Contributions of selected therapist and client nonverbal behavior to ratings of empathy

Maxwell Robert Knauss

The University of Montana
THE CONTRIBUTIONS OF SELECTED THERAPIST AND CLIENT NONVERBAL BEHAVIOR TO RATINGS OF EMPATHY

By
Maxwell R. Knauss
B.A., University of Nevada, Las Vegas, 1974
Presented in partial fulfillment of the requirements for the degree of Master of Arts UNIVERSITY OF MONTANA 1978

Approved by:

Chairman, Board of Examiners
Dean, Graduate School

Date 5/18/78
In this study, an attempt was made to examine the contributions of therapist verbal empathy and client and therapist nonverbal behavior to ratings of accurate empathy in a roleplayed counseling interview for two therapist-client dyads. Four clinical psychology graduate students participated in an improvised roleplaying situation with two of the students playing the therapist role and the other two roleplaying the client role. The clients were given a role description but were required to improvise the actual dialogue. Systematic manipulations were made during the roleplaying in the levels of therapist and client nonverbal behavior. The study employed a completely randomized analysis of variance design. During the roleplaying, a three minute video tape segment was filmed for each of the 16 stimulus conditions. Manipulation check ratings for the level of therapist verbal empathy, client nonverbal behavior, and therapist nonverbal behavior were completed to insure that the intended experimental manipulations had been effective. Eighty introductory psychology students were taught to rate accurate empathy with a brief training procedure. A different group of five subjects rated each of the videotape segments and these ratings served as the dependent variable in the experiment. The statistical analysis revealed two significant main effects and five significant interactions. However, the unreliability of the empathy ratings made it impossible to determine if the results supported the experimental hypotheses.

The results suggested that nonverbal behaviors had a significant impact on empathy as rated by the Accurate Empathy Scale. This pattern of results was discussed in light of recent criticisms of the Accurate Empathy Scale's reliability and construct validity. While further research on the impact of nonverbal behaviors upon rated empathy is indicated, it was concluded that this research should wait until these basic measurement problems are resolved. A multitrait-multimethod strategy for establishing the scale's reliability and construct validity was proposed.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF FIGURES AND TABLES</td>
<td>v</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Definition and History of Accurate Empathy</td>
<td>1</td>
</tr>
<tr>
<td>Description of the Truax Accurate Empathy Scale</td>
<td>3</td>
</tr>
<tr>
<td>Validity and Reliability of the Truax Accurate Empathy Scale</td>
<td>4</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>10</td>
</tr>
<tr>
<td>Accurate Empathy as a Dyadic Variable</td>
<td>17</td>
</tr>
<tr>
<td>Nonverbal Behavior</td>
<td>22</td>
</tr>
<tr>
<td>Experimental Design and Hypothesis</td>
<td>27</td>
</tr>
<tr>
<td>II. METHOD</td>
<td>30</td>
</tr>
<tr>
<td>Construction of Stimulus Material</td>
<td>30</td>
</tr>
<tr>
<td>Checks on the Manipulations of the Independent Variables</td>
<td>32</td>
</tr>
<tr>
<td>Subjects</td>
<td>35</td>
</tr>
<tr>
<td>Procedure</td>
<td>35</td>
</tr>
<tr>
<td>III. RESULTS</td>
<td>38</td>
</tr>
<tr>
<td>Manipulation Checks</td>
<td>38</td>
</tr>
<tr>
<td>Experimental Results</td>
<td>38</td>
</tr>
<tr>
<td>IV. DISCUSSION</td>
<td>44</td>
</tr>
</tbody>
</table>

iii
## Summary

### References

### Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. The Truax Accurate Empathy Scale</td>
<td>72</td>
</tr>
<tr>
<td>B. Melloh's Schematic Presentation of the Truax Accurate Empathy Scale</td>
<td>76</td>
</tr>
<tr>
<td>C. Accurate Empathy Rating Form</td>
<td>78</td>
</tr>
<tr>
<td>D. Client Role Description</td>
<td>80</td>
</tr>
<tr>
<td>E. Nonverbal Behavior Instructions</td>
<td>82</td>
</tr>
<tr>
<td>F. Overview of Accurate Empathy and Therapist Role Instructions</td>
<td>84</td>
</tr>
<tr>
<td>G. Nonverbal Behavior Rating Form</td>
<td>88</td>
</tr>
<tr>
<td>H. Authenticity Rating Form</td>
<td>90</td>
</tr>
<tr>
<td>I. Experimental Instructions</td>
<td>92</td>
</tr>
<tr>
<td>J. Manipulation Check Preratings for Final Videotape Segments</td>
<td>98</td>
</tr>
</tbody>
</table>
LIST OF FIGURES AND TABLES

Figures

1. Completely Randomized Factorial  
   2 X 2 X 2 X 2 Design .................. 29

2. Scatter Plot for the Audiotape  
   Empathy Ratings .......................... 43

3. Proposed Multivariate Experimental  
   Design .................................. 59

Table

1. Significant F-Tests in the  
   Analysis of Variance ....................... 41
CHAPTER I

INTRODUCTION

Definition and History of Accurate Empathy

In a 1957 article, "The Necessary and Sufficient Conditions of Therapeutic Personality Change," Carl Rogers discussed the factors which he felt were both necessary and sufficient to initiate the process of constructive personality change within the client. Although Rogers discussed six factors, three of which referred to characteristics of the client, this article stimulated a research focus on those characteristics of the therapist which would allow him to facilitate constructive personality change in his interactions with the client. Rogers felt that there were three vital conditions which were attitudinal in nature and which the therapist must possess if he were to facilitate this constructive personality change. These conditions were: 1) that the therapist be genuine and honest in his relationship with the client; 2) that the therapist experience an unconditional positive regard for the client, and 3) that the therapist experience an empathic understanding of the client's private world and convey something of that understanding to the client.
Rogers hypothesized that if the therapist possessed these three qualities and if the client perceived them, then an orderly process of client personality change would be initiated and maintained. Rogers advanced these hypotheses to stimulate experimental investigation of the causal factors of personality change. He stated emphatically that all of the constructs were amenable to operational definition and that it was necessary to test these constructs in a rigorous, experimental manner if a better understanding of the process of psychotherapy were to be gained. The therapist variables came to be known as congruence, unconditional positive regard, and empathy (Rogers, 1961, pp. 61-62). Rogers regarded these variables as the facilitative core conditions of personality change.

Rogers' theoretical formulations served as an impetus for investigators to construct scales which would permit valid and reliable measurement of these core conditions. Of particular interest to the present study, Truax (1961a, 1961b) developed individual scales to measure each core condition. Rogers (1957) characterized empathy as the ability

To sense the client's world as if it were your own without ever losing the 'as if' quality . . . . When the client's world is this clear to the therapist and he moves about in it freely, then he can both communicate his understanding of what is clearly known to the client and can also voice meanings in the client's experience of which the client is scarcely aware.
However, in attempting to operationalize the construct, Truax and Mitchell (1971) stated that,

When empathy is defined in terms of operational scales measuring the therapist's responses to the client it becomes clear that what is being measured is an interpersonal skill rather than simply an attitude or personality attribute, since a person can have an understanding, or empathic attitude, . . . without making an accurately empathic response.

Truax coined the term "accurate empathy" to emphasize the therapist's moment-to-moment accuracy in conveying his empathic understanding to the client. Truax (1961a) stated that,

Accurate empathy involves both 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.

Therefore, the construct "accurate empathy" encompassed both the therapist's sensitivity and his ability to accurately convey this sensitivity to the client. However, the depth of the therapist's empathic understanding can never be assessed directly but must be inferred from the accuracy with which he responds to the client's feelings.

Description of the Truax Accurate Empathy Scale

The scales developed by Truax to measure the core conditions are operational scales which are used by trained judges to rate the therapists' responses to their clients in either live or taped interactions. The Truax Accurate
Empathy Scale (the AE Scale) is a nine-point scale. It defines a continuum of accurate, empathic responding, specifying a low point in which the therapist appears totally unaware of even the most conspicuous feelings of the client and there is no evidence of empathic understanding in the therapist's responses. At the midpoint of the continuum, the therapist responds accurately to the client's more conspicuous feelings and conveys a concern for the client's deeper feelings. However, he is inaccurate in his understanding of these feelings. At the high point of the accurate empathy continuum the therapist not only accurately conveys his understanding of the client's current feelings, he also offers accurate additional understanding of the client's deeper feelings. This results in actual discussion of these deeper feelings and increases the client's self-understanding (Truax and Mitchell, 1971).

**Validity and Reliability of the Truax Accurate Empathy Scale**

After developing the scales to measure the facilitative core conditions, Truax engaged in an ambitious research program to establish their reliability and validity. He began by comparing four hospitalized patients who improved, with four hospitalized patients who deteriorated, on a number of personality tests. Two-minute segments which were taken from sessions over the course of therapy were rated by judges using the AE Scale for level
of accurate empathy offered by each patient's therapist. The results revealed that the improved patients consistently received a significantly higher level (p<.01) of accurate empathy from their therapists than did the deteriorated patients. Truax also noted little variability in the therapists' levels of accurate empathic responding over a six-month period (Truax, 1961b).

In an attempt to gain a better understanding of the relationship between level of accurate empathy and case outcome, Truax compared an entire early and late interview for fourteen schizophrenic patients. Five consecutive segments covering each entire therapy session were rated for level of accurate empathy. The results of the study indicated that the patients who received higher average levels of accurate empathy were the ones who showed improvement (p<.05). However, Truax found that the highest moments of empathy obtained throughout the interviews were more predictive (p<.01) of improvement on outcome measures. Although Truax stated that therapists varied in their level of accurate empathic responding, he did not discuss the nature or extent of this variation.

After establishing a relationship between level of accurate empathy and case outcome, Truax attempted to cross validate his findings with a new sample of patients (Truax, 1963). The level of accurate empathy offered to fourteen schizophrenic patients who had been in therapy
from six months to four and one-half years was rated from audiotape segments taken from every fifth interview. The correlation between case outcome, as determined by a number of criterion measures, and level of accurate empathy was significant (p<.01). Next, Truax attempted to extend these preliminary findings to an outpatient population (Truax, 1963). Fourteen hospitalized patients were compared with fourteen outpatients for the level of empathy which they received via tape segments taken from early and late psychotherapy. For both types of patients, Truax found that those cases with a favorable outcome received significantly higher levels of accurate empathy (p<.01) than those who deteriorated during therapy.

Truax et al. (1966a) attempted a cross validation of the findings with hospitalized schizophrenics and outpatients with a new therapist and patient sample. In this study the level of accurate empathy offered forty outpatients was studied in relation to two outcome measures. The results again indicated that the level of accurate empathy was significantly related to improvement on outcome measures. Moreover, in this study each therapist saw more than one patient and the therapists offering high levels of accurate empathy produced a 90% improvement rate while the therapists offering low levels of accurate empathy produced a 50% improvement rate. However, the possibility of systematic variation in the level of therapist accurate
empathy across different patients was not directly examined.

Truax, Carkhuff, and Kodman (1965) studied the relationship between the level of accurate empathy and outcome improvement. Forty hospitalized patients were given group therapy twice weekly over a three-month period. The patients in the groups receiving high levels of accurate empathy were compared with patients in the groups receiving relatively lower levels of accurate empathy. Pre- and post-therapy Minnesota Multiphasic Personality Inventories were administered to all patients. The results indicated statistically significant improvements (p<.05) on the 7 scale, 8 scale, and the Welsh Anxiety Index for the high empathy versus the low empathy groups. These preliminary findings suggested a similar relationship in group psychotherapy between level of accurate empathy and improvement on outcome measures as those established in individual psychotherapy.

Truax and Wargo (1966b) attempted to further confirm this relationship in group psychotherapy. Ratings of the three core conditions were correlated with nineteen outcome measures for eighty juvenile delinquents who had received three months of group psychotherapy. The results of the experiment indicated that patients receiving high levels of the core conditions showed significant improvement (p<.001) on eighteen outcome measures, while those receiving low
levels of the core conditions showed significant
deterioration (p<.001) on seventeen outcome measures and
nonsignificant improvement on the remaining two outcome
measures (p<.001). Truax, Wargo, and Carkhuff (1966)
replicated Truax and Wargo (1966b) on a sample of eighty
outpatients receiving group therapy. The results were
essentially similar. Patients receiving high levels of
all the core conditions showed above average improvement on
two outcome measures, while the reverse was true of
patients receiving low conditions (p<.001). However, when
the relationship between accurate empathy and the outcome
measures was analyzed separately, only six outcome measures
were related to high levels of empathy while seventeen were
related to low levels of empathy (p<.01). These findings
tend to equivocate the relationship between accurate empathy
and outcome improvement in group psychotherapy.

Truax, Wargo, and Silber (1966) attempted to
clarify this relationship. In an experiment involving
seventy institutionalized juvenile delinquents, forty were
involved in group therapy with therapists who through
previous research were known to provide high levels of the
three core conditions. The other thirty patients comprised
a control group and received no therapy. On all twelve
pre- and post-therapy measures, the patients receiving
group therapy showed significant increases (p<.10 to
p<.001) over the control group.
From the available experimental evidence, it can be concluded that Truax and others have established reasonable predictive validity for the AE scale. If a patient is involved in therapy with a therapist who offers high levels of rated accurate empathy, the chances are far greater that the patient will improve rather than remain unchanged or deteriorate as a result of therapy. This improvement has been determined by a number of outcome criterion measures including the Minnesota Multiphasic Personality Inventory (Truax, 1963; Truax and Wargo, 1966c; Truax, Wargo & Carkhuff, 1966; Truax, Carkhuff & Kodman, 1965), the Rorschach (Truax, 1963), the Minnesota Counseling Inventory (Truax & Wargo, 1966c; Truax, Wargo & Sibler, 1966), the Welsh Internalized Anxiety Scale (Truax, 1963; Truax & Wargo, 1966c; Truax, Carkhuff & Kodman, 1965), the Palo Alto Group Therapy Scale (Truax & Wargo, 1966c; Truax, Wargo & Carkhuff, 1966), time spent out of an institution on a twelve month follow-up (Truax, 1963; Truax & Wargo, 1966c; Truax, Wargo & Silber, 1966c), therapist and client rating of client improvement following psychotherapy (Truax et al., 1966a), and congruence between ideal and actual self concept as measured by the Q-sort technique. This relationship has been demonstrated with both institutionalized patients (Truax, 1961b; Truax, 1963; Truax, Carkhuff & Kodman, 1965; Truax & Wargo, 1966; Truax, Wargo & Silber, 1966) and outpatients (Truax, 1963;
Truax et al., 1966a; Truax, Wargo & Carkhuff, 1966). The diagnostic label given to these patients has ranged from juvenile delinquency (Truax & Wargo, 1966c; Truax, Wargo & Silber, 1966) to schizophrenia (Truax, Carkhuff & Kodman; Truax, 1963). The relationship between rated accurate empathy and positive outcome has been clearly demonstrated in individual psychotherapy (Truax, 1961b; Truax, 1963; Truax et al., 1966a). It has also been established albeit less clearly in group psychotherapy (Truax, Carkhuff & Kodman, 1965; Truax & Wargo, 1966c; Truax, Wargo & Silber, 1966; Truax, Wargo & Carkhuff, 1966). The interrater reliabilities in these studies have ranged from .42 to .79 as determined by average Pearson correlations, and from .50 to .95 as determined by Ebel intraclass reliabilities. However, this evidence does not establish the construct validity of the AE Scale.

**Construct Validity**

Several articles have criticized the AE Scale by stating that it measures something other than accurate empathy, as defined by Truax. Among the first investigators to make this criticism were Kiesler, Mathieu, and Klein (Rogers, Gendlin, Kiesler, Truax, 1967). After using the AE Scale in an intensive, long-term outcome study with hospitalized schizophrenics they stated that,

... rather than reflecting the accuracy and refinement of the therapist's responses, the
Accurate Empathy Scale seems to have been tapping a more global therapist quality—the therapist's communicated commitment to the therapy interaction and involvement in the problems of a specific patient in the interaction.

Muehlberg, Pierce and Drasgow (1969) arrived at a similar conclusion. In their study three successful, experienced therapists completed therapy sessions with a "standard" client. Tape segments from these sessions were rated for empathy, respect, genuineness, concreteness, and self-disclosure. The correlations between these ratings were then factor analyzed which revealed that a single factor accounted for nearly all the variance in the ratings. Collingwood, Hefele, Huehlberg and Drasgow (1970) named this single factor the "good guy" factor and stated that it reflected the therapist's consistency of effectiveness in successfully dealing with life. However, Muehlberg et al's findings should be regarded as tentative due to methodological weaknesses in their study. Two raters made each of the five therapist variable ratings in succession for each film segment. This brings the independence of these ratings into serious question. Also, the fact that each therapist saw only one client might tend to increase the correlations between the therapist variables and thus increase the loadings on the one factor, obscuring the existence of other possible factors.

Kurtz and Grummond (1972) completed one of the most methodologically sound studies to date aimed at establishing
construct validity for measures of therapist empathy. They correlated six measures of empathy with one client process measure and four outcome measures. The six different measures of therapist empathy were: The Affective Sensitivity Scale; ratings of three-minute audiotape segments by judges using Carkhuff's Empathic Understanding in Interpersonal Process Scale (a five-point derivation of the AE Scale); both the client's and the therapist's level of empathic understanding as assessed by the Barrett-Lennard Relationship Inventory; the therapist's attempt to reproduce his client's responses to a presession completion of the Interpersonal Checklist; and the therapist's attempt to predict his client's self-description on Langfield's Rep Test. The outcome measures included the Tennessee Self-Concept Scale, the MMPI, and both the clients' and the therapists' evaluations of outcome. The results of this study indicated that tape-judged empathy and the clients' perceptions of empathy were the only correlation approaching significance (p<.10). These were also the only measures of empathy that were positively and significantly related to the outcome measures. The other measures of empathy were generally negatively correlated with the outcome measures. Kurtz and Grummon stated that with the exception of the correlation between these two measures, their study was unable to establish construct validity. They closed by stating,
There is little doubt that there is a relationship, and perhaps a substantial relationship, between what has been called empathy and therapy outcomes. But the present empathy measures may in fact be tapping other aspects of the therapists' behavior and the therapeutic relationship which accounts for the findings.

Paré (1970) completed a study bearing on both the validity and the reliability of the AE Scale. This study attempted to test the validity of comparing results from separate research studies in which different judges rated accurate empathy. Paré hypothesized that such variables as socio-economic level, educational level, age, sex and degree of authoritarian would differentially influence raters' interpretations of the AE Scale and thus make comparisons of results from different studies invalid. Seventy-two raters differing on the above variables rated eighteen three-minute audiotape segments from psychotherapy interviews for level of accurate empathy. The experimental data supported none of Paré's hypotheses. In fact, there was a high degree of similarity rather than difference in accurate empathy ratings across different groups. The average correlation between different educational groups was .84. On the basis of these results, Paré concluded that it was valid to compare results between research studies using different judges to rate accurate empathy. Paré's study is also significant in that he was able to evolve a method for training subjects to rate accurate empathy after a one-half hour training session. Moreover,
these ratings were in close agreement with those made by highly trained raters at the University of Arkansas $r = .77$).

Chinsky and Rappaport question both the validity and the reliability of the AE Scale (Chinsky and Rappaport, 1970; Rappaport and Chinsky, 1972). They echo the criticisms of other investigators, stating that the scale appears to tap a more global therapist quality such as voice quality or "communicated commitment" to therapy. They also criticize the reliability of the scale by stating that the size of the reliability coefficients are related to the number of therapists rated. Chinsky and Rappaport support this argument by stating that the highest reliabilities ($r > .70$) reported in the research literature are achieved when the therapist N's are less than fifteen. They explain their contention that higher reliabilities are related to lower therapist N's by stating that nonindependent ratings yield spuriously high reliability coefficients. They reason that this is caused by judges rating more than one tape segment per therapist. To insure independent ratings in future research, they suggest that a given therapist should not be rated more than once by the same judge.

Bozarth and Krauft (1972) undertook an experimental examination of Chinsky and Rappaport's criticisms. The study involved a national sample of seventy-five experienced therapists who submitted audiotapes from sessions with a
total of 120 clients. A minimum of five three-minute segments per client were rated by three highly experienced judges using the AE Scale. The judges also rated the counselor characteristics of good therapist (GT) and likability (L) on five-point scales. The results showed that both L and GT were significantly (p<.01) correlated with accurate empathy. The results also indicated that when the variance of L is partialed out of GT, the correlation between GT and accurate empathy is still significant (p<.01). However, when the variance of GT is partialed out of L, the correlation between L and accurate empathy is negligible. These results suggest that perceptions of therapeutic competency tend to influence AE ratings much more than therapist likability. Bozarth and Krauft also computed Ebel intraclass reliabilities for twelve, 100-segment blocks (Ebel, 1951). The coefficients ranged from .64 to .85 with the majority being over .70. The average reliability of one randomly selected segment per therapist-client pair equalled .76. The average reliability for all of the segments per therapist-client pair was .68. Bozarth and Krauft's data does not support, and in fact provides contradictory evidence to, Chinsky and Rappaport's contentions that reliability is necessarily related to therapist N and that repeated ratings of a therapist-client pair results in spuriously inflated reliability coefficients.
Beutler, Johnson, Neville, and Workman (1973) also designed an experimental study to examine the criticisms of Chinsky and Rappaport. Their sample consisted of fifty-four inpatients and eight therapists. All AE ratings were made from typed transcripts to avoid contamination by therapist voice quality. Their results showed that when reliabilities were computed using multiple ratings of therapists the correlation coefficient equalled .79. However, when the reliability coefficients were based on single ratings of each therapist the reliability coefficient equalled .94. Again, Chinsky and Rappaport's criticisms were not upheld by the experimental data. More interestingly, Beutler et al. found that therapist intersession consistency of accurate empathic responding varied greatly (x value r = .08). They also found that therapists varied greatly in their level of accurate empathic responding from one patient to another. These results were corroborated by Gurman (1973). In this experiment, three previously identified high functioning therapists were compared with three previously identified low functioning therapists for level of empathic responding over several sessions. The results indicated that the high- and low-facilitative therapists tended to function at high and low empathic levels, respectively. However, both groups of therapists were inconsistent within and across different sessions in terms of the levels of their therapeutic
Accurate Empathy as a Dyadic Variable

The inconsistency in empathic therapist responding has led Gurman (1973), as well as other investigators, to question the conception of accurate empathy as a stable therapist quality. Recent research suggests that accurate empathy lacks the stability which Rogers and Truax assumed that it had. It is hard to conceive of accurate empathy as being either an attitude or a communicational skill which is not influenced by the patient, especially when it is so inconsistently manifested with different patients. However, the fact that rated levels of high empathy have been demonstrated to be positively and significantly related to positive therapy outcome leads to conceptual difficulties. If the AE Scale is not measuring a stable therapist quality which is uninfluenced by patient characteristics but which is responsible for positive change in the patient's behavior and personality, then what is it measuring? Several investigators (Van der Veen, 1965; Moos & MacIntosh, 1970; Beutler et al., 1973; Heck & Davis, 1973) have proposed that the AE Scale is measuring a quality of the patient-therapist interaction rather than a quality which resides solely with the therapist.

Van der Veen (1965) was the first experimenter to advocate this reconceptualization. He designed an
experiment in which each of three patients were seen by the same five therapists' and patients' consistency of functioning across different interactions. Four-minute tape segments from each dyadic interview were rated for accurate empathy and congruence (therapist variables), and for problem expression and immediacy of experiencing (patient variables). The results of the experiment revealed that both patient and therapist interview behavior was determined by the patient, the therapist and the particular patient-therapist pair.

Van der Veen not only found that therapist level of empathic responding varied differentially across clients; he also found that patients differentially influenced the AE ratings which therapists received. In Van der Veen's study, approximately 45% of the total variance of AE ratings was accounted for by the therapist alone. The other 55% of the variance was accounted for by the particular patient and by the interactions between the therapist, patient, and session. On the basis of his results Van der Veen argues for a methodological model for investigating accurate empathy, as well as therapeutic interactions in general. He feels that there are three possible determinants for the behavior of each person in the therapeutic interaction. These determinants are the person himself, the person with whom he interacts, and the particular combination of the two people. Van der Veen explicates
this model by stating:

Each of these sources of determinants is considered important for an understanding of what occurs in therapy. If the participant's behavior were completely independent of the other person, he would be totally uninvolved and mechanical in the relationship. If his behavior were totally dependent on the other, he would not exist as an independent influence. If what he did depended only on idiosyncratic elements in the particular relationship, the relationship events would be entirely unique and generalizations unfounded. While these alternatives taken singly are unreal, together they represent significant and complementary vantage points for considering the determinants of therapy behavior (Van der Veen, 1965, p. 19).

Moos and Clemes (1967) attempted to use the type of model proposed by Van der Veen to investigate various patient and therapist behaviors within a therapeutic interaction. Each of four therapists interviewed each of four patients for one session. Both the patients' and the therapists' behaviors were rated for the same five variables. Moos and Clemes chose variables which were more easily rated in an objective manner than the client-centered constructs which Van der Veen used. These variables included total activity, percentage of feeling words, percentage of action words, number of questions and number of reinforcements. The results of the experiment corroborated those of Van der Veen, indicating that both therapist and patient behaviors were influenced by the three determinants which Van der Veen specified in his methodological model. The results also indicated that
therapists showed much more behavioral variability across patients than did patients across therapists.

Heck and Davis (1973) completed an analogue experiment along this same vein. They investigated the effects of therapist cognitive style (concrete or abstract) and level of complexity of client response (concrete or abstract) on level of therapist accurate empathy. Forty counselors were classified as either high or low in cognitive style on the basis of the Paragraph Completion Test. They were then presented with twenty-four client responses varying in level of conceptual complexity and were required to make written responses to them. The results revealed that there was a significant interaction between counselors' cognitive styles and levels of client responses indicating that level of AE responding does not remain constant across different stimulus conditions. Heck and Davis regard their results as tentative because the written responses preclude nonverbal and other important factors. However, their results do suggest that client and therapist levels of cognitive organization interact to differentially effect the level of rated AE.

In an attempt to replicate Moos and Clemes (1967) and Van der Veen (1965), Moos and MacIntosh (1970) designed an experiment to assess both patient and therapist behavioral consistency along the behavioral dimensions utilized by Moos and Clemes, as well as on the
client-centered process variables utilized by Van der Veen. Six patients saw each of four therapists for two sessions. These sessions were then rated for the relevant variables. The results generally corroborated both Van der Veen's (1965) and Moos and Clemes' (1967) results, indicating that patient X therapist X session interactions accounted for large portions of the variance on most of the variables. For accurate empathy, in particular, there were significant between-patient effects, but no significant between-therapist effects or patient X therapist interaction effects. However, transposition of mean squares to variance components indicated that 29% of the total variance in AE ratings was accounted for by between-patient differences; 56% was accounted for by patient X therapist X session interactions; 15% was accounted for by patient X therapist interactions; and only 1% of the total variance was accounted for by between-therapist differences. Therefore, interactions accounted for large portions of the total variance in the AE ratings, even though the F ratios failed to reach significance.

Moos and MacIntosh state that their results, as well as the results of this whole research trend, require that the therapeutic relationship be viewed as an interdependent system. Adopting this view implies not only that the characteristics of the therapist influence the patient's behavior, but also that the characteristics of
the patient influence the therapist's behavior. In addition, the characteristics of both can interact to influence the behavior of each of them. Moos and MacIntosh feel that given this view, the question of whether or not accurate empathy is a therapist trait becomes irrelevant. They state that,

... the relevant research question is not whether accurate empathy is or is not a therapist trait; it is rather a further specification of the proportions of variance in accurate empathy accounted for by different sources in different settings (Moos and MacIntosh, 1970, p. 305).

Nonverbal Behavior

Although in his writings concerning the AE Scale Truax suggests that nonverbal as well as verbal behaviors are important in the communication of accurate empathy, he does not address the issue as to which nonverbal behaviors facilitate empathic communication. Truax does make nonspecific references to voice quality, posture and gestures. However, he does not explicitly discuss which of the specific nonverbal behaviors might be perceived as empathic or nonempathic. The AE Scale is defined almost exclusively in terms of the verbal components of accurate empathy. No provisions are made for rating nonverbal behaviors. Haase and Tepper (1972) state that verbal components are given the predominant emphasis in current conceptualization and measurement of accurate empathy, while nonverbal components are treated in a subordinate or
subsidiary manner. Although Truax directed his research toward the verbal components of accurate empathy he recognized that once the validity of these verbal components was established, the specification and validation of nonverbal components should be conducted. Truax states:

... since empathy seems to be of significance, it becomes important to know which specific behaviors among those now labeled as "empathic" or "warm" are doing the actual work; e.g., is the tonal quality of the voice a significant factor or only the understanding? At one point, parametric studies specifying exact functions must be carried out (Truax and Carkhuff, 1967, p. 141).

Recent research on nonverbal components of communication suggests that they may be as important or more important than verbal components in determining the nature of the perceived message. Mehrabian and Ferris (1967) varied three degrees of attitude in facial expression and three degrees of attitude in vocal expression. These facial-vocal combinations were then rated by subjects for attitude conveyed. The results showed that the nonverbal facial component accounted for one and one-half times as much variance in final ratings of attitude as did the vocal component.

In a study more relevant to discovering the nonverbal components of empathy, Shapiro, Foster and Powell (1968) designed an experiment in which trained judges rated still photographs for level of empathy, warmth, and genuineness. Untrained judges also rated the photographs
for the same variables. The results indicated that both trained and untrained judges were able to reliably rate empathy from still photographs. However, the results gave no indication of the cues which the judges relied upon to make their ratings. Fetz (1966) attempted to identify nonverbal behaviors correlated with the facilitative core conditions as measured by the Barrett-Lennard Relationship Inventory. The nonverbal behaviors which significantly correlated with clients' perceptions of empathy were forward body lean and horizontal and vertical hand gestures.

Shapiro (1968) investigated the relationship between judgments of the facilitative core conditions as rated via audio, video, and audio-video stimulus material. The results for the accurate empathy ratings produced significant correlations (p<.01) for the stimulus modes. These results suggested that nonverbal behaviors could be rated reliably for level of accurate empathy. These results were remarkable because the raters used the verbal-oriented AE Scale and were given no suggestions concerning the application of that scale in rating nonverbal behaviors for empathy.

Charney (1966) related postural congruence to patient and therapist verbalizations indicative of the quality of the therapeutic relationship. He found that high levels of postural congruence were associated with specific, positive, and inter-personal verbalizations.
Incongruent postures, on the other hand, were related to negational, nonspecific, and self-contradictory verbalizations. On the basis of these results, Charney suggested that postural congruence between the two members of a therapeutic interaction is an indication of rapport in psycho-therapy.

Strong, Taylor, Bratton and Loper (1971) investigated the impact of the frequency of counselors' nonverbal behavior on subjects' descriptions of the counselors using an adjective checklist. In a factorial experiment, Strong compared adjective ratings of counselor behaviors in a high and low nonverbal behavior condition and via audio or audio-video segments described the counselor more negatively than subjects rating audio segments. Strong et al. interprets these results as being due to visual cues which disrupted positive counselor stereotypes. Concerning the frequency of nonverbal behavior, the results indicated that the high frequency of nonverbal behavior was perceived as conveying warmth, friendliness, and casualness. The low frequency of nonverbal behavior resulted in perceptions of coldness, seriousness, and aloofness. Strong et al. conclude that the high frequency of nonverbal behavior increased the counselor's interpersonal attractiveness while the low frequency of nonverbal behavior decreased the counselor's interpersonal attractiveness.

In an attempt to determine the relative contribution
of verbal and nonverbal behaviors to rated accurate empathy, Haase and Tepper (1972) designed a repeated measures experiment in which judges rated film segments of a counselor roleplaying empathic responses. Twenty-six judges rated forty-eight film segments in which the therapist communicated either high, medium or low empathy as measured by the AE Scale. In each film segment, four nonverbal behaviors were systematically varied. These nonverbal behaviors included eye contact, trunk lean, body orientation and distance from the client. The results of the experiment suggest,

... that maintaining eye contact, forward trunk lean, close distance, and medium- and high-rated verbal empathy all independently contribute to higher levels of judged empathy (Haase and Tepper, 1973, p. 419).

In addition, transposition of mean squares to variance components revealed that nonverbal behaviors accounted for 66% of the variance in AE ratings. These results are in agreement with those of Mehrabian and Ferris (1967).

On the basis of several significant interactions between nonverbal behaviors and level of verbal empathy, Haase and Tepper propose a "compensatory model" of empathy. They suggest that engaging in certain nonempathic behaviors can be compensated for by engaging in other empathic behaviors. For instance, maintaining eye contact and leaning forward might improve the rating of a low verbal AE message. However, the reverse is also true. High levels
of verbal AE may deteriorate to nonempathic communications when the therapist does not maintain eye contact or leans back in his chair.

Haase and Tepper deserve much credit because they were among the first to conduct a parametric study to determine specific nonverbal behaviors which are perceived as empathic by judges. Their study does, however, contain two weaknesses which might be improved upon in future research. The first is that the content of the verbal empathic responses were taken verbatim from Truax and Carkhuff (1967) and Carkhuff and Barenson (1967). The judges were all trained therapists who had had previous exposure to the concept of accurate empathy. If the judges gained that exposure by reading either of the above books, there is a possibility of contamination of ratings. Another weakness is that the segments may be too short in length to adequately represent the quality of nonverbal behaviors as they occur in an actual therapist-client dyad.

**Experimental Design and Hypotheses**

The present study was designed to assess the effects of therapist verbal empathic behavior, and client and therapist nonverbal behavior as manifested in two therapist-client dyads upon judges' ratings of accurate empathy. The contributions of these selected behaviors to rated accurate empathy was investigated through the
construction of videotape segments of a simulated therapist-client interaction. This study conformed to a 2 (Therapist Verbal Empathy: high vs. low) X 2 (Therapist Nonverbal Behavior: high vs. low) X 2 (Client Nonverbal Behavior: high vs. low) X 2 (Dyad 1 vs. Dyad 2) analysis of variance design (see Figure 1). Combining all possible levels of the four independent variables yielded 16 stimulus conditions. Three-minute videotape segments exemplifying each stimulus condition were constructed. Each segment was rated for level of accurate empathy by a separate group of five judges. These final empathy ratings constituted the dependent variable in this experiment, and were analyzed to test the following hypotheses:

1) Minimally trained judges would be capable of reliably rating both audio and videotape segments for level of accurate empathy. Reliable rating was defined as an interrater reliability of greater than .60.

2) Therapists would receive significantly higher accurate empathy ratings when the level of their nonverbal behavior was congruent with the client's level of nonverbal behavior, as compared to when it was incongruent.

3) When the level of Therapist Verbal Empathy was low, an accompanying high level of Therapist Nonverbal Behavior would result in a significantly higher level of rated accurate empathy than when the accompanying level of Therapist Nonverbal Behavior was low.

4) The results of the experiment would be consistent across two therapist-client dyads.

5) Nonverbal Behavior would account for a significantly larger proportion of the total variance in the accurate empathy ratings than would Therapist Verbal Empathy.
<table>
<thead>
<tr>
<th>Therapist &amp; Nonverbal Behavior</th>
<th>Client &amp; Nonverbal Behavior</th>
<th>Dyads</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Verbal (a&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>High Emotion (c&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;1&lt;/sub&gt;b&lt;sub&gt;1&lt;/sub&gt;c&lt;sub&gt;1&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Low Nonverbal (b&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Low Behavior (c&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;1&lt;/sub&gt;b&lt;sub&gt;1&lt;/sub&gt;c&lt;sub&gt;1&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>High Verbal (a&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Low Emotion (c&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;1&lt;/sub&gt;b&lt;sub&gt;1&lt;/sub&gt;c&lt;sub&gt;2&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Low Nonverbal (b&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>High Behavior (c&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;1&lt;/sub&gt;b&lt;sub&gt;2&lt;/sub&gt;c&lt;sub&gt;2&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Low Verbal (a&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Low Emotion (c&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;2&lt;/sub&gt;b&lt;sub&gt;1&lt;/sub&gt;c&lt;sub&gt;1&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>High Nonverbal (b&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Low Behavior (c&lt;sub&gt;1&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;2&lt;/sub&gt;b&lt;sub&gt;1&lt;/sub&gt;c&lt;sub&gt;1&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Low Verbal (a&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>High Emotion (c&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;2&lt;/sub&gt;b&lt;sub&gt;2&lt;/sub&gt;c&lt;sub&gt;2&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Low Nonverbal (b&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Low Behavior (c&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Dyad 1 (a&lt;sub&gt;2&lt;/sub&gt;b&lt;sub&gt;2&lt;/sub&gt;c&lt;sub&gt;2&lt;/sub&gt;d&lt;sub&gt;1&lt;/sub&gt;)</td>
</tr>
</tbody>
</table>

Figure 1. Completely Randomized Factorial 2X2X2X2 Design
CHAPTER II

METHOD

Construction of Stimulus Material

The audiovideo stimulus material was videotaped at the University of Montana Clinical Psychology Center. Several filming sessions were required to complete the stimulus segments for each dyad. Two male graduate clinical psychology students experienced in conducting therapy roleplayed the therapists. Two male first-year graduate students roleplayed the clients. The role of the "client" was that of a first-quarter freshman experiencing interpersonal problems and having difficulty adjusting to the university environment (see Appendix D). The students playing the "client" were required to improvise actual dialogue. Both students were given the same role description and played that role across all the stimulus conditions. Both the "clients" and the "therapists" were requested to wear the same clothing to each taping session in an attempt to keep their physical appearance as constant as possible across the stimulus conditions.

The manipulation of both clients' and therapists' nonverbal behaviors involved two levels: 1) a high level
of nonverbal behavior which included forward trunk lean and high frequency of head nod, hand gesture, and eye contact, and 2) a low level of nonverbal behavior which included backward trunk lean, low frequency of head nod and hand gesture, and avoidance of eye contact. The level of both client and therapist nonverbal behaviors was manipulated through instructions from the experimenter indicating which behaviors to emphasize and which to suppress in each stimulus condition. The instructions were identical for both the clients and the therapists (see Appendix E).

The manipulations of the level of Therapist Verbal Empathy included two levels of empathy: a low level of empathy defined as one to three on the Truax AE Scale; and a high level of empathy defined as five to nine on the Truax AE Scale. In the low Therapist Verbal Empathy conditions, the therapists were instructed to approximate as closely as possible Level 2 AE responses. In the high Therapist Verbal Empathy conditions, the therapists were instructed to approximate Level 6 AE responses. The graduate students roleplaying the therapists were provided with a copy of the Truax AE Scale and with an overview and summary of the concept of accurate empathy. This overview included a description of the levels of AE they were asked to approximate in their roleplaying. The therapists were also given a list of specific behaviors and response styles
characteristic of either Level 2 or Level 6 accurate empathic responses to help them in their roleplaying (see Appendix F).

During the filming sessions, the experimenter instructed the clients as to the nonverbal behavioral level they should manifest, and the therapists as to the nonverbal behavioral and accurate empathy levels they should manifest before the particular film segments were taped. The clients and the therapists were kept blind as to each other's instructions.

**Checks on the Manipulations of the Independent Variables**

Before the videotape segments were presented to the groups of subjects for final AE ratings, they were subjected to ratings by three sets of judges to determine the success of the intended experimental manipulations of the independent variables. Two Psychology 235 students rated only the audio portion of the videotape segments for the level of Therapist Verbal Empathy. They received 10 hours of experimental credit for participating in the experiment.

These students were trained to use the Truax AE Scale by a procedure similar to that described in Rogers et al. (1967). They were given a copy of the Truax AE Scale (see Appendix A) and of Melloh's Schematic Presentation of the AE Scale (see Appendix B), and were also
presented with an overview of the concept of accurate empathy. The students were encouraged to ask questions and discuss the scale. Next, prerated audiotape segments were played to illustrate various levels of accurate empathic responding. The students then began to make practice ratings on audiotape segments of recorded therapy sessions. Intraclass reliability coefficients were computed for blocks of five segment ratings. When the reliability coefficient exceeded .60, the students began to rate the audio portion of the videotape segments for accurate empathy. These judges were frequently asked to define various levels of the AE Scale to insure that they were making valid ratings. Those segments in which the ratings of level of Therapist Verbal Empathy agreed with the intended experimental manipulation were further rated to check the successfulness of the nonverbal behavior manipulations and to check for the authenticity of the videotape segments.

Two doctoral candidates in clinical psychology with experience supervising psychotherapy and viewing videotaped therapy sessions rated the segments for both level of nonverbal behavior and authenticity. These students rated the segments for level of nonverbal behavior first. After the intended nonverbal manipulation was explained and the students understood the difference between the high and low levels of nonverbal behavior, they rated the segments using a five-point scale (see Appendix G). Only the video portions’
of each segment were presented to these judges for ratings. On each segment, the judges rated the level of client nonverbal behavior as well as the level of therapist nonverbal behavior. Those segments in which the judges' ratings of the levels of nonverbal behavior were not in agreement with the intended experimental manipulation were discarded. The remaining segments were rerandomized and rated for authenticity.

When making the authenticity ratings, the judges viewed both the audio and video portions of the tape segments. For the purposes of this study, authenticity was defined as the degree to which the roleplayed segments simulated an actual client-therapist interaction for level of authenticity on a five-point scale (see Appendix H). Only the overall authenticity ratings were used to determine if the film segments were used for the final ratings. The level of authenticity was held relatively constant across the final selected segments so that differential plausibility of tape segments would not result in confounding of the final AE ratings. Only the overall authenticity ratings were used to determine if the film segment was used for the final ratings. No segment receiving less than a 3 from either judge on overall authenticity was presented for final ratings. The client and therapist authenticity scales were used to identify sources of implausibility in those segments representing a
given stimulus condition which needed to be refilmed.

After the three sets of manipulation-check ratings were completed, a segment representing each of the 16 stimulus conditions meeting the criteria of each of the three manipulation-check ratings was edited onto a master tape. In the event that more than one segment met the manipulation-check criteria, the segment in which rated AE most closely approximated the desired level of therapist Verbal Empathy (Level 2 in the low Therapist Verbal Empathy conditions; Level 5 in the high Therapist Verbal Empathy conditions) was edited onto the master tape. When none of the segments representing a given stimulus condition met the criteria of the manipulation-check ratings, the segment was refilmed and subjected to the same manipulation-check ratings.

**Subjects**

The eighty male subjects who served as the final judges were solicited from the University of Montana psychology 110 and 235 classes during fall quarter, 1977. For their participation, subjects received one experimental credit toward the five required for their classes.

**Procedure**

The subjects were trained to rate accurate empathy in groups of three to five. Each videotape segment was rated by five subjects. The experimenter trained and ran
all of the subjects in one-hour experimental sessions following the procedure developed by Paré (1970).

The experimenter used a structured format in training subjects which could be easily repeated across training groups (see Appendix I). The subjects were given an overview of the concept of accurate empathic responding and its importance in counseling and psychotherapeutic relationships. Rating scales which incorporated Melloh's schema were distributed (see Appendix C). The full Truax AE Scale was not used because it was felt to be too wordy and detailed for subjects to adequately grasp in a one-hour training session. The format of Melloh's rating scale was explained and questions were answered. Two three-minute tape segments illustrating Level 2 and Level 6 empathic responding were played for the subjects. They were encouraged to discuss each rating and relate it to the scale criteria. The subjects were then required to rate two practice audiotape segments for level of accurate empathy. These segments had both been given ratings of Level 4 AE by Truax's highly trained raters. To insure independent ratings, the subjects were not allowed to discuss the segments while listening to them or making their ratings. After the subjects rated each of these segments, the experimenter encouraged a discussion of the ratings. No effort was made to force a majority opinion on all the group members. Instead, each subject was encouraged
to justify his rating on the basis of the scale criteria. After each subject had discussed his rating, the experimenter revealed the prerating and subjects were encouraged to adjust their conceptions of the scale criteria to more closely match it.

When the subjects completed rating the practice audiotape segments, they made final AE ratings on one of the videotape segments. No instructions were given concerning the use of the rating scale with videotaped material. The subjects were merely told to make their ratings on the basis of "how well the therapist understands the client's current feelings and conveys that understanding to the client." The videotape segments had been randomly assigned to each group before the experiment began. The experimenter was kept blind to this assignment during the training of the subjects to control for the possible influence of experimenter bias upon final AE ratings. After the subjects finished rating the video segment, they were informed that a handout detailing the nature and results of the experiment would be distributed in their psychology class in four weeks. The subjects were then thanked for their participation, given an experimental credit slip, and dismissed.
CHAPTER III

RESULTS

Manipulation Checks

The manipulation checks for the level of Therapist Verbal Empathy, level of Client Nonverbal Behavior, level of Therapist Nonverbal Behavior, and degree of authenticity indicated that the 16 final videotape segments met the established criteria (see Appendix J). Raters achieved interrater correlations of .89 for the AE ratings, 168 for the nonverbal behavior ratings, and .55 for level of authenticity.

Experimental Results

The interrater reliabilities of the minimally trained judges' ratings was determined for both the audio practice tapes and the final videotape segments. An intraclass correlation was computed for the 80 judges across the two audiotape segments and yielded a reliability coefficient of -.08. To corroborate this estimation of reliability, a Pearson product-moment correlation was computed for the same ratings. It also yielded a reliability coefficient of -.08. The overall means for these audio segments were 2.90 and 6.69. Since these segments had been
prerated as 4.00 by the Truax raters, it was expected that
the judges' ratings on these two segments would result in
similar means. An F-test indicated that the difference
between the means was significant at the p<.01 level. To
determine the effect of this large difference in the size
of the means upon the interrater reliability coefficient, a
corrected Pearson product moment correlation was computed
by subtracting the respective mean from each individual
rating before computing the correlation coefficient. The
correlated interrater reliability coefficient was -.02.
The interrater reliability for the final video segments was
estimated with an intraclass correlation procedure. It
resulted in an interrater reliability coefficient of -.07.
Thus, these two independent estimates of interrater
reliability revealed an almost total lack of reliability in
the final judges' accurate empathy ratings.

Despite this lack of interrater reliability, analysis
of variance indicated significant main effects for Therapist
Verbal Empathy, F(1, 64) = 60.313, p<.001, and for Client
Nonverbal Behavior F(1, 64) = 5.49, p<.02. There were
also five significant interactions between the four
independent variables: Therapist Verbal Empathy X Therapist
Nonverbal Behavior, Client Nonverbal Behavior X Dyad,
Therapist Nonverbal Behavior X Client Nonverbal Behavior,
Therapist Nonverbal Behavior X Client Nonverbal Behavior X
Dyad, and Therapist Verbal Empathy X Client Nonverbal.
Behavior X Dyad (see Table 1). Due to the unreliability of the AE ratings these interactions were not further analyzed.

Orthogonal F-tests were computed to test Hypothesis 3 which stated that when the level of Therapist Verbal Empathy is low, an accompanying high level of Therapist Nonverbal Behavior will result in significantly higher ratings of AE than when the accompanying level of Therapist Nonverbal Behavior is low. It was therefore hypothesized that $X_{a2b1c1d1} + X_{a2b1c2d1} > X_{a2b2c1d1} + X_{a2b2c2d1}$ and that $X_{a2b1c1d2} + X_{a2b1c2d2} > X_{a2b2c1d2} + X_{a2b2c2d2}$. Both hypotheses were supported by significant differences ($p < .001$) in the predicted directions.

Expected mean squares were derived for each independent variable as well as for the interaction terms. This allowed an estimation of each variable's contribution to the total variance in the dependent variable. Proportions were computed for the verbal and nonverbal variance components. Therapist Verbal Empathy contributed eleven times more variance to the final AE ratings than the Therapist and Client Nonverbal Behavior variables combined. Even when the Therapist Verbal Empathy variable was compared to Therapist and Client Nonverbal variables plus interaction terms containing nonverbal variables, it accounted for an equal amount of variance.

An a-posteriori examination of the audiotape AE ratings was undertaken to gain an understanding of how
<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Therapist Verbal Empathy</td>
<td>1</td>
<td>130.05</td>
<td>60.313***</td>
</tr>
<tr>
<td>Therapist Nonverbal Behavior</td>
<td>1</td>
<td>3.20</td>
<td>1.484</td>
</tr>
<tr>
<td>Client Nonverbal Behavior</td>
<td>1</td>
<td>12.80</td>
<td>5.936*</td>
</tr>
<tr>
<td>Dyad</td>
<td>1</td>
<td>0.20</td>
<td>0.093</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Therapist Nonverbal Behavior</td>
<td>1</td>
<td>16.20</td>
<td>7.513**</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Client Nonverbal Behavior</td>
<td>1</td>
<td>5.00</td>
<td>2.319</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Dyad</td>
<td>1</td>
<td>1.80</td>
<td>0.835</td>
</tr>
<tr>
<td>Therapist Nonverbal Behavior X Client Nonverbal Behavior</td>
<td>1</td>
<td>8.45</td>
<td>3.919*</td>
</tr>
<tr>
<td>Therapist Nonverbal Behavior X Dyad</td>
<td>1</td>
<td>1.25</td>
<td>0.580</td>
</tr>
<tr>
<td>Client Nonverbal Behavior X Dyad</td>
<td>1</td>
<td>26.45</td>
<td>12.267***</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Therapist Nonverbal Behavior X Client Nonverbal Behavior</td>
<td>1</td>
<td>0.05</td>
<td>0.023</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Therapist Nonverbal Behavior X Dyad</td>
<td>1</td>
<td>0.45</td>
<td>0.209</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Client Nonverbal Behavior X Dyad</td>
<td>1</td>
<td>8.45</td>
<td>3.919*</td>
</tr>
<tr>
<td>Therapist Nonverbal Behavior X Client Nonverbal Behavior X Dyad</td>
<td>1</td>
<td>12.80</td>
<td>5.936*</td>
</tr>
<tr>
<td>Therapist Verbal Empathy X Therapist Nonverbal Behavior X Client Nonverbal Behavior X Dyad</td>
<td>1</td>
<td>0.84</td>
<td>0.371</td>
</tr>
<tr>
<td>Error</td>
<td>64</td>
<td>2.16</td>
<td>---</td>
</tr>
</tbody>
</table>

*p < .05  **p < .01  ***p < .001
unreliable data could produce significant analysis of variance results. A scatter plot diagram was constructed for these ratings (see Figure 2). There appeared to be a core of agreement among the raters with a small number of outliers. It was hypothesized that there were enough outlying ratings to reduce what otherwise might have been a moderate interrater reliability to a very low one. To test this hypothesis, the standard deviation was determined for each tape segment ($\sigma_{\text{segment #1}} = 1.31, \sigma_{\text{segment #2}} = 1.35$). Any pair of ratings which contained a rating that was more than two standard deviations from the mean (13% of the total) was discarded. The recalculated interrater reliability yielded a .84 Pearson product moment correlation coefficient.
Figure 2. Scatter Plot for the Audiotape Empathy Ratings. The underlined points were dropped for the recomputed reliability coefficient.
CHAPTER IV

DISCUSSION

The first hypothesis predicted that minimally trained judges would be capable of rating both audio and videotape segments reliably ($r > .60$). The two independent estimates reveal an almost total lack of agreement between the judges and thus provided no support for this hypothesis. This is contrasted to the high interrater reliability ($r = .89$) for the judges who made the AE manipulation check ratings for the audio portion of the final video segments. Both sets of judges were trained by the same experimenter, but the manipulation check judges were given approximately eight hours of training before they began the experimental ratings, whereas the final judges were given only one-half hour of training. In addition, the manipulation check judges used both the Truax AE Scale and Melloh's Schematic Presentation of the AE Scale to make their ratings, while the final judges used only the Schematic Presentation of the AE Scale.

This pattern of results represents a failure to replicate the results obtained by Paré (1970). On the basis of his results, a high level of agreement was expected
on the AE ratings made by the highly trained and minimally trained judges. The present study used Paré's audiotape segments and duplicated his procedure for training minimally trained judges as closely as possible. Similar results were expected, but not obtained. A major difference between the two studies is that Paré's judges rated 18 tape segments while the judges in the present study rated only three tape segments. It is possible that Paré's judges attained a higher level of interrater reliability because they had more practice with the scale. Nevertheless, if Paré's training technique is valid and was properly employed in this study, at least a moderate level of interrater reliability should have been obtained.

Both of the audiotape segments used in this study had been prerated by Truax's highly trained judges at 4.00 on the AE Scale. It was expected that the judges in this study would give the segments similar ratings. An F-test indicated a highly significant difference in the AE ratings of these audio segments and did not support this expectation. This result revealed that this experimenter's minimally trained judges were using the AE Scale in a very different manner than were Truax's judges. Two kinds of unreliability are demonstrated by the results of this study. Interrater reliability is lacking due to an inability of the minimally trained judges to agree among themselves concerning the AE ratings. There is also a lack of intertrainer reliability.
because the minimally trained judges' ratings did not agree with the Truax judges' ratings.

The recomputed interrater reliability on the audiotape segments does suggest that a reliable core of raters accounted for the significant experimental results but that there were enough outlying ratings to reduce the reliability of the ratings to close to zero. These results are only suggestive because they were computed on the practice audiotape ratings and not the final videotape ratings. It was not possible to complete a similar analysis on the final videotape ratings because each judge rated only one videotape segment. These results do not support the use of minimally trained judges in future empathy research. When external judges' ratings of tape segments are used to measure empathy, the investigator should take the time to train his judges to a high level of interrater reliability. However, if an investigator does decide to use minimally trained raters, the procedure of discarding ratings more than two standard deviations from the mean might be employed to improve interrater reliabilities.

In regard to the other hypotheses, the data suggests that both therapist and client nonverbal behaviors influence rated accurate empathy. Factors unique to the individual dyads also appear to interact to influence empathy ratings. These results are also congruent with Haase and Tepper's (1972) compensatory model of empathic
functioning. In a study similar in design to Haase and Tepper, McMullan (1974) obtained results supporting the compensatory model and demonstrated that therapist nonverbal behavior is a powerful influence in determining ratings of AE.

In spite of this pattern of significant results, the data can only be considered as suggestive and cannot be interpreted to either support or fail to support Hypotheses 2-5. The reliability of a set of data must be demonstrated before conclusions regarding the validity of that data can be drawn. The lack of reliability of the present data makes it impossible to determine if the significant results represent real effects of the independent variables or are a function of the large amount of error variance in the ratings. Because of this lack of reliability, further data interpretation is not justifiable.

The pattern of results obtained in this study is reflective of methodological issues in the measurement of empathy in psychotherapy. Research attempting to determine the contribution of specific client and therapist nonverbal behaviors to ratings of empathy may be premature. This type of research presupposes the existence of a valid and reliable instrument to measure the theoretical construct. Enough challenges have been made to both the reliability (Avery and Danish, 1976; Bachrach, 1976; Thoresen, 1977) and the construct validity (Kurtz and Grummon, 1972;
Bachrach, 1976; Avery and Danish, 1976; Avery, D'Augelli, and Danish, 1976; Thoresen, 1977) of the Accurate Empathy Scale to indicate that these psychometric scale properties have not been adequately established.¹

Concluding that this type of research is premature does not imply that it is unimportant. It is likely that nonverbal behaviors will be empirically demonstrated to significantly effect both the process and outcome of psychotherapy. However, until the reliability and construct validity difficulties with the AE Scale are resolved, examining the influence of nonverbal behaviors upon AE ratings provides results with uncertain accuracy and questionable generalizability to actual therapeutic encounters. The obtained relationships may represent an artifact due to error variance rather than a true relationship between the experimental variables. The light that these results shed on the relationship between the nonverbal behaviors and the construct of empathy is also questionable. Thus, this important research is better delayed until basic scale validation has been accomplished.

¹The criticisms in some of these studies have been directed at Carkhuff's (1969) Empathic Understanding Scale. The Empathic Understanding Scale is a five-level modification of the Truax AE Scale. These two scales are similar in conception and format and have similar weaknesses. The focus of the critique is upon this particular approach to the operationalization and measurement of empathy.
One of the difficulties with the reliability of the AE Scale is that interrater reliabilities are generally in the moderate range. Interrater reliabilities in the .50 to .80 range are typical in the AE literature and are generally accepted as adequate. However, reliabilities of this magnitude leave 20% to 50% of the variance in the ratings unexplained. It is difficult to establish the construct validity of the scale with this much residual variance in the ratings.

One strategy for improving interrater reliability is to choose judges who are highly similar in some attribute or ability. Carkhuff argues that only individuals currently functioning at a high level of empathy can make valid and reliable ratings (Burnstein and Carkhuff, 1969; Cannon and Carkhuff, 1969). Since most clients do not function at a high level of empathy, they cannot make accurate judgments concerning the levels of empathic understanding offered by their therapists. Carkhuff concludes that clients should not rate their therapists for level of empathy. On the surface, this seems a reasonable strategy to improve the reliability of empathy ratings but its weaknesses become apparent upon closer examination. This strategy serves to shield the Truax-Carkhuff conceptualization and rating method from a theoretically justifiable source of criticism—the client's immediate perception of his therapist's behavior. Thoresen (1977)
points up the error of this strategy stating that,

By measuring empathic understanding in only one way and by requiring raters to already possess what is being measured . . . , we force the data to conform to the theory. There is little opportunity for empirical data to discredit beliefs about the theory. A theory is not scientific if it is logically immune to experiences that may contradict it (author's emphasis, p. 301).

Another source of unreliability in AE ratings, which has only recently been recognized and addressed, is intertrainer reliability (Avery and Danish, 1976). Consistency within teams of raters does not insure consistency between teams of raters. Intertrainer reliability refers to whether or not different trainers can teach raters to use the AE Scale in a similar manner. Without intertrainer reliability, the comparability of results between different studies is problematic. It also makes it impossible for investigators to replicate each other's results. Failures to replicate could be caused by raters using the scale differently rather than by true differences in the independent variables of the studies. Avery and Danish (1976) demonstrated that raters trained by Carkhuff-educated trainers made significantly different empathy ratings than judges trained by non-Carkhuff-educated raters. This brings the construct validity of the AE Scale into question. If this scale had construct validity, independent investigators should be able to train raters so that their ratings have both high interrater and intertrainer
reliability.

Greater specificity in both the definition of the scale levels and the description of the rater training process has been proposed as a solution to these reliability problems. Barrow (1977) suggests revising the scale and defining each level in terms of specific therapist behaviors as a solution to both the interrater and intertrainer problems. Avery and Danish (1976) recommend a greater specification of the guidelines for training raters and a more detailed description of the way in which the interrater reliability was obtained. The present methods of gauging interrater reliability, i.e., the Pearson product moment correlation and the intraclass correlation, provide a single estimate of reliability but give little understanding of the sources of instability or disagreement in the ratings. No clues are provided as to how the reliability can be improved in further studies.

Generalizability theory developed by Cronbach, Gleser, Nanda and Rajaratnam (1972) allows for the simultaneous examination of several sources of instability in ratings. This approach involves a multivariate design and can allow an assessment of instability in ratings due to such factors as raters, clients, repeated measures over time and experimental treatments (Thoresen, 1977). In essence, generalizability theory provides an analysis of error variance associated with the various experimental
components. This would provide an overview of the instability of the scale as well as giving investigators clues as to how the scale needs to be improved.

There have been other criticisms of the AE Scale's construct validity besides those involving intertrainer reliability. Construct validation is necessary when an investigator has no direct way of measuring the phenomena with which he is concerned and must use indirect measures. It involves an assessment of how adequately the measuring instrument reflects the theoretical construct (Cronbach and Meehl, 1955). Traditionally, empathy has been considered a private phenomenological event of the therapist (Rogers, 1957). Empathy is by definition the therapist's ability to assume the client's internal frame of reference and to see the world as he does. It is not possible to assess this type of therapist perception directly, and thus the measuring scales have focused upon therapist responses theorized to follow from empathic understanding (Truax and Carkhuff, 1967).

Testing the accuracy of this approach to the measurement of empathy requires the formulation of an explicit network of assertions concerning the meaning of empathy. This has been labeled the nomological network (Cronbach and Meehl, 1955). This network is a collection of interwoven logical, empirical and statistical relationships. It serves to relate observable properties to a
construct like empathy, as well as to relate independent observable properties of that construct to each other. The nomological network also specifies the relationship of the construct to other constructs. The establishment of the construct validity of a measure of empathy like the AE Scale requires both convergent and discriminant validation. The nomological network predicts positive relationships between empathy and other constructs, and between different methods of measuring empathy (convergent validity). It also predicts negative relationships between empathy and variables which should not theoretically be related to it (discriminant validity). The more of these convergent and discriminant relationships that are empirically confirmed, the more confidence we have that our scale is measuring the construct which we want to measure.

The construct validation of the AE Scale has not received systematic and sustained effort from investigators using it. It seems that some investigators have been unaware of the necessity of this type of validation and have confused the construct with the method of measuring it. In other words, empathy becomes the ratings on the AE Scale rather than the AE Scale being one method of measuring empathy. However, there are several research studies testing theoretical predictions involving the AE Scale which have implications for its convergent and discriminant validity.
For example, Kurtz and Grummon's (1972) research demonstrates a lack of convergence between the AE Scale and six other measures of empathy. The relationship between AE ratings and client perceptions of empathy was the only measure that approached significance. Theoretically, greater correlations with the other empathy measures was expected. Other research has failed to demonstrate the expected discriminant validity. Muehlberg, Pierce and Drasgow (1969) studied intercorrelations between ratings of AE, respect, genuineness, concreteness and self-disclosure. All of these measures were so highly correlated (r = .78 to .91) that they could be considered to measure a single common dimension. It is interesting to note that many of these correlations were higher than interrater reliabilities usually attained on the AE Scale.

Collingwood, Hefele, Muehlberg, and Drasgow (1970) named this single factor the "good guy" factor and saw it as reflecting the therapist's overall therapeutic competency. This pattern of results does not bolster the construct validity of the AE Scale. Positive correlations with these other constructs was expected. However, the correlations are higher than the interrater reliability of the test and this brings the existence of empathy into question. Results obtained by Bachrach, Luborsky, and Mechanick (1974) raise similar questions. These investigators found that skill and empathy ratings were so highly
related ($r = .94$) that they were essentially measuring the same quality. These results provide compelling evidence to suggest that what the scale is measuring is not in line with the definition of empathy on which the scale is based. Although the AE Scale is intended to measure a particular type of therapeutic process, these ratings are significantly influenced by the raters' cognitive evaluative dimensions. The raters using the AE Scale are required to make complex judgments about a very complex human interaction on the basis of a brief sample of taped verbal behavior. They are expected to do this with a scale that has no overt therapist behaviors anchoring the scale levels. This is a very difficult task, especially for raters with no clinical training. To meet the demands of this task, raters may rely on the semantic differential dimension of "nice/good" and make a global judgment for the tape as a whole. Raters may then make the inference that "goodness" is related to particular levels of empathy, warmth or skill. This explanation would account for the findings of a number of multivariate studies which show very high relationships between AE and a number of other "good" constructs. Avery, D'Augelli, and Danish (1976) contend that the raters' cognitive evaluative dimensions "might be more important than the client in judging the helper's response, despite the client's conceptual importance in client-centered therapeutic approaches" (p. 178).
It can be concluded that the AE Scale presently lacks both convergent and discriminant validity. There have been few properly designed studies executed to establish this type of validity. The majority of studies employing the AE Scale have been univariate studies. This strategy has succeeded in establishing the predictive validity of the scale. Accurate empathy ratings are related to outcomes in psychotherapy. There are an infinite number of alternative explanations of this relationship. The construct validity of the AE Scale is not developed to the point that these rival hypotheses can be ruled out and level of empathy can be pointed to as the only viable hypothesis.

Systematic multivariate research provides a promising strategy to develop the construct validity of an empathy measurement scale. Campbell and Fiske's (1959) multitrait-multimethod approach offers a viable multivariate technique to establish construct validity and to improve the Accurate Empathy Scale. It involves a correlational approach in which the construct is viewed in a number of different ways. This technique provides the investigator with a tool for systematically examining and establishing the reliability and convergent and discriminant validity of the scale.

One of the strengths of this approach is that it provides a method of determining the contributions of
method variance and content variance to the total variance. For example, accurate empathy, genuineness and warmth may be highly correlated not because they are truly related but because they were all measured in the same manner. The multitrait-multimethod approach requires the investigator to utilize more than one trait or construct and to measure each of them in more than one way. This results in a multitrait-multimethod correlation matrix summarizing all the intercorrelations achieved when each trait or construct is measured by each of several methods. An analysis of this matrix allows the investigator to assess both the convergent and discriminative validity of the measurement techniques which he employs. A logically and theoretically related series of multitrait-multimethod studies could be executed in a systematic manner to develop a reliable and valid scale to measure empathy.

An appropriate beginning point for this research program would be to attempt to improve existing empathy scales. Both the intertrainer and interrater reliability problems must be remedied before construct validation can be successful. If a standardized training method could be established which insured high interrater and intertrainer reliabilities, then research aimed at establishing the convergent and discriminant validity of the scale could be initiated. The following research design proposes a multitrait-multimethod approach aimed at developing such a
training procedure.

This proposed design (see Figure 3) would assess the influence of trainer training method and measurement method upon the reliability of empathy ratings. It would also allow an assessment of the influence of training method and measurement method upon intertrainer reliability. Two inexperienced trainers would learn to train different groups of raters to make empathy ratings using either the Accurate Empathy Scale (AE) or the Empathic Understanding Scale (EU). The trainers would teach the raters to make empathy ratings using one of two training methods: a didactic-programmed method, or a group discussion method. Therefore, each trainer would train four different groups of raters (didactic-programmed method for the AE Scale, didactic-programmed method for the EU Scale, group discussion method for the AE Scale, group discussion method for the EU Scale).

Both training methods would be based on a structured format which was easily repeated across the two scales and the two experimenters. The goal of both training methods would be teaching raters to rate segments in a highly similar manner to the highly trained Arkansas raters. Their ratings would serve as the criterion of empathy in this study. Using the didactic-programmed method, raters would be trained individually using a programmed text format. These raters would be given a
<table>
<thead>
<tr>
<th>Trainer</th>
<th>Training Method</th>
<th>Measurement Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Didactic-Programmed</td>
<td>AE Scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EU Scale</td>
</tr>
<tr>
<td></td>
<td>Group Discussion</td>
<td>AE Scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EU Scale</td>
</tr>
<tr>
<td>2</td>
<td>Didactic-Programmed</td>
<td>AE Scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EU Scale</td>
</tr>
<tr>
<td></td>
<td>Group Discussion</td>
<td>AE Scale</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EU Scale</td>
</tr>
</tbody>
</table>

Figure 3. Proposed Multivariate Experimental Design. 
AE Scale = the Truax Accurate Empathy Scale; 
EU Scale = the Empathic Understanding in Interpersonal Processes Scale.
written overview of the concept of empathy and the particular scale. They would be quizzed frequently and given immediate feedback on the correctness of their answers. Segments demonstrating different scale levels of empathic responding and the rationales for these ratings would be presented in a similar written format. The raters would then begin to make practice ratings. They would listen to a tape segment, make their rating, and write a short justification of that rating in terms of the scale criteria. The programmed text would then present the criterion prerating and the justification for it. Raters would be encouraged to adjust their scale conception to more closely approximate the criterion ratings. Using this approach, the trainer would have little responsibility for the direct training of the raters. He would basically be there to play the tapes and answer any questions not covered by the programmed text.

Using the group discussion format, raters would be trained as a group using the standard method (Rogers et al., 1967). Using a structured format, the trainer would begin training with a verbal explanation of empathy and the given scale. Demonstration tapes would be played for the group and the trainer would present the rationale for the preratings. The trainer would attempt to involve all the raters in the training process by encouraging group discussion. During the practice rating of segments subjects
would be required to announce their ratings and justify them in terms of the scale criteria.

After the five hours of structured training, all raters would rate the same 30 tape segments for empathy using the empathy scale they were familiar with. These ratings would also be prerated by the Arkansas raters. The ratings made by the raters using either the AE or the EU scale would constitute the dependent variable in this study.

The data analysis following generalizability theory guidelines would allow for an evaluation of many facets of the reliability of the empathy ratings at the same time. It could be determined which training method with which scale correlated most highly with the Arkansas rater's preratings. This would provide another indication of interrater reliability. If neither rating scale in conjunction with neither training produced high reliabilities or high correlations with the preratings, other training methods could be devised and systematically evaluated in this multivariate format. If the cumulative result of these varied training methods failed to yield a reliable training method, we would have to question the viability of the external rater approach to empathy measurement. We might conclude that future efforts were better spent devising another kind of measuring instrument. However, if we could establish a structured training method with which
independent investigators could train raters and obtain high intertrainer and interrater reliabilities, we could then begin to establish the construct validity of the scale using multivariate methods.
CHAPTER V

SUMMARY

This study was designed to assess the contributions of therapist nonverbal behavior to accurate empathy ratings of a roleplayed therapeutic interaction for two therapist-client dyads. The study utilized a 2 (Therapist Verbal Empathy: high vs. low) X 2 (Therapist Nonverbal Behavior: high vs. low) X 2 (Client Nonverbal Behavior: high vs. low) X 2 (Dyad 1 vs. Dyad 2) analysis of variance design. Four clinical graduate students participated in the roleplayed interaction. The two graduate students who roleplayed the clients were given descriptions of the roles they were to portray but were required to improvise the actual dialogue. The client role description depicted a college freshman having difficulty adjusting to the university environment and experiencing interpersonal conflicts with family and friends.

The two graduate students roleplaying the therapists altered their verbal empathic behavior by giving either high empathy responses (AE Scale levels 5-9) or low empathy responses (AE Scale levels 1-3). Both the client's and therapist's level of nonverbal behavior was manipulated.
through experimental instructions conveying which nonverbal behaviors to increase in frequency and which to decrease. The high levels of nonverbal behavior included forward trunk lean and a high frequency of head nod, hand gesture, and eye contact. The low levels of nonverbal behavior included backward trunk lean and a low frequency of head nod, hand gesture, and eye contact. Combining all levels of the four independent variables resulted in 16 stimulus conditions. Several three-minute videotape segments were filmed for each of the stimulus conditions during the roleplayed interactions.

Manipulation check ratings were made for level of therapist verbal empathy and level of client and therapist nonverbal behavior to determine if the intended experimental manipulations had been successful. The level of authenticity, which was defined as the degree to which the roleplayed interaction simulated an actual client-therapist interaction, was rated and held constant across tape segments chosen for the final ratings. This was done to control for the possible confounding effects of differing levels of authenticity upon the final empathy ratings. One videotape segment representing each of the 16 stimulus conditions was edited onto a master tape and presented to groups of judges for the final empathy ratings.

Eighty introductory psychology students served as the final judges. A brief training procedure was utilized
in which these judges learned to rate accurate empathy. The eighty judges were divided into 16 five-member groups. Each group rated one of the videotape segments and these ratings constituted the dependent variable in this experiment.

An analysis of variance indicated two significant main effects and five significant interactions. However, there was a total lack of reliability in the final empathy ratings. This lack of reliability made it impossible to determine if the data did or did not support the hypotheses concerning the relationship between nonverbal behavior and empathy. A post hoc analysis of the ratings revealed agreement among a core of raters. However, there were enough outlying ratings to reduce the interrater reliability significantly.

These results were discussed in light of recent criticisms of the reliability and construct validity of the AE Scale. Two types of unreliability in AE Scale ratings were identified. Lack of interrater reliability occurs when judges trained by the same trainer fail to agree in their ratings. Lack of intertrainer reliability occurs when judges trained by different trainers fail to agree in their ratings of the same stimulus material. It was also argued that the construct validity of the AE Scale has not been adequately established. Due to this weakness, it was concluded that assessment of the contributions of
nonverbal behaviors to the variance of empathy ratings was premature. It was suggested that experimental efforts should be focused on establishing adequate scale reliability and construct validity. A multitrait-multimethod strategy for developing both interrater and intertrainer reliability for the AE Scale was proposed.
REFERENCES


Bachrach, H. M. Empathy: We know what we mean but what do we measure? Archives of General Psychiatry, 1976, 33, 35-38.


APPENDIX A

THE TRUAX ACCURATE EMPATHY SCALE
Stage One

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 Two

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 two is distinguishable from Stage three in that the therapist ignores feeling rather than displaying an inability to understand them.

Stage Three

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 Four

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 Three 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 Five

Therapist accurately responds to all of the client's more readily discernable 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 Five 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 Six

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 Seven, 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 Seven

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 towards more emotionally significant material. Stage Seven is distinguishable from Stage Six in that often the
therapist's response is a kind of precise pointing of the finger toward emotionally significant material.

Stage Eight

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.

Stage Nine

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 (through 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 B

MELLOH'S SCHEMATIC PRESENTATION OF A SCALE
FOR THE MEASUREMENT OF ACCURATE EMPATHY
A Schematic Presentation of
A Scale for the Measurement of Accurate Empathy

<table>
<thead>
<tr>
<th>LEVEL OF CLIENT FEELINGS PERCEIVED AND REFLECTED BY THE THERAPIST</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
<th>Stage 5</th>
<th>Stage 6</th>
<th>Stage 7</th>
<th>Stage 8</th>
<th>Stage 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present obvious feelings</td>
<td>ignores</td>
<td>understands poorly</td>
<td>often accurate</td>
<td>usually accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>unhesitating flawless accuracy</td>
</tr>
<tr>
<td>Veiled feelings</td>
<td>ignores</td>
<td>senses but understands poorly</td>
<td>accuracy very low but trying</td>
<td>sensitive but somewhat inaccurate tentative interpretation</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
</tr>
<tr>
<td>Preconscious feelings</td>
<td>ignores</td>
<td>a precise &quot;pointing toward&quot;</td>
<td>sensitive trial and error exploration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. This schematic presentation of levels of accurate empathy, developed by Richard A. Mellish, University of Florida, has been found useful for both research raters and therapist trainees. It provides a brief summary of the table scale, and is intended to facilitate the training of raters in the use of the scale.
APPENDIX C

ACCURATE EMPATHY RATING FORM
### Accurate Empathy Rating Form

**Client Feelings Perceived & Reflected by the Therapist**

<table>
<thead>
<tr>
<th>Stage</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Present Obvious Feelings</strong></td>
<td>ignores</td>
<td>understands poorly</td>
<td>often accurate</td>
<td>usually accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>accurate</td>
<td>flawless accuracy</td>
</tr>
<tr>
<td><strong>Veiled Feelings</strong></td>
<td>ignores</td>
<td>ignores</td>
<td>senses but understands poorly</td>
<td>accuracy but trying</td>
<td>sensitive but somewhat inaccurate interpretation</td>
<td>accurate but not intensity</td>
<td>accurate</td>
<td>accurate</td>
<td>flawless accuracy</td>
</tr>
<tr>
<td><strong>Pre-conscious Feelings</strong></td>
<td>ignores</td>
<td>ignores</td>
<td>ignores</td>
<td>ignores</td>
<td>ignores</td>
<td>ignores</td>
<td>a precise &quot;pointing toward&quot;</td>
<td>sensitive trial and error exploration</td>
<td>flawless accuracy</td>
</tr>
</tbody>
</table>

---

**Segment #1**

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
---

**Segment #2**

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
---

**Segment #3**

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
You are a first-quarter freshman at the University of Montana. It is very important for you to attend college. Your parents, who did not go to college have never been very supportive of your going to school. They feel that going to college is a waste of time and money and are constantly trying to convince you to drop out of school and work with them in the family hardware business. You try to explain to your parents why it is important for you to go to school but you can't seem to make them understand. You also have difficulty communicating with many of your friends. You feel your friends often take advantage of you by making unreasonable requests. Although you don't feel like complying with their requests, you do so anyways because you don't know how to say "no." You are afraid that your friends would be mad at you if you did and that they would not like you any more. At the same time you are mad at yourself for not being able to express your feelings to your friends. More and more, it seems like your friends don't really "know" you at all and are just using you for their own purposes. You are beginning to feel very lonely. Nobody ever seems to listen to you and you're beginning to wonder if anybody really cares about you.
APPENDIX E

NONVERBAL BEHAVIOR INSTRUCTIONS
Nonverbal Behavior Instructions

In the role playing that you will be doing, your nonverbal behavior will be of as much importance as your verbal behavior. I am going to try to systematically alter the level of your nonverbal behaviors during the filming sessions. I will do this by asking you to either increase or decrease the frequency of certain nonverbal behaviors. The result of this will be to produce either a Low Level or a High Level of nonverbal behavior. The specific behaviors I want you to alter to produce these levels are:

**Low Level**

1. Try **not** to lean forward. Sit with your back against the chair.
2. Try **not** to nod your head when you speak.
3. Try **not** to gesture with your hands when you speak.
4. Avoid eye contact with the other person as often as possible. Look at the other person, but try not to look him in the eyes.

**High Level**

1. Lean forward in your chair more frequently.
2. Increase your frequency of head nodding when you speak to the other person.
3. Gesture with your hand when you speak. Be as active and animated as possible without acting unnaturally.
4. Increase the frequency with which you make eye contact with the other person.
APPENDIX F

OVERVIEW OF ACCURATE EMPATHY AND THERAPIST ROLE INSTRUCTIONS
You are about to participate in a research project concerning the effects of accurate empathic understanding in psychotherapy. Carl Rogers originally characterized empathic understanding as the ability, "To sense the client's world as if it were your own without ever losing the 'as if' quality" (1957). In further developing the construct, Charles Truax stated that, "Accurate empathy involves more than just the ability of the therapist to 'sense the client's or patient's 'private world' as if it were his own. Accurate empathy involves both 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" (Truax and Carkhuff, 1967). Thus accurate empathy refers to a therapist's ability to understand a client's feelings and also his capacity to communicate some part of this understanding to the client. Truax developed a nine point scale, The Truax Accurate Empathy Scale, which is used to rate a therapist's level of accurate empathic responding.

The Truax Accurate Empathy Scale defines a range of accurate empathic responding in terms of nine levels. Level 1 is the low end of the scale. A therapist displaying a low level of accurate empathy may be judging the client, giving advice, or moralizing. The therapist ignores or misunderstands the client's current feelings and inhibits the client's exploration and understanding of these feelings. Whereas, at the high end of the range of accurate empathic responding (Level 9), the therapist's responses indicate a clear and sensitive understanding of the client's obvious feelings. In addition, the therapist's responses serve to clarify and expand the client's own awareness of his current feelings and experiences. Many research
projects have been conducted to determine the relationship between the therapist's accurate empathy and client outcome in psychotherapy. On the whole, the results of this research demonstrate a positive relationship between level of therapist accurate empathy and the client's personal growth and positive personality change as a result of psychotherapy. At the present time accurate empathy is widely regarded as a necessary ability of an effective therapist.

In this research project, you will be roleplaying the part of the therapist in a therapeutic interaction. Your "client" will be provided with a general role description but will improvise actual dialogue. Your job will be to interact with the "client" in as realistic a therapeutic manner as possible. At different points during the filming sessions, you will be requested to respond to the client in a high or low empathic manner. When you are asked to respond in a low empathic manner, try to approximate Level 2 AE responses as closely as possible (see the Truax AE Scale). You can approximate Low Level responses by:

1. Offering "advice" or quick and easy solutions to the client's problems.
2. Moralizing or judging the client's behavior.
3. Ignoring obvious feelings or misjudging their significance or intensity.
4. Conveying boredom or non-caring in your responses.
5. Steering the client away from feelings toward the content of the responses.

When you are asked to respond in a highly empathic, try to approximate Level 7 (or higher) AE responses. You can approximate High Level responses by:

1. Not offering "advice" and allowing the client to be responsible for solving his problems.
2. Not moralizing or evaluating the client's behavior.

3. By conveying concern and warmth in your responses to the client.

4. Giving the best feeling-level feedback you can.

5. By moving the client away from content issues and encouraging him to explore his feelings.
APPENDIX G

NONVERBAL BEHAVIOR RATING FORM
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Therapist</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Low    Medium    High
APPENDIX H

AUTHENTICITY RATING FORM
<table>
<thead>
<tr>
<th>Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**AUTHENTICITY RATING FORM**

<table>
<thead>
<tr>
<th>Client</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>very unrealistic</strong></td>
<td>very realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Therapist</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>very unrealistic</strong></td>
<td>very realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>very unrealistic</strong></td>
<td>very realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Client</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>very unrealistic</strong></td>
<td>very realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Therapist</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>very unrealistic</strong></td>
<td>very realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>very unrealistic</strong></td>
<td>very realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX I

EXPERIMENTAL INSTRUCTIONS
EXPERIMENTAL INSTRUCTIONS

I. Introduction

You are going to be involved in an experiment today in which you will learn to judge psychotherapists' accurate empathic responding. I will begin by explaining to you what accurate empathy is and why it is important in psychotherapy. Then you will become familiar with a scale used to rate accurate empathy. I will play tape segments from actual therapy sessions to illustrate different levels of therapist empathic responding. We will then discuss the scale and these tape segments. Finally, I will play three more tape segments which you will rate for level of therapist accurate empathy.

Are there any questions about the experiment so far?

II. Overview of AE

Accurate empathy refers to a therapist's ability to understand his client's current feelings. Accurate empathy also involves the effective communication of that understanding to the client. In other words, when a therapist responds in an accurate empathic manner, he both understands the client's feelings and is able to share that understanding with the client. There is a continuum of empathic responding in psychotherapy. At the low end of that continuum, the therapist shows a complete lack of understanding of even the client's most obvious feelings. At the high end of the continuum, the therapist shows a complete understanding of the client's feelings and is very effective in conveying to the client that he understands. This high level of empathic understanding helps the client to better understand and to explore his feelings.

Many research projects have been conducted to determine the re-
relationship between the level of therapist empathic responding and
the client's progress as a result of psychotherapy. On the whole,
the results of this research show a positive relationship. This
means that if a client sees a therapist with high empathic ability
the chances are much greater that the client will experience positive
personality and behavioral change than if he sees a therapist with
low empathic ability. At the present time accurate empathy is widely
regarded as a necessary ability of an effective therapist.

Are there any questions over what I've covered so far?

III. Explaining the Rating Scale

Now I'm going to pass out the rating scale.

PASS OUT THE RATING SCALE

This will serve as both your rating scale and answer sheet. Before
we discuss this rating scale, I'd like you to fill in the information
at the top of the form.

PAUSE

Now take a minute to read the rating scale and become familiar with it.

PAUSE

Across the top of the scale you'll see that the continuum of empathy
responding which I described to you is represented. Stage 1 is the
low point of that continuum and Stage 9 is the high point with 7 levels
in between. In the column on the left margin you'll notice that the
client's current feelings, which are perceived and reflected by the
therapist, are divided into three levels: present obvious feelings,
veiled feelings, and preconscious feelings. Present obvious feelings
are those feelings which the client readily displays or speaks about
(give examples). Veiled feelings are those feelings which the client
doesn't readily display or refer to but which can be detected from the content of his speech or from his nonverbal behavior (give examples). Preconscious feelings are those feelings which are beyond the client's level of awareness but which he could become aware of with proper direction from the therapist (give examples). These preconscious feelings underlie the other two levels and are very important in psychotherapy. Many types of therapy attempt to bring these preconscious feelings into awareness and to help the client understand and integrate them.

Are there any questions on the format of the scale?

You'll notice that with each increasing stage, the therapist displays a more accurate understanding of these three levels of feelings.

**BRIEFLY EXPLAIN EACH STAGE**

In making your ratings you will circle the number on the scale below the stage which best represents the level at which the therapist is responding to the client's feelings. Notice that there is a separate scale for each tape segment.

Are there any other questions about this scale?

**IV. Demonstration Tape Segments**

Now I'm going to play two taped segments from actual therapy sessions which illustrate low and high accurate empathic responding. You will not rate these tapes. Listen to them and try to determine why they were rated the way they were in terms of the scale criteria. The first tape is an example of low empathic functioning and was rated 2 by a group of trained raters.

**PLAY FIRST # MIN. SEGMENT**

Why do you think this tape segment was given a rating of 2?
DISCUSSION

This second segment is an example of high empathic responding and was given a rating of 6.

PLAY SECOND 3 MIN. SEGMENT

Why do you think this segment was given a rating of 6?

DISCUSSION

V. Segment Ratings

Now you'll be listening to a tape segment and making your own ratings. After I play the segment, please circle the number on scale 1 which you feel best represents the level of therapist empathic responding in the tape segment. Please do not discuss the segments while you are listening to them or making your ratings.

PLAY AUDIO SEGMENT # 1
(point out therapist). What did you rate the segment and why (each subject is given a chance to speak and support his rating)? This segment was given a rating of 4 by the trained raters.

Now I'm going to play a second segment for you to rate. Again, please do not discuss the segment while you are making your ratings.

PLAY AUDIO SEGMENT # 2
(point out therapist)

DISCUSS

This segment was also given a rating of 4 by the trained raters.

The final segment I'd like for you to rate is a videotape segment. You'll use the same scale basing your ratings on how well the therapist understands the client's current feelings and conveys that understanding to the client.

PLAY VIDEO TAPE SEGMENT
(point out therapist)

This concludes the rating part of the experiment.

VI. CONCLUSION

OBTAIN FEEDBACK (clarity of training instructions, their ability to understand and use the scale, other comments regarding the experiment)

You will be given feedback on the purpose and results of this experiment in about 4 weeks. Your teacher will make an announcement when this description is available. You can then pick up a copy of it at the psychology building.

PASS OUT EXPERIMENTAL CREDIT FORMS

Thank you for participating.
APPENDIX J

MANIPULATION CHECK PRERATINGS FOR
FINAL VIDEOTAPE SEGMENTS
Manipulation Check Preratings for Final Video Tape Segments

<table>
<thead>
<tr>
<th>Segment</th>
<th>Experimental Manipulations</th>
<th>X AE</th>
<th>X T NV</th>
<th>X CNV</th>
<th>X Auth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hi AE, Hi T NV, Hi C NV, D$_1$</td>
<td>6.15</td>
<td>5.00</td>
<td>4.50</td>
<td>3.50</td>
</tr>
<tr>
<td>2</td>
<td>Hi AE, Lo T NV, Hi C NV, D$_1$</td>
<td>6.25</td>
<td>1.50</td>
<td>3.50</td>
<td>4.00</td>
</tr>
<tr>
<td>3</td>
<td>Hi AE, Hi T NV, Lo C NV, D$_1$</td>
<td>6.65</td>
<td>4.50</td>
<td>1.00</td>
<td>4.00</td>
</tr>
<tr>
<td>4</td>
<td>Hi AE, Lo T NV, Lo C NV, D$_1$</td>
<td>6.00</td>
<td>1.50</td>
<td>1.00</td>
<td>3.50</td>
</tr>
<tr>
<td>5</td>
<td>Lo AE, Hi T NV, Hi C NV, D$_1$</td>
<td>2.15</td>
<td>4.00</td>
<td>4.00</td>
<td>3.50</td>
</tr>
<tr>
<td>6</td>
<td>Lo AE, Lo T NV, Hi C NV, D$_1$</td>
<td>1.85</td>
<td>1.00</td>
<td>5.00</td>
<td>3.50</td>
</tr>
<tr>
<td>7</td>
<td>Lo AE, Hi T NV, Lo C NV, D$_1$</td>
<td>1.80</td>
<td>4.50</td>
<td>1.00</td>
<td>3.00</td>
</tr>
<tr>
<td>8</td>
<td>Lo AE, Lo T NV, Lo C NV, D$_1$</td>
<td>1.35</td>
<td>1.00</td>
<td>1.00</td>
<td>3.50</td>
</tr>
<tr>
<td>9</td>
<td>Hi AE, Hi T NV, Hi C NV, D$_2$</td>
<td>5.90</td>
<td>4.00</td>
<td>4.50</td>
<td>3.50</td>
</tr>
<tr>
<td>10</td>
<td>Hi AE, Lo T NV, Hi C NV, D$_2$</td>
<td>6.75</td>
<td>2.50</td>
<td>4.00</td>
<td>3.50</td>
</tr>
<tr>
<td>11</td>
<td>Hi AE, Hi T NV, Lo C NV, D$_2$</td>
<td>6.75</td>
<td>4.00</td>
<td>1.50</td>
<td>3.50</td>
</tr>
<tr>
<td>12</td>
<td>Hi AE, Lo T NV, Lo C NV, D$_2$</td>
<td>6.55</td>
<td>1.50</td>
<td>2.00</td>
<td>3.50</td>
</tr>
<tr>
<td>13</td>
<td>Lo AE, Hi T NV, Hi C NV, D$_2$</td>
<td>2.55</td>
<td>4.50</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>14</td>
<td>Lo AE, Lo T NV, Hi C NV, D$_2$</td>
<td>1.70</td>
<td>2.00</td>
<td>4.00</td>
<td>3.00</td>
</tr>
<tr>
<td>15</td>
<td>Lo AE, Hi T NV, Lo C NV, D$_2$</td>
<td>2.55</td>
<td>4.50</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>16</td>
<td>Lo AE, Lo T NV, Lo C NV, D$_2$</td>
<td>1.95</td>
<td>1.50</td>
<td>2.00</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Note. Hi AE = the high Therapist Verbal Empathy condition (5 - 9)  
Lo AE = the low Therapist Verbal Empathy condition (1 - 3)  
Hi T NV = the high Therapist Nonverbal Behavior condition (>)  
Lo T NV = the low Therapist Nonverbal Behavior condition (<)  
Hi C NV = the high Client Nonverbal Behavior condition (>3)  
Lo C NV = the low Client Nonverbal Behavior condition (<3)  
D = Dyad one  
D = Dyad two  
Auth = level of authenticity (>3 for all segments)