THE IMPLICATIONS FOR CONSTRUCT VALIDITY AND ROGERIAN THEORY IN THREE MEASURES OF THERAPIST EMPATHY

Thesis for the Degree of M. A.

MICHIGAN STATE UNIVERSITY
LAWRENCE LANGER

1972

ABSTRACT

THE IMPLICATIONS FOR CONSTRUCT VALIDITY AND ROGERIAN THEORY IN THREE MEASURES OF THERAPIST EMPATHY

By

Lawrence Langer

This study examined three measures of therapist empathy to (1) determine their construct validity and (2) provide additional information about the theoretical dispute between Rogers and Truax and Carkhuff regarding the necessity that the client actually perceive the therapist's empathy if therapy is to be successful.

Two of the therapist empathy measures required judges to rate tape-recorded interviews with (1) the Carkhuff Empathic Understanding in Interpersonal Processes Scale and (2) the Barrett-Lennard Relationship Inventory. The latter instrument was also used by the client to record his therapist's empathy. A previous study by Kurtz and Grummon (1972) failed to establish construct validity for several therapist empathy measures but could not determine if this was due to the empathy instrument or

to the different persons making the ratings, i.e., an objective judge or the client. The present study corrected this deficiency. It also examined the relationship between therapy outcome and the same measure of therapist empathy as recorded by the client and an objective judge, thereby adding to our knowledge about the theoretical dispute between Rogers and Truax and Carkhuff.

Twenty-four clients and their therapists at the Michigan State University Counseling Center served as the sample for this study. All the therapist empathy measures were correlated with each other and with six different measures of therapy outcome.

Clients completed the Relationship Inventory after the third therapy interview and again after the conclusion of therapy. One judge rated three tape-recorded segments of five different therapy interviews for each client to determine the therapist's level of empathy on the Carkhuff scale. Another judge made ratings on the same five therapy interviews to obtain the therapist's empathy score on the Relationship Inventory. The reliabilities for the tape-judged Relationship Inventory and Carkhuff empathy scale were low.

Therapeutic outcome was assessed by: 1) changes in the Number of Deviant Signs score of the Tennessee Self Concept Scale, 2) changes in the clinical judgments of improvement over therapy based upon MMPI profiles, 4) a four-point scale upon which the therapists made judgments as to the clients' progress in therapy, 5) a seven-point scale upon which clients recorded their judgments as to how helpful therapy had been to them, and 6) a composite outcome score obtained by converting the above measures into standard scores.

The results comparing the empathy scales show that: 1) there is no relationship between the tape-judged Carkhuff empathy scale and the client-perceived Relationship Inventory, 2) there is no relation between the tape-judged Relationship Inventory and the tape-judged Carkhuff scale, and 3) there is a positive relation between the client-perceived Relationship Inventory and the tape-judged Relationship Inventory. The results comparing the empathy scales to the outcome measures show that: 1) there is no relation between either the tape-judged Relationship Inventory or the tape-judged Carkhuff scale with any of the six outcome measures, and 2) there is a positive

relation between three of the outcome measures and the client-perceived Relationship Inventory.

The first three findings comparing the empathy scales to each other found no support for construct validity of therapist empathy. Further, the findings overall tended to support the Rogerian rather than Truax-Carkhuff position that the client must perceive the therapist's empathy. The relation of those findings to the findings of previous research was discussed. A possible explanation was offered for the: 1) random relationship found between the tape-judged Carkhuff empathy scale with both the tape-judged and client-perceived Relationship Inventories, and 2) the random relationship found between both the Carkhuff scale and the tape-judged Relationship Inventory with the many outcome measures. Implications for future research regarding the therapist empathy concept were discussed.

Thesis Committee

Dr. Donald L. Grummon, Chairman

Dr. Mary M. Leichty

Dr. Bill L. Kell

Approved: Donald L. Grummon

Date: 11/6/72

THE IMPLICATIONS FOR CONSTRUCT VALIDITY AND ROGERIAN THEORY IN THREE MEASURES OF THERAPIST EMPATHY

By

Lawrence Langer

A THESIS

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

MASTER OF ARTS

Department of Psychology

1972

To the memory of my father

ACKNOWLEDGEMENTS

I would like to thank the many people who helped me with this research. First, I would like to especially thank Dr. Donald L. Grummon for the patience, understanding, and support he has extended to me in this research and throughout my graduate career. Further thanks goes out to the other members of my committee, Dr. Bill L. Kell and Dr. Mary M. Leichty for their careful reading and constructive criticisms of this thesis.

Next, I would like to thank my rater, Maressa Gershowitz, who spent many long hours rating the therapy interviews. In addition, a word of thanks goes to Miriam Smith of the Michigan State University Counseling Center for her helpful assistance.

Finally, I would like to thank my wife, Ginette. It is impossible to really express my feelings toward you and to thank you enough for all your help and encouragement in a short statement. Yet, I'm sure you know how much you mean to me.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF TABLES

LIST OF APPENDICES

Chapter I

INTRODUCTION

General Definitions of Empathy

This research attempts to clarify the concept of therapist empathy and its relationship to psychotherapy outcomes. Before reviewing the literature specific to the concept of therapist empathy, we will briefly examine the ways that empathy, in general, has been defined.

The original use of the word "einfühling" or empathy was by Lipps (1909). Einfühling meant, to Lipps, a process wherein entering stimuli would be reintegrated by the receiver. In this definition, the word "stimuli" is a general, undifferentiated term. It remained for further researchers to specify the elements or makeup of a stimulus.

Following the original formulation of empathy, a variety of disciplines within psychology have employed, defined, and worked with the concept of empathy. Within the psychoanalytic school, Sullivan (1945) employed the term empathy as a part of his theory of personality.

Within the theory, empathy "refers to the peculiar emotional linkage that subtends the relationship with significant people--the mother and the nurse. Long before there are signs of any understanding of emotional problems there is evidence of this emotional contagion." Empathy is part of a learning process which occurs in association with significant others.

Fromm-Reichmann (1950) generally agreed with Sullivan's formulation of empathy. She further postulated that in a psychotherapeutic situation "success is dependent upon an empathic quality between the psychiatrist and patient."

Another psychoanalytic theorist, Fenichel (1945) further elaborated on the definition of empathy. He conceived of empathy as a process consisting of two acts: "an identification with the other person and an awareness of one's own feelings after the identification." Here Fenichel introduced the notion that one becomes aware of his own feelings, an active process which is more than identification. Fenichel's formulation of empathy has some important implications for both psychotherapy and research as it lends itself to operational definitions.

The position of both Sullivan and Fenichel give one an intuitive understanding of empathy and provide a base from which operational definitions may be evolved. There is, however, in some of the psychoanalytic writing, a tendency toward circular reasoning. Remembering Fenichel's (1945) definition of empathy as being composed of identification with the other person and self-awareness, we have Adler's (1929) definition of identification as "the ability to place oneself empathically in the situation."

Other researchers have used the concept of empathy. Within the area of industrial and social psychology, in a study of interpersonal harmony in an industrial setting, Speroff (1953) proposed a "quantitative definition of empathy designed for communicative interpersonal use." He defined empathy as "the ability to put yourself in the other person's position, establish rapport and anticipate his feelings, reactions, and behavior." This approach, then, is concerned with prediction and control of behavior. Another researcher, whose theory of sensitivity seems to be consistent with Speroff's definition of empathy, is Smith (1966). He sees sensitivity as "the ability to

predict what an individual will feel, say, and do about you, himself, and others.

Dymond (1948, 1949) who was among the first to conduct research studies on empathy defined empathy as "the imaginative transposing of oneself into the thinking, feeling, and acting of another." She further believed that the family atmosphere that surrounds a person during childhood is one of the major determinants in his future ability to empathize.

Therapist's Empathy and its Relationship to Counseling and Psychotherapy

As the main thrust of this paper is concerned with issues centered around therapist's empathy, it is appropriate to begin with the work of Rogers (1957, 1959, 1967). Empathy for him, "is to perceive the internal frame of reference of another with accuracy and with the emotional components and meanings which pertain thereto, as if one were the other person, but without losing the 'as if' condition. Thus it means to sense the hurt or pleasure of another as he senses it, and to perceive the causes thereof as he perceives them, but without ever losing the

recognition that it is as if I were hurt or pleased, etc. If this 'as if' quality is lost, then the state is one of identification." Rogers uses this definition of empathy as part of his general theory of therapy which states that the greater the therapist's genuineness, his warm acceptance and his empathic understanding of the client, the more likely it is that the therapeutic process will progress, which in turn is related to positive therapeutic outcomes. Thus, Rogers attaches considerable importance to the notion of therapist's empathy and his theory has stimulated much empathy research in the counseling and psychotherapy area.

Positive Results

According to Truax and Carkhuff (1967), Halkides (1958) was one of the earliest empathy researchers who tried to relate outcome with the therapist's level of empathic understanding, unconditional positive regard and self-congruence. Brief samples of early and late therapy interviews, taken from ten successful and ten unsuccessful therapy cases, were rated on a scale based upon Rogers' theory of therapist conditions. She found that the

successful cases showed significantly higher levels of the three conditions than did the unsuccessful cases.

Another study which related the three therapist conditions to the therapy process was done by Hountras and Anderson (1969). The subjects of the study were 27 male and 27 female undergraduates coming for help at a university counseling center. The subjects were assigned to one of nine counseling interns on the basis of whether their problems were educational, vocational, or personal-social. Conditions of empathy, respect, and genuineness were measured by the Carkhuff scales. As a measure of process, the Truax Depth of Self-Exploration Scale was given to the clients. They found a significant relationship between all three counselor-offered conditions and self-exploration in each of the problem categories.

In another study, Dickenson and Truax (1966) investigated the usefulness of time limited group counseling upon college underachievers, by comparing a group of underachievers receiving group therapy with a matched group receiving no group counseling. As part of the study, Dickenson and Truax hypothesized that of those receiving group therapy, the underachievers receiving the highest levels of accurate empathy, nonpossessive warmth and

genuineness would have the greatest improvement in grade point average. Those students in group therapy who received high levels of therapist conditions showed significantly more improvement than either the control students, or those students receiving only moderate levels of therapist conditions.

Although the preceding studies have shown a positive relationship between therapist variables of empathy, warmth, and genuineness and process and outcome, it is difficult to determine precisely what effect empathy alone had on outcome, as it was only one of three variables. The following studies were concerned specifically with the relationship of therapist empathy and therapy process and outcome.

A study by Dombrow (1966) was in part concerned with the relationship of therapist empathy and outcome. The subjects were 29 mothers at a Child Guidance Clinic who were seen once a week for six to ten months by 14 social workers. Empathy was determined by the therapists' sorting of Q sort statements as if they were the patients. Change or outcome was determined by the degree of the mothers' divergence or convergence of self and ideal self concepts during the course of therapy, as measured by

Q sortings. Dombrow found a significant positive relationship between empathy and change, although the actual change which occurred was small.

In a study by Cartwright and Lerner (1963), it was hypothesized that empathic understanding of the patient by the therapist is directly related to the degree of improvement in the patient during psychotherapy. The subjects included 28 clients at a university counseling center and their 16 client-centered therapists. The therapists predicted what their clients' responses would be on a scale developed from the Kelly Repertory Test. The measure of empathic understanding was the difference between the client's self-description and the therapist's prediction of the client's self-description. This measure was taken in the early part of therapy and later at post-therapy testing. Outcome or improvement was determined by judges' ratings of three components: 1) the patient's integration, 2) the kind of organization (defensive vs. open), and 3) present life adjustment. An additional measure of improvement was the therapists' post-therapy ratings of therapy outcome. They found that empathic understanding measured at post-therapy testing was positively related to client improvement. The relationship between

improvement and empathic understanding was not significant when the earlier measures of empathic understanding were correlated with outcome. From this, Cartwright and Lerner concluded that the magnitude of the therapist's empathy which occurs later in therapy, is more crucial for client change than the magnitude of empathy which occurs earlier in therapy.

Another researcher, Barrett-Lennard (1962), developed an empathy scale, the Relationship Inventory, based upon Rogers' theory (1957, 1959) that the client's experience of his therapist's empathy is necessary for therapy to progress. In the Barrett-Lennard study, the Relationship Inventory was administered after the first five therapy interviews, after 15, after 25, and at termination of therapy. The sample consisted of 42 clients at a university counseling center and their 21 therapists. The degree of client change during therapy was determined by scores on Q sorts, adjustment, therapist rating measures, the Taylor Manifest Anxiety and Minnesota Multiphasic Personality Inventory Depression scales. Two composite indices of change were used, one derived from therapist ratings, and the other from client

self-descriptive data. He found that the 16 cases in the upper category of change perceived their therapists as having greater empathic understanding than those falling in the lower category.

The psychotherapy research project with hospitalized schizophrenics at the University of Wisconsin, under the leadership of Rogers (1967) has provided data for a number of studies investigating the relationship between therapist empathy and its effect on therapy process and outcome. The overall design of the project is too extensive to report here. Therefore, only the findings and conclusions germane to this study will be summarized.

Therapist empathy in the study was measured in three ways. First, judges rated therapists on the Accurate Empathy Scale (Truax, 1961). Second, it was measured by clients' perceptions of their therapists' empathy on the Barrett-Lennard Relationship Inventory. Finally, therapists rated their own empathy on the Relationship Inventory.

The major process scale used was designed by Gendlin (1962) in terms of the experiencing dimension. This seven-point rating scale is operationalized as follows: At the lowest levels of experiencing, the patient

makes no personal reference and has little or no expression of feeling. At the higher levels, the patient experiences his feelings, understands them, and integrates them into a meaningful framework.

A great number of outcome indices were employed in this study. Among them were the Minnesota Multiphasic Personality Inventory, the Thematic Apperception Test, the Butler-Haigh Q sort, etc.

In general it was found that: 1) Judges' ratings of therapist empathy (Accurate Empathy Scale--AE) were significantly correlated to the level of process the patient engaged in. 2) Patient perceptions of the therapist empathy were moderately associated with process level. 3) Judges' ratings (AE) and patient perceptions were related (initial, $r = .71$; terminal, $r = .38$). 4) Therapists' perceptions of their empathy were negatively correlated with judge ratings (initial, $r = -.53$; terminal, $r = .02$). 5) Therapists' perceptions of their own empathy were negatively correlated with the level of process the patient was engaged in (initial, $r = -.47$; terminal, $r = -.37$). 6) Patient perceptions of therapist empathy tended to remain stable over therapy. 7) The process level exhibited by patients was positively related with

many measures of outcome, although patients did not move on the process dimension as a group to greater experiencing. 8) The patients receiving the highest level of accurate empathy (tape ratings) in their relationships, showed the greatest reduction in schizophrenic tendencies as measured by the MMPI.

Truax and Carkhuff (1967) reported a series of studies which were designed to investigate the relationship between therapist empathy and therapy outcome. Therapist empathy was measured by the Accurate Empathy Scale (Truax, 1961). In one of the first studies, Truax (1961) compared the levels of accurate empathy, after six months of psychotherapy, for four hospitalized patients who showed a considerable improvement on a variety of personality scales, with four who showed personality deterioration. A total of 384 two-minute segments from the middle third of the therapy sessions, were randomly assigned to raters who made judgments using the Accurate Empathy Scale. He found that those therapists whose clients improved on the personality scale were rated consistently higher on Accurate Empathy than those therapists whose clients had deteriorated.

This finding relating therapist empathy to therapy outcome was extended and refined in a study using 14 schizophrenic patients who had been in intensive psychotherapy from six months to four and one half years (Truax, 1963). Five naive undergraduates having no knowledge of the clients, therapists, or therapy cases were the raters for the study. A four-minute tape segment from every fifth interview was selected for rating. The correlation between the mean level of Accurate Empathy in each case and a composite outcome score for each client was $r = .77$. A correlation between Accurate Empathy and a second outcome criterion, consisting of blind ratings of degree of personality change based upon the Rorschach and MMPI, was $r = .48$. Both of these correlations were significant.

After having shown a positive relationship between therapist empathy and outcome with a sampling of psychotic patients, Truax (1963) sought to extend his research to an outpatient population by the use of an additional 14 cases (seven successful and seven unsuccessful). Scoring tape segments from the beginning and end of therapy with the 14 hospitalized and 14 outpatient cases, he found that the level of Accurate Empathy was significantly higher for successful cases than for the failures, for both populations.

Table 1 presents a summary of the studies comparing the Accurate Empathy Scale to therapy outcome. Many of the studies support the notion of a relationship between therapist empathy and therapy outcome.

Negative Results

Not all research has shown a positive relationship between therapist empathy and outcomes. Lesser (1958) investigated the relationship between counseling progress and empathic understanding. The therapist's empathic understanding was measured by means of the Empathic Understanding Scale, a 12 statement scale derived from Fiedler's (1950) statements of therapeutic relationships. This scale was administered to both therapist and client. In addition, the investigator had the therapists predict what the clients' initial q sorts would be as a measure of the counselor's ability to predict his client's self-concept.

Differences between pre- and post-Q sorts were used as an index of change. Lesser found no significant relationship between empathy as measured by the Empathic

TABLE 1.--Summary of therapy outcome studies using accurate empathy scale.*

Study	Treatment (Ind. or Group)	Type of Client	N	Overall Combined Outcome Measures Favoring Hyp. (p .05)	Overall Combined Outcome Measures Against Hyp. (p .05)	Types of Measures
Truax (1961)	Ind.	Hospital	8	1	0	Combination Tests, MMPI, Rorschach, Diagnostic Evaluation
Truax (1963)	Ind.	Hospital	14	2	0	
Truax (1963)	Ind.	Outpatient	14	1	0	
Truax (1962)	Ind.	Hospital	14	1	0	
Truax, Wargo, Frank, Imber, Battle, Hoehm-Saric, Nash, and Stone (1966)	Ind.	Outpatient	40	2	0	
Truax, Carkhuff and Kodman (1965)	Group	Hospital	40	0	0	
Truax and Wargo (1966)	Group	Hospital	160	1	0	
Truax and Wargo (1966)	Group	Delinquent	80	1	0	
Truax, Wargo and Carkhuff (1966)	Group	Outpatient	80	0	0	
Truax (1966)	Group	Hospital	40	1	0	
		Delinquent	40			

*Table taken from C. B. Truax & R. R. Carkhuff. Toward effective counseling and psychotherapy: training and practice. Chicago: Aldine Publishing, 1967. P. 124.

Understanding Scale and the counselor's prediction of his client's Q sort and neither measure correlated with outcome.

Katz (1962) investigated the relationship between two different measures of empathy and client change in short term counseling. Using 21 high school students and their 21 counselors-in-training as subjects, Katz took measures of predictive and behavioral empathy. Predictive empathy was measured by the counselor's ability to predict how his client would arrange the self-concept Q sort. Behavioral empathy was measured by independent judges' ratings of the counselor's behavior within the interview on the Rating Scale of Empathic Behavior.

Two measures of client change were used. The first was the difference in the scores on the self-concept Q sort taken by the client before and after counseling. The second measure was the difference in the scores at the beginning and end of therapy on the Rating Scale of Counselor Behavior which independent judges rated. All clients had at least six interviews. However, Katz found no relationship between either predictive or behavioral empathy and outcome.

Gonyea (1963) studied the relationship between the quality of therapeutic relationship provided by eight counseling interns and counseling outcome. The counseling supervisors were asked to sort statements (Q sort) describing the therapeutic relationship established by each of the supervisees. The q sorts were then correlated with Fiedler's (1950) descriptions of experienced therapists to provide a measure of therapeutic relationships offered by each counselor.

As a measure of outcome, the Gregg and Kelly Self-Description Form, a measure of improvement, was used. The 208 clients seen by the eight interns in the study took the measure before and after counseling. These outcomes were correlated with the supervisor's ratings of the counselors.

No significant relationships were found between the ratings of the counselors and counselor outcome. One aspect of the study worth adding is that it was not possible to isolate the counselor empathy variable from other parts as influencing the supervisors' ratings. For example, Fiedler's statements, while containing references to counselor's empathic understanding, includes statements of other variables as well.

The literature relating therapist empathy to therapy process and outcome, therefore, has been mixed. There are, of course, many differences among the various studies: population, treatment modalities, empathy instruments, outcome and process measures, etc. However, two issues have compicated the research investigating the relationship of therapist empathy to counseling and psychotherapy. The first of these is a methodological issue, the problem of construct validity of the empathy measures. The second is a theoretical issue, that of whether the client's perception of his therapist's empathy or the level of therapist empathy as determined by an independent judge, is the more crucial measure of empathy in terms of the relationship to therapy process and outcome.

Construct Validity

Investigators have used four general approaches in operationalizing and studying therapist empathy: situational, predictive, tape-judged and client-perceived. The situational approach (Astin, 1957) involves presenting

subjects with some standardized situation to which the subjects are to respond empathically. The predictive approach asks one to make predictions about another's behavior. The judge may know the subject or is given data about him. The judge is then asked to make predictions of the subject's performance on test items or his responses to personality inventories (Astin, 1957). In the tape-judged approach, trained judges either listen to or observe interactions between people. On the basis of some defined criteria, the judges rate the amount of empathy displayed by the subject being rated. Finally, the client-perceived approach consists of having the client rate how empathic he perceived his counselor had been during the previous interview. This is typically done by the use of a questionnaire, however, post-therapy interviews or q sorts may also be used for this purpose.

Kurtz and Grummon (1972), two researchers interested in the problem of construct validity of the empathy measures, used six different measures of therapist empathy which corresponded to the four general approaches to measuring empathy, previously discussed. As a measure of therapy process the Carkhuff Self Exploration in Interpersonal Processes scale (1967) was used. For their

outcome measures, the following were used: Number of Deviant Signs and Total Positive of the Tennessee Self Concept Scale (Fitts, 1965), changes in the clinical judgments of improvement over therapy based upon MMPI profiles, client and counselor judgments of the client's progress, and a composite outcome measure based upon the above five measures.

The subjects of the study consisted of 31 clients and their counselors where the experience of the counselors ranged from one year of doing therapy to over twenty years of experience. The therapy ranged from 4 to 27 one-hour interviews, with a mean of 12 interviews.

Kurtz and Grummon found that the only significant intercorrelation among the empathy measures was between the client-perceived empathy after the third and after the final therapy interviews. A positive relationship between tape-judged empathy and client-perceived empathy taken after the third interview approaches significance. With regard to the relationship between empathy and therapy process, only tape-judged empathy was significantly related to the client's Depth of Self Exploration. Finally, Kurtz and Grummon generally found non-significant correlations between the empathy measures and outcome.

An exception to this was that client-perceived empathy after the third interview showed strong and mostly significant relationships with outcome measures. The correlations between tape-judged empathy and outcome were all positive, but only one of these six correlations was statistically significant.

Therefore, the several different ways of conceptualizing and measuring empathy failed to correlate with each other, thus casting considerable doubt upon the construct validity of all the operational measures of empathy employed in previous research. The only relationship that would tend to support construct validity would be that between client-perceived and tape-judged empathy measures. As mentioned, the Rogers et al. (1967) study with hospitalized schizophrenics found a positive relationship between client-perceived and tape-rated empathy.

However, the issue is not clear, as somewhat different results were obtained in a study by Burstein and Carkhuff (1968). Their results indicated that among moderate to low functioning therapists, there were no positive relationships between objective ratings and client-perceived ratings of empathy. It is debatable

as to how much consideration should be given to this study as they only sampled moderate to low functioning therapists.

In the Kurtz and Grummon (1972) study, a relationship approaching significance was found between tape-judged empathy and client-perceived empathy after the third interview. These researchers felt that a sampling error may have been committed which may have some bearing on the results. In their study, the rater scored segments of the first or second interviews, as well as segments of the next-to-last interview for tape-judged empathy. The mean of these two ratings was used as a measure of tape-judged empathy for all therapy interviews. The possibility exists therefore, that if more tapes were sampled of the middle therapy interviews, this might have bearing on the relationship between tape-judged and client-perceived empathy and thus on the construct validity of these two measures. Thus, one part of the present study involved making additional ratings of the therapy interviews to obtain an overall measure of tape-judged empathy which has a better sampling, in order that its relationship to client-perceived empathy be re-examined.

Additional studies (Hanson, Moore and Carkhuff, 1968; Truax, 1966) were also concerned with the relationship between client perceived and tape-rated empathy measures and, therefore, have a bearing on the issue of construct validity. One study (Hanson, Moore, and Carkhuff, 1968) was designed to compare correlations of the raters' judgments with clients' perceptions, as related to change in self concept. The client population was composed of students in the eighth to twelfth grades who had behavioral problems. The counselors were all school counselors who had completed an intensive course in multiple counseling. Client perceptions were obtained by administering the Barrett-Lennard Relationship Inventory to the clients. Tape-judged therapist conditions were obtained by having Truax-trained raters score four-minute tape segments for Accurate Empathy, Unconditional Positive Regard, and Genuineness. No significant correlations existed between any of the client's ratings and the judge's ratings of the counselor-offered therapeutic conditions. Further, the judge's ratings of each variable was positively related to client changes in self-concept while the clients' ratings were not.

An additional study related to this issue was done by Truax (1966). The predictive value for outcome of measures of Accurate Empathy, Nonpossessive Warmth, and Genuineness taken from tape recordings of therapy, were compared with measures taken from patient perceptions. Three approaches to tape ratings were used: therapist-patient-therapist (TPT) samples, patient-therapist-patient (PTP) samples and group time samples. In order to measure patient perceptions, the Barrett-Lennard Relationship Inventory was administered to the clients after the twelfth group session. Half the groups were composed of hospitalized psychiatric in-patients, while the rest of the groups were composed of institutionalized juvenile delinquents. Each group met twice a week.

The results indicated that for all three therapist-offered conditions, the TPT, PTP, and group time sampling procedures tended to be sampling the same aspect of psychotherapy, while the Relationship Inventory appeared to be measuring some different aspect of the therapy process. Further, the three sample procedures yielded approximately equivalent predictive power (average r with patient outcome for all three conditions of AE,

UPR, and Gen. of .40, .37, and .42, respectively). The Relationship Inventory tended to be less predictive of outcome (average r = .25).

One of the difficulties in all these studies is the confounding of variables. When the two measures of empathy have been compared, they have differed in two ways simultaneously--(1) the measuring instrument (e.g. the Carkhuff Empathic Understanding in Interpersonal Processes Scale and the Relationship Inventory), and (2) the person doing the rating (the client or an impartial judge).

This study then eliminated the confounding of the variables and thus the issue of construct validity was examined without this influence. Therefore, therapy interviews were scored by independent judges using the Relationship Inventory. These empathy ratings were compared to the client-perceived Relationship Inventory and tape-judged Carkhuff empathy ratings.

Theoretical Issue

The confounding of variables also has some bearing on a theoretical dispute between the positions of Rogers

and Barrett-Lennard vs. Truax and Carkhuff. According to Rogers (1957, 1959) and Barrett-Lennard (1962), in order for therapy to progress, the client must perceive the therapist's empathy. The Truax and Carkhuff (1967) position, in effect, states that the actual amount of therapist empathy influences the outcome, even though the client may not be sufficiently aware to perceive it. They further doubt that clients, particularly disturbed clients such as hospitalized psychotics, are good judges of therapist empathy.

The research which has come from Truax and Carkhuff (1967) has attempted to demonstrate that the client-perceived measure is not as effective as the tape-rated measure in predicting outcome. Their emphasis, therefore, has been primarily on the differences in the two measures. Rogers (1957, 1959), in his work, has been more concerned with the idea that the client must perceive the empathy, regardless of the instrument used to measure therapist empathy.

The fact that the two variables are confounded makes it difficult to draw any conclusions with respect to whether or not the client must perceive empathy. What this research attempted to do then was to try to eliminate

the confounding so that the Rogers and Barrett-Lennard positions could be compared with the Truax and Carkhuff position.

Statement of the Problem

This research, a follow-up to the Kurtz and Grummon (1972) study, investigates two issues: 1) a methodological one--construct validity of two empathy measures and 2) a theoretical dispute between the positions of Truax and Carkhuff (1967) and Rogers (1957, 1959) as to whether or not clients must perceive their therapist's empathy in order for therapy to be successful.

The construct validity issue was looked at in two different ways. First unlike the Kurtz and Grummon (1972) study, a more complete sampling of the therapy tapes was rated on the Carkhuff empathy scale (a tape-judged measure). These ratings were correlated with the client's ratings made on the Relationship Inventory.

Second, it is felt that a correlation of the Carkhuff empathy scale and the client-perceived Relationship Inventory is confounded by two variables varying

simultaneously--1) by the different instruments used, and 2) by the person making the ratings--a client or an objective judge. To control for this, the Relationship Inventory, a client-perceived empathy measure, was used by objective judges to make ratings on the therapy tapes. These ratings were then correlated with those made on the client-perceived Relationship Inventory and the tape-judged Carkhuff empathy scale.

The other issue, the theoretical dispute, was also investigated in light of the confounding of variables operating in the two-empathy measures--the client perceived R.I. and the tape-judged Carkhuff scale. As in the construct validity issue, the relationship between both the tape-judges Carkhuff empathy scale and the client-perceived Relationship Inventory were examined in relation to the tape-judged Relationship Inventory. Further, both the ratings made on the Carkhuff empathy scale and the tape-judged RI were correlated with several empathy measures.

Research Questions

To summarize, this research attempts to answer the following questions:

1) Is there a relationship between the tape-rated Relationship Inventory and the client-perceived Relationship Inventory measures of therapist empathy?

2) Is there a relationship between the tape-rated Relationship Inventory and the tape-rated Carkhuff empathy measures?

3) Is there a relationship between the tape-rated Relationship Inventory and the outcome measures?

4) Is there a relationship between the tape-rated Carkhuff and client-perceived Relationship Inventory measures of therapist empathy?

5) Is there a relationship between the tape-rated Carkhuff scale and the outcome measures?

6) Is there a relationship between the client-perceived Relationship Inventory and the outcome measures?[1]

---

[1]This question was investigated by Kurtz and Grummon (1972) but is reported here in order to present the results in a consistent and systematic way. The sample used in the present study is basically the same one used by Kurtz and Grummon (1972) with the exception of two or three clients.

Chapter II

METHOD

Data Collection

The data for this research were part of a larger research project conducted at the Michigan State University Counseling Center during 1967 through 1969. All therapy interviews were taped. In addition, there were about four hours of pre- and post-therapy testing of the clients and their counselors. Participation in this research project was voluntary and neither the therapist nor the client had access to the test data until after completion of therapy.

Subjects

The 24 clients that provide the data for this research consisted of eight undergraduate males and 16 undergraduate females who were self-referred for a variety of personal and emotional problems. To be included

in the overall research design clients had to present some emotional problem (rather than vocational or educational) that in the judgment of the intake interviewer would involve several therapy sessions. The clients also had to agree to donate their time for the testing.

In the research project, many clients were asked to participate. However, several of them failed to complete the in- and post-therapy testing after formerly agreeing to do so. As the 24 clients in the present research completed all of the testing, some bias may exist in this sample. Impressionistically, however, the clients included in this study appear no different from the usual run of clients bringing personal-emotional problems to the MSU Counseling Center.

About half of the 24 therapists in this study were senior staff counselors at the Michigan State University Counseling Center who held their Ph.D. in either clinical or counseling psychology. The other half were second year interns in either clinical or counseling psychology and had at least one year of experience doing therapy before participating in this study. The range of experience for all therapists was one to twenty years. The therapists tended to emphasize self-exploration, insight,

interpersonal personality theory, and a here and now rather than a historical approach to treatment, although the therapists did hold diverse theoretical orientations within this general framework.

The number of therapy interviews ranged from three to 27 with a mean number of 12.2 interviews per client. In some cases where the end of the school year was drawing near and the counselor would not be seeing the client until the next fall, post-testing occurred prior to termination.

Description of Empathy Measures

Empathic Understanding in Interpersonal Processes Scale

This scale (see Appendix B) was derived from the Truax Accurate Empathy Scale (Truax and Carkhuff, 1967) which was based on earlier work by Rogers. Carkhuff (1969) employs a five-point rating scale of which the following is illustrative: Level 1--the therapist's responses either do not attend to or detract from the expression of the client. Level 3--the therapist's responses

are essentially interchangeable with the client's in that they express the same affect and meaning. Level 5--the therapist's responses add significantly to the feeling and meaning of the client such that they accurately express those feelings which the client himself is unable to express, or, in the event of the client's ongoing self-exploration, the therapist is fully with him in his deepest moments.

The Carkhuff scale was chosen instead of the Truax Accurate Empathy Scale since it tends to get better reliability. Further, the person who trained the one judge in this study had himself been trained by Carkhuff.

Five tapes per client were chosen for rating in an attempt to sample different segments of the therapy process. The first, third, and next to last interviews were rated except in one case where there was only three sessions. In addition, two more tapes were selected to sample the middle segments of the therapy process. For each therapy interview three, three-minute segments selected from the beginning, middle, and end of the tape were rated.

The order in which the therapy tapes were rated was randomized in two ways. First, the order of the

clients was randomized by a table of random numbers. Then, one of the five tapes was randomly chosen, by the same procedure for each client and the ratings of therapist empathy made. This procedure was repeated four more times with the same order or clients, until all the tapes were rated.

In order to accurately make ratings on the Carkhuff scales, the writer met with an expert rater for two sessions. During this time, the two listened to three, three-minute segments of therapy tapes and made their judgments of the level of empathy. The expert rater, in each instance, required the rater to give a rationale for each rating based upon those definitions of the levels of empathy established by Carkhuff. If there were no discrepancies, the two would proceed to the next tape. However, if the expert rater differed, he would explain why he made the ratings and would illustrate this by replaying the segment of the tape pertaining to that rating.

The reliability computed after the training sessions was $r = .74$ over 19 ratings. At the conclusion of the study, the judge rerated those tapes originally used to establish reliability. The correlation between the judge's new ratings and the ratings originally made by

the other judge was r = .95 over 12 ratings. Another method used to compare the judge's ratings to a standard was to match the mean score of the first and next-to-last Carkhuff tape ratings of the Kurtz and Grummon (1972) study with the same scores of the present study. A correlation of r = .51 for 24 ratings was found. A statistical procedure was then used to determine if a correlation of r = .51 for 24 ratings was from the same population as the reliability r = .74, or whether these two scores were from different populations (Walker and Lev, 1953). The two correlations were found to be from the same population but nevertheless are both low.

Barrett-Lennard Relationship Inventory (RI)

To measure the client's perceptions of the therapist's empathy, the Barrett-Lennard Relationship Inventory (1962) was administered to the client after the third interview and after the conclusion of therapy (see Appendix B). Only that part of the measure dealing with empathic understanding was used in this study. This scale consists of 16 statements such as: "He tries to see things through my eyes," and "Sometimes he thinks a

certain way, because he feels that way." The client indicates three levels of agreement or disagreement, with no neutral position provided.

Barrett-Lennard reports a split-half reliability on the Relationship Inventory of .86 for 42 clients. He reports no test-retest reliability for clients because the therapist-client relationship can be expected to vary from interview to interview. However, he did report a test-retest value of .89 on 45 college students describing their friendship relationships.

Relationship Inventory--Tape Rated

In this approach, an independent judge listened to the tapes and made a judgment of the therapist's empathy on the Barrett-Lennard Relationship Inventory items. The Barrett-Lennard items were slightly modified from the original as noted in the parentheses as: "He (the therapist) tries to see things through my (the client's) eyes," and "Sometimes he (the therapist) thinks that I (the client) feel (feels) a certain way because he (the therapist) feels that way."

The same interviews used to make the tape-rated Carkhuff scale ratings also provided the raw data source for the Relationship Inventory tape judged ratings and a similar procedure was used for randomizing the order of the ratings. The actual therapeutic interactions rated within these interviews differed, however, in that instead of three three-minute segments, the R.I. tape-ratings were made on one 10-minute segment of the interview. This was done because pilot studies on these ratings suggested that in order to be able to score 16 statements pertaining to the therapist's empathy more information was needed than merely a three-minute segment. A ten-minute section of each of the interviews was randomly selected for ratings except that the judge was instructed to vary her selections between the first, middle, and last thirds of the interview in a non-systematic way.

Two different judges were used in making the empathy ratings, one for Carkhuff empathy scale (the writer) and a second judge for the tape-judged Relationship Inventory so as to control for bias which might have occurred had the same judge scored both empathy scales. In order to begin to establish reliability on the tape-judged R.I., these two judges met together for about five

training sessions. During these sessions, they listened to therapy tapes together, then each made ratings and the two discussed their differences. As a basis for making ratings, Barrett-Lennard's definition of therapist empathy was used.

Two judges established reliability of r = .99 using tape material not connected with the present study. However, this correlation is inflated because there was some discussion between the judges during the time they were establishing reliability. A post-study check on reliability was then made where a sample of the tapes rated in the study was re-rated by the writer. When these were correlated with those ratings made by the R.I. judge, the reliability was r = .63. The interrater reliability of r = .63 is pro-ably the more accurate measure.

Computation of the Empathy Scores

To obtain a therapist empathy score on the Carkhuff Scale which would reflect the whole therapy process for each client, a mean score was derived. Three, three-minute tape segments of five tapes were rated for each

client and the ratings of the fifteen tape segments were then added together and divided by fifteen. This procedure was done for each of the clients with the necessary adjustments made if all five tapes were not scored for a particular client.

Computation of the client-perceived empathy scores followed directly from the scoring system of Barrett-Lennard and is therefore straightforward. The procedure used to derive the tape-judged RI was similar with some additions. After following the Barrett-Lennard method and obtaining a measure of empathy for each tape segment rated, a mean of these rating scores was derived. Since there were five tapes per client, and one 10-minute segment rated from each tape, the five scores were added together and divided by five. It was found that for both R.I. scores, it was necessary for purposes of eliminating negative numbers, to add a constant to each score. This is noted in Tables 4, 5, and 6.

Description of the Outcome Measures

Tennessee Self Concept Scale

The Tennessee Self Concept Scale (Fitts, 1965) provided two measures of client change: "the Number of Deviant Signs" (NDS) score and the "Total Positive" (P) score. Ashcraft and Fitts (1964) demonstrated the scale's sensitivity to changes in the client resulting from therapy. According to Fitts, the Total Positive score, which reflects the client's self-esteem, is the more important score derived from the test. The other measure, the Number of Deviant Signs is a measure derived from the deviance features of the other sub-scales. Fitts has stated that, "this score alone identifies deviant individuals with about 80% accuracy. The NDS is the best index of psychological disturbance." The outcome measure used for both of these scores was the difference between pre- and post-therapy scores.

Reliability

Reliability based upon test-retest with college students over a two-week period was r = .92 for the "P"

score, and r = .90 for the NDS score. Fitts reports a correlation of r = -.68 between the "P" score and the NDS score.

Validity

Ashcraft and Fitts (1964) compared changes on the TSCS between 30 patients who had been in therapy for an average of six months and a no-therapy control group (N = 24) who had been waiting for therapy for an average of 6.7 months. Using a test-retest procedure, the therapy group changed significantly in the expected direction on 18 of the 22 variables (including the P score and NDS scores) while the control group changed on only two of the variables. This suggests that this instrument is able to detect changes on the different scales as a result of psychotherapy.

Clinical Judgments of the MMPI

Another outcome measure involved clinical judgments based upon the Minnesota Multiphasic Personality Index. Three judges who had a great deal of experience

with the MMPI compared pre- and post-therapy MMPI profiles for each client who had completed the post-test. Each pair of profiles was rated on a five-point scale for client changes as follows: 5...satisfactory, 4...partially satisfactory, 3...no change, 2...partially unsatisfactory, 1...unsatisfactory.

Kurtz reported intra-judge reliability which involved having each profile rated twice, one week apart. The intra-judge reliability index of the three judges was: $r = .68$. The reliability of average ratings was $r = .81$. The reliability of inter-judge ratings was $r = .74$ and the reliability of average ratings between judges was $r = .90$.

Therapist Ratings of Therapy Outcome

At the conclusion of therapy, each therapist completed a five-point scale which attempted to assess his client's progress as follows: 4...successful, 3...partially successful, 2...partially unsuccessful, 1...unsuccessful.

Client Ratings of Therapy Outcome

Also after termination, the client made a judgment as to how helpful the therapeutic experience had been to him. This was based on a seven-point scale such that: 1...extremely harmful, 2...harmed me a lot, 3...harmed me somewhat, 4...indifferent, 5...helped me somewhat, 6...helped me quite a lot, 7...extremely helpful.

Composite Outcome

Finally a composite outcome was used. This was derived by converting all five measures to standard (t) scores and dividing by five. In this way, each outcome measure had an equal weight in determining the composite score.

Chapter III

RESULTS

In this section, the experimental questions have been restated. This is followed by an analysis of the results and conclusions.

1) Is there a relationship between the tape-rated Relationship Inventory and the client-perceived Relationship Inventory?

The correlation of $r = .46$ ($p < .05$) between the tape-rated Relationship Inventory and the client-perceived Relationship Inventory, after the third interview was found. The corresponding correlation of $r = .42$ ($p < .05$) was found with the client-perceived Relationship Inventory at post-therapy testing. Thus, the client and an independent observer agree about therapist empathy when they record their judgments on the same instrument.

2) Is there a relationship between the tape-rated Relationship Inventory and the tape-rated Carkhuff scale?

These two measures of therapist empathy correlate r = .11 which is nonsignificant. Therefore, independent observers do not agree on the amount of therapist empathy when they record their judgments on different instruments. Construct validity has not been established.

3) Is there a relationship between the tape-rated Relationship Inventory and the outcome measure?

Table 2 shows that none of the correlations between therapist empathy as measured by the tape-rated Relationship Inventory and the several outcome measures were significant.

4) Is there a relationship between the tape-rated Carkhuff scale and the client-perceived Relationship Inventory?

The correlation of r = -.04 between the tape-rated Carkhuff scale and the client-perceived Relationship Inventory after the third interviews was nonsignificant. The corresponding correlation of r = -.13 with the Relationship Inventory at post-therapy testing was also nonsignificant. Thus even with a larger sampling of tapes scored on the Carkhuff empathy scale, no relationship has

been found to exist between that measure and the client-perceived Relationship Inventory.

5) Is there a relationship between the tape-rated Carkhuff scale and the outcome measures?

Table 2 shows that none of the correlations between the tape-judged Carkhuff scale and the several outcome measures were significant.

6) Is there a relationship between the client-perceived Relationship Inventory and the outcome measures?

The results, summarized in Table 2, indicate that of the six measures of outcome, three of them were positively correlated ($p < .01$) with the empathy scores from the client-perceived Relationship Inventory administered after the third therapy interview. In addition, a correlation approaching significance was found between the Total Positive scale of the Tennessee Self-Concept Scale and the client-perceived Relationship Inventory.

Similar results were obtained between the outcome measures and the ratings obtained by the client-perceived empathy measure when administered after the conclusion of

therapy. The same three outcome measures were positively correlated with this measure of therapist empathy ($p < .01$). In addition, a correlation approaching significance was found between the empathy scale and the client's judgment of therapy outcome measure.

TABLE 2.--Correlations between the client-perceived Relationship Inventory after the third interview, at post-therapy testing, the tape-judged Relationship Inventory, and the Carkhuff Empathic Understanding Interpersonal Processes Scale with outcome measures.

Outcome Measures	Client-perceived Relationship Inventory--After Third Interview	Client-perceived Relationship Inventory--Post	Tape-judged Relationship Inventory	Tape-judged Carkhuff Empathy Scale
Tennessee Self-Concept Scale				
Total P	.38	.11	.09	.22
NDS	.25	.01	.25	.18
Clinical judgments based upon the MMPI	.63**	.49*	-.04	.11
Therapist judgments of outcome	.54**	.79**	-.04	-.05
Client judgments of outcome	.26	.41	.19	-.02
Composite outcome	.65**	.60**	.14	.05

*Significant at .02 level.
**Significant at .01 level.

Chapter IV

DISCUSSION

In their discussion of construct validity, Cronbach and Meehl (1955) pointed out that when two tests are purported to measure the same construct, a correlation between them is expected. However, if no correlation is obtained, one can only speculate as to whether the error lies in one test, or the other, or in the formulation of the construct.

In the study by Kurtz and Grummon (1972) they generally found no evidence in support of construct validity. The possible exception they noted might be a relationship between the Carkhuff empathy scale (the tape-judged measure) and the Barrett-Lennard Relationship Inventory (the client-perceived measure). However, the research comparing the empathy scales of these two approaches has been mixed. In the Wisconsin project with hospitalized schizophrenics, Rogers et al. (1967) found there was a relationship between the client-perceived Relationship Inventory and the tape-judged Truax Accurate

Empathy Scale. Kurtz and Grummon (1972) found a positive correlation approaching significance between the Relationship Inventory and Carkhuff empathy scales.

On the other hand, some researchers using either the Truax Accurate Empathy scale or the Carkhuff empathy scale (Burstein and Carkhuff, 1968; Hansen, Moore and Carkhuff, 1968; Truax, 1966) have found no relationship between the tape-judged and client-perceived approaches. It is difficult to know how to interpret these conflicting findings.

In the present study, the relationship between the client-perceived Relationship Inventory and the tape-judged Carkhuff empathy scale was considered in terms of the construct validity issue. One part of the study was concerned with correcting a possible sampling error committed by Kurtz and Grummon (1972). In their research the empathy score for each client was determined by the mean of the ratings of the first and next-to-last interviews. In that way, none of the tapes were sampled from the middle of the therapy process. However it was found that with the addition of more data, there was a lower correlation between the Carkhuff empathy scale rated by judges and the client-perceived Relationship Inventory,

than in the Kurtz and Grummon study. On the basis of the present research, there does not seem to be any evidence in support of construct validity for the measures studied.

Another way in which the relationship between the client-perceived Relationship Inventory and the tape-judged Carkhuff scale was examined was by controlling for confounding of variables. In this study the Relationship Inventory, though usually a client-perceived measure, was scored by an objective judge who rated the therapist's empathy. The correlation between the tape-judged Relationship Inventory and the client-perceived Relationship Inventory was significant. Thus the findings show that both the client and an independent judge are in fair agreement about the amount of therapist empathy offered when they both employ the same instrument to record their judgments. This is true whether one compares the judge's average ratings over therapy with the client's ratings after the third interview or at the end of therapy. It must, of course, be kept in mind that with the low correlations found between these two instruments, much of the variance is unaccounted for. However, Rogers et al. (1967) reported an unpublished study by Van der Veen with similar results, but no details of the study are available.

The relationship between the tape-judged Relationship Inventory and the tape-judged Carkhuff scale was also examined. In this case the same type of rater, i.e. objective judges, made ratings on two different empathy scales. No relationship between these two measures was found.

At this juncture, it is appropriate to draw together all the findings of this study relating to construct validity. The results indicated that by controlling for confounding of variables, when two types of judges made ratings on the same scale, they were in agreement. However, when the same type of rater, i.e., two objective judges, made ratings on two different empathy scales, they did not agree. Finally when more data were included in the Carkhuff empathy scores and it was correlated with the client-perceived Relationship Inventory, no relationship was found. All of this demonstrates that there is no support of construct validity for these two measures. When these results are considered along with those of Kurtz and Grummon (1972), the evidence is even stronger that the operational measures of therapist empathy lack construct validity.

Future research should re-examine the whole concept of therapist empathy in order to determine whether this concept is either too broad or is perhaps so embedded in other concepts that, by itself, it cannot be measured. In a discussion of the results of the Wisconsin project, Keisler, Mathieu, and Klein (Rogers, et al., 1967) had some interesting insights on the problem of construct validity. They proposed that the Truax Accurate Empathy scale may be tapping a more global quality which they perceive as the therapist's communicated commitment to the therapy interaction and involvement in the problems of a specific patient in the interaction. Although their comments were directed to this one scale, perhaps they could be applied more generally to other empathy measures.

Before considering the rest of this research, some statement about the inter-rater reliabilities is necessary. The reliabilities obtained on both the tape-judged Relationship Inventory and the tape-judged Carkhuff Empathy scale were low. The results and conclusions of this study must therefore be evaluated in light of this.

Another part of this study was concerned with a theoretical dispute between Rogers and Barrett-Lennard vs. Truax and Carkhuff. Rogers (1957, 1959) believes

that in order for change to occur in therapy the client must perceive the therapist empathy. The Truax and Carkhuff (1967) position is that the therapist-offered empathy, whether perceived by the client or not, is what helps to determine success in therapy. Therefore the Truax and Carkhuff position is concerned with how well the instrument can predict outcome without regard to who is making the ratings.

To recapitulate the results of this study, a significant positive relationship was found between the tape-judged and client-perceived Relationship Inventory measures of therapist empathy. No relationship was found between the tape-judged Relationship Inventory and the tape-judged Carkhuff scale. Further, a significant positive relationship was found between the client-perceived Relationship Inventory measured either after the third therapy interview or at the conclusion of therapy, and three of the outcome measures. Finally, no relationship was found between either the tape-judged Relationship Inventory or tape-judged Carkhuff empathy scale with any of the outcome measures.

The fact that the two tape-judged instruments were unrelated and that neither of these instruments was

significantly related to outcome, contradicts the Truax and Carkhuff position. Although previous research (Kurtz and Grummon, 1972; Carkhuff, 1969; Hansen, Moore, and Carkhuff, 1968) found that the Carkhuff empathy scale was correlated with some outcome measures, this study found no such evidence. The earlier study by Kurtz and Grummon (1972) used a smaller sample which may have accounted for their findings. However, the low inter-rater reliability on the Carkhuff empathy, in the present study, may have some bearing on why no relationship was found between this scale and outcome.

The results of this study indicate some support for the Rogerian position. This study demonstrated a relationship between the client-perceived empathy scale taken at two different points in therapy with three of the outcome measures. Using the same data, the Relationship Inventory, when scored by an objective judge, was unrelated to any of the outcome measures. This suggests, then, that clients must experience the therapist's empathy in order for therapy to be successful.

Chapter V

SUMMARY

This study examined three measures of therapist empathy to (1) determine their construct validity and (2) provide additional information about the theoretical dispute between Rogers and Truax and Carkhuff regarding the necessity that the client actually perceive the therapist's empathy if therapy is to be successful.

Two of the therapist empathy measures required judges to rate tape-recorded interviews with (1) the Carkhuff Empathic Understanding in Interpersonal Processes Scale and (2) the Barrett-Lennard Relationship Inventory. The latter instrument was also used by the client to record his therapist's empathy. A previous study by Kurtz and Grummon (1972) failed to establish construct validity for several therapist empathy measures but could not determine if this was due to the empathy instrument or to the different persons making the ratings--i.e. an objective judge or the client. The present study corrected this deficiency. It also examined the

relationship between therapy outcome and the same measure of therapist empathy as recorded by the client and an objective judge, thereby adding to our knowledge about the theoretical dispute between Rogers and Truax and Carkhuff.

Twenty-four clients and their therapists at the Michigan State University Counseling Center served as the sample for this study. All the therapist empathy measures were correlated with each other and with six different measures of therapy outcome.

Clients completed the Relationship Inventory after the third therapy interview and again after the conclusion of therapy. One judge rated three tape-recorded segments of five different therapy interviews for each client to determine the therapist's level of empathy on the Carkhuff scale. Another judge made ratings on the same five therapy interviews to obtain the therapist's empathy score on the Relationship Inventory. The reliabilities for the tape-judged Relationship Inventory and Carkhuff empathy scale were low.

Therapeutic outcome was assessed by: 1) changes in the Number of Deviant Signs score of the Tennessee Self Concept Scale, 2) changes in the Total Positive score

of the Tennessee Self Concept Scale, 3) changes in the clinical judgments of improvement over therapy based upon MMPI profiles, 4) a four-point scale upon which the therapists made judgments as to the clients' progress in therapy, 5) a seven-point scale upon which clients recorded their judgments as to how helpful therapy had been to them, and 6) a composite outcome score obtained by converting the above measures into standard scores.

The results comparing the empathy scales show that: 1) there is no relationship between the tape-judged Carkhuff empathy scale and the client-perceived Relationship Inventory, 2) there is no relation between the tape-judged Relationship Inventory and the tape-judged Carkhuff scale, and 3) there is a positive relation between the client-perceived Relationship Inventory and the tape-judged Relationship Inventory. The results comparing the empathy scales to the outcome measures show that: 1) there is no relation between either the tape-judged Relationship Inventory or the tape-judged Carkhuff scale with any of the six outcome measures, and 2) there is a positive relation between three of the outcome measures and the client-perceived Relationship Inventory.

The first three findings comparing the empathy scales to each other found no support for construct validity of therapist empathy. Further, the findings overall tended to support the Rogerian rather than Truax-Carkhuff position that the client must perceive the therapist's empathy. The relation of these findings to the findings of previous research was discussed. A possible explanation was offered for the: 1) random relationship found between the tape-judged Carkhuff empathy scale with both the tape-judged and client-perceived Relationship Inventories, and 2) the random relationship found between both the Carkhuff scale and the tape-judged Relationship Inventory with the many outcome measures. Implications for future research regarding the therapist empathy concept were discussed.

REFERENCES

REFERENCES

Adler, A. The science of living. New York: Greenberg Publishers, 1929.

Ashcraft, C. & Fitts, W. W. Self concept change in psychotherapy. Psychotherapy, 1964, 1, 115-118.

Astin, H. A comparative study of the situational and predictive approaches to the measurement of empathy. Unpublished doctoral dissertation, University of Maryland, 1957.

Barrett-Lennard, G. T. Dimensions of therapist response as causal factors in therapeutic change. Psychological Monographs, 1962, 76 (43 Whole No. 562).

Buchheimer, A. The development of ideas about empathy. Journal of Counseling Psychology, 1963, 10, 61-70.

Burstein, J. W. & Carkhuff, R. R. Objective therapist and client ratings of therapist offered facilitative conditions of moderate to low functioning therapists. Journal of Clinical Psychology, 1968, 24, 240-249.

Carkhuff, R. R. Helping and human relations, Vol. 11, Practice and research. New York: Holt, Rinehart and Winston, 1969.

Cartwright, R. D. & Lerner, B. Empathy, need to change, and improvement with psychotherapy. Journal of Consulting Psychology, 1963, 27, 138-144.

Cronbach, L. J. & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Dickenson, W. & Truax, C. B. Group counseling with underachievers. Personnel and Guidance Journal, 1966, 45, 243-247.

Dombrow, R. A. A study of the relationship between therapist's empathy for patients and changes in patient's self concepts during therapy. Dissertation Abstracts, 1966, 27 (1-B), 301-302.

Dymond, R. F. A preliminary investigation of the relation of insight and empathy. Journal of Consulting Psychology, 1948, 12, 228-223.

Dymond, R. F. A scale for the measurement of empathic ability. Journal of Consulting Psychology, 1949, 13, 127-133.

Fenichel, O. The psychoanalytic theory of neurosis. New York: W. W. Norton, 1945.

Fiedler, F. C. A comparison of therapeutic relationships in psychoanalytic, nondirective, and Adlerian therapy. Journal of Consulting Psychology, 1950, 14, 436-445.

Fitts, W. W. Tennessee self concept scale: Manual. Nashville: Counselor Recordings and Tests, 1965.

Fromm-Reichmann, F. Principles of intensive psychotherapy. Chicago: University of Chicago Press, 1950.

Gendlin, E. T. Experiencing and the creation of meaning. New York: Free Press of Glencoe, 1962.

Gonyea, G. G. The "ideal therapeutic relationship" and counseling outcome. Journal of Clinical Psychology, 1963, 19, 481-487.

Hansen, J. C., Moore, G. D. & Carkhuff, R. R. The differential relationship of objective and client perceptions of counseling. Journal of Clinical Psychology, 1968, 24, 244-246.

Hountras, P. & Anderson, D. Counselor conditions for self-exploration of college students. Personnel and Guidance Journal, 1969, 48(1), 45-48.

Katz, B. Predictive and behavioral empathy and client change in short-term counseling. Unpublished doctoral dissertation, New York University, 1962.

Kurtz, R. R. & Grummon, D. L. Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. Journal of Consulting and Clinical Psychology, 1972, 39, 106-115.

Lesser, W. M. The relationship between counseling progress and empathic understanding. Journal of Counseling Psychology, 1969, 8, 330-336.

Rogers, C. R. The necessary and sufficient conditions of therapeutic personality change. Journal of Consulting Psychology, 1957, 21, 95-103.

Rogers, C. R. A theory of therapy, personality, and interpersonal relationships, as developed in the client centered framework. In S. Koch (Ed.), Psychology: a study of a science. Vol. III. Formulations of the person and the social context. New York: McGraw-Hill, 1959, 184-258.

Rogers, C. R. (Ed.) The therapeutic relationship and its impact: a study of psychotherapy with schizophrenics. Madison: University of Wisconsin Press, 1967.

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

Speroff, B. J. Empathy and role reversal as factors in industrial harmony. Journal of Social Psychology, 1953, 37, 117-120.

Sullivan, H. S. The interpersonal theory of psychiatry. New York: W. W. Norton, 1945.

Truax, C. B. Therapist empathy, warmth and genuineness, and patient personality change in group psychotherapy: A comparison between Interaction unit measures, time sample measures, patient perception measures. Journal of Clinical Psychology, 1966, 22, 225-229.

Truax, C. B., Wargo, D. C., and Silber, L. D. Effects of group psychotherapy with high accurate empathy and non-possessive warmth upon female institutionalized delinquents. Journal of Abnormal Psychology, 1966, 71, 267-274.

Truax, C. B. & Carkhuff, R. R. Toward effective counseling and psychotherapy. Chicago: Aldine, 1967.

Walker, H. W. and Lev, J. Statistical inference. New York: Herry Hold, 1953.

APPENDICES

APPENDIX A

NUMBER OF INTERVIEWS PER CLIENT

TABLE 3.--Number of interviews per client.

Client Number	Counselor Number	Number of Interviews*
832	02	3I
848	43	12M
812	12	12M
801	05	24M
834	35	7I
835	29	5M
805	22	5M
831	04	24M
818	26	17M
830	27	9M
808	08	9M
867	10	20M
829	25	6M
858	09	6M
815	18	7M
847	45	15M
845	44	15M
855	24	6M
828	15	13M
838	03	16M
849	49	17M
850	40	12M
825	28	7M
823	19	22M
		E=289
		$\bar{X}$=12.04

*I = Independent Termination.
M = Mutual Termination.

APPENDIX B

EMPATHY MEASURES USED IN THIS STUDY

EMPATHIC UNDERSTANDING IN INTERPERSONAL PROCESSES. II

A Scale for Measurement[1]

Robert R. Carkhuff

Level I

The verbal and behavioral expressions of the first person either do not attend to or detract significantly from the verbal and behavioral expressions of the second person(s) in that they communicate significantly less of the second person's feelings than the second person has communicated himself.

---

[1]The present scale "Empathic Understanding in Interpersonal Processes" has been derived in part from "A Scale for the Measurement of Accurate Empathy" by C. B. Truax which has been validated in extensive process and outcome research on counseling and psychotherapy (summarized in Truax and Carkhuff, 1967) and in part from an earlier version which has been validated in extensive process and outcome research on counseling and psychotherapy (summarized in Carkhuff and Berenson, 1967). In addition, similar measures of similar constructs have received extensive support in the literature of counseling and therapy and education. The present scale was written to apply to all interpersonal processes and represent a systematic attempt to reduce the ambiguity and increase the reliability of the scale. In the process many important delineations and additions have been made, including in particular the change to a systematic focus upon the additive, subtractive or interchangeable aspects of the levels of communication of understanding. For comparative purposes. Level 1 of the present scale is approximately equal to Stage 1 of the Truax scale. The remaining levels are approximately correspondent: Level 2 and Stages 2 and 3

Examples: The first person communicates no awareness of even the most obvious, expressed surface feelings of the second person. The first person may be bored or disinterested or simply operating from a preconceived frame of reference which totally excludes that of the other person(s).

In summary, the first person does everything but express that he is listening, understanding or being sensitive to even the feelings of the other person in such a way as to detract significantly from the communications of the second person.

Level 2

While the first person responds to the expressed feelings of the second person(s), he does so in such a way that he subtracts noticeable affect from the communications of the communications of the second person.

Examples: The first person may communicate some awareness of obvious surface feelings of the second person but his communications drain off a level of the affect and distort the level of meaning. The first person may communicate his own ideas of what may be going on but these are not congruent with the expressions of the second person.

In summary, the first person tends to respond to other than what the second person is expressing or indicating.

Level 3

The expressions of the first person in response to the expressed feelings of the second person(s) are essentially interchangeable with those of the second person in that they express essentially the same effect and meaning.

---

of the earlier version; Level 3 and Stages 4 and 5; Level 4 and Stages 6 and 7; Level 5 and Stages 8 and 9. The levels of the present scale are approximately equal to the levels of the earlier version of this scale.

Example: The first person responds with accurate understanding of the surface feelings of the second person but may not respond to or may misinterpret the deeper feelings.

The summary, the first person is responding so as to neither subtract from nor add to the expressions of the second person; but he does not respond accurately to how that person really feels beneath the surface feelings. Level 3 constitutes the minimal level of facilitative interpersonal functioning.

Level 4

The responses of the first person add noticeably to the expressions of the second person(s) in such a way as to express feelings of level deeper than the second person was able to express himself.

Example: The facilitator communicates his understanding of the expressions of the second person at a level deeper than they were expressed, and thus enables the second person to experience and/or express feelings which he was unable to express previously.

In summary, the facilitator's responses add deeper feeling and meaning to the expressions of the second person.

Level 5

The first person's responses add significantly to the feeling and meaning of the expressions of the second person(s) in such a way as to (1) accurately express feelings levels below what the person himself was able to express or (2) in the event of ongoing deep self-exploration on the second person's part to be fully with him in his deepest moments.

Examples: The facilitator responds with accuracy to all of the person's deeper as well as surface feelings. He is "together" with the second person or "tuned in" on his wave length. The facilitator and the other person might proceed

> together to explore previously unexplored areas of human existence.

In summary, the facilitator is responding with a full awareness of who the other person is and a comprehensive and accurate empathic understanding of his most deep feelings.

RELATIONSHIP INVENTORY--CLIENT FORM

(Please do not write your name on this form. It will be coded anonymously and your answers used for research purposes only.)

Below are listed a variety of ways that one person could feel or behave in relation to another person. Please consider each statement with respect to whether you think it is true or not true in your present relationship with your therapist. Mark each statement in the left margin according to how strongly you feel it is true or not true. Please mark every one. Write in +1, +2, +3; or -1, -2, -3, to stand for the following answers:

+1: I feel that it is probably true, or more true than untrue.

+2: I feel it is true.

+3: I strongly feel that it is true.

-1: I feel that it is probably untrue, or more untrue than true.

-2: I feel it is not true.

-3: I strongly feel that it is not true.

1. _____ He tries to see things through my eyes.

2. _____ He understands my words but not the way I feel.

3. _____ He is interested in knowing what my experiences mean to me.

4. _____ He nearly always knows exactly what I mean.

5. _____ At times he jumps to the conclusion that I feel more strongly or more concerned about something that I actually do.

6. _____ Sometimes he thinks that I feel a certain way, because he feels that way.

7. _____ He understands me.

8. _____ His own attitudes toward some of the things I say, or do, stop him from really understanding me.

9. _____ He understands what I say, from a detached, objective point of view.

10. _____ He appreciates what my experiences feel like to me.

11. _____ He does not realize how strongly I feel about some of the things we discuss.

12. _____ He responds to me mechanically.

13. _____ He usually understands all of what I say to him.

14. _____ When I do not say what I mean at all clearly he still understands me.

15. _____ He tries to understand me from his one point of view.

16. _____ He can be deeply and fully aware of my most painful feelings without being distressed or burdened by them himself.

Relationship Inventory--Judges Form

Client_____

Code number_____

Rater #_____

_____ 1. The therapist tries to see things through the client's eyes.

_____ 2. The therapist understands the client's words but not the way he feels.

_____ 3. The therapist is interested in knowing what the client's experiences mean to the client.

_____ 4. The therapist nearly always knows what the client means.

_____ 5. At times the therapist jumps to the conclusion that the client feels more strongly or more concerned about something than the client actually.

_____ 6. Sometimes the therapist thinks that the client feels a certain way because he (the therapist) feels that way.

_____ 7. The therapist understands the client.

_____ 8. The therapist's own attitudes toward some of the things the client says or does stop him from really understanding the client.

_____ 9. The therapist understands what the client says from a detached, objective point of view.

_____10. The therapist appreciates what the client's experiences feel like to him (the client).

_____11. The therapist does not realize how strongly the client feels about some of the things they discuss.

_____12. The therapist responds to the client mechanically.

_____13. The therapist usually understands all of what the client says to him.

_____14. When the client does not say what he means at all clearly, the therapist still understands him.

_____15. The therapist tries to understand the client from his (the therapist's) own point of view.

_____16. The therapist can be deeply and fully aware of the client's most painful feelings without being distressed or burdened by them himself.

_____ Total

APPENDIX C

RAW SCORES OBTAINED BY CLIENTS AND OBJECTIVE JUDGES ON THE EMPATHY MEASURES

TABLE 4.--Barrett-Lennard Relationship Inventory, Client Perceptions after 3rd interview.*

Client	Counselor	Score	# Converted Score
832	02	-4	10
848	43	34	48
812	12	29	43
801	05	23	37
834	35	33	47
835	29	32	46
804	22	8	22
831	04	16	30
818	26	25	39
830	27	24	38
808	08	09	33
867	10	05	19
829	25	4	18
858	09	25	39
815	18	15	29
847	45	22	36
845	44	22	36
855	24	18	32
828	15	8	22
838	03	21	35
849	49	25	39
850	40	8	22
825	28	25	39
823	19	4	18
N = 24		$\bar{X}$ = 32.37	SD = 10.12

*Constant of 14 added to avoid negative numbers.

TABLE 5.--Barrett-Lennard Relationship Inventory, Client Perceptions, Post.*

Client	Counselor	Score	# Score Converted
832	02	-9	5
848	43	27	41
812	12	36	50
801	05	34	48
834	35	35	49
835	29	30	44
804	22	8	22
831	04	3	17
818	26	34	48
830	27	26	40
808	08	-12	2
867	10	4	18
829	25	9	23
858	09	26	40
815	18	16	30
847	45	21	35
845	44	27	41
855	24	16	30
828	15	23	37
838	03	29	43
849	49	40	54
850	40	7	21
825	28	27	41
823	19	-1	13
N = 24	$\bar{X}$ = 33	SD = 14.39	

*Constant of 14 added to avoid negative numbers.

TABLE 6.--Barrett-Lennard Relationship Inventory, Tape-rated* Average ratings.

Client	Counselor	Mean Score
832	02	29.67
848	43	46.00
812	12	36.20
801	05	39.40
834	35	42.00
835	29	46.00
804	22	14.00
831	04	49.20
818	26	34.00
830	27	47.40
808	08	30.40
867	10	24.80
829	25	42.80
858	09	45.40
815	18	17.60
847	45	39.60
845	44	46.00
855	24	36.80
828	24	36.80
838	03	46.00
849	49	46.00
850	40	7.60
825	28	49.20
823	19	46.00

N = 24 $\quad \bar{X} = 33.95 \quad$ SD = 19.42

*Constant of 14 added to avoid negative numbers.

TABLE 7.--Judges ratings of counselor empathy average ratings, Carkhuff Empathic understanding in Interpersonal Processes Scale.

Client	Counselor	Average Ratings
832	02	3.17
848	43	1.90
812	12	2.00
801	05	3.10
834	35	2.80
835	29	2.30
804	22	1.75
831	04	2.20
818	26	2.00
830	27	2.40
808	08	2.60
867	10	2.00
829	25	2.00
858	09	2.70
815	18	1.70
847	45	1.20
845	44	1.70
855	24	2.10
828	15	2.40
838	03	2.20
849	49	1.30
850	40	1.90
825	28	2.20
823	19	1.88
N = 24	$\overline{X}$ = 2.15	SD = .37

APPENDIX D

SCORES ON OUTCOME MEASURES

TABLE 8.--Tennessee Self Concept difference scores.

Client	Counselor	Total "p" Gain	NDS Fewer
832	02	36	10
848	43	55	33
812	12	40	2
801	05	44	17
834	35	34	8
835	29	-2	2
804	22	-9	-17
831	04	16	2
818	26	12	17
830	27	-3	0
808	08	103	40
867	10	2	13
829	25	-23	10
858	09	53	14
815	18	0	-4
847	45	6	-7
845	44	68	22
855	24	-9	-9
828	15	10	-5
838	03	34	11
849	49	60	20
850	40	38	13
825	28	42	37
823	19	-20	11

TABLE 9.--MMPI profile analysis--average ratings.

Client	Counselor	Average Ratings
832	02	2.17
848	43	5.00
812	12	3.50
801	05	3.67
834	35	5.00
835	29	3.67
804	22	
831	04	1.67
818	26	5.00
830	27	3.00
808	08	5.00
867	10	2.17
829	25	1.33
858	09	4.67
815	18	3.33
847	45	3.33
845	44	5.00
855	24	3.33
828	15	4.17
838	03	
849	49	4.33
850	40	4.83
825	28	5.00
823	19	1.67
N = 22	$\overline{X}$ = 3.67	SD = 1.228

1--Unsatisfactory
2--Partly unsatisfactory
3--No change
4--Partly satisfactory
5--Satisfactory

TABLE 10.--Clients and counselors evaluation of counseling scores.

Client	Counselor	Client Rating	Counselor Rating
832	02	4	3
848	43	6	3
812	12	7	4
801	05	7	4
834	35	7	4
835	29	6	3
804	22	5	4
831	04	5	3
818	26	7	4
830	27	6	4
808	08	5	3
867	10	6	3
829	25	6	4
858	09	6	2
815	18	-	3
847	45	6	4
845	46	6	3
855	24	5	2
828	15	7	3
838	03	7	3
849	49	6	4
850	40	6	3
825	28	7	4
823	19	5	2

Client rating:

1--extremely harmful
2--harmed me a lot
3--harmed me somewhat
4--indifferent
5--helped me somewhat
6--helped me quite a lot
7--extremely helpful

Counselor rating:

1--unsuccessful
2--partially unsuccessful
3--partially successful
4--successful

TABLE 11.--Post t scores--for combined outcome score.

Client	Counselor	MMPI	Ten "p"	Ton NDS	CO EV	Cl EV	Ave.
832	02	38	54	50	49	27	44
848	43	61	61	70	49	51	58
812	12	49	56	48	59	64	55
801	05	54	61	52	59	64	58
834	35	61	54	50	59	64	58
835	29	50	42	53	49	51	49
804	22	--	38	34	59	39	
831	04	34	45	73	49	39	46
818	26	61	45	59	59	64	58
830	27	45	42	55	59	51	50
808	08	61	75	63	49	39	57
867	10	38	42	47	49	51	45
829	25	31	34	40	59	41	46
858	09	59	60	56	38	51	53
815	18	48	42	40	49	51	46
847	45	48	43	42	59	51	49
845	44	61	65	68	49	51	59
855	24	48	42	40	38	39	41
828	15	55	45	43	49	64	51
838	03	--	53	42	59	39	
849	49	66	61	61	59	51	60
850	40	60	55	56	49	51	54
825	28	61	56	67	59	64	61
823	19	34	40	40	38	39	38