Subtest	Mean Score	Standard Deviation	95 Percent Confidence Interval
Females (N = 43)			
Subtest 1	22.05	1.74	21.51 to 22.58
Subtest 2	8.98	2.04	8.35 to 9.60
Subtest 3	11.72	2.28	11.02 to 12.42
Males (N = 86)			
Subtest 1	21.12	2.25	20.64 to 21.59
Subtest 2	7.88	2.39	7.38 to 8.40
Subtest 3	9.98	3.21	9.29 to 10.66

Approximate Overall F Value = 1.360

Significance of F Value = .258

Group by Sex Interaction Effects

The interaction effect between group and sex for all subtest scores was examined using a multivariate analysis of variance to determine if there were differences in mean scores for men and women in one group but not in the other, or differences in mean scores in opposing directions for each group. The results of the multivariate analysis for interaction effects between group and sex were not significant at the .05 level, however. Since the multivariate results were not significant, univariate analyses of variance were not performed for individual subtest socres. Results of the multivariate analysis are summarized in Table 4.5.

TABLE 4.5.--Summary Data for Multivariate Analysis of Group by Sex Interaction Effects.

Dependent Variable	Group and Sex	N	Mean Score	Standard Deviation	95 Percent Confidence Interval
Subtest 1	Group I:				
	Females	29	22.41	1.45	21.86 to 22.97
	Males	35	21.28	1.95	20.62 to 21.96
	Group II:				
	Females	14	21.14	1.96	20.01 to 22.27
	Males	51	21.04	2.47	20.35 4o 21.73
Subtest 2	Group I:				
	Females	29	10.07	.70	9.80 to 10.34
	Males	35	9.97	1.38	9.50 to 10.45
	Group II:				
	Females	14	6.57	1.87	5.49 to 7.65
	Males	51	6.52	1.89	5.99 to 7.05
Subtest 3	Group I:				
	Females	29	12.96	1.08	12.55 to 13.38
	Males	35	12.77	1.52	12.25 to 13.29
	Group II:				
	Females	14	9.14	1.92	8.04 to 10.25
	Males	51	10.55	3.04	10.02 to 11.08

Approximate Overall F Value = 1.370

Significance of F Value = .255

Analysis of Intelligence Effects

Although it is unlikely that two groups of graduate students would differ significantly in academic ability or intelligence since they have already been selected as graduate students largely on measures of intellectual aptitude and academic performance, further analysis of the data was done to determine if intelligence, as measured by undergraduate and graduate grade-point average, had an effect on scores for the Affect Recognition and Response Scale. The first analysis that was done was a univariate analysis of variance between groups, using undergraduate grade-point average as the dependent variable. A second univariate analysis of variance was then performed, using graduate grade-point average as the dependent variable. The results of the two analyses are summarized in Table 4.6.

There was no significant difference between subjects in the two groups for mean undergraduate grade-point average ($F = .8626$, $p < .3550$). For graduate grade-point average, however, results were significant at the .05 level ($F = 16.543$, $p < .00001$). Although the differences between the two groups was not large (the mean grade-point average for Group I was .24 larger than the mean grade-point average for Group II), subjects in Group I did have significantly higher graduate grade-point averages than subjects in Group II.

Regression analyses were performed to determine if there was a relationship between grade-point average and subtest scores for each subtest, using all subjects together. As in the test for

TABLE 4.6.--Univariate Analyses of Variance for Groups on Undergraduate and Graduate Grade-Point Average.

Dependent Variable	Mean Group I	Mean Group II	Mean Squares Between Groups	F Ratio	Probability of F Ratio
Undergraduate Grade-Point Average	3.29	3.36	.1507	.8626	.3550
Graduate Grade-Point Average	3.80	3.64	.7212	16.5430	.0001*

*Significant at the .05 level.

Hypothesis II, stepwise multiple regression analyses were performed with the significance level set at .05. Three separate analyses were done, one with each subtest score as the dependent variable. The results of these analyses are summarized in Table 4.7.

As can be seen in Table 4.7, the relationship between graduate grade-point average and subtest scores was significant at the .05 level for all three subtests. Graduate grade-point average accounted for 3.8 percent of the variance in scores for Subtest 1, 9.6 percent of the variance in scores for Subtest 2, and 3.7 percent of the variance in scores for Subtest 3. Thus, while the effects of graduate grade-point average are significant at the .05 level, the amount of variance accounted for is still small for scores on each of the three subtests. The relationship between undergraduate grade-point average and subtest scores was not significant at the .05 level for scores on Subtest 1 and Subtest 3. It was significant, however, for scores on Subtest 2, although accounting for less than one percent additional variance.

Since there was some relationship found between both graduate and undergraduate grade-point average and scores on the three subtests, a second multivariate analysis of variance for group effects was performed to test Hypothesis I, this time including undergraduate and graduate grade-point average as covariates to control for their effrects in the analysis. Although the F value calculated with the adjustments made for undergraduate and graduate grade-point average was slightly lower (73.028 as compared to 77.133), the results were still significant at the .00001 level.

TABLE 4.7.--Stepwise Multiple Regression Summary Table for Prediction of Subtest Scores from Undergraduate and Graduate Grade-Point Average (G.P.A.)

Dependent Variable	Independent Variable	Beta	Standard Error of Beta	R^2	Overall F Value	Significance of F Value
Subtest 1						
	Graduate G.P.A.	1.84	.89	.039	4.64	.033*
	Undergraduate G.P.A.	.11	.48	.039	2.33	.102
Subtest 2						
	Graduate G.P.A.	3.37	.94	.096	12.25	.001*
	Undergraduate G.P.A.	-.44	.50	.102	6.50	.002*
Subtest 3						
	Graduate G.P.A.	2.83	1.29	.037	4.42	.038*
	Undergraduate G.P.A.	-.52	.69	.042	2.48	.088

*Significant at the .05 level.

Univariate analyses for individual subtest scores, again using undergraduate and graduate grade-point average as covariates, also produced results similar to the original analyses. Although adjusting for the effects of undergraduate and graduate grade-point average did result in slightly lower F values, the results did not change the outcome of the hypothesis tests for group effects. Table 4.8 presents a comparison of the results of the univariate analyses of variance originally performed with the results of the analyses adjusted for the effects of graduate and undergraduate grade-point average.

Summary

Three hypotheses were tested to examine the construct validity of the Affect Recognition and Response Scale. A multivariate analysis of variance was used to test for group and sex main effects and to examine group by sex interaction effects. Since group differences were significant at the .05 level, additional univariate analyses of variance were performed for each subtest. Stepwise multiple regression analysis was used to test the predictive ability of the subtest scores, with supervisor's ratings of affective skills as the independent variable.

The following is a summary of the results for each hypothesis test:

1. Hypothesis I predicted differences in subtest scores for subjects in Group I and Group II, with subjects in Group I scoring higher on each of the subtests. The null hypothesis was not rejected

TABLE 4.8.--Comparison of Univariate Analyses of Variance for Group Effects on Subtest Scores with Undergraduate and Graduate Grade-Point Average as Covariates and without Covariates.

Variable	Hypothesis Mean Squares	F Ratio	Significance of F Ratio
Subtest 1			
With Covariates	20.09	4.63	.03358
Without Covariates	17.61	4.04	.04666
Subtest 2			
With Covariates	325.07	150.55	.00001*
Without Covariates	394.70	161.86	.00001*
Subtest 3			
With Covariates	606.01	150.18	.00001*
Without Covariates	670.07	165.10	.00001*

*Significant at the .017 level.

at the .05 level for Subtest 1 but was rejected for Subtests 2 and 3 in favor of the alternative hypothesis. Subjects in Group I did have higher mean scores than subjects in Group II for the second and third subtest. Mean score differences were 3.49 for Subtest 2 and 4.54 for Subtest 3.

2. Hypothesis II predicted a positive relationship between subtest scores and supervisor's ratings of affective skills for subjects in Group I. Results of the stepwise multiple regression analysis were not significant at the .05 level either for all subtest scores taken together in the regression equation or for each subtest score entered individually into the equation. All subtest scores

together accounted for only two percent of the variance in supervisor's ratings. The null hypothesis was not rejected for Hypothesis II.

3. Hypothesis III predicted differences in subtest scores for male and female subjects, with female subjects scoring higher on each subtest. The null hypothesis was not rejected at the .05 level of significance.

4. An examination of group and sex interaction effects was done, and results were not significant at the .05 level.

An additional analysis of the effects of intelligence, as measured by undergraduate and graduate grade-point average, was also performed. Subjects in Group I were found to have significantly higher graduate grade-point averages than subjects in Group II, but no significant differences were found for undergraduate grade-point average. Graduate grade-point average was also found to account for a small but significant amount of variance in scores for each of the three subtests, while undergraduate grade-point average accounted for a significant amount of variance only for Subtest 2. Multivariate and univariate analyses of variance using undergraduate and graduate grade-point average as covariates produced slightly lower F values for group effects, but had the same significance levels and therefore did not alter the results of the original hypothesis tests.

In Chapter V a summary of the study will be presented. The findings of the study will be discussed and conclusions presented. Limitations of the study and implications for future research will be discussed.

CHAPTER V

SUMMARY AND CONCLUSIONS

The major purpose of this study was to provide information on the Affect Recognition and Response Scale, a measure of empathic skills, including the ability to recognize emotions and the ability to respond appropriately. In this chapter a summary of the study will be presented. Conclusions based on the results of the analysis of the data will be included, as well as a discussion of the results of the study. Limitations of the present study and implications for future research, including further suggestions for validation of the scale, will be given.

Summary

The present study was based on two related areas of theory and research. The first of these is the study of emotions and the recognition of emotions, particularly in the use of facial expressions of emotion. The second of these areas is the study of empathy and the measurement of empathic ability. Although these areas are theoretically related, little has previously been done to integrate them in a research study. A discussion of these two theoretical dimensions was presented, with a focus on methodology which has been used in previous attempts at measurement. The lack of a currently available, adequately validated measure of empathic ability

was also discussed, as well as the need for such a measure in admissions programs in psychology and related areas.

The literature related to the current study was reviewed in two areas: studies related to the emotions and the recognition and labeling of facial affect and studies related to the measurement of empathic ability. The review of related research indicated that while many instruments have been used in attempts to measure empathic ability, there is currently no instrument available which has been adequately or extensively validated and none which includes the aspect of recognition of nonverbal expressions of emotion through facial affect.

There is support in the literature for the concept of discrete emotion categories. Posed stimulus situations have also proven better than spontaneous expressions for accuracy of judgments, are more easily standardized, and may help to increase comparability of research results in the future. While research has been conducted in the area of recognition and labeling of facial affect, there has been no attempt to study these abilities as an aspect of empathy or to compare high and low empathy criterion groups in their ability to accurately recognize or label emotion expressions.

The use of criterion groups as a means of determining construct validity has extensive support in past research, although the exact definition of high and low empathy groups varies from study to study. The use of criterion groups in previous research has focused on distinctions of training and experience or has used some measure of performance, such as supervisors' ratings, to

designate high and low empathy groups. In this study, therefore, the criterion groups were designated by both training and experience, and supervisors' ratings were used as an additional validity check.

Research findings on sex differences both on measures of empathic ability and in recognition and labeling of facial affect have been consistently inconclusive, showing either no differences or higher ability for women.

The Affect Recognition and Response Scale was developed for use in this study. This scale is a revised form of the Empathy Skills Rating Scale (Krupka and Parsons, 1978), which was developed for use in medical school admissions. Items for Subtest 1, a test of recognition of emotions using slides of facial expressions of affect, were taken from the Pictures of Facial Affect developed by Ekman and Friesen (1976). Items for Subtest 2 and Subtest 3, measures of the ability to respond appropriately to expressions of emotion, were taken from an item pool of six hundred items developed for use in the Empathy Skills Rating Scale. An initial form of the Affect Recognition and Response Scale was administered to a group of expert judges. Based on their responses items with low inter-judge agreement were dropped from the scale.

An analysis of the scale properties of the Affect Recognition and Response Scale was performed. This included estimates of reliability, expert judges' agreement for the final form of the scale, indices of item difficulty and discrimination, and factor analyses of the items. The overall percentage of agreement among the expert judges was found to be 94 percent. Scale reliability

was calculated using the Kuder-Richardson Formula #20, a measure of reliability which uses all the possible split-half combinations for the scale. Overall scale reliability was estimated at .853, with a range of reliability estimates from .416 for Subtest 1 to .804 for Subtest 3. Mean item difficulty for the total scale was 18.6, ranging from 13.6 for Subtest 1 to 24.6 for Subtest 3. Mean item discrimination was 31.4 for the total scale, ranging from 13.6 for Subtest 1 to 60.3 for Subtest 3.

Three separate factor analyses were performed, once for all items, with the number of factors extracted by varimax rotation determined by a minimum eigen value of 1.00, once for all items with the number of factors preset to three to correspond to the number of subtests in the scale, and once for the items in Subtest 1 (slides of facial affect), with the number of factors preset to five to correspond to the number of emotion categories used with the slides of facial affect. The factor analysis results indicated one main factor corresponding to empathic responding, regardless of the stimulus situation format. The factor structure did not correspond to the subtest structure of the scale, and the factor structure for the slides of facial affect did not correspond to the emotion categories used.

Sixty-five subjects were obtained for each of two criterion groups. Subjects in Group I were drawn from a population of graduate students majoring in counseling and clinical psychology

at Michigan State University and Central Michigan University. Subjects in Group II were drawn from a population of graduate students majoring in engineering, mathematics, and the physical sciences at Michigan State University. All subjects were paid volunteers, and no attempt was made to randomize selection from either population. Because of skewed distributions in the populations sampled, it was not possible to obtain equal numbers of female and male subjects. It was also impossible to obtain enough minority subjects to test any hypotheses about differences in scale scores due to race.

The Affect Recognition and Response Scale was administered to all subjects in small groups. Supervisors' ratings were obtained for subjects in Group I who were currently involved in clinical work, using the Supervisory Rating Scale, an eight-point scale measuring affective ability which was developed for the study. Data on sample characteristics was obtained using a Biographical Data Sheet, also developed for the study.

A multiple measures design with two crossed factors of group and sex was used to test the main hypotheses. Scores on the three subtests of the Affect Recognition and Response Scale were the dependent variables.

Statistical hypotheses were formulated to test the differences between subjects in Group I and subjects in Group II, the differences between male and female subjects, and the relationship between subtest scores and supervisors' ratings of affective skills. A two-way multivariate analysis of variance for multiple measures was used to

test for group and sex differences. Since this test was significant at the .05 level for group differences, additional univariate analyses of variance were carried out to determine which subtest scores were significantly different for the two groups. For the univariate analyses the significance level was set at .017 by distributing the .05 alpha level equally among the three subtests.

A summary of the results of the hypothesis tests is presented below:

1. Significant differences were found between groups for subtest scores for Subtest 2 and Subtest 3, but not for Subtest 1. The mean scores for Subtest 1 were 21.80 for Group I and 21.06 for Group II. The mean scores for Subtest 2 were 10.02 for Group I and 6.53 for Group II. The mean scores for Subtest 3 were 12.85 for Group I and 8.32 for Group II.

2. There was no relationship found between subtest scores and supervisors' ratings of affective skills either for all subtest scores taken together or for each subtest score taken individually. All subtest scores together accounted for only two percent of the variance in supervisors' ratings.

3. There were no significant interaction effects for group and sex.

An additional analysis of the effects of intelligence, as measured by undergraduate and graduate grade-point average, was also performed. Subjects in Group I were found to have significantly higher graduate grade-point averages than subjects in Group II, but no significant differences were found for undergraduate grade-point

average. Graduate grade-point average was also found to account for a small but significant amount of variance in scores for each of the three subtests, while undergraduate grade-point average accounted for a significant amount of variance only for Subtest 2. Multivariate and univariate analyses of variance using undergraduate and graduate grade-point average as covariates produced slightly lower F values for group effects, but had the same significance levels and therefore did not alter the results of the original hypothesis tests.

Conclusions

Assessment of the Scale

The assessment of the Affect Recognition and Response Scale was encouraging. Reliability for the total scale and for the second and third subtests was within acceptable limits, as was the percentage of agreement among expert judges on responses for scale items. However, reliability for the first subtest was quite low. The factor analysis results indicate more support for the measurement of the ability to respond, rather than the ability to recognize affect. A further refinement of the scale might include developing a more complex task for measuring recognition of facial affect, since the slides used in the current form of the scale were not found to be highly reliable and did not seem to adequately measure the emotion categories included. As will be discussed in the following section, a different measure may be necessary to discriminate actual differences in the ability to

recognize expressions of emotion between high and low empathy groups, assuming that such differences do exist between the groups.

In addition, there appears to be no difference between items using color videotape vignettes as stimulus situations and items using written stimulus situations in the ability to discriminate between the criterion groups. The factor analyses and subtest correlation matrix also support the concept of these items measuring the same construct. This finding has important implications for refinement of the test. It would be much easier to administer the scale without the need for videotape equipment. In addition, the cost of producing such a videotape is prohibitive and makes it difficult to add new items or lengthen this subtest. The written stimulus situations have many advantages in terms of ease of item construction and ease of administration. Written stimulus situations are easier to develop and add to the scale and increase only minimally the time required to administer the scale. Adding items to the third subtest would also be likely to increase the reliability of the scale. The main advantage of the color videotape is in face validity and maintaining subject interest. Many subjects commented during test administrations that the videotape was "more realistic," easier to respond to and more interesting than other parts of the scale.

Validation of the Scale

The use of criterion groups as a means of construct validation for a measure of empathic ability has a long history

in the literature. The definition of high and low empathy groups has varied from study to study, but in general groups have been designated on some aspect of training or experience. The criterion groups used in this study differed along both of these dimensions. Graduate students in Group I had all received at least one year of formal graduate training in clinical or counseling psychology and most had been graduate students for several years. Almost all of these subjects, in addition, had experience in actual clinical practice, either through a practicum, internship, job experience, or some combination of these. Graduate students in Group II, on the other hand, had no formal graduate training in clinical or counseling psychology and were in fields where they would not gain clinical experience. Thus, the assumption that subjects in Group I were higher in empathic ability than subjects in Group II, due to training and experience, seems a reasonable one. In addition, the average supervisors' rating on the measure of affective skills for subjects in Group I was 6.0, indicating more than a minimal achievement of affective skills necessary to recognize a client's affect and respond appropriately to that affect.

Subjects in Group I did significantly better on the second and third subtests of the scale than subjects in Group II, not only in terms of statistically significant differences but in terms of meaningful differences. These subtests do seem to be measuring some aspect of empathic ability. Differences in scores for the recognition of facial affect on the first subtest were not significantly different. These slides were developed and tested on a

sample of college freshmen, and only slides that could be correctly identified and agreed on by 70 percent or more of these students were included in the original Pictures of Facial Affect. Given the high percentage of correct identification for a sample of college freshmen, it is not surprising that the level of agreement for graduate students, in both groups, is similarly high. In fact, it may be more surprising that there was not an even greater agreement for some of the slides. For example, only one slide in the "pleasure" category was identified correctly by all subjects, even though previous research has shown this to be the easiest emotion to identify.

It seems likely that while the ability to accurately identify emotions is a necessary precondition for the ability to respond appropriately to an expression of emotion, this ability is quite widespread. It is also likely that the ability to identify expressions of primary emotions is a fairly universal skill. Expressions of primary emotions are generally uncommon in actual experience. The cultural overlay of display rules serves to mask such expressions. Part of the training and experience necessary for clinical work is the ability to recognize, or uncover, the actual emotion being experienced, often when a client is denying the emotion or masking its expression. The task of picking correct labels for relatively unambiguous expressions of primary emotions may have been too simplified to explain the real differences in the ability to recognize emotions between the high and low empathy groups. More study of these simpler tasks of emotion recognition

and labeling is necessary to adequately provide a baseline of data before the more complex expression of emotion in actual experience, particularly clinical experience, can be adequately studied.

The experience in the present study of trying unsuccessfully to develop slides from videotapes of actual experiences points up the problems in this area. For example, while it may be easy to study the expression of pleasure, even in a primary form, this is not the emotion most frequently encountered in clinical experience, nor most likely to present a problem for the client. The emotion of anger is a much more common clinical experience, but less likely to be expressed as a primary emotion without being masked in some way.

The second means of construct validation used with the scale was a test of predictive validity using supervisors' ratings of affective skills. Supervisors' ratings have the advantage of being behaviorally-oriented and measuring the actual use of the skills being tested by the Affect Recognition and Response Scale with clients in a clinical therapy setting. Results for this test were disappointing, however. No relationship was shown between scores for the subtests and supervisors' ratings. Since supervisors' ratings were returned for only 37 of the 65 subjects in Group I, it may be that there was some significant difference between this subset of Group I and the entire group, although this seems unlikely. The scale was able to discriminate among subjects in the criterion groups which differed in empathic skills but not among subjects within the high empathy group. It is likely that the scale is able

to discriminate at a gross level between levels of empathic ability but is not capable of making the finer distinctions between those already performing at a high level of empathic ability.

Another difficulty inherent in the test of predictive validity was the lack of reliability or validity data for the Supervisory Rating Scale. Although the rating scale is similar to those actually used to rate performance in clinical settings, there is no objective data available with which to judge it. Since the ultimate value of a measure of empathic ability is how well it can predict the actual use of empathic skills in clinical settings, additional validity tests of predictive power are certainly needed. This will be discussed further in the section on implications for future research.

Differences Between Men and Women

The question of differences between men and women in both the ability to recognize and label emotions and the ability to respond empathically has been included in the majority of studies in these areas. There is consistent research support for the concept of no differences between men and women, as well as another large group of studies which reports consistently better performance for women. The research hypothesis used in this study was that there would be no differences between men and women in the abilities measured by the Affect Recognition and Response Scale. Results of the study were consistent with this prediction. The number of women, unfortunately, was less than the number of men

for both groups. In addition, the differences in the number of men and women were not proportional for each group. Thus, the number of men and women were nearly equal in Group I, but only 22 percent of the subjects in Group II were women. While the analysis used was able to adjust for the unequal cell sizes, it would have been preferable to have a larger proportion of women. The proportion of women in the samples, does, however, reflect the proportion of women in the populations from which the sample were drawn. Thus, including more women in the samples might have biased the generalizability of group differences to the populations involved. For an instrument to be used in actual admissions decisions, however, the question of whether or not men and women score differently must be answered conclusively. It may be that with the changing of sex role stereotypes the findings of no differences between men and women in the area of affective skills will become consistent over time and from study to study.

Interaction Effects Between Group and Sex

Although little was actually known about differences between men and women in Group I and men and women in Group II, it was possible that there would be an interaction effect for these two variables, such that the sex differences for each group would cancel each other out. As was seen in Chapter IV, there were small differences in scores on each subtest for men and women in each group, but none of these differences were statistically significant. Again, it may be that these results were affected by the smaller number of

women in Group II, but the findings are consistent with a large body of other research findings.

Intelligence Effects

The effects of intelligence, as measured by undergraduate and graduate grade-point average, did not significantly affect differences between the two criterion groups. Given that both groups were from populations of graduate students, this result is not surprising. The fact that grade-point average did account for a small percentage of the variance in subtest scores, even in a group of graduate students who should be fairly homogeneous on this dimension, suggests that differences due to intelligence or educational level should be explored further and that caution would certainly be needed in comparing scores for groups with different levels of education. The development of norms for groups of graduate students, college undergraduate students, or high school students would prove useful in comparing persons from different educational backgrounds.

Future Use of the Scale

The results of the present study indicate that the Affect Recognition and Response Scale is a reliable measure, with some validation support for its further use. Some changes in format would be desirable, particularly in either eliminating the slides of facial affect or in refining this task to make it able to discriminate finer distinctions in recognition ability. Further validation of the scale is necessary before it can be used for

actual selection procedures. Norms for groups differing in educational level need to be established, as well as normative data on differences between men and women and between minorities and nonminorities on the scale.

The scale may prove more useful for gross discriminations than for prediction of performance among students already screened for empathic ability. For example, the scale might more accurately discriminate between those accepted into a graduate psychology program and those rejected, or between those to be admitted to a training program and those to be rejected. The scale may also prove useful as a measure of training outcome for those in crisis intervention or other interpersonal skills training programs at the paraprofessional, rather than the graduate or professional, level.

Limitations of the Study

There are several limitations to the study in terms of providing validation data for the Affect Response Scale. The use of criterion groups for construct validation has been discussed, and the most important limitation remains the necessity of assuming that the two criterion groups chosen do in fact differ in the expected direction on the dimension of empathic ability. While this assumption seems to be a reasonable one, based on the differences between the groups in terms of training and clinical experience, it is nevertheless an assumption, rather than an observable or measurable difference.

actual selection procedures. Norms for groups differing in educational level need to be established, as well as normative data on differences between men and women and between minorities and nonminorities on the scale.

The scale may prove more useful for gross discriminations than for prediction of performance among students already screened for empathic ability. For example, the scale might more accurately discriminate between those accepted into a graduate psychology program and those rejected, or between those to be admitted to a training program and those to be rejected. The scale may also prove useful as a measure of training outcome for those in crisis intervention or other interpersonal skills training programs at the paraprofessional, rather than the graduate or professional, level.

Limitations of the Study

There are several limitations to the study in terms of providing validation data for the Affect Response Scale. The use of criterion groups for construct validation has been discussed, and the most important limitation remains the necessity of assuming that the two criterion groups chosen do in fact differ in the expected direction on the dimension of empathic ability. While this assumption seems to be a reasonable one, based on the differences between the groups in terms of training and clinical experience, it is nevertheless an assumption, rather than an observable or measurable difference.

The use of volunteer subjects, rather than random sampling from the populations involved, is another limitation of the study. Particularly in the measurement of empathic ability, it would be useful to know if those who volunteered differed significantly from those who did not. However, it seems a reasonable assumption that those who volunteered would be more likely to be higher in empathy than those who did not. This would more likely have affected the sampling for Group II, since the sample for Group I represented a large proportion of the actual population. In addition, subjects in Group I who did not volunteer mentioned conflicts of time or scheduling almost entirely, while subjects in Group II who did not volunteer were more likely to mention reasons of not wanting to participate in a study or not wanting to participate in a "psychological" study.

The lack of minority subjects and the smaller number of women, particularly in Group II, has been noted. While increasing the number of minority and female subjects would have increased the power of the statistical tests of the hypothesis on sex differences and provided data on race differences, it would at the same time have made the samples unrepresentative of their respective populations. If the numbers of both minority and female students admitted into graduate programs increase, it will be possible to more adequately study the effects of sex and race without distorting the characteristics of the populations being sampled.

The lack of validity and reliability data on the Supervisory Rating Scale is another limitation. It seems unlikely, given the

small amount of variance in supervisors' ratings accounted for by the subtest scores, that increasing either the reliability or the validity of the Supervisory Rating Scale would have had a significant effect on the outcome of the regression analysis. It seems more probable that the analysis was reflecting a true lack of relationship between the subtest scores and the supervisors' ratings, due to the inability of the scale to make finer discriminations among those already chosen as being high in empathic ability.

Implications for Future Research

The process of validating an instrument for use in making admissions decisions is a long and complex one. While this study provides some initial data on the scale properties of the Affect Recognition and Response Scale and some initial positive validation findings, much more validation will need to be done before the scale can be used as part of an actual decision-making process for admissions. Further validation can be done following the methodology used in this study, using different criterion groups and different performance measures. Possible criterion groups include counseling staff and professionals not engaged in counseling, experienced counseling staff and beginning graduate students in psychology programs, or high and low empathy subjects as designated by peers or supervisors.

Other performance measures might include peer ratings of empathic ability, client's ratings, or ratings of actual therapy sessions by trained judges. Since the initial findings of group

differences were significant, it would be useful to begin to examine the scale for use as an outcome measure in training programs, particularly paraprofessional programs such as crisis intervention training programs.

Further research is particularly needed in the study of the relationship between recognition of facial affect, and the recognition of emotions in general, and the ability to respond empathically. Initial results from this study suggest that the recognition of expressions of primary emotions may be a necessary but not sufficient condition for the ability to respond empathically, and that the ability to recognize emotions for the highly empathic person is more complex than can be measured by a simple labeling task. The area of display rules, particularly, has not been studied in depth but may well hold the key to the more complex recognition tasks used in the clinical therapy session.

The results of the present study suggest that it is possible to develop an instrument to measure at least some aspects of empathic ability, although caution must be used in interpreting the results of the study since no differences were found in the ability to recognize expressions of emotion. The task is complex and will require more extensive validation studies.

APPENDICES

APPENDIX A

SAMPLE TEST PACKET

APPENDIX A

SAMPLE TEST PACKET

A copy of the introduction, instructions, and sample items for the Affect Recognition and Response Scale are provided on the following pages. Due to copyright restrictions on the Pictures of Facial Affect and the Empathy Skills Rating Scale, the Affect Recognition and Response Scale is not reproduced here. The Pictures of Facial Affect are available from:

Consulting Psychologists Press, Inc.
577 College Avenue
Palo Alto, California 94306

The Empathy Skills Rating Scale is available from:

Marketing Dept.
Instructional Media Center
Michigan State University
East Lansing, Michigan 48824

PARTICIPANT CONSENT FORM

1. I have freely consented to take part in a scientific study being conducted by Margaret Parsons as part of the research requirement for her doctoral dissertation under the supervision of Dr. William Hinds, Department of Counseling and Personnel Services, Michigan State University.

2. The study has been explained to me, and I understand the explanation that has been given and what my participation will involve.

3. I understand that I am free to discontinue my participation in the study at any time without penalty.

4. I understand that the results of the study will be treated in strict confidence and that I will remain anonymous. Within these restrictions, results of the study will be made available to me at my request.

5. I understand that my participation in the study does not guarantee any beneficial results to me.

6. I understand that, at my request, I can receive additional explanation of the study after my participation is completed.

Signed _____

Date _____

AFFECT RECOGNITION AND RESPONSE SCALE

INTRODUCTION

In taking this test you will be asked to respond to situations requiring that you recognize another person's emotional state or feeling, even if he or she has not precisely identified it for you. There are three subtests in the scale, providing a number of different situations in which you can identify and respond to emotions.

AFFECT RECOGNITION AND RESPONSE SCALE

SUBTEST 1

INSTRUCTIONS

The person in each of the following slides is expressing a certain emotion. Look at each slide as it is presented and decide how the person is feeling. You may decide that more than one emotion is present. Choose the clearest and strongest feeling that the person in the slide is expressing.

Select one emotion from those provided in the Table of Emotions. Note that there are four feelings listed under each major emotion category. Indicate your choice on the answer sheet by marking the letter which corresponds to the major emotion category which you have chosen. Mark your answers for this subtest in numbers 1-25 on the answer sheet.

Each slide will be presented for thirty seconds.

TABLE OF EMOTIONS

A	B	C
ANGER	PLEASURE	DISTRESS
Annoyed Irritated Hostile Enraged	Contented Happy Glad Joyful	Sad Unhappy Anguished Depressed
D		E
DISGUST		SURPRISE
Dislike Scorn Contempt Revulsion		Startled Astonished Alarmed Shocked

AFFECT RECOGNITION AND RESPONSE SCALE

SUBTEST 2

INSTRUCTIONS

In the next few minutes you will be viewing a series of short vignettes. As you observe the vignettes, assume that the person speaking is speaking directly to you and has come to you for help. Select a response from the ones provided which is closest to the one you would make. Indicate your choice on the answer sheet by marking the corresponding letter. Mark your answers for this subtest in numbers 26-40 on the answer sheet.

You will have one minute to respond after each vignette.

Subtest 2

Sample Item

Videotape Script: "My employer thinks that just because I never finished high school, I'm stupid. He has a lot of nerve talking down to me like that."

VIGNETTE A

1. What would you say to this person?

   a. It makes you mad when someone doesn't judge you on your merits, but on their prejudices.

   b. That's really unfair. You shouldn't let him get away with that.

   c. Don't let it get to you. If you do your best, it'll pay off.

   d. What does he do that makes you feel stupid?

(Sample item written by Dr. Judith Krupka)

AFFECT RECOGNITION AND RESPONSE SCALE

SUBTEST 3

INSTRUCTIONS

Two types of items comprise this subtest:

1. You are asked to assume that an individual has come to you asking for your help. You will read a brief statement which the individual is making to you.

2. You are asked to imagine that you are an individual in a situation which is briefly described.

Select a response from the ones provided which is closest to the one you would make. Indicate your choice on the answer sheet by marking the corresponding letter. Mark your answers for this subtest in numbers 41-54 on the answer sheet.

Subtest 3

Sample Item

A. My children want to put me in an old age home because they don't think that I can take care of myself. I don't want to go into a home. I can take care of myself.

What would you say to this person?

a. It's good to see someone so independent at your age!

b. You know, that's really their way of saying they love you. You shouldn't feel bad about that.

c. It's hard for someone who's as independent as you to have to think of having someone else taking over your life.

d. Would you feel better if they were asking you to come live with them, instead of putting you in the home?

(Sample item written by Dr. Judith Krupka)

Code Number: _____

BIOGRAPHICAL DATA SHEET

AGE _____ CURRENT LEVEL: ___M.A.___Ph.D.

SEX _____ UNDERGRAD. G.P.A. _____

NUMBER OF CHILDREN _____ GRADUATE G.P.A. _____

ETHNIC ORIGIN _____ MAJOR _____

How many courses in psychology have you taken?

___0-5

___6-10

___11-15

___16+

In which of the following areas?

___General ___Personality

___Experimental ___Clinical

___Industrial ___Social

___Other

Comments about the test: _____

_____

_____

_____

_____

SUPERVISORY RATING FORM

As part of the research for my dissertation, I am investigating the relationship between the ability to recognize and respond to affect, as measured by a test which I have developed, and supervisors' ratings of affective skills. Your supervisee has already taken the Affect Recognition and Response Scale. Your cooperation in completing the Supervisory Rating Form and returning it to me in the attached envelope will be greatly appreciated.

Thank you,

Margaret Parsons

Margaret Parsons
Graduate Student
Dept. of Counseling Psych.

SUPERVISORY RATING FORM

AFFECTIVE SKILLS

8	7	6	5	4	3	2	1
Demonstrates mastery of this behavior		Demonstrates minimal achievement of these behaviors.		Is actively working to achieve these behaviors.		Demonstrates ineffective use of these behaviors or has shown no improvement.	

Empathic. Accurately recognizes client's affect. Responds to client's affect without avoidance. Recognizes and responds to both surface expressions and deeper levels of affect.

Judgmental or indifferent. Unable to accurately label client's feelings. Avoids responding to client's affect. Unable to recognize or respond to even the client's obvious expressed surface feelings.

Code Number:_____

APPENDIX B

PERCENT OF AGREEMENT FOR PICTURES OF FACIAL AFFECT

TABLE B.1.--Percent of Judgments of Each Emotion for Each Photograph.

Photo No. (Ekman)	Item No. (Subtest 1)	Emotion Category					
		Happiness (Pleasure)	Sadness (Distress)	Fear*	Anger	Surprise	Disgust
7	8	99	0	0	1	0	0
35	7	97	0	0	0	3	0
66	11	92	0	4	0	4	0
84	1	96	0	0	0	4	0
100	24	97	0	0	0	3	0
15	16	0	97	0	0	3	0
23	6	0	71	3	13	0	13
58	5	0	87	4	0	0	8
75	14	0	74	16	3	0	6
103	2	0	88	0	4	0	8
10	3	3	0	0	74	3	19
25	23	0	0	4	70	0	26
52	19	0	3	3	84	6	3
89	13	0	0	0	79	0	21
106	15	0	0	2	96	0	2
19	18	3	0	0	0	91	3
39	25	0	0	3	0	97	0
70	4	0	0	16	0	81	3
81	9	0	0	23	0	74	3
197	17	0	0	9	0	91	0
27	12	0	3	0	13	0	84
40	10	0	12	0	0	0	88
71	22	0	0	0	17	0	83
98	20	0	0	0	6	0	94
109	21	0	0	0	20	0	80

*Fear category was eliminated from Subtest 1.

APPENDIX C

INTER-ITEM CORRELATION MATRIX

TABLE C.1.--Inter-Item Correlation Matrix.

	S1	S2	S3	S4	S5	S6	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	T26	T28	T29	T30	T31	T32	T33	T34	T37	T38	T40	W41	W42	W43	W44	W45	W46	W47	W48	W49	W50	W51	W52	W53	W54
S 1	1.00																																																
S 2	-.10	1.00																																															
S 3	-.06	-.16	1.00																																														
S 4	-.02	.06	.00	1.00																																													
S 5	-.01	-.07	.11	.36	1.00																																											
S 6	.13	-.06	.03	-.12	-.10	1.00																																										
S 8	-.01	-.14	.05	-.03	-.02	.05	1.00																																									
S 9	-.02	.10	.10	-.07	-.04	.04	-.03	1.00																																									
S10	-.02	-.03	.10	-.07	-.04	.24	-.03	.20	1.00																																								
S11	.26	-.12	.10	-.09	.12	.11	-.04	-.09	.34	1.00																																								
S12	.12	-.03	.02	-.07	.00	-.06	.17	-.12	.02	.03	1.00																																						
S13	-.03	-.18	.02	-.08	-.05	.04	.19	.15	.27	.08	.07	1.00																																					
S14	-.04	.03	-.02	-.03	.08	-.15	-.05	-.11	-.11	-.07	.19	-.05	1.00																																				
S15	-.03	-.07	.06	.00	.10	-.09	-.05	.01	.01	-.04	-.06	.15	.11	1.00																																				
S16	.24	-.02	.16	.00	-.06	.18	-.04	.01	.11	.20	.10	-.02	.18	.09	1.00																																			
S17	-.01	.14	.00	-.04	-.02	.01	-.02	.17	.17	-.05	-.11	-.05	-.07	-.06	.10	1.00																																		
S18	-.02	.03	-.11	-.07	-.04	.04	-.03	.07	.07	-.08	.02	.04	.07	.21	.01	.17	1.00																																		
S19	.24	-.02	.06	.00	-.06	.12	.15	.11	.11	.12	.04	.24	-.16	.09	.09	.10	.11	1.00																																		
S20	-.02	.16	-.03	-.05	-.03	-.02	-.02	.14	-.05	.09	.15	.10	-.08	-.06	.07	-.03	-.05	.07	1.00																																	
S21	-.04	.01	.22	.14	.08	-.04	-.05	-.02	.07	.15	-.07	.11	-.06	.19	.12	.08	.07	.12	.05	1.00																															
S22	-.04	.00	.19	-.12	-.07	.11	-.06	.15	.15	.13	.17	.17	-.07	.10	.17	-.07	-.03	.23	.04	.23	1.00																													
S23	.11	-.23	.13	-.16	.09	.02	.03	.06	-.01	.10	.06	.10	.05	.16	.21	-.02	.12	.16	.13	-.11	.04	1.00																												
S24	.70	-.02	-.09	-.03	-.02	.05	-.02	-.03	-.03	.17	.04	-.04	.12	-.04	.15	-.02	-.03	.15	-.02	-.05	-.06	.03	1.00																									
S25	-.01	-.10	.13	-.02	-.01	-.06	-.01	-.02	-.02	-.03	.12	-.03	-.04	.24	-.03	-.01	-.02	-.03	-.02	-.04	.20	.11	-.01	1.00																								
T26	-.06	-.09	.07	.06	.03	.02	-.07	.00	.08	.05	-.02	.02	-.14	-.03	.02	-.09	-.07	-.09	.00	.02	-.05	.04	-.07	-.05	1.00																								
T28	.14	-.20	.14	-.11	-.10	.24	.17	.12	.19	.24	.12	.11	-.08	.14	.14	-.10	-.02	.25	.08	.12	.18	.09	.06	-.06	.16	1.00																							
T29	-.06	-.11	.12	-.07	-.11	.14	.11	.08	.22	.13	.16	.12	-.09	.19	.14	-.11	.01	.24	.06	-.03	.16	.08	-.09	.12	.12	.54	1.00																						
T30	-.04	-.02	-.01	-.04	-.07	.19	.09	.05	.14	.12	.06	.01	.09	.22	.16	-.07	-.03	.03	.16	-.13	.08	.10	-.06	-.04	.13	.35	.45	1.00																					
T31	-.05	-.14	.12	.00	-.08	.13	.16	.02	.10	.07	-.02	-.03	-.07	.16	.04	-.08	.18	.04	.01	.10	.07	.12	-.07	-.05	.23	.16	.44	.21	1.00																				
T32	.25	-.25	.08	-.10	-.05	.14	-.05	-.09	-.09	.13	.27	.07	-.01	.18	.26	-.05	.01	.03	.08	-.07	.05	.18	.16	.25	.04	.33	.32	.17	.18	1.00																			
T33	-.04	.08	-.04	.15	-.06	.02	.08	.08	-.01	.01	.04	-.12	.01	.06	.06	-.06	.17	.06	-.07	.02	-.06	.00	-.05	-.04	.19	.19	.27	.11	.28	.08	1.00																		
T34	-.05	-.13	.04	.06	.03	.05	.16	.00	.08	-.01	.02	.02	.11	.08	.02	-.09	.08	.02	-.10	-.03	.00	.04	-.07	-.05	-.06	.08	.16	-.06	.06	-.02	.09	1.00																
T37	.11	-.19	.11	.09	-.02	.25	-.06	.12	.19	.26	.05	.09	-.08	.05	.20	-.12	.12	.15	.13	.11	.12	.07	.03	-.07	.21	.61	.57	.35	.46	.28	.36	.10	1.00															
T38	-.04	-.04	.05	.04	.07	.02	-.06	.15	-.11	.06	.08	-.13	-.01	.10	.10	-.07	.15	-.03	.04	.00	.20	.17	-.06	.20	.10	.32	.29	.19	.33	.32	.31	.05	.42	1.00														
T40	-.05	-.05	.01	-.03	.02	.07	-.08	.05	.05	.19	.17	.00	-.08	.26	.04	-.10	.13	.04	.09	.03	.19	.25	-.08	.14	.14	.41	.52	.36	.47	.39	.30	.10	.56	.52	1.00													
W41	.28	-.10	.00	.02	-.05	.19	.18	.03	.14	.16	.11	.09	-.06	-.03	.13	-.05	.14	.30	-.06	-.06	.08	.18	.18	-.03	.13	.27	.27	.07	.28	.32	.18	.07	.34	.22	.28	1.00												
W42	-.05	-.15	.03	-.02	-.09	.17	.21	.14	-.01	.03	.05	.01	-.06	.17	.00	-.09	.14	.17	.00	.00	.07	.17	-.08	-.05	.12	.40	.41	.38	.24	.30	.11	.12	.38	.31	.37	.29	1.00											
W43	-.04	-.12	.15	.03	-.07	.04	-.06	.05	.13	.11	.01	.08	-.14	.15	.02	-.07	-.04	.15	.03	-.02	.12	.21	-.06	.19	.16	.24	.47	.16	.30	.23	.22	.12	.41	.34	.43	.20	.27	1.00									
W44	-.02	-.02	.07	-.05	-.03	-.02	-.02	.14	-.05	.09	.15	-.05	.17	.21	.07	-.03	.32	-.06	-.03	.18	.04	-.05	-.02	-.02	.11	.08	.06	.04	.34	.08	.32	.11	.22	.16	.19	.25	.10	-.08	1.00							
W45	.10	-.13	.09	.02	.08	.16	.02	.24	.18	.19	.20	.08	.03	.19	.14	-.02	.11	.24	.12	.05	.23	.21	.02	.10	.23	.48	.48	.33	.29	.32	.44	.08	.55	.36	.50	.38	.36	.31	.21	1.00						
W46	-.03	.06	.12	-.02	-.06	.08	.14	.09	-.10	.02	.06	.13	-.10	.07	-.07	-.06	.09	.07	.19	.03	.08	.02	-.05	-.03	.16	.26	.34	.25	.24	.16	.11	.10	.34	.20	.42	.27	.24	.12	.19	.37	1.00					
W47	.15	-.22	.15	.11	.02	.21	.07	.14	.07	.21	.05	.01	-.06	.00	.11	.02	.06	.22	.10	.15	.26	.10	.07	.15	.12	.40	.45	.29	.28	.30	.22	.12	.60	.45	.45	.23	.32	.36	.10	.47	.39	1.00				
W48	-.04	-.24	.17	-.07	-.08	.19	.24	.02	.10	.14	.15	.12	-.06	.11	.11	-.08	.02	.11	-.10	.00	.19	.10	-.06	-.04	.24	.39	.42	.27	.42	.37	.29	.02	.56	.34	.57	.36	.30	.36	.24	.46	.36	.34	1.00			
W49	.10	-.14	.01	.01	.08	.08	.15	.04	.04	.18	.12	.02	-.02	.13	.04	-.13	-.02	.18	.12	.05	.22	.16	.02	.10	.11	.40	.34	.32	.28	.31	.11	.15	.44	.31	.45	.27	.24	.30	.03	.50	.32	.38	.41	1.00			
W50	-.04	-.08	.19	-.13	.06	.00	.10	-.04	.05	.18	.18	.15	.02	.08	.08	-.07	.13	.02	.15	.04	.07	.17	-.06	-.04	.16	.15	.08	.11	.15	.16	.10	.02	.29	.01	.21	.13	.04	.10	.38	.35	.30	.09	.41	.30	1.00		
W51	-.05	.01	-.03	.12	.03	.07	.07	.14	.14	-.08	.06	.08	-.05	.06	.12	.14	.29	.12	.10	.11	.27	.23	-.07	-.05	.13	.14	.13	.25	.34	.08	.18	-.04	.29	.27	.27	.12	.13	.14	.10	.30	.09	.29	.27	.29	.28	1.00	
W52	.14	-.06	.06	-.10	.02	.16	-.08	.20	-.01	.03	.08	-.12	-.02	.16	.10	.02	.20	.10	-.01	-.01	.16	.16	.06	.14	.10	.38	.43	.32	.27	.29	.16	.06	.36	.58	.55	.22	.31	.22	.19	.41	.33	.43	.37	.30	.03	.24	1.00	
W53	.08	-.01	-.18	.05	.03	-.04	-.02	.09	-.04	.02	.01	-.13	-.03	-.04	.06	-.08	.09	-.04	.06	-.18	.03	.09	-.02	-.10	.18	.09	.06	.09	.18	-.01	.25	-.04	.20	.12	.18	.16	.15	.07	.06	.23	.05	-.06	.20	.24	.19	.17	.13	1.00	
W54	-.03	.00	-.06	.08	-.06	-.05	-.05	.10	.10	.03	.03	.14	-.03	-.06	.01	.10	-.10	.08	-.07	.04	.09	-.06	-.05	-.03	.06	.28	.12	.08	-.03	.10	.12	.01	.13	.15	.08	.20	.15	-.05	.07	.16	.13	.15	.15	.20	.07	.10	.09	-.06	1.00

S = Slides (Subtest 1)
T = Tape (Subtest 2)
W = Written (Subtest 3)

APPENDIX D

ITEM-TOTAL RELIABILITY DATA

TABLE D.1.--Reliability Analysis for Total Scale.

Item	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Alpha If Item Deleted
S1	39.70	40.35	.09	.853
S2	40.26	41.52	-.20	.864
S3	40.02	39.39	.14	.855
S4	39.76	40.48	-.02	.855
S5	3.71	40.46	-.01	.854
S6	40.00	39.17	.18	.854
S7	39.69	40.46	0	.854
S8	39.70	40.30	.09	.853
S9	39.75	39.89	.17	.853
S10	39.75	39.83	.19	.852
S11	39.80	39.49	.23	.852
S12	40.03	39.23	.17	.854
S13	39.78	40.02	.10	.854
S14	39.84	40.63	-.06	.858
S15	39.81	39.48	.22	.852
S16	39.81	39.38	.24	.852
S17	39.71	40.58	-.07	.855
S18	39.75	39.91	.16	.853
S19	39.81	39.33	.26	.852
S20	39.72	40.17	.12	.853
S21	39.84	39.93	.09	.855
S22	39.85	39.00	.29	.851
S23	40.08	38.78	.24	.853
S24	39.70	40.46	.00	.854
S25	39.70	40.38	.06	.854
T26	39.93	39.19	.20	.853
T28	39.98	37.02	.58	.855
T29	40.04	36.66	.62	.843
T30	39.86	38.32	.43	.848
T31	39.91	37.90	.47	.857
T32	39.80	38.77	.41	.850
T33	39.83	38.83	.35	.850
T34	39.93	39.71	.11	.855
T37	40.08	36.00	.72	.840
T38	39.85	38.03	.50	.847
T40	39.97	36.58	.68	.842
W41	39.78	38.76	.44	.848
W42	39.95	37.79	.46	.847
W43	39.87	38.29	.43	.848
W44	39.72	39.81	.28	.852
W45	40.10	35.94	.72	.840
W46	39.82	38.54	.43	.848
W47	39.95	37.11	.59	.844
W48	39.90	37.23	.62	.844
W49	40.11	36.88	.56	.840
W50	39.87	38.69	.34	.850
W51	39.94	38.08	.40	.848
W52	39.96	37.38	.53	.845
W53	40.27	39.15	.17	.854
W54	39.81	39.65	.17	.853

Scale	Mean	Variance	Alpha
Total	40.69	40.46	.853

S = Slides
T = Tape
W = Written

TABLE D.2.--Reliability Analysis for Subtest 1 (Slides of Facial Affect).

Item	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Alpha If Item Deleted
1	20.70	3.92	.16	.408
2	21.26	4.04	-.15	.489
3	21.02	3.47	.16	.391
4	20.76	4.03	-.11	.441
5	20.71	3.92	.06	.414
6	21.00	3.67	.05	.425
7	20.69	3.98	0	.417
8	20.70	3.96	.01	.418
9	20.75	3.81	.12	.405
10	20.75	3.69	.25	.384
11	20.79	3.67	.19	.389
12	21.03	3.58	.10	.412
13	20.78	3.71	.18	.393
14	20.84	3.88	-.02	.436
15	20.81	3.69	.15	.396
16	20.81	3.45	.35	.353
17	20.71	3.92	.06	.414
18	20.75	3.83	.10	.408
19	20.81	3.52	.30	.365
20	20.72	3.86	.13	.405
21	20.84	3.59	.20	.386
22	20.85	3.42	.31	.357
23	21.08	3.48	.14	.398
24	20.70	3.93	.08	.412
25	20.70	3.95	.08	.414

Subtest	Mean	Variance	Alpha
Total	21.69	3.98	.416

TABLE D.3.--Reliability Analysis for Subtest 2 (Videotape Vignettes).

Item	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Alpha If Item Deleted
26	7.65	6.37	.21	.809
28	7.70	5.59	.57	.771
29	7.76	5.31	.67	.758
30	7.58	6.12	.40	.789
31	7.63	5.84	.50	.779
32	7.52	6.34	.36	.793
33	7.55	6.23	.38	.791
34	7.65	6.64	.09	.821
37	7.81	5.17	.72	.751
38	7.57	5.98	.49	.781
40	7.69	5.40	.67	.759

Subtest	Mean	Variance	Alpha
Total	8.41	7.01	.799

TABLE D.4.--Reliability Analysis for Subtest 3 (Written Stimulus Situations).

Item	Scale Mean If Item Deleted	Scale Variance If Item Deleted	Corrected Item-Total Correlation	Alpha If Item Deleted
41	9.68	8.33	.44	.793
42	9.85	7.91	.42	.792
43	9.77	8.21	.36	.797
44	9.62	8.86	.25	.804
45	10.00	7.11	.68	.767
46	9.72	8.16	.45	.791
47	9.85	7.68	.52	.784
48	9.80	7.58	.62	.776
49	10.01	7.35	.58	.778
50	9.77	8.20	.37	.796
51	9.84	8.02	.38	.796
52	9.86	7.76	.48	.787
53	10.17	8.27	.22	.812
54	9.71	8.67	.19	.807

Subtest	Mean	Variance	Alpha
Total	10.59	9.15	.804

APPENDIX E

ITEM ANALYSIS DATA

TABLE E.1.--Percent of Agreement for Slides of Facial Affect for All Subjects.

Anger	Pleasure	Distress	Disgust	Surprise
(3)* 67.4	(1) 99.2	(2) 43.4	(10) 93.8	(4) 93.0
(13) 91.5	(7) 100.0	(5) 97.7	(12) 65.9	(9) 93.8
(15) 88.4	(8) 98.4	(6) 69.0	(20) 96.9	(17) 97.7
(19) 88.4	(11) 89.9	(14) 84.5	(21) 85.3	(18) 93.8
(23) 61.2	(24) 98.4	(16) 88.4	(22) 83.7	(25) 99.2

*Slide Number

TABLE E.2.--Raw Score Distribution for Total Scale.

Raw Score	Frequency	Cumulative Frequency	Percentile Rank
49	5	5	98
48	9	14	93
47	13	27	84
46	11	38	75
45	10	48	67
44	10	58	59
43	4	62	54
42	6	68	50
41	5	73	46
40	7	80	41
39	4	84	37
38	7	91	33
37	3	94	29
36	3	97	27
35	7	104	24
34	3	107	19
33	5	112	16
32	4	116	12
31	6	122	8
29	2	124	5
27	1	125	4
26	1	126	3
25	2	128	2
24	1	129	1
22	1	130	0

TABLE E.3.--Item Difficulty and Discrimination Indices.

Item	Item Difficulty	Item Discrimination
S1	2	3
S2	55	-11
S3	35	32
S4	7	9
S5	3	3
S6	31	26
S7	0	0
S8	2	3
S9	6	11
S10	6	11
S11	11	20
S12	35	28
S13	9	17
S14	15	0
S15	12	23
S16	12	26
S17	2	0
S18	7	11
S19	12	29
S20	4	9
S21	15	23
S22	17	31
S23	39	31
S24	2	0
S25	2	6
T26	25	34
T28	28	66
T29	36	77
T30	18	46
T31	22	46
T32	12	34
T33	14	34
T34	24	26
T37	42	88
T38	16	43
T40	28	74
W41	10	34
W42	26	51
W43	18	43
W44	3	11
W45	41	91
W46	14	43
W47	26	66
W48	22	60
W49	42	80
W50	18	31
W51	27	48
W52	28	63
W53	58	25
W54	12	17

S = Slides
T = Tape
W = Written

TABLE E.4.--Item Response Summary Data.

Item Difficulty	Number of Items	Percentage	Item Discrimination	Number of Items	Percentage
91 - 100	0	0	91 - 100	1	2
81 - 90	0	0	81 - 90	1	2
71 - 80	0	0	71 - 80	3	6
61 - 70	0	0	61 - 70	3	6
51 - 60	2	4	51 - 60	2	4
41 - 50	3	6	41 - 50	6	12
31 - 40	5	10	31 - 40	8	16
21 - 30	10	20	21 - 30	8	16
11 - 20	15	30	11 - 20	7	14
00 - 10	15	30	00 - 10	10	20
			Less than 00	1	2

APPENDIX F

ITEM-FACTOR CORRELATIONS

TABLE F-1.--Item-Factor Correlations for All Items with No Preset Factors.

Item	Factor									
	1	2	3	4	5	6	7	8	9	10
S1	.03	.95	-.02	.01	-.02	-.01	-.04	.00	.00	.07
S2	-.16	-.06	-.08	-.05	.19	-.16	-.16	.05	-.03	-.37
S3	.10	-.08	.10	.06	-.07	.00	.48	.07	.02	.17
S4	.01	-.02	-.13	-.07	-.02	.04	.05	-.05	.65	-.16
S5	-.04	.00	.07	.02	-.04	-.08	.06	-.03	.59	.13
S6	.25	.14	-.08	.19	.05	.08	.08	.04	-.19	.01
S8	.09	-.02	.04	-.04	-.02	.67	-.02	-.03	-.02	.02
S9	.13	-.06	-.03	.08	.32	-.12	.00	.11	-.07	.07
S10	.07	-.02	-.01	.98	.18	.01	.04	.06	-.05	-.01
S11	.16	.29	.16	.35	-.19	-.10	.19	.04	.03	.07
S12	.09	.08	.18	.02	-.09	.21	-.07	.12	-.06	.00
S13	-.03	-.02	.09	.24	.01	.33	.12	.10	-.09	.13
S14	-.09	.00	.05	-.06	-.02	-.04	-.05	-.03	.06	.02
S15	.17	-.04	.05	-.02	.06	-.04	.09	.02	.05	.08
S16	.12	.24	-.03	.11	.05	-.05	.25	.08	-.08	.20
S17	-.11	.00	-.10	.11	.44	-.02	.08	-.06	-.02	-.01
S18	.08	.01	.18	-.04	.55	.03	-.06	-.01	-.05	.02
S19	.13	.29	-.13	.06	.16	.28	.10	.25	-.05	.14
S20	.08	.00	.06	-.02	.00	.01	.03	.02	-.02	.06
S21	.00	.01	.09	.02	.10	-.04	.52	.19	.16	-.15
S22	.18	-.03	.02	.08	.00	-.02	.18	.88	-.10	-.01
S23	.15	.07	.06	.00	.12	.02	-.03	.01	.00	.65
S24	-.03	.70	-.04	.00	-.02	-.01	-.05	-.03	-.01	-.02
S25	.07	-.02	-.08	-.02	-.03	-.05	.06	.09	.00	.07
T26	.22	-.08	.22	.07	-.11	-.10	-.01	-.08	.06	.06
T28	.62	.14	.01	.15	-.14	-.08	.15	.07	-.15	.11
T29	.71	-.08	-.10	.18	-.11	.17	.04	.00	-.17	-.01
T30	.56	-.09	-.08	.14	-.06	.06	-.14	-.02	-.06	-.01
T31	.50	-.06	.24	.03	.10	.07	.05	.00	-.03	-.02
T32	.43	.25	.09	-.08	-.10	.19	.08	-.13	-.14	.18
T33	.33	-.03	.15	-.04	.10	-.13	-.10	-.05	.06	-.09
T34	.12	-.05	.00	.04	.00	.08	.03	-.01	.07	.04
T37	.75	.10	.16	.14	-.04	.09	.15	.00	.06	.02
T38	.61	-.05	.01	-.19	.16	-.23	.00	.09	.09	.08
T40	.74	-.05	.20	.02	.01	-.11	-.08	.04	.01	.03
W41	.36	.31	.20	.05	.10	.21	-.07	.02	-.07	.09
W42	.54	-.05	-.06	-.09	.04	.16	-.05	-.02	-.11	.14
W43	.47	-.08	-.08	.13	-.12	-.01	.04	.02	.00	.22
W44	.15	.02	.59	-.10	.26	-.09	.16	-.03	-.09	-.20
W45	.61	.10	.25	.13	.07	.00	-.05	.13	.10	.16
W46	.48	-.02	.30	-.14	.04	.14	.04	-.04	-.03	-.16
W47	.67	.14	-.06	.00	.11	.06	.27	.10	.16	.00
W48	.60	-.05	.39	.06	-.09	.20	.02	.08	-.09	.07
W49	.54	.09	.19	.04	-.13	.12	-.13	.20	.19	.12
W50	.14	-.05	.69	.08	-.03	.13	.05	.04	.02	.18
W51	.31	-.10	.14	.05	.44	.14	-.10	.29	.22	.14
W52	.66	.11	.01	-.11	.26	-.22	-.05	.05	-.06	.03
W53	.17	.03	.22	-.04	-.04	-.12	-.44	.12	.07	.16
W54	.16	-.03	.04	.03	.05	.01	.00	.03	.03	-.10

S = Slides (Subtest 1)

T = Tape (Subtest 2)

W = Written (Subtest 3)

TABLE F.2.--Item-Factor Correlations for All Items with Three Preset Factors.

Item	Factor		
	1	2	3
S1	-.14	.90	-.12
S2	-.11	-.17	-.16
S3	.07	.00	.26
S4	.04	-.08	-.12
S5	.01	-.02	-.10
S6	.09	.20	.26
S8	.05	.06	.16
S9	.12	-.04	.18
S10	-.02	.04	.50
S11	.06	.32	.28
S12	.14	.18	.01
S13	-.07	.02	.48
S14	.00	.04	-.24
S15	.24	-.01	.04
S16	.08	.31	.10
S17	-.12	-.04	.07
S18	.21	-.02	-.08
S19	.04	.28	.37
S20	.06	.03	.11
S21	.01	-.05	.22
S22	.16	.06	.34
S23	.20	.19	.07
S24	-.16	.67	-.16
S25	.09	.04	-.02
T26	.27	-.04	.06
T28	.46	.29	.40
T29	.59	.11	.39
T30	.42	.07	.19
T31	.56	.00	.08
T32	.38	.40	.05
T33	.47	-.01	-.11
T34	.12	-.02	.06
T37	.67	.23	.32
T38	.64	.08	-.10
T40	.78	.09	.08
W41	.35	.37	.14
W42	.48	.09	.18
W43	.44	.06	.22
W44	.34	.00	-.10
W45	.65	.24	.25
W46	.45	.03	.15
W47	.53	.25	.28
W48	.63	.12	.25
W49	.51	.22	.20
W50	.31	.04	.14
W51	.39	-.02	.14
W52	.60	.21	-.02
W53	.29	.05	-.16
W54	.14	.03	.17

S = Slides
T = Tape
W = Written

TABLE F.3.--Item-Factor Correlations for Subtest 1 with Five Preset Factors.

Item[a]	Factor				
	1	2	3	4	5
P1	.94	.01	-.03	-.04	.07
P8	-.03	.07	-.01	-.11	.21
P11	.29	.37	-.12	.20	.18
P24	.73	-.09	-.03	-.06	.00
Di2	-.05	-.09	-.10	-.06	-.40
Di5	.01	-.13	.04	.48	.07
Di6	.13	.32	-.12	-.12	.00
Di14	.06	-.26	.21	.00	.08
Di16	.28	.21	.16	.08	.03
A3	-.06	.26	.10	.26	.18
A13	-.04	.38	.15	-.06	.13
A15	-.02	.02	.61	.17	.00
A19	.23	.36	.16	-.03	-.03
A23	.12	.12	.33	-.07	.24
S4	-.02	-.20	-.10	.49	-.12
S9	-.05	.29	.07	-.11	-.23
S17	.02	.15	.04	-.05	-.40
S18	.01	.04	.35	-.13	-.26
S25	-.04	.00	.25	.02	.19
D10	.02	.55	-.06	.02	-.09
D12	.08	.02	.10	-.10	.30
D20	.00	.12	.00	-.06	.02
D21	-.01	.26	.14	.41	-.14
D22	-.03	.41	.18	.06	.12

[a]Item P7 was omitted due to zero variance.

P = Pleasure
Di = Distress
A = Anger
S = Surprise
D = Disgust

BIBLIOGRAPHY

BIBLIOGRAPHY

Adler, L. and Enelow, A. "An Instrument to Measure Skill in Diagnostic Interviewing: A Teaching and Evaluation Tool." Journal of Medical Education 41 (1966), 281-288.

Allen, T. W. "Effectiveness of Counselor Trainees as a Function of Psychological Openness." Journal of Counseling Psychology 14 (1967), 35-41.

Anastasi, A. Psychological Testing. 4th ed. New York: Macmillan Publishing, 1976.

Anthony, W. and Wain, H. "Two Methods of Selecting Prospective Helpers." Journal of Couseling Psychology 18 (1971), 155-156.

Ashby, J. D.; Ford, D. J.; Guerney, B. G., Jr.; Guerney, L. F.; and Snyder, W. U. "Effects on Clients of a Reflective and Leading Type of Psychotherapy." Psychological Monographs 71 (1957), No. 453, 1-32.

Astin, H. "A Comparison Study of the Situational and Predictive Approaches to the Measurement of Empathy." Ph.D. dissertation, University of Maryland, 1957.

Austin, D. "Is the Interview an Analog?" The Personnel Administrator 14 (1972), 13-15.

Bandura, A. "Psychotherapist's Anxiety Level, Self-Insight and Psychotherapeutic Competence." Journal of Abnormal and Social Psychology 52 (1956), 333-337.

Barbara, D. "The Value of Nonverbal Communication in Personality Understanding." Journal of Nervous and Mental Diseases 123 (1956), 286-291.

Barrett-Lenhard, G. T. "Dimensions of the Client's Experience of his Therapist Associated with Personality Change." Genetic Psychology Monograph (1962), 76.

Bath, K. and Calhoun, R. "The Effects of Professional Counselor Training on Empathy: Continued Cause for Concern." Counselor Education and Supervision (December 1977), 98-106.

Bellucci, J. "The Development and Validation of a Measure of Counseling Empathy." Ph.D. dissertation, Southern Illinois University, 1971.

Bender, I. E. and Hastorf, A. G. "On Measuring Generalized Empathic Ability (Social Sensitivity)." Journal of Abnormal and Social Psychology 48 (1953), 503-506.

Berger, M. "Nonverbal Communications in Group Psychotherapy." International Journal of Group Psychology 8 (1958), 161-178.

Bergin, A. and Jasper, L. "Correlates of Empathy in Psychotherapy: A Replication." Journal of Abnormal Psychology 74 (1969), 447-481.

Bergin, A. and Solomon, S. "Personality and Performance Correlates of Empathic Understanding in Psychotherapy." American Psychologist 18 (1963), 393.

Bernstein, L.; Brophy, M.; McCarthy, M.J.; and Roepe, R. "Teaching Nurse-Patient Relationships: An Experimental Study." Nursing Research 3 (1954), 80-84.

Bernstein, L. "Evaluating the Teaching of Physician-Patient Relationships." Psychotherapeutic Action of the Physician. Proceedings of the 4th International Congress on Psychosomatic Medicine. Psychother. Psychosom. 21 (1972/1973), 214-219.

Bernstein, L.; Headlee, T.; and Jackson, B. "Changes in 'Acceptance of Others' Resulting from a Course in the Physician Patient Relationship." British Journal of Medical Education 4 (1970), 65-66.

Blumstein, E. "A Predictive Measure of Empathy." Ph.D. dissertation, Northwestern University, 1972.

Boles, B. "The Effect of Personality Factors and Training on the Development of Interpersonal Skills by Freshman Medical Students." Ph.D. dissertation, Ohio State University, 1975.

Boomer, D. and Dittman, A. "Speech Rate, Filled Pause, and Body Movement in Interviews." Journal of Nervous and Mental Diseases 139 (1964), 324-327.

Boyd, R. E. "Whitehorn-Betz A-B Score as an Effector of Client-Counselor Interaction." Journal of Consulting Psychology 17 (1970), 279-283.

Brams, J. M. "Counselor Characteristics and Effective Communication in Counseling." Journal of Counseling Psychology 8 (1961), 25-30.

Breisinger, G. "Sex and Empathy, Reexamined." Journal of Counseling Psychology 23 (1976), 289-290.

Campbell, R. "Counselor Personality and Background and His Interview Subrole Behavior." Journal of Counseling Psychology 9 (1962), 329-334.

Cantrell, M. H. "An Investigation of Adult Empathy for Children as Measured by the Dawe-Jones Test Including Its Relation to Sibling Position." Ph.D. dissertation, Ohio State University, 1967.

Carkhuff, R. "Critical Variables in Effective Counselor Training." Journal of Counseling Psychology 16 (1969a), 238-245.

Carkhuff, R. "Helper Communication as a Function of Helpee Affect and Content." Journal of Counseling Psychology 16 (1969b), 126-131.

Carkhuff, R. Helping and Human Relations: A Primer for Lay and Professional Helpers, Vols. 1 and 2. New York: Holt, Rinehart, Winston, 1969c.

Carkhuff. R. "The Prediction of the Effects of Teacher-Counselor Education: The Development of Communication and Discrimination Selection Indexes." Counselor Education and Supervision VIII (Summer 1969d), 265-272.

Carkhuff, R. "Principles of Social Action in Training for New Careers in Human Services." Journal of Counseling Psychology 18 (1971), 147-151.

Carkhuff, R. and Banks, G. "Training as a Preferred Mode of Facilitating Relations between Races and Generations." Journal of Counseling Psychology 17 (1970), 413-418.

Carkhuff, R. and Bierman, R. "Training as a Preferred Mode of Treatment of Parents of Emotionally Disturbed Children." Journal of Counseling Psychology 17 (1970), 157-161.

Carkhuff, R.; Kratochvil, D.; and Friel, T. "Effects of Professional Training: Communication and Discrimination of Facilitative Conditions." Journal of Counseling Psychology 15 (1968), 68-74.

Carkhuff, R. and Griffin, A. "The Selection and Training of Human Relations Specialists." Journal of Counseling Psychology 17 (1970), 443-450.

Chandler, D. R. "Construct Validity of a Scale of Accurate Empathy in Psychotherapists." Ph.D. dissertation, University of Michigan, 1970.

Chapin, F. S. "Preliminary Standardization of a Social Insight Scale." American Sociological Review 7 (1942), 214-225.

Chapin, J. L. "The Development and Validation of a Scale to Measure Empathy." Ph.D. dissertation, Michigan State University, 1966.

Charles, C. "Correlates of Adjudged Empathy as Early Identifiers of Counseling Potential." Ph.D. dissertation, Michigan State University, 1973.

Chinsky, J. and Rappaport, J. "Brief Critique of the Meaning and Reliability of 'Accurate Empathy' Ratings." Psychological Bulletin 73 (1970), 379-382.

Cohen, E. C. "Empathy, Awareness of Interpersonal Responsibility and Consideration for Others in Young Children." Ph.D. dissertation, Michigan State University, 1973.

Cohen, J. and Struening, E. L. "Opinions About Mental Illness in the Personnel of Two Large Mental Hospitals." Journal of Abnormal and Social Psychology 63 (1962), 349-360.

Coleman, J. "Facial Expressions of Emotion." Psychological Monographs 63 (1949), 1-36.

Conklin, R. C. and Hunt, A. S. "An Investigation of the Validity of Empathy Measures." Counselor Education and Supervision (December 1975), 119-127.

Craig, A. "A Test of Empathy for Public School Administrators." Experimental Edition, 1959.

D'Augelli, A. and Danish, S. "Evaluating Training Programs for Paraprofessionals and Nonprofessionals." Journal of Counseling Psychology 23 (1976), 247-253.

Davitz, J. and Davitz, L. "Correlates of Accuracy in the Communication of Feelings." Journal of Communication 9 (1959a), 110-117.

Davitz, J. and Davitz, L. "The Communication of Feelings by Content-Free Speech." Journal of Communication 9 (1959b), 6-13.

Dawes, R. M. "A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making." American Psychologist (1970), 180-188.

Dilley, J. S. and Tierney, D. E. "Counselor Candidate Verbal Behavior and Relationship Orientation." Counselor Education and Supervision VIII (1969), 93-98.

Ditiberio, J. "The Strength of Sensing-Intuition Preference on the Myers-Briggs Type Indicator as Related to Empathic Discrimination of Overt or Covert Feeling Messages of Others." Ph.D. dissertation, Michigan State University, 1976.

Dittman, A. "The Relationship Between Body Movements and Moods in Interviews." Journal of Consulting Psychology 26 (1962), 480.

Dittman, A. and Wynne, L. "Linguistic Techniques and the Analysis of Emotionality in Interviews." Journal of Abnormal and Social Psychology 66 (1961), 201-204.

Dymond, R. F. "A Scale for the Measurement of Empathic Ability." Journal of Consulting Psychology 13 (1949), 127-135.

Ekman, P. "Body Position, Facial Expression and Nonverbal Behavior During Interviews." Journal of Abnormal and Social Psychology 68 (1964), 295-301.

Ekman, P. "Differential Communication of Affect by Head and Body Cues." Journal of Personality and Social Psychology 2 (1965), 726-735.

Ekman, P. "Movements with Precise Meanings." Journal of Communication 26 (1976), 14-26.

Ekman, P. and Friesen, W. "Constants across Cultures in the Face and Emotion." Journal of Personality and Social Psychology 17 (1971), 124-129.

Ekman, P. and Friesen, W. "Detecting Deception from the Body or Face." Journal of Personality and Social Psychology 29 (1974), 288-298.

Ekman, P. and Friesen, W. "Measuring Facial Movement." Environmental Psychology and Nonverbal Behavior 1 (1976), 56-75.

Ekman, P. and Friesen, W. Pictures of Facial Affect. Palo Alto, Calif.: Consulting Psycholgists Press, Inc., 1976.

Ekman, P. and Friesen, W. "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding." Semiotica 1 (1969), 49-98.

Ekman, P. and Friesen, W. Unmasking the Face. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.

Ekman, P., Friesen, W., and Ellsworth, P. Emotion in the Human Face. New York: Pergamon Press, 1972.

Ekman, P., Friesen, W., and Scherer, K. "Body Movement and Voice Pitch in Deceptive Interaction." Semiotica 16 (1976), 23-27.

Ekman, P., Friesen, W., and Tomkins, S. "Facial Affect Scoring Technique: A First Validity Study." Semiotica 3 (1971), 37-58.

Eldred, S. and Price, D. "A Linguistic Evaluation of Feeling States in Psychotherapy." Psychiatry 21 (1958), 115-121.

Feshbach, N. and Feshbach, S. "The Relationship between Empathy and Aggression in Two Age Groups." Developmental Psychology 1 (1969), 102-107.

Frois-Wittmann, J. "The Judgment of Facial Expression." Journal of Experimental Psychology 13 (1930), 113-151.

Gazda, G. Human Relations Development. Boston: Allyn and Bacon, Inc., 1974.

Gormally, J. and Hill, C. "Guidelines for Research on Carkhuff's Training Model." Journal of Counseling Psychology 21 (1974), 539-547.

Gough, H. "Nonintellectual Factors in the Selection and Evaluation of Medical Students." Journal of Medical Education 42 (1967), 642-650.

Gough, H. "The Adjective Check List as a Personality Assessment Research Technique." Psychological Reports 6 (1960), 107-122.

Greenberg, B., Kagan, N., and Bowes, J. "Dimensions of Empathic Judgment of Clients by Counselors." Journal of Counseling Psychology 16 (1969), 303-308.

Grunberg, R. "A Significant Counselor Personality Characteristic: Tolerance of Ambiguity." Counselor Education and Supervision, VIII (1696), 119-124.

Guerney, B., Jr., Stover, L. and DeMeritt, S. "A Measurement of Empathy in Parent-Child Interaction." Journal of Genetic Psychology 112 (1968), 49-55.

Halpern, M. "Some Factors Involved in Empathy." Ph.D. dissertation, Columbia University, 1954.

Hambleton, R. and Novick, M. "Toward an Integration of Theory and Method for Criterion-Referenced Tests." Journal of Educational Measurement 10 (1973), 159-170.

Hastorf, A. H. and Bender, I. E. "A Caution Respecting the Measurement of Empathic Ability." Journal of Abnormal and Social Psychology 47 (1952), 574-576.

Hayden, S. "Some Behavioral Correlates of Empathy." Ph.D. dissertation, Columbia University, 1955.

Hebb, D. "Emotion in Man and Animal: An Analysis of the Intuitive Processes of Recognition." Psychology Review 53 (1946), 88-106.

Heilman, K. "Empathy: The Construct and Its Measurement." Ph.D. dissertation, Purdue University, 1972.

Helfer, R. "An Objective Comparison of the Pediatric Interviewing Skills of Freshman and Senior Medical Students." Pediatrics 45 (1970), 623-627.

Hersch, P., Kulik, J., and Scheibe, K. "Personal Characteristics of College Volunteers in Mental Hospitals." Journal of Consulting and Clinical Psychology 33 (1969), 30-34.

Hill, G. "The Selection of School Counselors." Personnel and Guidance Journal 39 (1961), 355-360.

Hogan, R. "Development of an Empathy Scale." Journal of Consulting and Clinical Psychology 33 (1969), 307-316.

Horwitz, M. "A Comment on 'An Empirical Investigation of the Construct Validity of Empathic Understanding Ratings.'" Counselor Education and Supervision (June 1977), 292-295.

Huber, Harold W. "Investigation of the Effects of Selected Simulated Classroom Situations on Student Teacher Attitude and Empathy." Ph.D. dissertation, Michigan State University, 1972.

Hulin, W. and Katz, D. "The Frois-Wittmann Pictures of Facial Expression." Journal of Experimental Psychology 18 (1935), 482-498.

Hurst, M. "An Investigation of Empathy and Its Measurement: The Development of an Integrative Systems Model." Ph.D. dissertation, Boston University, 1973.

Hurst, M. and Shatkin, S. "Relationship between Standardized Admissions Variables and Certain Interpersonal Skills." Counselor Education and Supervision (1974), 22-33.

Inman, D. "Investigation of Affective Facial Expressions through Slow Motion and Normal Speed Videotape Techniques." Ph.D. dissertation, Michigan State University, 1976.

Izard, C. The Face of Emotion. New York: Appleton-Century Crofts, 1971.

Jarrett, F. J.; Waldron, J. J.; Burra, P. and Handforth, J. R. "Measuring Interviewing Skill: The Queen's University Interviewer Rating Scale (QUIRS)." Canadian Psychiatric Association Journal 17 (1972), 183-188.

Johnson, D.; Shertzer, B.; Linden, J.; and Stone, S. "The Relationship of Counselor Candidate Characteristics and Counseling Effectiveness." Counselor Education and Supervision VI (Summer 1967), 297-304.

Johnson, H., Ekman, P., and Friesen, W. "Communicative Body Movements: American Emblems." Semiotica 15 (1975), 335-353.

Jones, L. K. "Toward More Adequate Selection Criteria: Correlates of Empathy, Genuineness, and Respect." Counselor Education and Supervision (1974), 13-21.

Kagan, N. and Krathwohl, D. Studies in Human Interaction. Educational Publication Services, College of Education, Michigan State University, East Lansing, 1967.

Kagan, N., Krathwohl, D., and Farquhar, W. Developing a Scale to Measure Affective Sensitivity. Bureau of Educational Research Services, College of Education, Michigan State University, East Lansing, Educational Research Series, No. 30, 1965.

Kandler, H. and Hyde, R. "Changes in Empathy in Student Nurses During the Psychiatric Affiliation." Nursing Research (1953).

Kerr, W. and Speroff, B. "Validation and Evaluation of the Empathy Test." Journal of General Psychology 50 (1954), 269-276.

Krupka, J. and Parsons, M. The Empathy Skills Rating Scale. University Press, Michigan State University, 1978.

Kurtz, R. "A Comparison of Different Approaches to the Measurement of Counselor Empathy in Personal Counseling." Ph.D. dissertation, Michigan State University, 1967.

Kurtz, R. and Grummon, D. "Different Approaches to the Measurement of Therapist Empathy and Their Relationship to Therapy Outcomes." Journal of Consulting and Clinical Psychology 39 (1972), 106-115.

Lawton, M. P. "Personality and Attitudinal Correlates of Psychiatric-Aide Performance." The Journal of Social Psychology 66 (1965), 215-226.

Lee, D.; Zingle, H.; Patterson, J.; Ivey, A.; and Haase, R. "Development and Validation of a Microcounseling Skill Discrimination Scale." Journal of Counseling Psychology 23 (1976), 468-472.

Linden, J., Stone, S., and Shertzer, Bruce. "Development and Evaluation of an Inventory for Rating Counseling." Personnel and Guidance Journal (1965), 267-276.

McGreevy, C. P. "Factor Analysis of Measures Used in the Selection and Evaluation of Counselor Education Candidates." Journal of Counseling Psychology 14 (1967), 51-56.

McWhirter, J. J. "Two Measures of the Facilitative Conditions: A Correlation Study." Journal of Counseling Psychology 20 (1972), 317-320.

Mehrens, W. and Lehmann, I. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, Inc., 1973.

Merenda, P. and Reilly, R. "Validity of Selection Criteria in Determining Success of Graduate Students in Psychology." Psychological Reports 28 (1971), 259-266.

Mickelson, D. and Stevic, R. "Differential Effects of Facilitative and Nonfacilitative Behavioral Counselors." Journal of Counseling Psychology 18 (1971), 314-319.

Milliken, R. and Patterson, J. "Relationship of Dogmatism and Prejudice to Counseling Effectiveness." Counselor Education and Supervision VI (1967), 125-129.

Morris, W. W. "Does the Transcript Tell Enough?" ED 049, 488 (1971).

Nie, N.; Hull, C.; Jenkins, J.; Steinbrenner, K.; and Bent, H. Statistical Package for the Social Sciences. New York: McGraw-Hill, 2nd ed., 1975.

O'Hern, J. and Arbuckle, D. "Sensitivity: A Measurable Concept?" Personnel and Guidance Journal (1964), 572-576.

O'Hern, J. "A Study of Sensitivity as a Measurable Concept for the Screening and Evaluation of Counselors." Ph.D. dissertation, Boston University, 1962.

Passons, W. and Olsen, L. "Relationship of Counselor Characteristics and Empathic Sensitivity." Journal of Counseling Psychology 16 (1969), 440-445.

Patterson, C. H. "Selection of Rehabilitation Counseling Students." Personnel and Guidance Journal 41 (1962), 318-324.

Plutchik, R. The Emotions: Facts, Theories and a New Model. New York: Random House, 1962.

Plutchik, R., Conte, H., and Vandler, H. "Variables Related to the Selection of Psychiatric Residents." American Journal of Psychiatry 127 (1971), 79-84.

Porter, E. Jr. Therapeutic Counseling. Boston: Houghton, Mifflin and Co., 1950.

Rappaport, J. and Chinsky, J. "Accurate Empathy: Confusion of a Construct." Psychological Bulletin 77 (1972), 400-404.

Rasche, L., Veenhis, P., and Bernstein, L. "Training Non-Counselor Professionals in the Application of Basic Counseling Principles." ED 086, 909 (1973).

Rawls, J., Rawls, D., and Harrison, C. W. "An Investigation of Success Predictors in Graduate School in Psychology." The Journal of Psychology 72 (1969), 125-129.

Resnikoff, A. "A Measure of Discrimination for the Prediction of Counselor Effects." Journal of Counseling Psychology 19 (1972), 464-465.

Rogers, C. Client-Centered Therapy. Boston: Houghton, Mifflin Co., 1951.

Rogers, C. On Becoming a Person. Boston: Houghton, Mifflin Co., 1961.

Rogers, C. "The Necessary and Sufficient Conditions of Therapeutic Personality Change." Journal of Consulting Psychology 21 (1957), 95-103.

Ruckmick, C. "A Preliminary Study of the Emotions." Psychology Monographs 30 (1921), 30-35.

Sandler, I. "Characteristics of Women Working as Child Aides in a School-based Preventive Mental Health Program." Journal of Consulting and Clinical Psychology 39 (1972), 56-61.

Santavicca, G. "Supervised Experience and Selection of Counselor Trainees." Personnel and Guidance Journal 38 (1959), 195-197.

Sax, G. "Obtaining Information from Respondents: I--the Interview and the Questionnaire." In Empirical Foundations of Educational Psychology, pp. 201-240. New York: Prentice Hall, Inc., 1968.

Schlosberg, H. "Three Dimensions of Emotion." Psychological Review 61 (1954), 81-88.

Schofield, W. "The Use of Scholastic Aptitude, Personality and Interest Test Data in the Selection of Medical Students." Journal of Medical Education 41 (1966), 504-509.

Schwab, D. "Relationship between Interview Structure and Inter-interviewer Reliability in an Employment Situation." Journal of Applied Psychology 53 (1969), 214-217.

Schwab, D. "Why Interview? A Critique." Personnel Journal (February 1969), 126-129.

Scott, R. and Kemp, D. "The A-B Scale and Empathy, Warmth, Genuineness, and Depth of Self-Exploration." Journal of Abnormal Psychology 77 (1971), 49-51.

Sidman, J. "Empathy and Helping Behavior in College Students." Ph.D. dissertation, University of Colorado, 1968.

Smith, H. Sensitivity to People. New York: McGraw-Hill, 1966.

Smith, L. "The Effect of an Immediate Client-Activated Empathy Measure on Counselor Behavior and its Relationships to Other Empathy Measures." Ph.D. dissertation, University of Maryland, 1971.

Spadone, A. "An Investigation of Raters Levels of Accurate Empathy as a Function of Training Method." Ph.D., University of Southern California, 1974.

Starkweather, J. "The Communication Value of Content-Free Speech." American Journal of Psychology 69 (1956), 121-123.

Stefflre, B., King, P., and Leafgren, F. "Characteristics of Counselors Judged Effective by their Peers." Journal of Counseling Psychology (1962).

Stotland, E, and Dunn, R. "Empathy, Self-Esteem, and Birth Order." Journal of Abnormal and Social Psychology 66 (1963), 532-540.

Stotland, E. and Walsh, J. "Birth Order and an Experimental Study of Empathy." Journal of Abnormal and Social Psychology 66 (1963), 610-614.

Stripling, R. and Lester, J. "Selection, Preparation, and Professionalization of Specialists." Review of Educational Research 33 (1963), 171-178.

Strunk, O. Jr. "Empathy: A Review of Theory and Research." Psychological Newsletter 9 (1957), 47-57.

Strupp, H. and Jenkins, J. "The Development of Six Sound Motion Pictures Simulating Psychotherapeutic Situations." The Journal of Nervous and Mental Disease 136 (1963), 317-328.

Sub-Committee on Counselor Trainee Selection, American Psychological Association. "An Analysis of Practices in Counselor Trainee Selection." Journal of Counseling Psychology 1 (1954), 174-179.

Taft, R. "The Ability to Judge People." Psychological Bulletin 52 (1952), 1-23.

Tagiuri, R. "Person Perception." In Handbook of Social Psychology, 2nd ed., vol. 3, Ch. 23. Edited by G. Lindzey. Reading, Mass.: Addison-Wesley, 1965.

Thoresen, C. "Constructs Don't Speak for Themselves." Counselor Education and Supervision (June 1977), 296-303.

Tomkins, S. and McCarter, R. "What and Where are the Primary Affects? Some Evidence for a Theory." Perceptual and Motor Skills 18 (1964), 119-158.

Truax, C. "The Meaning and Reliability of Accurate Empathy Ratings: A Rejoinder." Psychological Bulletin 77 (1972), 397-399.

Truax, C. and Carkhuff, R. Toward Effective Counseling and Psychotherapy: Training and Practice. Chicago: Aldine Publishing Co., 1967.

Truax, C. "Therapist Empathy, Warmth, and Genuineness and Patient Personality Change in Group Psychotherapy: A Comparison Between Interaction Unit Measures, Time Sample Measures, Patient Perception Measures." Journal of Clinical Psychology 22 (1966), 225-229.

Turner, E., Helper, M., and Kriska, D. "Predictors of Clinical Performance." Journal of Medical Education 49 (1974), 338-342.

Veeser, W. "The Development of a Measure of the Verbal, Contextual, and Nonverbal, Vocal Components of Empathy." Ph.D. dissertation, Georgia State University, 1974.

Vesprani, G. "Personality Correlates of Accurate Empathy in a College Companion Program." Journal of Consulting and Clinical Psychology 33 (1969), 722-727.

Walker, R. "Client Perception of Counselor Empathy: Measurement and Use." Ph.D. dissertation, University of Tennessee, 1969.

Whitehorn, J. "Studies of the Doctor as a Crucial Factor for the Prognosis of Schizophrenic Patients." International Journal of Social Psychiatry 6 (1960), 71-77.

Whitehorn, J. and Betz, B. "Further Studies of the Doctor as Crucial Variable in the Outcome of Treatment of Schizophrenic Patients." American Journal of Psychiatry 117 (1960), 215-223.

Wiggins, N. and Blackburn, M. "Prediction of First-Year Graduate Success in Psychology: Peer Ratings." Journal of Educational Research, 63 (1969), 81-85.

Winborn, B. and Rowe, W. "Self-Actualization and the Communication of Facilitative Conditions--A Replication." Journal of Counseling Psychology 19 (1972), 26-29.

Winich, C. and Holt, H. "Seating Position as Nonverbal Communication in Group Analysis." Psychiatry 24 (1961), 171-182.

Wolf, R. and Murray, H. A. "An Experiment in Judging Personalities." Journal of Psychology 3 (1936), 345-365.

Zerba, M. "Labeling and Recognition of Izard's Facial Expressions by Three Age Groups." Ph.D. dissertation, Michigan State University, 1977.