HUGHES, Anita Esther Roberg, 1939-

STUDY I: THE ACCURATE EMPATHY RATINGS OF THERAPISTS IN TELEPHONE AND FACE-TO-FACE INTERVIEWS. STUDY II: THE EFFECT OF GROUP SENSITIVITY-TRAINING ON THE ACCURATE EMPATHY RATINGS OF THERAPISTS.

University Microfilms, Inc., Ann Arbor, Michigan
STUDY I:
THE ACCURATE EMPATHY RATINGS OF THERAPISTS
IN TELEPHONE AND FACE-TO-FACE INTERVIEWS

STUDY II:
THE EFFECT OF GROUP SENSITIVITY-TRAINING
ON THE ACCURATE EMPATHY RATINGS OF THERAPISTS

DISSER TATION

Presented in Partial Fulfillment of the Requirements for the
Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Anita Esther Hughes, B.A., M.A.

* * * * *

The Ohio State University

1969

Approved by

[Signature]
Adviser
Department of Psychology
ACKNOWLEDGMENTS

I wish particularly to record my indebtedness to the following people who have provided support and assistance throughout the various stages of this dissertation. Their generous help has made the task less arduous and the experiences gained more rewarding.

To Dr. Philip A. Marks I owe a lasting debt not only for his consultation with respect to this dissertation, but for his confidence in me, for his patience and encouragement, and for providing a model of clinical psychology of which I had been previously unaware.

I am grateful to Dr. Saul Siegel and Dr. Dennis Nolan for their continued support and their many helpful suggestions since the first expression of the ideas that formed this work.

This study would not have been possible without the combined efforts of the staff of the Columbus Area Community Mental Health Center. Not only did they so willingly give of their time in spite of frequent intrusions in their busy schedules, but as colleagues and friends they lent an esprit de corps to the task of data collection. I am also indebted to the Center for allowing me to use their facilities and for supplying the recording equipment and other material assistance.

To Dr. Clovis Shepherd for his participation in the study, and to Raymond Beyer whose resourcefulness helped to expedite matters, I owe special thanks. I am particularly grateful to Dr. Frank Lawlis, who provided ratings of the interviews and many helpful suggestions regarding the segmenting and editing of the tapes.
The excellent clerical assistance of Florence Minnich has been invaluable.

And finally to my husband, Howard--for the benefit of his scholarly and scientific acumen; for his keen eye in detecting unclarities and inconsistencies...and for his gentleness in conveying them; for his constant encouragement, untiring humor, and assistance throughout--I acknowledge much more than gratitude.
VITA

September 27, 1939 ............................................. Born - Stamford, Connecticut

1961 ......................................................... B.A., University of Wisconsin, Madison, Wisconsin

1961 - 1963 .................................................. Research Assistant, Department of Psychology, University of Wisconsin, Madison, Wisconsin

1963 ......................................................... M.A., University of Wisconsin, Madison, Wisconsin

1963 - 1964 .................................................. Teaching Assistant, Department of Educational Psychology, University of Wisconsin, Madison, Wisconsin

1964 - 1965 .................................................. Teaching Assistant, Department of Psychology, The Ohio State University, Columbus, Ohio

1965 - 1966 ................................................ United States Public Health Fellow, Department of Psychology, The Ohio State University, Columbus, Ohio

1966 - 1967 ................................................ Psychology Internship, University Hospitals, The Ohio State University, Columbus, Ohio

1967 - 1968 ................................................ United States Public Health Fellow, Department of Psychology, The Ohio State University, Columbus, Ohio

1968 - 1969 ................................................ Staff Psychologist, Columbus Area Community Mental Health Center, Columbus, Ohio

1969 ......................................................... Director, Community Service Unit, Dayton State Hospital, Dayton, Ohio
PUBLICATIONS


# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>I. STUDY I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Purpose</td>
<td>3</td>
</tr>
<tr>
<td>Background</td>
<td>4</td>
</tr>
<tr>
<td>II. METHODOLOGY</td>
<td>11</td>
</tr>
<tr>
<td>Subjects</td>
<td>11</td>
</tr>
<tr>
<td>Procedure</td>
<td>11</td>
</tr>
<tr>
<td>Raters</td>
<td>13</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>15</td>
</tr>
<tr>
<td>IV. DISCUSSION</td>
<td>19</td>
</tr>
<tr>
<td>V. STUDY II INTRODUCTION</td>
<td>25</td>
</tr>
<tr>
<td>Purpose</td>
<td>25</td>
</tr>
<tr>
<td>Background</td>
<td>25</td>
</tr>
<tr>
<td>VI. METHODOLOGY</td>
<td>41</td>
</tr>
<tr>
<td>Subjects</td>
<td>41</td>
</tr>
<tr>
<td>Procedure</td>
<td>41</td>
</tr>
<tr>
<td>Raters</td>
<td>45</td>
</tr>
<tr>
<td>VII. RESULTS</td>
<td>47</td>
</tr>
<tr>
<td>VIII. DISCUSSION</td>
<td>54</td>
</tr>
<tr>
<td>IX. SUMMARY AND CONCLUSIONS</td>
<td>57</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>60</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>80</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Means and Standard Deviations of Mean Accurate Empathy Ratings</td>
<td>16</td>
</tr>
<tr>
<td>2.</td>
<td>Analysis of Variance of Mean Accurate Empathy Ratings</td>
<td>17</td>
</tr>
<tr>
<td>3.</td>
<td>Analysis of Variance for Simple Effects of Raters</td>
<td>18</td>
</tr>
<tr>
<td>4.</td>
<td>Analysis of Variance for Simple Effects of Mode</td>
<td>18</td>
</tr>
<tr>
<td>5.</td>
<td>Means and Standard Deviations of Mean Accurate Empathy Ratings by Trained Raters</td>
<td>48</td>
</tr>
<tr>
<td>6.</td>
<td>Analysis of Variance of Mean Accurate Empathy Ratings by Trained Raters</td>
<td>49</td>
</tr>
<tr>
<td>7.</td>
<td>Means and Standard Deviations of Mean Accurate Empathy Ratings by Untrained Raters</td>
<td>50</td>
</tr>
<tr>
<td>8.</td>
<td>Analysis of Variance of Mean Accurate Empathy Ratings by Untrained Raters</td>
<td>51</td>
</tr>
<tr>
<td>9.</td>
<td>Means and Standard Deviations of Relationship Questionnaire Scores</td>
<td>52</td>
</tr>
<tr>
<td>10.</td>
<td>Analysis of Variance of Relationship Questionnaire Scores</td>
<td>53</td>
</tr>
</tbody>
</table>
STUDY I:
The Accurate Empathy Ratings of Therapists
In Telephone and Face-To-Face Interviews

I. INTRODUCTION

There has been a growing trend in the mental health field toward brief psychotherapy and short-term crisis intervention. Although the mental health professions have long been aware of the occurrence of psychological crises, it is only recently that they have focused their attention on effective intervention during this crucial period. Since World War II, there has been an increasing trend to bring psychodiagnostic and therapeutic activities out from the hospitals and offices and into the community; and with this movement have come new procedures. Usually crisis implies an early phase as well as an emotional peak and therefore a time when prompt intervention can be most effective. Promptness of intervention depends on accessibility and availability. Some agencies have attempted to meet this need by offering immediate personal interviews, by establishing walk-in centers, by sending teams to the home, or by utilization of the telephone for contact during crises (Farberow, Shneidman, Litman, Wold, Heilig, and Kramer, 1966). The telephone, in the past few years, has been assuming an ever increasing role among suicide prevention services and other help-givers (Bartholomew and Kelley, 1963; Kaplan and Litman, 1962; Koumans, Muller, and Miller, 1967; Litman,
Farberow, Shneidman, Heilig, and Kramer, 1965; Waltzer and Hankoff, 1965). Often "crisis therapy" will be limited to a telephone interaction between the patient and a mental health professional. Many mental health facilities operate 24-hour emergency telephone services, whereby individuals may seek immediate help for their problems without having to make specific arrangements to get to a specific office at a specific time. While there are many advantages to dealing with clients over the telephone, the very nature of the interaction presents certain limitations (Farberow, 1962). First of all, the patient has more control of the conversation and thus can determine and direct the therapeutic process. For example, he can easily terminate an interview at any point he chooses. Secondly, in the use of the telephone, both therapist and patient miss the non-verbal cues of facial expression, mannerisms, posture and stance which can be observed in a face-to-face encounter and which aid in transmitting warmth and understanding. Wachtel (1967) maintains that body positions and movements plus verbal material provide a richer understanding of what is happening than would either alone.

These limitations suggest that the task of a telephone interview is quite different from that of a face-to-face interview and that a therapist may respond quite differently to a patient over the telephone than he would to that patient in his office.

As the telephone comes to play an increasingly important role in crisis intervention, examination must be made of the variables related to telephone procedures. Improvement of our present methods of crisis
intervention begins with carrying out research designed to explicate and evaluate what it is we are now doing. With respect to conducting interviews over the telephone, it would be important to know if therapists who are effective in face-to-face encounters with patients are also effective with them over the phone. Perhaps there are people who are more effective in telephone interaction than when directly confronted with a person. The answers to these questions would have definite implications for the selection and training of volunteers for a suicide prevention service, for example. It may be that the most efficient method for selecting volunteers is to have them interview someone on the telephone and to objectively evaluate their effectiveness. An objective measure of effectiveness of such an interview would be invaluable in the selection and training of personnel to conduct therapy on the telephone and in the quality-control of personnel on the job.

Purpose

The purpose of the first study is to compare the effectiveness of therapists in the two types of interview situations: telephone and in-person. More specifically, it is proposed that therapeutic effectiveness be measured in terms of accurate empathy. The difficulties associated with measuring effectiveness of therapy are particularly apparent in the telephone situation, since the telephone does not lend itself well to patient follow-up and assessment of outcome. Many crisis calls are "one-time" events, and in many cases the caller refuses to
identify himself, thereby cutting off even the possibility of obtaining follow-up information. Even those cases that can be followed up represent a biased sample of the total population of original callers. Furthermore, telephone contact does not permit use of the traditional assessment procedures upon which prediction of outcome is usually based. How, in the light of these limitations can the effectiveness of telephone therapy be evaluated? It would seem that when outcome variables are very difficult to assess, it would be desirable to focus on variables that have been shown to be related to outcome in a general sense.

Background

Of the variables that can be assessed from listening to an interview, one that has been related to positive patient outcome has been that which Truax and Carkhuff (1967, p. 46) call "accurate empathy—the therapist's sensitivity to current feelings and his verbal facility to communicate this understanding in a language attuned to the client's current feelings." Theoretical formulations of psychotherapy as divergent as psychoanalysis (Fliess, 1942; Fromm - Reichman, 1950; Levine, 1961) and client-centered therapy (Rogers, 1957) have stressed the importance of empathy in bringing about a successful therapeutic outcome. Cottrell (1942) in his work on the analysis of situational fields, holds that empathy is the basic process in all social interaction. Hoskins (1946) speculated that inadequate empathy may be the primary defect in schizophrenia from which the remainder of the
symptomatology stems. Not only have theorists posited the importance of empathy, but a great deal of empirical evidence has accumulated regarding the importance of this variable in successful client treatment (Banks, Berenson, and Carkhuff, 1966; Bergin and Solomon, 1963; Combs and Soper, 1963; Lesser, 1961; Truax, 1961, 1962, 1963; Truax and Carkhuff, 1967; Truax, Carkhuff, and Kochman, 1965; Truax and Wargo, 1966a, 1966b; Truax, Wargo, and Carkhuff, 1966; Whitehorn and Betz, 1954). Studies of many types of help-seekers, including college underachievers, juvenile delinquents and hospitalized schizophrenics show that all respond positively to therapeutic encounters where the therapist offers a high level of Accurate Empathy, and even more significantly, may react unfavorably when Accurate Empathy is absent or when it is offered at only a low level (Truax and Carkhuff, 1967). These effects hold not only for many types of patients, but also for a wide variety of therapists, regardless of their training or theoretical orientation, and regardless of the type of therapy (viz. individual or group, long-term or time-limited - Truax and Carkhuff, 1967). Many research studies offer evidence of the critical importance of Accurate Empathy to human encounters intended to change human behavior. The most commonly used measure of empathic ability of therapists is the Accurate Empathy Scale developed in 1961 by Charles B. Truax (Appendix II). An early study (Truax, 1961) compared levels of empathy in four hospitalized patients who showed clear improvement on a variety of personality tests and four who showed clear deterioration
after six months of intensive psychotherapy. Level of empathy was assessed by rating two-minute segments from the middle-third of the therapy sessions. It was found that the therapists whose patients improved on the tests were rated consistently higher on Accurate Empathy than those with test deteriorated cases. In a later study (Truax, 1963), 14 schizophrenic patients were seen in intensive therapy for periods ranging from six months to four and a half years. High levels of empathy were associated with positive change in psychological test data and more time out of the hospital since the initiation of therapy. The positive relationship between Accurate Empathy and outcome of therapy still held when out-patients seen in counseling were added to the sample (Truax, 1963): successful cases received fewer moments of superficial understanding from their therapists, while failure cases received far fewer moments of deeply empathic understanding. More evidence for the importance of empathy comes from Combs and Soper (1963). They found that effective counselors tended to assume the internal rather than the external frame of reference with others, and to see people as able, dependable and friendly, rather than as lacking in these qualities. Lesser (1961) reported that the ability to predict accurately the degree of similarity between his own and his patient's Q-sort was significantly related to the patient's progress. This finding again suggests that the sensitive, empathic therapist who is able to assess accurately the patient as well as himself is perhaps the most effective (Truax and Carkhuff, 1967).
While the above studies suggest that therapists who are accurately empathic are indeed effective, there still remains the question of to what extent the therapist's level of Accurate Empathy is a function of the particular patient he is interviewing. In their book, Truax and Carkhuff (1967) discuss a study (no reference cited) in which the same eight therapists saw the same eight patients in a balanced incomplete block design. Analysis of the Accurate Empathy ratings indicated that different therapists did produce different levels of accurate empathy, even when interacting with the same set of patients \((p < .01)\) and, in contrast, different patients did not receive significantly different levels of Accurate Empathy when interacting with the same therapists. Van der Veen (1965) in a study of three patients from the same population as the above study, found that patients did have significant effects on the level of Accurate Empathy expressed by therapists. Truax and Wargo (1966) re-analyzed Van der Veen's data and concluded that patients did not significantly affect level of Accurate Empathy. In a study of hospitalized schizophrenics (Truax, 1962) a single standard interviewer saw each patient who was in therapy every three months throughout the course of therapy. There was no significant correlation between the levels of Accurate Empathy offered throughout therapy and the levels occurring in the sampling interview, as would be expected if the patient affected the level of Accurate Empathy. There was evidence to indicate, however, that patients did affect the level of non-possessive warmth offered by the
therapist and the sampling interviewer. In a study by Truax, Wargo, Frank, Imber, Battle, Hoehn-Saric, Nash and Stone (1966) 40 patients were randomly assigned to each of two screening interviewers and then assigned randomly to one of four different therapists. Levels of Accurate Empathy were obtained for each interviewer and for each therapist, based on samples of interviews with these 40 patients. Analysis indicated that the two interviewers differed from each other significantly on Accurate Empathy and that the different therapists also offered different levels of Accurate Empathy to their patients even though the patients themselves had been randomly assigned. Other studies also confirm the original findings which suggest that level of Accurate Empathy is due to the counselor rather than the client. In one study (Martin, Carkhuff, and Berenson, 1966) 16 college students were interviewed by both their "best available friend" and a professional counselor. The level of empathy offered by the counselor as indicated by objective tape ratings and inventories which the counselors filled out were significantly higher than those offered by the "best available friends." All of these studies taken together would indicate that empathy is an interviewer variable which operates independently of the interviewee. One important implication of this finding is that it makes possible studies utilizing small numbers of interviewees. In other words, if empathy remains stable across clients, it is not necessary to have a large number of clients to get an adequate sample of interviewers' levels of empathy.
The format of the present study permitted a comparison between two modes of interview: telephone and face-to-face. More specifically, it was intended to assess the relationship between mode of interview and empathy, as measured by the Truax Accurate Empathy Scale. Several alternative predictions are possible. Shapiro, Foster and Powell (1968) found that therapeutic attitudes are communicated through non-linguistic behavior, particularly facial cues. It would appear that interviewing a person whom one could see would facilitate empathic understanding by virtue of the many non-verbal cues provided in such a situation, and that a telephone interview would not lend itself as well to an empathic understanding of either party. On the other hand, Wachtel (1967) maintains that our ability to understand the language of the body is rather limited and that we have considerable difficulty understanding, and often do not make use of, the non-verbal communications of others. If this is so, then the essential difference between the two modes of interview is negligible and one would not expect the difference to be reflected in ratings of empathy. Shapiro (1968a) found that judges of therapy tapes were more predisposed to verbal behavior than to verbal and visual cues combined. Or, it could be argued that visual cues serve primarily as distractors and that the empathic process is facilitated when these cues are kept at a minimum as in the case of the telephone interview. Some indirect evidence for this position comes from a study by Maier and Thurber (1968) who tested the accuracy of judgments of deception attempts under three conditions: (1) watching and hearing the interview, (2) listening to a tape
recording, and (3) reading a transcript. Since the listeners and readers did better than the watchers, it was concluded the visual cues interfered with the ability to make accurate decisions. With no direct evidence appearing to the contrary, a statement of no difference would best reflect our present knowledge of empathy in various communication modes. It is hypothesized that there is no difference between the Accurate Empathy ratings of therapists in telephone and face-to-face interviews.
II. METHODOLOGY

Subjects

The Ss for this experiment were 10 members of the staff of the Columbus (Ohio) Area Community Mental Health Center. They consisted of one psychiatrist, one psychologist, five psychiatric social workers, two psychiatric nurses and one graduate student in social work. Five of the Ss were male and five were female. All staff members had received printed material on the use of telephone techniques, had participated in discussions of handling problem situations over the phone and made use of "telephone therapy" in the normal course of their jobs as well as conducting face-to-face interviews with patients several times per week.

Procedure

All Ss conducted and tape recorded five interviews, three over the telephone and two in person. Since tape recording calls had not been the practice at the Center, the first phone call was regarded as "practice" and was not included in the data analysis. No differentiation in this respect was made to the subjects however. Furthermore, they were completely naive as to the purpose of the study. All of the interviewees were female and had no prior contact with the S. The order of the interviews was counterbalanced such that half of the Ss conducted the telephone interview first and half the Ss conducted the face-to-face interview first. The recording equipment was a Stenorette dictating machine with a telephone adapter.
Telephone interview. Telephone interviews were calls initiated by the _S_ to females who called the Suicide Prevention Service during 1968 and the first two and-a-half months of 1969 and for whom a name and telephone number had been recorded. Of the callers meeting these criteria, names were assigned to subjects randomly. Callers were not assigned to the staff person who took the original call. In the event that a _S_ was unable to reach the caller assigned to him, or if the call was of less than six minutes duration, another name was randomly selected from the pool of remaining callers. _Ss_ were instructed to begin their interviews as follows:

"This is________from the Suicide Prevention Service. We are attempting to evaluate our services by calling those people who have called us in the past and seeing how they are doing now, and if we can be of any further help. Do you have time now to talk?" If they said "no," the _S_ asked, "When may I call you back?" If they said "yes," the _S_ continued by saying, "Let's see, you called the Suicide Prevention Service last________. How have things been going since then?"

_Ss_ were further instructed that although the rationale given to callers was one of evaluating the Service, "...the focus of the interview should be on the person--how she is doing and how she is feeling now."

Face-to-face interview. Interviews were conducted with females who were currently in treatment at the Center. Each staff member submitted to the _E_ the names of his current female clients. From this pool of clients, names were assigned to _Ss_ randomly. No _S_ was assigned the name of a person with whom he had had prior contact. The rationale given
to clients for coming in to see another staff member was as follows:

The Mental Health Center is always looking for ways to evaluate and improve the service they give to the people who come to them for help. We would appreciate your help in this. I would like you to come in once and talk to another person on our staff. The person with whom you talk will want to know about what brought you to the Center and how you are being helped with your problem.

Further instructions to the S were, "Although the rationale given to clients for coming in is one of evaluating the Center, the focus of your interview should be on the person - how she is doing and how she is feeling now." In the event that a S was unable to schedule an appointment for the interviewee assigned to him, another name was randomly selected from the pool of remaining interviewees. As in the case of the telephone interview, the only criterion for inclusion in the study was that the interview be of at least six minutes duration.

One four-minute segment was selected from the middle third of each interview, yielding four excerpts for each S. Any identifying information such as persons' names was deleted from the tapes. No segment contained content that could identify it as a telephone call. These 40 segments were then re-recorded in random order on a master tape.

Raters

Three individuals who have been trained at the Arkansas Rehabilitation Research and Training Center and routinely rate psychotherapy tapes served as raters in the present study. These individuals were part-time, undergraduate students who had been trained to rate
the dimension of Accurate Empathy with interrater reliabilities of no less than .50. (See Appendix I for the training procedure.) The range of reported interrater reliabilities for raters trained at Arkansas is .50 to .99. In addition to the Arkansas-trained raters, three additional judges rated tapes in the present study. These individuals were undergraduates at the University of Dayton who were untrained in the theory or practice of psychotherapy, as were the Arkansas raters. The untrained raters were given a copy of the Accurate Empathy Scale to read before making their ratings of the tapes. The two groups of raters, blind as to the purpose of the study and to the mode of interview, evaluated each excerpt using the Accurate Empathy Scale as a rating device. The Accurate Empathy Scale is a nine-point scale ranging from Stage 1 in which the therapist seems completely unaware of even the most conspicuous of the client's feelings, to Stage 9 in which the therapist unerringly responds to the client's full range of feelings in their exact intensity (see Appendix II). Each rater independently rated every segment. The ratings of the three raters in each group (trained and untrained) were average for each segment and these averages constituted the dependent measure.

---

1 Personal communication with Dr. Frank Lawlis, April, 1969.
III. RESULTS

Table 1 presents the means and standard deviations of the mean Accurate Empathy ratings by trained and untrained raters for the face-to-face and telephone interview conditions. An analysis of variance of these mean ratings (Table 2) indicated that the main effect for raters was significant \( p < .01 \) and that the interaction of rater and mode of interview was also significant \( p < .01 \). Analyses of variance for the simple effects of raters (Table 3) and for the simple effects of mode of interview (Table 4) were performed. The results indicated that the difference between the trained and untrained raters was significant \( p < .01 \) for both modes of interview. However, the modes of interview were significantly different \( p < .01 \) only for the untrained raters. The untrained raters found therapists to be more empathic in the face-to-face situation.

The correlation among the Accurate Empathy ratings of the three raters over the interview segments was .534 for the trained raters and .503 for the untrained raters. These interrater reliabilities were both significant at \( p < .005 \).

In order to determine the relationship between empathy in a face-to-face interview and empathy in a telephone interview, the correlation between the means of the two modes was determined. For the trained raters \( r = .066 \) and for the untrained raters \( r = -.181 \), indicating no significant correlation between the two modes of interview.
<table>
<thead>
<tr>
<th>Raters</th>
<th>N</th>
<th>Face-to-face</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean  SD</td>
<td>Mean  SD</td>
</tr>
<tr>
<td>Trained</td>
<td>20</td>
<td>3.12  .51</td>
<td>3.38  .53</td>
</tr>
<tr>
<td>Untrained</td>
<td>20</td>
<td>5.78  1.41</td>
<td>4.60  1.40</td>
</tr>
</tbody>
</table>
**TABLE 2**

**ANALYSIS OF VARIANCE OF MEAN ACCURATE EMPATHY RATINGS**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters (A)</td>
<td>1</td>
<td>75.41</td>
<td>67.33**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss within groups</td>
<td>38</td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>Mode (B)</td>
<td>1</td>
<td>4.20</td>
<td>3.39</td>
</tr>
<tr>
<td>A x B</td>
<td>1</td>
<td>10.52</td>
<td>8.48**</td>
</tr>
<tr>
<td>B x Ss within groups</td>
<td>38</td>
<td>1.24</td>
<td></td>
</tr>
</tbody>
</table>

**p .01**
### TABLE 3

**ANALYSIS OF VARIANCE OF SIMPLE EFFECTS OF RATERS**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater for face-to-face (A for b₁)</td>
<td>1</td>
<td>71.13</td>
<td>63.51**</td>
</tr>
<tr>
<td>Rater for telephone (A for b₂)</td>
<td>1</td>
<td>14.80</td>
<td>13.21**</td>
</tr>
<tr>
<td>Ss within groups</td>
<td>38</td>
<td>1.12</td>
<td></td>
</tr>
</tbody>
</table>

** p .01

### TABLE 4

**ANALYSIS OF VARIANCE FOR SIMPLE EFFECTS OF MODE**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode for untrained (B for a₁)</td>
<td>1</td>
<td>14.01</td>
<td>11.30**</td>
</tr>
<tr>
<td>Mode for trained (B for a₂)</td>
<td>1</td>
<td>0.71</td>
<td>0.57</td>
</tr>
<tr>
<td>B x Ss within groups</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p .01
IV. DISCUSSION

The results indicate that the Accurate Empathy ratings of the face-to-face interviews were significantly higher than those of the telephone interviews in the case of the untrained raters and that no significant difference was found when the segments were rated by trained raters. The interrater reliability of the untrained raters ($r=.503$) was at a level comparable to that of the trained raters (.534), and both are acceptable according to the standards set forth by Truax and Carkhuff (1967). Using their standard of acceptance, then, the present reliabilities are in keeping with the findings of Arnhoff (1954), Greenwood and McNamara (1967) and Shapiro (1968b), all of whom found that adequate reliabilities can be obtained from untrained evaluators. Cannon and Carkhuff (1969) found that experience level was not as critical an index of ability to discriminate levels of facilitative conditions (empathy, warmth and genuineness) as was the rater's level of facilitative interpersonal functioning (rating of simultaneous presence of facilitative conditions).

The difference in results between the two groups of raters may be viewed in one of two ways: (1) the two modes of interview do elicit differences in empathy level and these differences were not reflected in the trained rater's ratings or (2) there is actually no difference in empathy in the two situations and the untrained raters were in fact rating differentially some other aspect of the interviews. Shapiro (1968b) found that untrained raters were able to differentiate high
and low levels of psychotherapeutic conditions in a manner similar to that of trained raters. However, the naive raters did not show that they were able to differentiate among the therapeutic conditions of empathy, warmth and genuineness, as their ratings of empathy (for example) correlated more highly with trained raters' ratings of warmth than of empathy. It would be of interest to have the trained raters, in the present study, rate the segments on the dimensions of warmth and genuineness in order to assess the possibility of a difference between interview modes with respect to these variables. It is apparent that while the two modes may not differ with respect to empathy, as defined by the Accurate Empathy Scale and the trained raters, they do in fact influence untrained raters differentially. Even though the raters were blind with respect to interview mode, it is possible that their differential ratings reflect some uncontrolled aspect of the interview. For example the difference in length of interview (a mean of 29 minutes for the face-to-face and 11 minutes, 45 seconds for the telephone) or the possible difference in sound quality of the tapes could have influenced the raters. Perhaps this possibility could be ruled out by submitting the tapes to judges and asking them to discriminate between telephone and face-to-face interviews. If the discrimination could not be made when judges are provided with a set to look for differences, then it is unlikely that raters given no set would discriminate the two types of interviews on a spurious variable.

The alternate viewpoint regarding the interpretation of the data was that there is a difference in the level of Accurate Empathy across
interview modes and this difference was not reflected in the ratings of the trained raters. The finding of a difference by the untrained raters between the two modes of interview raises a question regarding the assumption of empathy as a stable therapist variable, as does the finding of a lack of correlation between the telephone and face-to-face interviews. Truax and Carkhuff (1967) maintain that Accurate Empathy is a very stable variable which operates independently of the client. The present findings suggest that it is not stable across interview modes. Since the same interviewees were not used for both modes, there does exist the possibility that the client does in fact influence the level of Accurate Empathy being expressed. Some recent investigations (Carkhuff and Alexik, 1967; Shapiro, 1968c; Vesprani, 1968) have also questioned the stability of this variable. In the Carkhuff and Alexik (1967) study it was found that level of functioning was influenced by the client's depth of self-exploration but that this was true only for low-functioning counselors (those who achieved ratings of 1.7 to 2.3 on a five-point scale of Empathy) and not for counselors functioning at a high level of therapeutic conditions (those who achieved ratings of 3.0 to 4.5 on a five-point scale of Accurate Empathy). In the present study, the therapists were consistently judged by the trained raters to be functioning at the lower end of the nine-point empathy continuum. It is possible that the relationships that hold

---

2A low level of functioning is here defined by the range given by Carkhuff and Alexik (1967), using a five-point scale, for their low-level functioning subjects and interpolating to find an equivalent range on a nine-point scale. The corresponding range on a nine-point scale would be values from 3.06 to 4.14. The therapists in the
well for therapists who function at high levels of facilitative conditions are not maintained among low-level therapists. According to Cannon and Carkhuff (1969), "...in general, the evidence indicates that nothing relates very much to anything else among low-level functioning subjects, while there are very consistent and stable relationships between relevant variables among high-level functioning subjects (p. 193)." In order to evaluate the probability of a chance finding in the present study, further research is needed. First the ratings should be cross-validated by having new groups of trained and untrained raters rate the segments. Differential findings (from present findings) in the cross-validation groups would suggest that a chance finding had occurred originally. A clearer picture of the relative difference of the two types of interview might be obtained by including both high and low functioning therapists in the subject population and by holding the interviewee constant across modes. If the cross-validation study yielded results similar to the present ones, a rather different design is suggested. Consistent findings of a discrimination between interview modes among untrained raters and of no discrimination among trained

present study achieved mean Accurate Empathy ratings (by the trained raters) of 3.12 and 3.38 for the face-to-face and telephone interviews, respectively.

3With respect to the implications of the low levels of ratings for the therapists involved in the study, no valid generalization should be implied with respect to their normal therapeutic contacts with patients. That is, the interviews utilized were analogues of, and not samples of actual therapeutic interviews. As such they were generated by the experimental design and confounded with such variables as time limitations, tape recording and knowledge that their interviews were to be used in a study.
raters would raise a serious question about the effect of training on the sensitivity of raters. It is conceivable that training desensitizes individuals to differences among therapists that are in fact important, i.e., related to therapeutic outcome. The design would compare different raters at varying levels of training, or the same raters as they progress through the training experience, to assess the relationship between amount of training and the ability to discriminate.

One of the findings of the present study was that telephone and face-to-face interviews were uncorrelated with respect to Accurate Empathy. That is, therapists who are empathic in one mode of interview are not necessarily empathic in the other. This finding would indicate that therapist's level of functioning should be assessed with respect to type of interview situation and that they should be trained and assigned according to their empirical level of functioning.

The major finding of the present study was that when measured by untrained raters therapists function more empathically in a face-to-face interview than they do over the telephone. A possible variable that could have contributed to this finding is the greater availability of cues in the face-to-face situation. This interpretation is consistent with the findings of Shapiro, Foster, and Powell (1968) that therapeutic attitudes are communicated through nonlinguistic behavior. A confounding variable in this study was interview length. The longer face-to-face interviews may have facilitated the therapists' expression of empathy.
Furthermore, it should be remembered the interviewees in the face-to-face situation were current clients at the Mental Health Center and the interviewees in the telephone situation were previous callers of a Suicide Prevention Service. These two groups of individuals may constitute different populations in terms of the degree of empathy that is manifested in their interactions with others. Any implications of this finding rest on the assumption made at the outset of this experiment that therapist empathy is positively related to client outcome and that this relationship would hold for telephone therapy as well as for therapy with clients sitting in the therapist's office. It remains an empirical question whether or not empathy in telephone contacts is indeed related to positive outcome and future research should be directed toward establishing the validity of, or disconfirming, the initially stated assumption of a positive relationship between empathy and outcome. If the positive relationship is supported by future research then methods of training should be developed which would elevate the level of empathy expressed over the telephone. An alternative to this approach would be to orient telephone interactions toward establishing face-to-face contact, either by getting the client to come in or by getting the therapist to go out.
STUDY II:
The Effect of Group Sensitivity-training on the Empathy Ratings of Therapists

V. INTRODUCTION

Purpose

There is a large body of research which suggests that therapists who possess a high level of Accurate Empathy are more effective with their patients than are therapists who possess only a low level (Truax and Carkhuff, 1967). If Accurate Empathy is related to positive outcome, as the research suggests, then the degree to which Accurate Empathy can be learned is of central importance to the aim of enhancing human relationships that change people. The purpose of this second study is to assess the influence of group sensitivity training on the Accurate Empathy ratings of its members.

Background

Truax and Carkhuff (1967) review several studies which suggest that Accurate Empathy is indeed "teachable" and that even non-professional persons lacking expert knowledge of psychopathology and personality dynamics can produce positive change in patients after training in the communication of Accurate Empathy. Their approach to training is aimed at integrating didactic and experiential elements. Traditionally, the didactic approach represents a conscious effort to provide therapists with repertoires of correct responses. Truax and Carkhuff (1967)
characterize didactic supervision as a behavior modification technique, whereby "the supervisory process may be seen as involving overt reinforcement to produce performance of desired therapist behaviors, and lack of reinforcement or even punishment to produce extinction of undesirable therapist behaviors (Truax and Carkhuff, 1967, p. 235)." They point out the implicit paradox involved in this type of supervision—that often the conditions, techniques, or attitudes that the future therapist is "taught" to employ are not being offered by the supervisor.

The experiential approach to training or supervision emphasizes the importance of genuine acceptance and respect for individuality. Unlike the didactic approach, the experiential approach attempts to influence behavior by offering certain attitudinal conditions which will promote openness to experiences and a willingness to experiment. The skill of the trainee will then develop from the trainee's own experience. In the integrated approach, the therapist-supervisor brings to bear a knowledge of therapy accumulated from his own experience and from those of others and communicates this knowledge within the context of a therapeutic relationship, offering to the trainee those conditions which research and clinical learning suggest are essential for personality change. The three elements central to the integrated approach of Truax and Carkhuff (1967) are: (1) a therapeutic context in which the supervisor himself provides high levels of therapeutic conditions (empathy, warmth and genuineness); (2) a highly specific didactic training in the implementation of the therapeutic conditions;
and (3) quasi-group therapy experience where the trainee can explore his own existence.

Three studies relating to such training are those of Carkhuff and Truax (1965), Truax and Silber (1966), and Baldwin and Lee (1965). In the first of these studies both graduate students and lay hospital personnel ($N = 17$) were given identical training programs over a period of approximately 100 hours. Following this each trainee had a single interview with each of three hospitalized patients. From the three tapes of each trainee, six 4-minute excerpts were randomly selected, two from each tape. Excerpts of therapy interviews were similarly selected from the tapes of eleven experienced therapists.

Four naive undergraduates who were not psychology majors were trained in the use of the Accurate Empathy Scale to a degree of intrarater reliability of not less than .50. There was not a significant difference in degree of Accurate Empathy between the groups of trainees and experienced therapists. The mean Accurate Empathy Scale values of the lay hospital personnel, the graduate students and the experienced therapists were 4.58, 5.14 and 5.22 respectively. The level attained by the graduate students compared very favorably with mean value of 2.50 for advanced trainees in counseling programs found in earlier studies (Bergin and Solomon, 1963; Melloh, 1964). Thus it was concluded that in a 100-hour training period, both graduate students and lay hospital personnel can be brought to display empathic understanding at a level commensurate with that of experienced therapists. The
authors failed to indicate whether or not the subjects' levels of empathy were rated before training and so it is possible that the groups of therapists were not significantly different from each other before training began. Truax and Carkhuff (1967) report an unpublished study by Truax and Silber (1966) who did a cross-validation on a group of 16 graduate student trainees. Measures of empathy were obtained by having ratings of "one-shot therapeutic interviews" early in the training program and again late in the program. Before the early interview the trainees had received 14 class hours of training and had spent an average of 11 hours listening to experienced therapists. The late interviews were made after an additional 34 to 36 hours of training. The analysis indicated significant gains in the level of Accurate Empathy (p < .01) over the 9-week period of additional training. Again, it is not known whether any pretraining measures were obtained. It is possible that maximum gain occurs in the initial stage of training. Also, since the exact nature of the additional training was not specified, there also exists the possibility that type, rather than quantity of training enhances empathy. In another study concerning the teaching of empathy, Baldwin and Lee (1965) compared the use of programmed teaching machines with an informal, didactic group therapy experience in undergraduate college students. The Accurate Empathy Scale was used to evaluate two-person interactions of the students before and after training. Those who participated in the teaching-machine approach to training in interpersonal relationships showed no change in measured empathic ability. However, the students who received informal didactic
group therapy showed significant improvement in empathic ability. This study suggests that an experiential and didactic approach focusing on interpersonal relating can produce significant increments in empathic skill. Also, as mentioned earlier, an element in the Truax training program for therapists is a "quasi-group therapy experience" (Truax and Carkhuff, 1967). It may be that within an approach which combines or integrates didactic and experiential elements, the experiential component alone contributes significantly to increases in empathic skill.

The increasing manpower shortage in the mental health fields and the increasing acceptance and use of non-professional and volunteer personnel suggests a need for more economical training procedures. Unfortunately there have been almost no well designed, controlled and implemented studies assessing the efficacy of even current training programs. Most lack any systematic attempt to provide appropriate training control groups and pre- and post-training measures (Carkhuff, 1966). Brief training programs may facilitate the use of experimental methods to evaluate their effectiveness both in terms of accomplishing the desired training and in terms of ultimate benefit to the client or patient. Some recent studies of brief training procedures, utilizing control groups and pre- and post-training measures, have been those of Delaney and Heimann (1966), Herron (1967), and Reddy (1969). In an attempt to determine what effect didactic training and experiential training had on the perception of non-verbal communication, Delaney and
Heimann (1966) had subjects in both groups rate 10 filmed responses (emotions) of actors three weeks prior to treatments, at the beginning of treatments and following the three-week treatment period. Treatment sessions were held twice per week for three weeks, each lasting 60 minutes. The didactic group changed in their perceptions of the person communicating non-verbal cues, while the experiential group changed in their perceptions of the emotions being communicated.

Herron (1967) compared an experimental group who received therapy supervision once per week with a control group receiving no supervision for a period of one semester. There were no significant changes over the semester for either group in the attitudinal areas of openmindedness, tolerance, acceptance of self and others, general adjustment or openness to experience. An effective and brief method of empathy training was described by Reddy (1969). Thirty-six undergraduates were assigned to one of three feedback groups: immediate, delayed, or control. They were instructed briefly in the concept of empathy and in the use of the Accurate Empathy Scale. They were then asked to assume the role of therapist and to respond aloud empathically to a simulated psychotherapy film. Ratings of their responses served as a pre-measure. They then viewed four more films and were given appropriate feedback (their level on the Accurate Empathy Scale) as to their empathic performance. The immediate feedback group received feedback after each statement, the delayed group after each film and the control group received no feedback. On a
sixth film subjects responded but received no feedback. Their
responses to the sixth film served as the post-measure. A comparison
of the pre- and post-responses, rated independently on a five-stage
Accurate Empathy Scale, showed that both of the experimental groups
had made significant gains. Analysis of the post-measure indicated
superiority of the immediate feedback method over the delayed method
and the control group.

One particular experiential approach to interpersonal relating
which has implications for use in the training of empathy is known as
the T-group method of sensitivity training. Briefly described, the
T-group learning experience has as its focal point a small, un-
structured, face-to-face group. A trainer is present, but usually
does not accept a leadership role. The participants are to discuss
themselves and the way they portray themselves in the group. Emphasis
is on behavior emitted in the group rather than behavior involving
past experiences or future problems. In reviewing the literature,
Campbell and Dunnette (1968) have listed as desired outcomes whose
which they find are common to most discussions of T-groups. One of
the goals of sensitivity training is "increased sensitivity to the
behavior of others...a goal very similar to the concept of empathy as
it is used by clinical and counseling psychologists, that is, the
ability to infer correctly what another person is feeling (Campbell
and Dunnette, 1968, p. 75)." The present discussion attempts to
delineate those measures which have been used to assess the
effectiveness of T-groups in terms of increased empathy and to point out the difficulties with the particular criterion measures used. Several studies (Boyd and Elliss, 1962; Bunker, 1965; Miles, 1965; Valiquet, 1964) used a "perceived change" measure as the basic criterion. This measure is an open-ended question asking a superior, subordinate, or peer of the subject to report any changes in the subject's behavior in the job situation during some specified period of time. In the Boyd and Elliss (1962) study, 42 managers were selected from three different T-groups to compose the experimental group. The two control groups consisted of 12 individuals who received no training and 10 managers who received a conventional human-relations training program employing lectures and conference techniques. Perceived change scores were obtained from interviews with each manager's superior, two of his peers, and two of his subordinates. The percentages of observers reporting changes were 65% for the laboratory-trained group, 51% for the conventionally trained group, and 34% for the control group. Of the 351 total statements of perceived change, only 137 changes were agreed upon by two or more observers. The percentage of subjects showing agreed upon changes were 64% for the experimental group and 23% for the two control groups combined. Six percent of the reported changes were of an unfavorable nature (e.g., an increase in irritability or loss of tolerance) and all but two of these 22 negative changes were attributed to members of the laboratory-trained group. The authors concluded that the behavioral outcomes of the trainees were very variable and that no particular pattern could be regarded as a typical training outcome.
Thirty-four high-school principals participating in a National Training Laboratory program were studied by Miles (1965) and compared with two control groups - a matched (by nomination) group of 29 principals and a randomly selected group of 148 principals. A variety of measures was administered before, during, and eight months following the laboratory. Three instruments were used to assess changes in job performance: the Leadership Behavior Description Questionnaire (Stogdill and Coons, 1957), the Group Participation Scale (Pepinsky, Siegel, and Van Alta, 1952), and an open-ended perceived-change measure. In addition to these external criterion measures, a large number of other measures were also utilized. These included ratings of various training behaviors and also five measures of the subject's organizational situation: (a) security, as measured by length of tenure in present job; (b) power, as measured by the number of teachers in the subject's school, (c) autonomy, as measured by length of time between required reports to the immediate superior; (d) perceived power, as measured by a Likert-type scale. In addition, a number of personality measures were administered and participants were also asked to rate their "desire for change" before starting training. Of the external criterion measures, only the results obtained with the perceived-change measure were statistically significant. The observers reported perceived behavioral changes for 30% of the experimentals, 10% of the matched controls and 12% of the randomly selected controls. The corresponding percentages for self-reported changes were 82%, 33%, and 21% for the three groups. The nature of the changes reported
included increased sensitivity to others, heightened equalitarian attitudes, greater communication and leadership skills, and patterns of increased consideration and relaxed attitudes on the job. There was a significant correlation between the perceived-change measure and trainer ratings of amount of change during the T-group. It was also found that more changes in job behavior tended to be observed for the high school principals with longer tenure and more subordinates.

The subjects in Bunker's (1965) study were 229 individuals from six different training laboratories and 112 matched (by nomination) controls. Open-ended behavior-change descriptions were obtained from five to seven co-workers and the self for each subject approximately one year after the training period. Eleven of the 15 subcategories derived for content analyzing the perceived-change data yielded statistically significant differences between groups with the trained group showing greater change in each category than the control group. The greatest differences were in the areas related to increased openness, receptivity, tolerance of differences, increased operational skill in interpersonal relationships, and improved understanding of self and others. Like Boyd and Elliss (1962), Bunker (1965) emphasized that changes among the trainees differed greatly from person to person and that actually there was no systematic outcome.

Valiquet's (1964) subjects were 34 participants from an ongoing laboratory-type training program conducted in a large corporation and 15 matched controls. Each subject nominated approximately five observers. Statistically significant differences were obtained between
experimentals and controls on total number of changes observed, total number of changes agreed upon by two or more observers, and total number of changes reported by the subjects themselves. A content analysis revealed that the categories of difference were much the same as in the Bunker (1965) study, i.e., increased openness and receptivity, increased interpersonal skill and improved understanding of self and others.

In all four studies reported above, between two and three times as many "changes" were reported for the experimental groups as for the control groups. The types of perceived changes which seemed to discriminate best between the experimentals and the controls had to do with increased sensitivity, more open communication, and increased flexibility in role behavior. Campbell and Dunnette (1968) maintain that these four studies seem to form the backbone of the evidence used to support the utility of the T-group method for the development of individuals in organizations. They go on to point out the limitations of the studies:

1. The observers responding to the criterion measures apparently knew whether or not the individual they were describing had been through T-group training.

2. The multiple observers for each subject were nominated by the subject and probably had varying degrees of interaction with each other.

3. No before measures were used and the estimation of change depended solely on recollection by the observers.
4. There was no evidence of a relationship between positive change and increased job effectiveness.

In a study by Bennis, Burke, Cutter, Harrington, and Hoffman (1957) a measure of "social sensitivity" was derived by first computing the discrepancy between an individual's prediction of another subject's response and the subject's actual response. While there was a slight tendency for the accurate predictors to be predicted more accurately themselves, no changes occurred in this measure over the course of the T-group. Gage and Exline (1953) also attempted to assess how well T-group participants could predict the questionnaire responses of the other group members. Two National Training Laboratories groups responded to equivalent 50-item questionnaires before and after a three-week laboratory. The items were opinion statements concerning group processes, leadership styles, and the scientific study of human relations. The subjects were asked to give their own opinions and also to predict how they thought the group as a whole would respond. There was no significant change in the accuracy of the predictions over the course of the training. The questionnaire method was also used by Lohman, Zenger, and Weschler (1959) who found a slight increase in the degree of agreement between members predictions and the trainer's responses. No control group was used, and no attempt was made to account for the effects of answering the same items twice.

In general, the studies attempting to measure how well an individual can predict the attitudes and values of others before and after
T-group training have yielded largely negative results. However, a recent study by Myers, Myers, Goldberg and Welch (1969) on the effect of feedback on interpersonal sensitivity reported more positive findings. The subjects for this study were 69 participants in a three-day workshop in group development. Subjects in experimental groups made sociometric ratings at the end of each of the eight group sessions and anticipated their own ratings by other group members. In each session after the first, subjects were given feedback information regarding their own ratings, ratings of other members, and information regarding the accuracy of their predictions. Control groups of subjects provided ratings but received no feedback. It was found that subjects in the experimental groups showed a significantly greater increase in sensitivity (operationally defined as "the degree of discrepancy between ratings received and anticipated) during the three-day period than control subjects. Control subjects did not show a significant increase in sensitivity between the beginning and the end of the workshop.

In contrast to the many negative findings regarding perceptual accuracy scores, several studies establish fairly well that people who have been through a T-group describe other people and situations in more interpersonal terms. Harrison (1962) found that T-group participants used a larger number of interpersonal terms in describing other T-group members than did controls. In a later study (Harrison, 1966) it was found that there were significant increases in the frequency of subjects' use of interpersonal concepts to describe
associates three months after training, where there had been no such effect three weeks after training. Bass (1962) found that training resulted in participants becoming more sensitive to interpersonal relationships exhibited in a film. Stock (1964) reported an unpublished study by Miles, Cohen, and Whitam (1959) in which participants' responses regarding the trainer's behavior and group interactions were compared with the trainer's diagnosis of the group's difficulties. The authors reported some change by the group members on a variable labeled "sensitivity to feelings." Clark and Culbert (1965) analyzed the content of nine college students' verbalizations before and after participating in a T-group and concluded that four of the subjects were better perceivers of group processes at the end of the training than they had been at the beginning.

These studies reflect the tendency of people who have been in a T-group to use more interpersonally oriented words to describe certain situations. However, as Campbell and Dunnette (1968) point out, this may reflect merely the acquisition of a new vocabulary rather than increased sensitization to interpersonal events. Some evidence bearing on this point comes from a study by Culbert, Clark and Bobele (1968) who attempted to determine whether changes in questionnaire indexes of self-actualization after a T-group experience correlated with changes in self-actualizing behavior. Sensitivity training was found to bring about increased Personal Orientation Inventory (Shostrom, 1963) scale means, but there was no correlation between that measure and scores on the Problem Expression Scale (Van der Veen and Tomlinson, 1967), which
measures increases in self-aware verbal behavior. The authors concluded that while sensitivity training seems to support and perhaps promote self-actualizing values, concepts, and percepts of its participants, the holding of such constructs does not correlate with changes in self-actualizing verbal behavior. Bunker and Knowles (1967) compared training laboratories of three-weeks' and two-weeks' duration. A questionnaire was used eight to ten months after training to elicit descriptions of the subject's post-laboratory behavior changes as seen by himself and seven of his co-workers. A "verified" change score was derived by summing those changes mentioned by two or more persons. While both laboratory trained samples differed from a matched control group on both "perceived" and "verified" change measures, they also differed significantly from each other. There were more changes of both types in the three-week sample. A content category analysis suggested that the three-week laboratory group made more overt, proactive changes, as opposed to the more passive, attitudinal changes made by the two-week group. However, there was a confounding of the duration variable with the training design. The authors suggested that the greater emphasis on back-home application of laboratory training in the three-week program may have significantly contributed to the qualitative differences in the kinds of changes made. The evidence is, however, that certain outcomes of sensitivity training reflect primarily attitudinal rather than behavioral change.

The number of studies on sensitivity-training, in general, is very limited. Miles (1965) claimed that 95 per cent of all treatment efforts
go unstudied. He referred to a report by Johnson (1964) which showed that only half of one per cent of over 1,500 National Defense Education Act grants made in a large state for experimental education programs were evaluated in any systematic manner. Those that are studied suffer serious defects in design, measurement or data analysis (Miles, 1965). The major criticisms of the studies of T-group effectiveness concern the lack of control groups, the possibility of Hawthorne-type effects, the unreliability of self-report and the frequent lack of pre-sensitivity training measures. With respect to the criterion of increased interpersonal sensitivity or empathy, the research methodology appears generally inadequate in that the measures are not relevant to the phenomenon under investigation. It would seem that the previous operational measures of empathy are neither reliable nor valid and that the Accurate Empathy Scale may provide a more meaningful criterion for use in T-group research. It is hypothesized that participating in a sensitivity-training group increases the level of one's measured empathy, and that this holds true both for interviews conducted in person and for telephone interviews.
VI. METHODOLOGY

Subjects

The Ss for this experiment were 10 members of the staff of the Columbus (Ohio) Area Community Mental Health Center, all of whom indicated a willingness to participate in a marathon sensitivity-training session as part of the experiment. The Ss consisted of one psychiatrist, one psychologist, five psychiatric social workers, two psychiatric nurses and one graduate student in social work. Five of the Ss were male and five were female. Of the 10 Ss, two (a psychiatric social worker and a psychiatric nurse) had participated in previous sensitivity-training groups. All staff members were in the practice of conducting interviews with clients several times per week, both in their offices and over the telephone.

Procedure

The 10 Ss were assigned to two groups, such that there were five Ss in the experimental group and five Ss in the control group. The groups were matched as closely as possible with respect to sex and profession. The two individuals with experience in sensitivity-training were not in the same group. All Ss were informed of the date of the forthcoming sensitivity-training session (T-group) and asked to clear their calendars for that day. Thus no S knew before the pre-measures were collected, to which group he belonged. Each subject initially
conducted and tape recorded three interviews, two over the telephone and one in person. Since tape recording calls had not been the practice at the Center, the first phone call was regarded as a "practice" call and is not included in the data analysis. No differentiation in this respect was made to the Ss, however. Furthermore, they were completely naive as to the purpose of the study. All of the interviewees were female and had no prior contact with the S. The order of the interviews was counterbalanced such that half of the Ss conducted the telephone interview first and half of the Ss conducted the face-to-face interview first. The recording equipment was a Stenorette dictating machine with a telephone adapter.

**Telephone interview.** The telephone interviews were calls initiated by the S to females who called the Suicide Prevention Service during 1968 and the first two-and-a-half months of 1969, and for whom a name and telephone number had been recorded. Of the callers meeting these criteria, names were assigned to Ss randomly. Callers were not assigned to the staff person who took the original call. In the event that a S was unable to reach the caller assigned to him, or if the call was of less than six minutes duration, another name was randomly selected from the pool of remaining callers. Ss were instructed to begin their interviews as follows:

"This is________ from the Suicide Prevention Service. We are attempting to evaluate our services by calling those people who have called us in the past and seeing how they are doing now, and if we can be of any further help. Do you have time now to talk?" If they said
"no," the subject asked, "When may I call you back?" If they said "yes," the subject continued by saying, "Let's see, you called the Suicide Prevention Service last (name of month). How have things been going since then?" Ss were further instructed that although the rationale given to callers was one of evaluating the Service, "...the focus of the interview should be on the person – how she is doing and how she is feeling now."

Face-to-face interview. These interviews were conducted with females who were currently in treatment at the Center. Each staff member submitted to the E the names of his current female clients. From this pool of clients, names were assigned to Ss randomly. No S was assigned the name of a person with whom he had had prior contact. The rationale given to clients for coming in to see another staff member was as follows:

The Mental Health Center is always looking for ways to evaluate and improve the service they give to the people who come to them for help. We would appreciate your help in this. I would like you to come in once and talk to another person on our staff. The person with whom you talk will want to know about what brought you to the Center and how you are being helped with your problem.

Further instructions to the S were, "Although the rationale given to clients for coming in is one of evaluating the Center, the focus of your interview should be on the person – how she is doing and how she is feeling now." In the event that a S was unable to schedule an appointment for the interviewee assigned to him, another name was randomly selected from the pool of remaining interviewees. As in the case of the telephone interview, the only criterion for inclusion in
the study was that the interview be of at least six minutes duration. At the end of the interview, interviewees were requested to fill out a modified form of the Relationship Questionnaire (see Appendix IV) developed by Truax (1963) which is designed to assess the client's experience of his therapist's empathy. The original questionnaire was modified in the present study by including only those items scored for empathy and excluding items scored on the other five scales. The wording of items was changed to assess the interviewees feelings about the person he had just talked with, rather than about an ongoing relationship. Items #87 and #94 of the Accurate Empathy scale were omitted on the basis of being non-applicable to the experimental situation. Six buffer items from other scales were included in the questionnaire but were not scored. The total questionnaire consisted of 50 items, 44 of which were scored for Accurate Empathy.

After each S had conducted and tape recorded two interviews (in addition to the practice interview) the Ss in the experimental group took part in a sensitivity-training session lasting 12 hours. All of the pre-training interviews were completed within two weeks prior to the T-group. Ss did not know until after the pre-treatment measures were obtained to which group they belonged. The 12-hour sensitivity-training session was conducted by a trainer affiliated with the National Training Laboratories - National Education Association. Since there were only five Ss in the experimental group, additional individuals were recruited to participate in the sensitivity session at the
trainer's request. These individuals were two social workers and two nurses who did volunteer work for the Suicide Prevention Service. The control group did not receive this sensitivity-training procedure.

Following completion of sensitivity-training for the experimental group, Ss in both the experimental and control groups repeated the initial procedure of tape recording two interviews, one on the telephone and one in person. The experimental group completed their interviews within 48 hours of the sensitivity-training session and the control group within 96 hours. The Relationship Questionnaire was again given to all interviewees in the face-to-face situation. When all of the data had been collected, one four-minute segment was selected from the middle third of each interview, yielding four excerpts for each S. Any identifying information such as persons' names was deleted from the tapes. No segment contained content that could identify it as a telephone call. These 40 segments were then re-recorded in random order on a master tape.

Raters

Three individuals who have been trained at the Arkansas Rehabilitation Research and Training Center and routinely rate psychotherapy tapes served as raters in the present study. These individuals were part-time, undergraduate students who had been trained to rate the dimension of Accurate Empathy with intrarater reliabilities of no less than .50. (See Appendix I for the training procedure.) The range of reported interrater reliabilities for raters trained at
Arkansas is .50 to .99. In addition to the Arkansas-trained raters, three additional judges rated tapes in the present study. These individuals were undergraduates at the University of Dayton who were untrained in the theory or practice of psychotherapy, as were the Arkansas raters. The untrained raters were given a copy of the Accurate Empathy Scale to read before making their ratings of the tapes. The two groups of raters, blind as to the purpose of the study and to the mode of interview, evaluated each excerpt using the Accurate Empathy Scale as a rating device (see Appendix II). Each rater independently rated every segment. The ratings of the three raters in each group (trained and untrained) were averaged for each segment and these averages constituted the dependent measure.

---

1 Personal communication with Dr. Frank Lawlis, April, 1969.
VII. RESULTS

Table 5 presents the means and standard deviations of the average Accurate Empathy ratings by trained raters for the experimental and control groups under conditions of face-to-face and telephone interviews, before and after the T-group. An analysis of variance (Table 6) revealed no significant differences for the trained raters. Table 7 presents the average Accurate Empathy ratings for the untrained raters. Although there were no significant differences at the p = .05 level of confidence (Table 8), the main effect of interview mode approached significance, for the untrained raters. The means and standard deviations of the Relationship Questionnaire scores are presented in Table 9. An analysis of variance (Table 10) revealed no significant differences between the experimental and control groups, before and after the T-group experience.

The correlation among the Accurate Empathy ratings of the three raters over the interview segments was .534 for the trained and .503 for the untrained raters. Both of these coefficients are significantly greater (p < .005) than zero.

47
<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Before T-group</th>
<th>After T-group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Face-to-face</td>
<td>Telephone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean SD</td>
<td>Mean SD</td>
</tr>
<tr>
<td>Experimental</td>
<td>10</td>
<td>2.93 .25</td>
<td>3.20 .58</td>
</tr>
</tbody>
</table>
TABLE 6

ANALYSIS OF VARIANCE OF MEAN ACCURATE
EMPATHY RATINGS BY TRAINED RATERS

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training conditions (A)</td>
<td>1</td>
<td>.10</td>
<td>.36</td>
</tr>
<tr>
<td>Subjects within groups</td>
<td>8</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of interview (B)</td>
<td>1</td>
<td>.28</td>
<td>.76</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>.10</td>
<td>.27</td>
</tr>
<tr>
<td>B x subjects within groups</td>
<td>8</td>
<td>.37</td>
<td></td>
</tr>
<tr>
<td>Mode of interview (C)</td>
<td>1</td>
<td>.71</td>
<td>1.87</td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>.05</td>
<td>.13</td>
</tr>
<tr>
<td>C x subjects within groups</td>
<td>8</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>ABC</td>
<td>1</td>
<td>.04</td>
<td>.17</td>
</tr>
<tr>
<td>BC x subjects within groups</td>
<td>8</td>
<td>.24</td>
<td></td>
</tr>
</tbody>
</table>
**TABLE 7**

**MEANS AND STANDARD DEVIATIONS OF MEAN ACCURATE EMPATHY RATINGS OF UNTRAINED RATERS**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Before T-group</th>
<th>After T-group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Face-to-face</td>
<td>Telephone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean     SD</td>
<td>Mean     SD</td>
</tr>
<tr>
<td>Experimental</td>
<td>10</td>
<td>5.46  1.11</td>
<td>4.40  1.62</td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>6.73  1.14</td>
<td>4.93  1.32</td>
</tr>
</tbody>
</table>
### TABLE 8

ANALYSIS OF VARIANCE OF MEAN ACCURATE EMPATHY RATINGS BY UNTRAINED RATERS

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training conditions (A)</td>
<td>1</td>
<td>1.80</td>
<td>0.660</td>
</tr>
<tr>
<td>Subjects within groups</td>
<td>8</td>
<td>2.74</td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of interview (B)</td>
<td>1</td>
<td>1.45</td>
<td>1.260</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>2.36</td>
<td>2.050</td>
</tr>
<tr>
<td>B x subjects within groups</td>
<td>8</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>Mode of interview (C)</td>
<td>1</td>
<td>14.01</td>
<td>5.020*</td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>.02</td>
<td>.007</td>
</tr>
<tr>
<td>C x subjects within groups</td>
<td>8</td>
<td>2.79</td>
<td></td>
</tr>
<tr>
<td>BC</td>
<td>1</td>
<td>.62</td>
<td>.267</td>
</tr>
<tr>
<td>ABC</td>
<td>1</td>
<td>.6</td>
<td>.285</td>
</tr>
<tr>
<td>BC x subjects within groups</td>
<td>8</td>
<td>2.32</td>
<td></td>
</tr>
</tbody>
</table>

* Approached significance p = .06
### TABLE 9

**MEANS AND STANDARD DEVIATIONS OF RELATIONSHIP QUESTIONNAIRE SCORES**

<table>
<thead>
<tr>
<th>Group</th>
<th>Before T-Group</th>
<th>After T-Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Experimental</td>
<td>10</td>
<td>33.2</td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>36.8</td>
</tr>
</tbody>
</table>
TABLE 10

ANALYSIS OF VARIANCE OF RELATIONSHIP QUESTIONNAIRE SCORES

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training conditions (A)</td>
<td>1</td>
<td>.05</td>
<td>.0008</td>
</tr>
<tr>
<td>Subjects within groups</td>
<td>8</td>
<td>64.88</td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time of interview (B)</td>
<td>1</td>
<td>6.05</td>
<td>.2008</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>54.45</td>
<td>1.8075</td>
</tr>
<tr>
<td>B x subjects within groups</td>
<td>8</td>
<td>30.13</td>
<td></td>
</tr>
</tbody>
</table>
VIII. DISCUSSION

The results indicate no significant differences between the experimental and control groups after the T-group experience on either measure (Relationship Questionnaire or interview segment-ratings) of empathy. It would appear that the T-group method is ineffective in producing empathy change. However, there are a number of possible factors which may have contributed to this finding. The procedure for conducting interviews was somewhat artificial in that the interviewees were not initiating interview contact, but rather the therapist initiated contact within a context of evaluation of services. This variable could be assessed by sampling actual interviews initiated by a client. Another variable that has been posited to influence training is the trainer (Pierce, Carkhuff, and Berenson, 1967; Pollack, 1968). The present study utilized only one trainer. A procedure utilizing multiple trainers and varying trainer variables differentially would be necessary to assess the influence of this variable. In addition, the technique utilized by the trainer was not varied in the present study. Different trainer techniques have been noted to affect level of empathy of the trainees in several studies (Baldwin and Lee, 1965; Bunker and Knowles, 1967; Delaney and Heimann, 1966; Payne and Gralinski, 1968; Truax and Carkhuff, 1967). One technique that could be utilized would be to provide the trainer with the specific task of increasing empathy. A fourth variable is
the length of training. The training time in the present study was 12 consecutive hours which may be too abbreviated to have a significant effect for this particular type of training. (Most T-groups are longer than 12 hours.) Kassarjian (1965) and Herron (1967) question the validity of brief training procedures, while Myers, Myers and Goldberg (1969) and Reddy (1969) achieved positive results with relatively brief training procedures. Another variable of importance is the time at which the post-training measures were collected. Harrison (1966), Shutz and Allen (1966) and Pilkey (1967) all found that the effects of a training experience were greater when measured at a later point in time rather than immediately after the training experience. In the present study, the post-training measures were collected within 48 hours of the T-group experience which may not have allowed sufficient time for the effects of training to be manifested.

The validity of the ratings obtained in this study (or of the construct of Accurate Empathy) may be questioned. The validity of the Accurate Empathy Scale was based on its correlation with therapeutic outcome (Truax and Carkhuff, 1967). However, in the present study the validity of the ratings is unknown, since the ratings have not been correlated with therapeutic outcome of the interviewees utilized in this study. Nor has there been any demonstration of agreement between the present raters and the original raters. That is, the present raters have not rated the original criterion tapes (those tapes known to be correlated with outcome). Such ratings could be correlated with the ratings of the original raters. Furthermore, since reliability
sets the upper limits that validity can obtain, the reliability coefficients of .534 and .503 in the present study, although of a magnitude acceptable in other studies in this area, could be viewed as less than adequate. Techniques should be developed to increase the reliability with which therapy tapes can be rated. Until greater interjudge reliability can be shown for the Accurate Empathy Scale, it is difficult to justify its use in further research.
IX. SUMMARY AND CONCLUSIONS

With the growing trend in the mental health field toward brief-psychotherapy and short-term crisis intervention, the telephone has come to assume an increasingly important role in communication. While there are many advantages to dealing with clients over the telephone, this type of interaction also presents many limitations, such as the unavailability of non-verbal cues and greater difficulty of outcome assessment. Furthermore, it is not known to what extent the results of outcome studies for face-to-face therapy interviews may be generalized to telephone therapy interviews. An objective measure is needed to evaluate the effectiveness of the therapists conducting telephone interviews. Such a measure would be invaluable in the selection, training and quality-control of personnel. Since often the caller refuses to identify himself, it is impossible in many cases to obtain follow-up information. Also, telephone contact does not permit use of the traditional assessment procedures upon which prediction of outcome is usually based. When outcome variables are very difficult to assess, one may focus on variables that have been shown to be related to outcome.

One measure that has been related to positive patient outcome is empathy as measured by the Truax Accurate Empathy Scale. Studies of many types of help-seekers, show that all respond positively to therapeutic encounters where the therapist offers high levels of Accurate Empathy, and many react unfavorably when Accurate Empathy is
offered at only low levels. These effects reportedly hold for a wide variety of therapists, regardless of training or theoretical orientation, and also for different types of therapy. The degree to which Accurate Empathy can be learned is central to the aim of effecting therapeutic change. Several studies suggest that Accurate Empathy is indeed "teachable" and that even persons untrained in psychopathology and personality dynamics can produce change in patients after training in the communication of Accurate Empathy. The T-group was considered as a method of training and the studies relating sensitivity training to outcome measures of increased interpersonal sensitivity were reviewed.

The present studies compared the Accurate Empathy ratings of therapists in face-to-face and telephone interviews, and examined the effect of group sensitivity-training on therapists' Accurate Empathy. The 10 Ss, all mental health professionals, conducted and tape recorded four interviews, two over the telephone and two in person. After completing the first two interviews (one of each mode) the experimental group participated in a sensitivity-training session lasting 12 hours. The control group received no such training. Following the training session two more interviews were completed. One four-minute segment was selected from the middle of each interview, yielding four excerpts for each S. These 40 segments were then re-recorded in random order on a master tape. Two groups of raters (trained and untrained) evaluated each excerpt using the Accurate Empathy Scale as a rating device.

Interrater reliabilities of .534 and .503 were achieved for the trained and untrained raters, respectively. It was found that the
two groups of raters differed significantly in their ratings of the interviews. The untrained raters rated the therapists as significantly more empathic in the face-to-face than in the telephone interviews, while there was no significant difference between modes for the trained raters. Sensitivity-training was not found to affect the Accurate Empathy ratings of therapists by either group of raters.

These results were discussed in terms of the difference between trained and untrained raters, the relative influence of verbal and non-verbal cues and variables which may modify the effectiveness of sensitivity-training. Caution was advised in the use of a scale which reports such low levels of interrater reliabilities.
APPENDIX I

DESCRIPTION FOR TRAINING RATERS
CURRENT PRACTICES IN TRAINING TAPE-RATERS
AT THE ARKANSAS REHABILITATION RESEARCH
AND TRAINING CENTER

University of Arkansas

by

G. Frank Lawlis, Ph. D.

The Arkansas Rehabilitation Research and Training Center has adopted the ratings of psychotherapy tapes as part of a continuing research program. The major source of these tapes have been from psychiatrists and psychologists who are participating in the NIMH Project; however, there are other projects that contribute to the files. At present, it is estimated that there are more than 1,000 in the library, which is more than all the other collections in the nation combined.

At present, most of the tapes are rated by three independent raters on the 3-dimensions of Accurate Empathy, Non-Possessive Positive Regard, and Genuineness of the therapist and the dimension of Depth of Exploration of the client. All the scales have been developed by Dr. Charles B. Truax, and are discussed in his book, *Toward Effective Counseling and Psychotherapy: Training and Practice*.

Since part-time, undergraduate students are used as judges, they must be trained to rate dimensions with reliability. There are three important aspects of this particular program that must be kept in mind when attempting to use students in this way: (1) the raters are not to be trained as psychologists, but rather as judges rating one, and only
one, aspect of the interview. Such references as to "good" or "bad" responses or therapy merely confuse students. There is also the danger that such evaluations will cause the students to create their own internal set of psychotherapy. (2) Each student brings into the situation his own sensitivity of what is said and how it was meaning. We have found that students who have poor interpersonal relationships, as well as poor grades, have the poorest reliability ratings. (3) The tapes themselves can be of such quality so as to cause poor ratings. This may be due to technical reasons such as the tape recorder, and it can be due to the therapist with the client not speaking clearly. Another difficulty encountered in terms of poor reliability scores is what is known as "flat data" which means that there is such a low amount of variance or range in the ratings that relationship coefficients are significantly reduced. For example, if all interactions are basically stage 3, and all raters agree, the reliability would be .00 because of such little variance. The real danger in "flat data" is that the raters will begin to make discriminations in the data where there should be any to change the monotony. Thus, raters begin to have a set that is erroneous and confuses them on other data. If possible, such data should be intraspersed with other data if detected in time.

The actual training takes three main phases: definition, clarity, and practice. Definition refers to having the student understand the meaning of the terms used to depict the specific dimension. This is usually done in group discussion. During the group discussion period, the three therapists variables are focused upon by having the students
themselves respond each other and being empathic, warm, and genuine. The therapist dimension of exploration is also discussed after the therapist variables. In this way, the student can better understand the concept they will be rating, and better discriminate one dimension from the others. For example, a most common error is the confusion between empathy and warmth. This phase usually takes approximately a week.

Clarity refers to the student clarifying the stages in his respective scale. In this phase, the students do not study any skill except his own. This is accomplished by listening to training tapes. The training tapes have several example segments of the different stages and a student is told what the example stage is. This can be done in a group or individually, and usually takes a week.

Practice refers to the students in actual rating of criterion tapes. These are tapes that have been reliably rated in the past and the student does not know the ratings. Some measure of confidence can be evaluated as to whether the rater is reliable enough to begin rating "real data." If the reliabilities are above .50, then a group of raters are permitted to begin judging research data. If the coefficient is poor, the group begins the clarity stage again and proceeds through the series. This phase takes approximately one and one-half weeks. In our experience approximately 70% of the students reach this stage. Evaluation of the raters continue after the last phase. After each research project, or within certain time periods, reliabilities are checked and data is evaluated as being flat. Inner
correlations between raters of the same scale are continuously run for a check of each judge with his associates, as well as occasional checks by the student supervisor.

Generally, we tried to make the students rely as much as possible on the scales with little interpretation as possible from other references. External influences or extreneous variances cannot be controlled and ultimately contaminate reliability.

In summary, the training of a student rater becomes quite complex with the consideration of the specific levels of instruction with hopes to achieve, the different sensitivity levels of each student, and the condition of the data itself. By progressing to the stage of (1) definition of the scale dimension, (2) clarity of stages, and (3) practice, the student learns not only what to listen for, but also on what intervals to judge the therapeutic situation. A word of caution is given about too much structure in explanation because the tendency to contaminate the scale by over-interpreting it.

The faculty of the Arkansas Rehabilitation Research and Training Center are interested and willing to help other institutions who might be interested in the rating of psychotherapy. We would like anyone to feel free to contact any one of the following persons if they are interested:

Charles B. Truax, Ph.D.
Area Code: 501 575-4358  Fayetteville, Arkansas

G. Frank Lawlis, Ph.D.
Area Code: 501 575-4358  Fayetteville, Arkansas

Vernon H. Glenn, Ed.D
Area Code: 501 575-3656  Fayetteville, Arkansas
APPENDIX II

ACCURATE EMPATHY SCALE
ACCURATE EMPATHY SCALE

Stage 1:

Therapist seems completely unaware of even the most conspicuous of the client's feelings; his responses are not appropriate to the mood and content of the client's statements. There is no determinable quality of empathy, and hence no accuracy whatsoever. The therapist may be bored and disinterested or actively offering advice, but he is not communicating an awareness of the client's current feelings.

Stage 2:

Therapist shows an almost negligible degree of accuracy in his responses, and that only toward the client's most obvious feelings. Any emotions which are not clearly defined he tends to ignore altogether. He may be correctly sensitive to obvious feelings and yet misunderstand much of what the client is really trying to say. By his response he may block off or may misdirect the patient. Stage 2 is distinguishable from Stage 3 in that the therapist ignores feelings rather than displaying an inability to understand them.

Stage 3:

Therapist often responds accurately to client's more exposed feelings. He also displays concern for the deeper, more hidden feelings, which he seems to sense must be present, though he does not understand their nature or sense their meaning to the patient.
Stage 4:

Therapist usually responds accurately to the client's more obvious feelings and occasionally recognizes some that are less apparent. In the process of this tentative probing, however, he may misinterpret some present feelings and anticipate some which are not current. Sensitivity and awareness do exist in the therapist, but he is not entirely "with" the patient in the current situation or experience. The desire and effort to understand are both present, but his accuracy is low. This stage is distinguishable from Stage 3 in that the therapist does occasionally recognize less apparent feelings. He also may seem to have a theory about the patient and may even know how or why the patient feels a particular way, but he is definitely not "with" the patient. In short, the therapist may be diagnostically accurate, but not empathically accurate in his sensitivity to the patient's current feelings.

Stage 5:

Therapist accurately responds to all of the client's more readily discernible feelings. He also shows awareness of many less evident feelings and experiences, but he tends to be somewhat inaccurate in his understanding of these. However, when he does not understand completely, this lack of complete understanding is communicated without an anticipatory or jarring note. His misunderstandings are not disruptive by their tentative nature. Sometimes in Stage 5 the therapist simply communicates his awareness of the problem of understanding another person's inner world. This stage is the midpoint of the continuum of accurate empathy.
Stage 6:

Therapist recognizes most of the client's present feelings, including those which are not readily apparent. Although he understands their content, he sometimes tends to misjudge the intensity of these veiled feelings, so that his responses are not always accurately suited to the exact mood of the client. The therapist does deal directly with feelings the patient is currently experiencing although he may misjudge the intensity of those less apparent. Although sensing the feelings, he often is unable to communicate meaning to them. In contrast to Stage 7, the therapist's statements contain an almost static quality in the sense that he handles those feelings that the patient offers but does not bring new elements to life. He is "with" the client but doesn't encourage exploration. His manner of communicating his understanding is such that he makes of it a finished thing.

Stage 7:

Therapist responds accurately to most of the client's present feelings and shows awareness of the precise intensity of most of the underlying emotions. However, his responses move only slightly beyond the client's own awareness, so that feelings may be present which neither the client nor therapist recognizes. The therapist initiates moves toward more emotionally laden material, and may communicate simply that he and the patient are moving toward more emotionally significant material. Stage 7 is distinguishable from Stage 6 in that often the therapist's response is a kind of precise pointing of the finger toward emotionally significant material.
Stage 8:

Therapist accurately interprets all the client's present, acknowledged feelings. He also uncovers the most deeply shrouded of the client's feelings, voicing meanings in the client's experience of which the client is scarcely aware. Since the therapist must necessarily utilize a method of trial and error in the new uncharted areas, there are minor flaws in the accuracy of his understanding, but these inaccuracies are held tentatively. With sensitivity and accuracy he moves into feelings and experiences that the client has only hinted at. The therapist offers specific explanations or additions to the patient's understanding so that underlying emotions are both pointed out and specifically talked about. The content that comes to life may be new but it is not alien.

Although the therapist in Stage 8 makes mistakes, these mistakes are not jarring, because they are covered by the tentative character of the response. Also, this therapist is sensitive to his mistakes and quickly changes his response in midstream, indicating that he has recognized what is being talked about and what the patient is seeking in his own explorations. The therapist reflects a togetherness with the patient in tentative trial and error exploration. His voice tone reflects the seriousness and depth of his empathic grasp.

Stage 9:

The therapist in this stage unerringly responds to the client's full range of feelings in their exact intensity. Without hesitation, he recognizes each emotional nuance and communicates an understanding
of every deepest feeling. He is completely attuned to the client's shifting emotional content; he senses each of the client's feelings and reflects them in his words and voice. With sensitive accuracy, he expands the client's hints into a full-scale (though tentative) elaboration of feeling or experience. He shows precision both in understanding and in communication of this understanding, and expresses and experiences them without hesitancy.

APPENDIX III

RAW DATA
## ACCURATE EMPATHY RATINGS OF THREE TRAINED RATERS

<table>
<thead>
<tr>
<th></th>
<th>Before T-Group</th>
<th></th>
<th>After T-Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Face-to-face</td>
<td>Telephone</td>
<td>Face-to-face</td>
<td>Telephone</td>
</tr>
<tr>
<td></td>
<td>$R_1$ $R_2$ $R_3$</td>
<td>$R_1$ $R_2$ $R_3$</td>
<td>$R_1$ $R_2$ $R_3$</td>
<td>$R_1$ $R_2$ $R_3$</td>
</tr>
<tr>
<td>Experimental</td>
<td>3 3 2</td>
<td>4 3 2</td>
<td>3 3 2</td>
<td>3 3 5</td>
</tr>
<tr>
<td></td>
<td>3 3 2</td>
<td>3 4 6</td>
<td>2 2 2</td>
<td>3 4 2</td>
</tr>
<tr>
<td></td>
<td>3 3 3</td>
<td>3 3 3</td>
<td>3 3 2</td>
<td>4 4 5</td>
</tr>
<tr>
<td></td>
<td>3 3 3</td>
<td>2 3 3</td>
<td>4 4 4</td>
<td>3 3 3</td>
</tr>
<tr>
<td></td>
<td>3 4 3</td>
<td>3 3 3</td>
<td>4 4 5</td>
<td>4 4 3</td>
</tr>
<tr>
<td>Control</td>
<td>3 3 3</td>
<td>4 3 4</td>
<td>3 3 3</td>
<td>3 3 3</td>
</tr>
<tr>
<td></td>
<td>3 3 2</td>
<td>3 3 4</td>
<td>3 3 4</td>
<td>3 3 5</td>
</tr>
<tr>
<td></td>
<td>4 4 3</td>
<td>3 3 3</td>
<td>3 3 4</td>
<td>3 4 3</td>
</tr>
<tr>
<td></td>
<td>3 3 3</td>
<td>3 3 3</td>
<td>3 3 4</td>
<td>3 3 2</td>
</tr>
<tr>
<td></td>
<td>3 3 4</td>
<td>3 4 5</td>
<td>3 4 3</td>
<td>3 3 7</td>
</tr>
</tbody>
</table>
## ACCURATE EMPATHY RATINGS OF THREE UNTRAINED RATERS

<table>
<thead>
<tr>
<th></th>
<th>Before T-Group</th>
<th></th>
<th>After T-Group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Face-to-face</td>
<td>Telephone</td>
<td>Face-to-face</td>
<td>Telephone</td>
</tr>
<tr>
<td>R1 R2 R3</td>
<td>R1 R2 R3</td>
<td>R1 R2 R3</td>
<td>R1 R2 R3</td>
<td>R1 R2 R3</td>
</tr>
<tr>
<td>4 3 3</td>
<td>2 3 2</td>
<td>5 7 6</td>
<td>3 3 2</td>
<td></td>
</tr>
<tr>
<td>2 7 7</td>
<td>5 8 9</td>
<td>4 4 3</td>
<td>4 8 7</td>
<td></td>
</tr>
<tr>
<td>5 5 8</td>
<td>5 3 4</td>
<td>3 3 6</td>
<td>6 2 3</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>6 5 8</td>
<td>5 3 5</td>
<td>5 7 7</td>
<td>8 4 8</td>
</tr>
<tr>
<td>6 7 6</td>
<td>5 3 4</td>
<td>7 8 8</td>
<td>2 4 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 2 7</td>
<td>6 4 8</td>
<td>8 3 6</td>
<td>5 3 5</td>
</tr>
<tr>
<td></td>
<td>6 8 9</td>
<td>7 7 5</td>
<td>8 7 8</td>
<td>3 2 6</td>
</tr>
<tr>
<td>Control</td>
<td>8 9 6</td>
<td>6 1 6</td>
<td>7 1 3</td>
<td>5 6 7</td>
</tr>
<tr>
<td></td>
<td>6 7 6</td>
<td>2 3 3</td>
<td>5 5 3</td>
<td>3 4 7</td>
</tr>
<tr>
<td></td>
<td>7 8 7</td>
<td>8 3 5</td>
<td>8 2 7</td>
<td>6 4 1</td>
</tr>
<tr>
<td>Subject</td>
<td>Before T-Group</td>
<td>After T-Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL</td>
<td>2 44</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 29</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 40</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 29</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONTROL</td>
<td>7 36</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8 32</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 41</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 39</td>
<td>41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX IV

RELATIONSHIP QUESTIONNAIRE
Please fill out the following questionnaire and return it to the secretary. Do not put your name on it. Your answers will be confidential and will not be seen by the person with whom you just had an interview.
RELATIONSHIP QUESTIONNAIRE

People feel differently about some people than they do about others. There are a number of statements below that describe a variety of ways that one person may feel about another person. Consider each statement carefully and decide whether it is true or false when applied to your feelings about the person with whom you have just had an interview. If the statement seems to be mostly true, then circle the T. If it is mostly false, then circle the F.

1. He understood my words, but did not know how I felt.  T  F
2. He understood me.  T  F
3. He understood exactly how I see things.  T  F
4. He was impatient with me.  T  F
5. He may have understood me but he did not know how I felt.  T  F
6. He often misunderstood what I was trying to say.  T  F
7. Sometimes he argued with me just to prove he was right.  T  F
8. He could read me like a book.  T  F
9. He was not very interested in what I had to say.  T  F
10. He acted too professional.  T  F
11. He ignored some of my feelings.  T  F
12. He knew more about me than I do about myself.  T  F
13. Sometimes he was so much "with me," in my feelings, that I was not at all distracted by his presence.  T  F
14. He sure made me think hard about myself.  T  F
15. Even when I could not say quite what I meant, he knew how I felt.  T  F
16. He helped me to know how I was feeling by putting my feelings into words for me.  

17. He must have understood me, but I often thought he was wrong.  

18. He seemed to follow almost every feeling I had while I was with him.  

19. He usually used just the right words when he tried to understand how I was feeling.  

20. Sometimes he was so much "with me" that with only the slightest hint he was able to accurately sense some of my deepest feelings.  

21. His voice usually sounded very serious.  

22. I often could not understand what he was trying to tell me.  

23. Whatever he said usually fit right in with what I was feeling.  

24. He sometimes seemed more interested in what he himself said than in what I said.  

25. He sometimes pretended to understand me when he really did not.  

26. He usually knew exactly what I meant, sometimes even before I finished saying it.  

27. He often led me into talking about some of my deepest feelings.  

28. He completely understood me so that he knew what I was feeling even when I was hiding my feelings.  

29. He helped me know myself better by pointing to feelings within me that I had been unaware of.  

30. I could learn a lot about myself from talking with him.  

31. He is the kind of person who might lie to me if he thought it would help me.  

32. When he saw me he seemed to be "just doing a job."
33. He never knew when to stop talking about something which was not very meaningful to me.  
   T  F

34. He sometimes cut me off abruptly just when I was leading up to something  
   T  F

35. There were lots of things I could have told him, but I was not sure how he would have reacted to them, so I kept them to myself.  
   T  F

36. He sometimes tried to make a joke out of something I felt really upset about.  
   T  F

37. He often pointed out what a lot of help he was giving me even though it didn't feel like it to me.  
   T  F

38. He used the same words over and over again, until I was bored.  
   T  F

39. I could lie to him and he would never know the difference.  
   T  F

40. I think he is dumb.  
   T  F

41. He probably laughs about the things I said to him.  
   T  F

42. I don't think he knew what is the matter with me.  
   T  F

43. There were times when I didn't have to speak, he knew how I felt.  
   T  F

44. He knows what it feels like to be ill.  
   T  F

45. He interrupted me whenever I was talking about something that really meant a lot to me.  
   T  F

46. There were times when he was silent for long periods, and then said things that didn't have much to do with what we had been talking about.  
   T  F

47. Often he made me feel stupid the way he used strange or big words.  
   T  F

48. He must think life is easy the way he talked about my problems.  
   T  F

49. He would talk to me, but otherwise he seemed pretty far away from me.  
   T  F

50. Even though he paid attention to me, he seemed to be just another person to talk with, an outsider.  
   T  F
REFERENCES


Fliess, R. The metapsychology of the analyst. The Psychoanalytic Quarterly, 1942, 11, No. 2.


Shapiro, J. G. Relationships between expert and neophyte ratings of therapeutic conditions. *Journal of Consulting Psychology, 1968, 32*, 87-89. (b)


Truax, C. B. *A scale for the measurement of accurate empathy.* Psychiatric Institute Bulletin, Wisconsin Psychiatric Institute, University of Wisconsin, 1961, 1, 12.


Truax, C. B. and Wargo, D. G. *Antecedents to outcome in group psychotherapy with hospitalized mental patients: Effects of therapeutic conditions, alternate sessions, vicarious therapy pretraining, and patient self-exploration.* Unpublished manuscript. University of Arkansas, 1966. (a)

Truax, C. B. and Wargo, D. G. *Antecedents to outcome in group psychotherapy with juvenile delinquents: Effects of therapeutic conditions, alternate sessions, vicarious therapy pretraining, and patient self-exploration.* Unpublished manuscript. University of Arkansas, 1966. (b)
Truax, C. B. and Wargo, D. G. Comment on Van der Veen's study "Effects of the therapist and the patient on each other's therapeutic behavior." Unpublished manuscript, University of Kentucky, 1966. (c)


