THE RELATIONSHIP BETWEEN THERAPIST ACCURATE EMPATHY AND THE SUCCESS OF PSYCHOTHERAPY

A THESIS

PRESENTED TO THE

FACULTY OF GRADUATE STUDIES

UNIVERSITY OF MANITOBA

IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

by

LYNN MEDA STOFFMAN

FEBRUARY, 1972

ACKNOWLEDGEMENTS

The author wishes to express her appreciation to Dr. D. G. Martin for providing the taped therapy sessions and other data which served as the basis for this research. She is indebted to Dr. D. G. Martin for his assistance offered throughout all phases of her work. The author also wishes to express her appreciation to Dr. J. Shallow, Dr. M. Wright and Mr. K. Alyluia for their critical evaluation of this thesis.

ABSTRACT

Prior research has shown that the three therapist qualities of unconditional positive regard, accurate empathy, and genuineness, in combination, tend to be positively related to therapeutic outcome. A number of studies (Truax, Wargo and Carkhuff, 1966; Truax, Wargo, Frank, Imber, Battle, Hoehn-Saric, Nash, and Stone, 1966; Vander Veen, 1967; Mullen and Abeles, 1971) attempted to analyze separately the three therapeutic conditions; the results of these studies were not consistent with regards to accurate empathy. The present study was an attempt to find out whether accurate empathy in and of itself is significantly related to patient outcome.

It was hypothesized that those clients who received a relatively high level of empathic behavior from their therapists, would have a greater increase in reported self-esteem (Total Positive Score) on the Tennessee Self Concept Scale from pre-therapy to post-therapy than would those clients who received a relatively low level of therapist empathic behavior.

It was also hypothesized that those therapists who provided their clients with a relatively high level of accurate empathy, would rate the success of the therapy higher than would those therapists who provided a relatively low level of empathy.

Similarly, it was hypothesized that those clients who received a relatively high level of accurate empathy, would rate the outcome of their therapy as having been more successful than would those who received a relatively low level of accurate empathy.

Four graduate students, trained in accordance with Carkhuff's (1969) five level scale of accurate empathy, rated the therapist accurate empathy found in samples taken from taped therapy sessions. The results of the Tennessee Self Concept Scale, administered both before and after therapy, served as one measure of therapeutic outcome. The scores obtained on a questionnaire that was also administered before and after therapy, served as the measure of felt success.

Overall, the results of the study did not confirm the main hypotheses. The clients who received a relatively high level of empathy did not have a greater increase in reported self-esteem (Total P Score) on the TSCS from pre to post-therapy than the clients who received a relatively low level of empathy. Contrary to the hypotheses, the therapists who displayed a relatively low level of empathy tended to evaluate the overall success of their clients significantly higher ($p < .05$) than did those who displayed a relatively high level of empathy. All the other results dealing with "felt success" did not prove to be significant at the .05 level.

A number of ancillary analyses were also performed to further explore the data available.

Several possible explanations for the results were discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (CONT'D)

CHAPTER I

INTRODUCTION

In recent years, much work has gone into the search for those therapeutic variables which contribute to the constructive change of the therapy patient. Rogers (1957), has put forth three therapist conditions as being necessary and sufficient for the client to achieve a therapeutic change: unconditional positive regard (nonpossessive warmth), accurate empathy, and genuineness (self-congruence).

The first therapist condition, unconditional positive regard, is based upon the premise that a human being is of inherent value or worthy, and ought to be positively or warmly regarded, regardless of the goodness or badness of the acts that he performs. Accurate empathy can be defined as the therapist's communicated understanding of the client's intended message; the therapist must be able to sense the "inner world" of his client as if it were his own and then be able to communicate this awareness to the client. In order for the therapist to be genuine or self-congruent, what he is communicating must be consistent with his own feelings and attitudes.

Several investigators have attempted to find a relationship between the three therapeutic conditions and patient outcome.

Halkides, as discussed by Barrett-Lennard (1962), found that the most successful of twenty outpatients, as defined on multiple criteria, had therapists who showed significantly higher

levels of the three conditions, than did the least successful patients' therapists. The scales used to arrive at the therapist levels of the three conditions were based upon Rogers' 1957-article.

Hart, as discussed by Truax (1963), replicated the Halkides study, and used the same data and similar methodology but different raters of the three conditions. Halkides' results were not confirmed by Hart's findings. Truax and Carkhuff (1967) suggest that thisdiscrepancy may be due to the global and implicitly defined scales that were used in these two studies. Also, the ratings in Halkides' study were made by experienced therapists, while Hart used less experienced raters who appear to have exercised some judgement independently of the scales used.

Barrett-Lennard (1962), designed a questionnaire to tap the client's perception of his therapists's empathy, warmth, and congruence. He found these variables to be positively related to therapeutic outcome.

Truax (1961(a)) found that therapists of four patients who improved on a variety of personality tests, were rated higher on the Accurate Empathy Scale (Truax, 1961(b)) than were the therapists of four patients who had deteriorated. The samples to be rated were taken from the middle third of the therapy sessions; the samples were two minutes in length and there were 384 samples in total.

A number of studies (Truax 1963; Truax and Carkhuff 1963; Truax and Carkhuff 1967; Rogers, Gendlin, Kiesler, and Truax, 1967) that were based upon Rogers' work with hospitalized schizophrenic

patients at the University of Wisconsin which began in 1958, indicate that the therapist's level of functioning on the three variables is related to the degree of patient improvement or deterioration.

The results of a study (Truax, 1963) involving 14 schizophrenic patients found accurate empathy, as rated by trained undergraduate students on the Accurate Empathy Scale, to be significantly related to psychological test change data, diagnostic evaluations of personality change, time spent in the hospital since initiation of therapy, and to degree of change in personality functioning as measured by pre and post-measures on the Rorschach and MMPI. The samples to be rated were chosen from every fifth interview and were four minutes in length. Once again, the improved group had higher therapist levels of empathy than did the deteriorated group.

Truax (1963), also found a positive relationship between accurate empathy as measured on the Accurate Empathy Scale and outcome of therapy for 14 outpatient cases. Those patients who were relatively successful were found to have therapists rated higher in accurate empathy. The patients who were relatively unsuccessful tended to have therapists with low and moderate levels of accurate empathy.

One study, (Truax, Wargo, Frank, Inber, Battle, Hoehn-Saric, Nash and Stone, 1966), served to cross-validate the results from the studies on individual psychotherapy with hospitalized schizophrenics and with outpatients from college counselling centres.

In this study 40 outpatients were treated by resident psychiatrists of a psychiatric clinic. The analysis indicated greater improvement for patients of therapists providing high levels of accurate empathy, non-possessive warmth, and genuineness combined than for patients receiving lower levels of these combined conditions. Secondly, those patients who received high conditions tended to show positive change, while a higher percentage of patients receiving low conditions tended to show negative change. Thirdly, those therapists who provided high levels of conditions had patients with a 90% improvement rate, while those providing low levels of conditions had patients with only a 50% rate of improvement. A separate analysis of the specific conditions revealed identical findings for empathy, and genuineness, but the opposite trend for nonpossessive warmth.

The study by Truax, Carkhuff, and Kodman (1965) illustrated the importance of the three therapeutic conditions to 40 hospitalized chronic mental patients receiving group psychotherapy. The patients were divided according to whether they were receiving high levels of the three conditions or relatively low levels. Those receiving high levels of accurate empathy or nonpossessive warmth showed improvement on the MMPI subscales equal to, or greater than that of the patients receiving relatively low levels. The data on genuineness was in the opposite direction to the prediction.

The findings of the above two mentioned studies suggest that when two therapeutic conditions are highly related but the

third is negatively related, the prediction of outcome should be based on the two that are most highly related. Truax and Carkhuff (1967) point out that Gendlin and Geist's suggestion may be correct - "that when any one of the therapeutic conditions is sufficiently low, it will interfere with therapy regardless of the level of the remaining two therapeutic conditions (p.92)."

The study by Truax and Wargo (1966(a)) involved a more heterogeneous patient and therapist population. The study involved 160 hospitalized patients receiving group psychotherapy. The results indicated a significant relationship between the levels of the three conditions with different measures of positive patient outcome - Q-sort measures of self concept, the MMPI subscales of Mf and Sc, the Welsch Anxiety Index, and time spent out of the hospital during one year followup. Similar results were obtained from an institutionalized juvenile delinquent population (80 patients) receiving group counselling and resulting behavior and personality change (Truax and Wargo, 1966 (b)).

Truax, Wargo, and Carkhuff (1966) carried out another study on 80 outpatients receiving group psychotherapy and also obtained similar findings for all three conditions combined. On 23 measures of outcome examined, the patients who were offered high levels on all conditions showed above-average improvement on 21 measures and below-average improvement on two measures. The opposite occurred for patients who were offered relatively low

conditions. An analysis of the three conditions separately, suggested warmth to be most related to improvement, genuineness less related, and empathy the least related.

Mullen and Abeles (1971) conducted a study in which raters scored interview samples for accurate empathy and "liking"; the Accurate Empathy Scale (Truax and Carkhuff, 1967) and the Nonpossessive Warmth Scale (Truax and Carkhuff, 1967) were the scales used for rating. In this study, there were 36 clients and 36 different therapists. The clients were divided into a successful and an unsuccessful therapy outcome group based on changes of pre to post MMPI clinical scales. The results indicated that "high liking and high empathy together did not predict successful outcome, though a post hoc analysis showed a positive relationship between high empathy alone and successful outcome (p. 39)."

The studies as discussed above deal with relatively successful and relatively unsuccessful cases, and, as a whole, provide support for the hypothesis that the level of therapist accurate empathy, nonpossessive warmth and genuineness are related to constructive change in patients, whether they be diagnosed as schizophrenic, neurotic, or delinquent. This seems to hold for both individual and group psychotherapy.

There have been a number of studies conducted which have made comparisons between control groups receiving no psychotherapy and the experimental groups receiving therapy.

The results from the work conducted in Wisconsin with 16 schizophrenic patients receiving individual psychotherapy and

16 matched control patients (Truax 1963, Truax and Carkhuff 1963) revealed a significant difference in psychological functioning among patients receiving high conditions, patients receiving low conditions and control patients. All patients in the "low group" were below the median change in psychological functioning; six of the eight patients receiving high conditions showed positive change; whereas half in the control group were above the median and half below. Also, it was found that "the high group" spent significantly more time out of the hospital than either the control group or "the low group", while "the low group" did not differ from the control population.

Truax (1970), conducted a nine-year follow up on the 16 schizophrenic patients from the original Wisconsin Schizophrenic Project. He examined the effects of empathy, warmth, and genuineness on hospitalization. The records of several hospitals were checked over the nine years preceeding therapy and the nine years after the initiation of therapy for the therapy and control patients. Truax found a significant difference in the linear trend of getting out of the hospital across time between patients receiving control conditions versus those receiving high therapeutic conditions from their therapists. Patients receiving high conditions showed significantly greater linear trend of getting out of the hospital than those receiving low conditions. Also, patients receiving high conditions tended to get out of the hospital quicker at the outset than did those receiving the control conditions. No difference was found

in the High, Low, and Control Groups in the nine years prior to treatment. Also, in the second nine year period, patients receiving low conditions tended not to get out of the hospital and, if they did get out, tended to come back to hospital. Thus, this study provides evidence for the long-term effects of empathy, warmth, and genuineness.

Truax, Wargo, and Silber (1966), conducted a study involving 40 female juvenile delinquents in group counselling. The results indicated that on several personality and self-concept measures obtained pre- and post-therapy, the delinquents receiving high conditions in group psychotherapy showed improvement significantly beyond that seen in the control group. Also, those receiving the high conditions spent significantly more time out of the institution than those in the low group and control group; after a one year follow-up the control group had spent more time in the institution than the experimental groups.

The studies involving control groups provide some evidence of the superiority of treatment to no treatment, especially when those treated receive high levels of accurate empathy, non-possessive warmth, and genuineness.

The studies discussed above were mainly concerned with the effect of the three conditions in combination on therapeutic outcome.

The question arises as to whether it is possible to predict accurately the success of the therapy by an examination of only one

of these qualities. As discussed previously, the study by Truax, Wargo, Frank, Imber, Battle, Hoehn-Saric, Nash, and Stone, 1966, did separately analyze the three therapeutic conditions. They found for both empathy and genuineness separately, that when the therapists provided high levels on one of these conditions, there was greater patient improvement than when the therapists provided a low level. Those patients who received a high condition (on empathy or genuineness) tended to show positive change, while a higher percentage of patients receiving the low condition tended to show negative change. The opposite trend was found for non-possessive warmth.

Truax, Wargo, and Carkhuff (1966) also conducted an analysis of the effects of the three conditions separately. This time the results suggested warmth to be most related to improvement, genuineness to a lesser degree, and empathy to be the least related.

Ferdinand van der Veen (1967), also examined the relationship of the three conditions, one at a time, to therapeutic outcome. This study involved 15 hospitalized schizophrenic patients who had undergone individual therapy with one of ten different therapists. The results of the therapy were based upon changes on the Rorschach, a 10-card TAT, the MMPI, a self-concept Q sort, the Wechsler Adult Intelligence Scale, and a 246-item anxiety scale. The combined outcome score was also based upon the percentage of time hospitalized. The samples rated were four minutes in length which were randomly selected from the first third and last third of five interviews; the five interviews selected from each case were taken from the beginning

of therapy, the 25% point, the 50% point, the 75% point, and the end of therapy. The ratings were based upon The Congruence Scale, The Positive Regard Scale, and the Accurate Empathy Scale.

The results revealed that the level of the therapist conditions, as examined separately, was related positively to case outcome.

> Empathy showed strong relationship with outcome for the overall mean as well as the initial, 25%, 50%, and 75% points, particularly with the combined outcome score. Congruence had borderline ($p < .10$) relationships with the combined outcome score for the overall mean and for the 25% ($r=.47$) and 75% ($r=.48$) points of therapy. The overall mean for positive regard did not reach significance, though the 25% point correlated significantly with the combined outcome score (.55) and with MMPI change (.59). All p values in the study are for two-tailed tests of significance (pp. 298, 299).

As discussed earlier (see p. 6), Mullen and Abeles (1971) found that high liking (as measured on the Nonpossessive Warmth Scale) and high empathy "together did not predict successful outcome, though a post hoc analysis showed a positive relationship between high empathy alone and successful outcome (p. 39)." A closer look at the results revealed that when both liking and empathy

were at a high level and occurred at different times in therapy, they contributed to successful outcome. When the prediction of success was based on the occurrence at any time in therapy of high conditions of both liking and empathy, actual successes were correctly predicted in 18 out of 20 cases. However, when the outcome was unsuccessful, correct prediction occurred only 5 out of 16 times. Mullen and Abeles explained that their findings support the theory that high levels of empathy are necessary but not necessarily sufficient for successful therapeutic outcome. It appears that "a high level of empathy typically occurs before a therapist experiences caring for a client in a nonjudgmental appreciative way, and indeed is a prerequisite of it (p. 43)."

The results of these four studies are not consistent with regards to accurate empathy. In two of the four studies, accurate empathy alone was found to be positively related to therapeutic outcome; in two studies, it was found to be relatively unrelated. The question remains - Is the therapist's level of accurage empathy in and of itself significantly related to patient outcome?

The major criticism levied against accurate empathy as being a meaningful variable which may influence the outcome of psychotherapy, is very well explained by Chinsky and Rappaport (1970). They suggest that the reliability estimates of the accurate empathy ratings are greatly dependent upon the number of therapists being rated. Reliability coefficients may be spuriously inflated if the

raters are responding to an identifiable therapist; for example, the therapist may be identified by his voice or through his language style. Instead of rating the amount of accurate empathy displayed by the therapist, another characteristic of the therapist may in fact be what is being rated. Whenever the rater recognizes a particular therapist he assigns a particular rating so that the ratings of each sample are not independent of each other. Recognition of the therapist would more likely occur when there are only a small number of different therapists being rated.

Although this does seem to be a valid criticism, the bulk of the experimental evidence does give support to accurate empathy as being a meaningful and significant variable in the therapy process. As discussed in the Discussion section of this paper, there are many methodological and theoretical issues which greatly complicate the investigation of the role that accurate empathy plays in therapy, and for this reason repeated studies dealing with accurate empathy are warranted.

It was hypothesized that those clients who receive a relatively high level of empathic behavior from their therapists, will have a greater increase in reported self-esteem (Total Positive Score) on the Tennessee Self Concept Scale from pre-therapy to post-therapy, than will those clients who receive a relatively low level of therapist empathic behavior.

In addition, it was hypothesized that those therapists

who provide their clients with a relatively high level of accurate empathy, will rate the success of the therapy higher than will those therapists who provide a relatively low level of empathy.

Similarly, it was hypothesized that those clients who receive a relatively high level of accurate empathy, will rate the outcome of their therapy as more successful than will those who receive a relatively low level of accurate empathy.

CHAPTER II

METHOD

Four graduate students attending the University of Manitoba, were trained to rate therapy samples on a scale of accurate empathy. Each rater was provided with a copy of Carkhuff's (1969) five level scale of accurate empathy. In addition, in order to facilitate rating, a number of conventions were agreed upon in the training sessions. (See Appendix A).

The first two training sessions were spent in discussion of the levels, and in training the raters to independently rate four minute practice samples taken randomly from a number of different taped therapy session. (The clients involved in these practice samples were not those involved in the experimental samples.) Each rater listened to one sample at a time, made an independent rating, and the instructor discussed the ratings of each sample with the group of raters as a whole. The instructor explained why a particular rating was appropriate or inappropriate for that sample.

In order to provide the raters with rating practise, each rater independently and in physical isolation from the others, rated 20 four minute samples similarly derived as the above-mentioned samples used for the training sessions. Half of the 20 samples were rated in one sitting.

In order to determine inter-rater reliability, the raters once again in physical isolation from one and other, rated 10 other four minute samples. These samples were taken randomly from the

middle third of the ninth interview of the experimental clients. (See below). Each rater rated the samples in a different order chosen randomly from a random numbers table. It was decided that a reliability coefficient of r=+.7 or above would be the cut-off point for adequate inter-rater reliability. Because none of the reliability coefficients met the designated criterion of r=+.7, (see Table 1), a third training session was held in which the ratings of the ninth interview samples were discussed. Although the ratings from the ninth interview were not used in the later experimental analysis, the possibility existed that the discussion of the ninth interview could have biased the rating of the experimental samples if the raters remembered specific voices of the client and/or therapist. However, the E felt that the procedure would not bias the results any more than the bias introduced by voice recognition from one interview to the others. (See following information on Method).

Because of limited resources and time, further training sessions and testing for adequate inter-rater reliability (r=+.7) could not be carried out.

The experimental samples consisted of four minute segments, taken randomly from the middle third of the second, fourth and eighth taped therapy interviews of twenty-two clients attending the University of Iowa Counselling Center. All of the clients were self-referred for help with personal problems (rather than vocational or educational problems). They were seen in an intake

TABLE I

Reliability Coefficients of Samples Taken From Ninth Interview

Raters (Coded by First Initial)	r
G-K	.054
K-S	.130
K-D	.000
G-S	.476
G-D	.466
S-D	.442

interview in which they agreed to be research clients. Also, each client participated in three hours of testing prior to the first interview with the therapist. The clients were retested after the tenth session with the therapist. All of the therapists were advanced graduate students. Each client was assigned to a therapist by the therapist's advisor on a non-random basis but without reference to any of the research data. The therapist was assigned a client if he had time available for him.

One sample was taken from each client for each of the three interviews. Each sample contained at least two therapist and two client comments, in order to facilitate the rating of accurate empathy. Each rater rated the samples in a random order according to a table of random numbers.

It was decided that three out of four raters would need to have the same rating before a rating could be determined for that sample. The E's advisor, Dr. D. Martin, allotted the final rating to a sample if there was agreement by only two raters or if there was no agreement. The E was fully aware that this criterion was not very stringent, but limited resources and time did not permit higher standards to be used.

Each client-therapist dyad was given an overall accurate empathy rating by averaging the ratings obtained from the second, fourth, and eighth interviews.

The measure of therapeutic outcome used in the experimental analysis was based upon the pre- and post-Total Positive Scores (Total P)

of the Tennessee Self Concept Scale. The Tennessee Self Concept Scale was developed by Fitts (1965). It is a paper and pencil test which measures self-reported self-esteem, and is scored on empirical scales. The Total P Score, as explained by Fitts (1965), "reflects an overall level of self esteem. Persons with high scores tend to like themselves, feel that they are persons of value and worth, have confidence in themselves, and act accordingly. People with low scores are doubtful about their own worth, see themselves as undesireable; often feel anxious, depressed, and unhappy; and have little faith or confidence in themselves (p. 2)." The Total P Score is made up of a number of sub-scores: Identity, Self-Satisfaction, Behavior, Physical Self, Moral-Ethical Self, Personal Self, Family Self, and Social Self.

The measure of therapeutic outcome was also based upon data on how successful the experimental clients felt their therapy had been for them after having had ten therapy sessions, and also on how successful their therapists felt their clients had been. These data on"felt success" were available for only seventeen of the twenty-two experimental dyads. The client (and separately the therapist) was given 11 different statements regarding success, and marked his goal for each statement and his felt achievement for each statement on an ascending scale from one to nine. (See Appendix B). A Goal-Achieved (G-A) score was derived by multiplying the Goal by the Achieved for each statement and then by finding an average G-A for that client or therapist.

Also, after the tenth interview, each client and therapist independently made an overall evaluation of felt progress, on an ascending scale from one to nine. (See Appendix B). These data were also utilized in the final analysis.

CHAPTER III

RESULTS

The main analysis of the relationship between therapist empathic behavior and increase in reported self-esteem was carried out with a mixed design of analysis of variance. There was one between subjects factor, being high versus low level of therapist accurate empathy; the one within subjects factor was pre-therapy versus post-therapy Total P Score.

In order to perform the analysis, it was first necessary to divide the client-therapist dyads into High and Low Accurate Empathy Groups. Each dyad was given an overall accurate empathy rating by averaging the ratings obtained from the second, fourth, and eighth interviews. Those dyads obtaining an overall rating of 3.000 or above were designated as the High Accurate Empathy Group, while those obtaining an overall rating less than 3.000 were designated as the Low Accurate Empathy Group. A cut-off of 3.000 was chosen since levels one and two, by definition, describe non-empathic therapist behavior. (See Appendix A). It was thus thought most reasonable to assign dyads with therapists who functioned at a level of 3.000 or higher to the High Empathy Group. Of the 22 dyads assigned to these groups, nine of the dyads were in the High Accurate Empathy Group, and 13 were in the Low Accurate Empathy Group.(See Table 2).

Contrary to the hypothesis, clients belonging to the High Accurate Empathy Group did not have a greater increase in

reported self-esteem (Total P Score) from pre- to post-therapy (F=1.249, df=1,20, $p > .05$). See Table 3. There was a nonsignificant trend toward a decrease in self-esteem from pre- to post-therapy for both groups. (Mean of High Empathy Group = 335.778 to 293.111, Mean Low Group = 325.077 to 316.154). The High Group began with a higher self-esteem score than did the Low Group.

No other main effects were found. A significant difference between the High Accurate Empathy Group and Low Accurate Empathy Group was not found (F=0.090, df=1,20, $p = > .50$). See Table 3. The mean Total P Score for the High Empathy Group was 314.444 while the mean for the Low Empathy Group was 320.615. Both means fall at approximately one standard deviation below the mean.

No significant difference on the self-esteem score was found between pre and post-therapy (F=3.021, df=1,20, $p = > .05$), (See Table 3), although there is some trend towards a lowering of self-esteem post-therapy (pre-mean = 330.427, post-mean = 304.632).

It was also hypothesized that the level of accurate empathy would be related to the therapist's rating of the success of the therapy. Data on felt success of therapy was available for only 17 of the 22 dyads. Of the 17, seven belonged to the High Empathy Group, and 10 to the Low Empathy Group. As mentioned in the Method, both an overall evaluation and a more detailed evaluation of success were available.

The therapists' overall evaluation of the clients' success

TABLE 2

DIVISION OF DYADS INTO HIGH AND LOW ACCURATE EMPATHY GROUPS ACCORDING TO OVERALL EMPATHY RATING

		Code Name For Dyads	Overall Empathy Rating
High Accurate Empathy Group	1.	Mahbaf	3.000
	2.	Mahbeb	3.333
	3.	Mahbew	3.000
	4.	Mahbor	3.000
	5.	Mahbuc	3.333
	6.	Mahcab	3.333
	7.	Mahcot	3.000
	8.	Mahdem	3.000
	9.	Mahduf	3.000
Low Accurate Empathy Group	1.	Mahbac	2.667
	2.	Mahbam	1.333
	3.	Mahbap	2.000
	4.	Mahbat	1.333
	5.	Mahbaw	2.333
	6.	Mahbof	2.667
	7.	Mahceb	2.333
	8.	Mahcec	2.333
	9.	Mahdot	2.667
	10.	Mahduc	1.333
	11.	Mahfof	2.333
	12.	Mahfuc	2.333
	13.	Mahguf	1.667

TABLE 3

SUMMARY OF MIXED ANALYSIS OF VARIANCE PERFORMED WITH EMPATHY GROUPS AND PRE-POST TOTAL P SCORE

Source	SS	DF	MS	F
Empathy Groups	405.5112	1	405.5112	0.090
Error 1	89736.000	20	4486.7969	
Pre-Post	7319.8125	1	7319.8125	3.021
Interaction	3026.7102	1	3026.7102	1.249
Error 2	48458.000	20	2422.8999	
Total	147308.000	43		

TABLE 4

SUMMARY OF SIMPLE ANALYSIS OF VARIANCE

PERFORMED WITH EMPATHY GROUPS AND THERAPISTS'

OVERALL EVALUATION OF SUCCESS

Source	SS	DF	MS	F	t
Empathy Groups	9.9845	1	9.9845	5.17	2.27
Success	28.9563	15	1.9304		
Total	38.9402	16			

TABLE 5

SUMMARY OF SIMPLE ANALYSIS OF VARIANCE

PERFORMED WITH EMPATHY GROUPS AND THERAPISTS'

DETAILED EVALUATION OF SUCCESS

Source	SS	DF	MS	F	t
Empathy Groups	53.8398	1	53.8398	0.28	0.53
Success	2899.6350	15	193.3090		
Total	2953.4749	16			

was first examined. A t-test was used and revealed that the difference between the means of the High and Low Empathy Group was significant at the .05 level but not at the .01 level (t=2.27, df=15, $p < .05 > .01$). See Table 4.

An Examination of the means (High Empathy Groups = 5.14, Low Empathy Group 6.70) reveals that the therapists in the Low Empathy Group tended to evaluate the success of their clients higher than did those in the High Empathy Group.

A t-test was also used to examine the therapists' more detailed evaluations of their clients' success (G-A Score). A significant difference between the groups on this measure was not found (t=.53, df=15, $p > .05$) See Table 5. As for the overall evaluation, there was also a tendency for the Low Empathy Group therapists to evaluate the success of their clients higher than did the High Empathy Group therapists.

A t-test performed in order to analyze the difference between the groups according to the clients' overall evaluation of the success in therapy, reveals no significant difference between the High and Low Empathy Groups (t=.3, df=15, $p > .05$). See Table 6. The means for the two groups were very similar (7.29 and 7.40 respectively), and so were their standard deviations (.76, and .84 respectively).

The more detailed evaluation (G-A Score) performed by the clients also resulted in a nonsignificant difference between

TABLE 6

SUMMARY OF SIMPLE ANALYSIS OF VARIANCE

PERFORMED WITH EMPATHY GROUPS AND

CLIENTS' OVERALL EVALUATION OF SUCCESS

Source	SS	DF	MS	MF	t
Empathy Groups	0.0568	1	.0568	0.09	0.3
Success	9.8279	15	.6552		
Total	9.8816	16			

TABLE 7

SUMMARY OF SIMPLE ANALYSIS OF VARIANCE

PERFORMED WITH EMPATHY GROUPS AND

CLIENTS' DETAILED EVALUATION OF SUCCESS

Source	SS	DF	MS	F	t
Empathy Groups	70.3230	1	70.3230	0.49	0.7
Success	2148.4487	15	143.2299		
Total	2218.7708	16			

the High and Low Empathy Groups (t=.7, df=15, p=> .05). See Table 7. Although the means and variances were very close (High mean = 29.97, S.D.=12.33, Low mean = 25.84, S.D. = 11.72), this was the only time that an evaluation made by the High Empathy Group exceeded the evaluation of the Low Empathy Group, if even only by chance.

It is interesting to note that these four measures - the therapists' and clients' overall and detailed evaluation of success - were not found to be significantly correlated with one another except for the therapists' own overall and detailed evaluation (r=0.61). See Table 8. These measures did not significantly correlate with the level of empathy. See Table 8.

Because of the exploratory nature of the study, a number of ancillary analyses were carried out in an attempt to fully explore the data.

Data was available for 17 of the 22 experimental dyads on the following scales of the Tennessee Self Concept Scale: The Defensive Positive Scale (DP), The Psychosis Scale (Psy), The Personality Integration Scale (PI), and The Number of Deviant Signs Score (NDS).

The analyses of the relationships between therapist accurately empathic behavior and the above-mentioned scales of the Tennessee Self Concept Scale were carried out with the use of a mixed design of analyses of variance. For each analyses, there was one between subjects factor - High Accurate Empathy Group versus

TABLE 8

CORRELATION COEFFICIENTS (r) BETWEEN

FELT SUCCESS VARIABLE AND LEVEL OF EMPATHY

Variable	1 (clients' overall)	2 (therapists' overall)	3 (therapists' detailed)	4 (clients' detailed)	5 (empathy)
1					
2	0.236884				
3	0.0184244	0.6155767			
4	0.3704436	-0.1394757	-0.1683481	0.9999999	
5	-0.2326718	-0.4285187	-0.0681111	0.1979633	

Low Accurate Empathy Group, and one within subjects factor - pre-therapy versus post-therapy score on the particular scale of the Tennessee Self Concept Scale being examined in the particular analysis.

A significant interaction effect was not found (at the .05 level) between high versus low empathy and pre versus post scores on any of the scales. See Table 9. There is a trend toward an interaction effect on the Psychosis Scale (F=4.132, df=1,20, $p > .05$) The High Accurate Empathy Group tended to have a decrease in deviant signs from pre to post-therapy (pre-mean = 48.444, post-mean = 47.889) while the Low Empathy Group tended to have an increase in deviant signs (pre-mean = 44.077, post-mean = 47.615).

The High Accurate Empathy Group differed significantly from the Low Accurate Empathy Group at the .05 level only on the Personality Disorder Scale (F=5.341, df=1,20, $p < .05$). See Table 9. An examination of the means reveals that the Low Empathy Group had a significantly higher mean on the PD Scale than did the High Empathy Group (mean = 69.615, mean = 59.889), respectively); this difference means that the Low Empathy Group had the less deviant scores on the PD Scale than the High Empathy Group both before and after 10 sessions.

The analyses reveal that within the empathy groups, there was little difference between pre and post-therapy on any of the scales examined. See Table 9.

TABLE 9

RESULTING F VALUES FROM A MIXED ANALYSIS OF VARIANCE ON HIGH AND LOW EMPATHY GROUPS SCORES OF TSC SUBSCALES GIVEN PRE AND POST THERAPY

	DP	PSY	PD	N	PI	NDS
Empathy Groups	1.083	0.588	5.341*	0.925	0.977	0.799
Pre-Post	0.996	2.269	1.469	0.332	0.0	0.015
Inter-Action	0.439	4.132	0.174	0.001	0.774	0.037

* $p < .05$

CHAPTER IV

DISCUSSION

Overall, the results were not confirmatory of the hypotheses. Attempting to interpret this fact involves some very complex substantive and methodological issues.

The sample of therapists rated in this study exhibited a relatively low level of empathic behavior. Of the 66 ratings made, only seven of these ratings reached level 4; no sample received a rating as high as level 5. In the High Accurate Empathy Group, only three of the overall ratings (averaged over three interviews) reached a score of 3.333; the remaining six in this group were rated at level 3.000. It also should be pointed out that nine of the 22 dyads were designated as having High Empathy; more than half - 13 dyads were placed in the Low Accurate Empathy Group. It is quite likely that this homogeneity at least partly accounts for the difficulty had in getting high inter-rater reliability. For future studies in this area, a more heterogeneous sample in terms of empathy, is recommended.

The results of the Mullen and Abeles study (1971), indicated that "inexperienced therapists were generally less empathic than experienced therapists (p. 39)." The therapists involved in our study were students and hence relatively inexperienced. This relative lack of therapist experience helps to explain why the ratings on

empathy were so low. It is interesting to note that a significant relationship was not found between accurate empathy and liking (warmth) for experienced therapists; the relationship was found to be significant for experienced therapists. Future researchers might keep in mind that it may be easier to get a more heterogeneous sample of empathy by taking samples from experienced therapists.

Chinsky and Rappaport (1970), suggest that reliability estimates are greatly dependent upon the number of therapists being rated. Reliability coefficients may be spuriously inflated if the raters are responding to an identifiable therapist; for example, the therapist may be identified by his voice or through his language style. Instead of rating the amount of accurate empathy displayed by the therapist, another characteristic of the therapist may in fact be what is being rated. Whenever the rater recognizes a particular therapist he assigns a particular rating so that the ratings of each sample are not independent of each other. Recognition of the therapist would more likely occur when there are only a small number of different therapists being rated.

In the experimental samples used for our analysis, there were 13 different thereapists for 22 clients. Therefore, the rater often heard the same therapist speak more than once in one rating session, and the reliability coefficients may have been spurious. Also, therapist recognition may have occurred from one interview to the other, so that the interviews were not independently rated.

Further, during one training session, one particular therapist was singled out as being noticeably non-empathic, which also likely biased the ratings. This therapist was involved in two of the 22 dyads. Although he was very non-empathic, he rated himself very highly as a therapist. Before this training session, some of the raters took his very confident but highly directive responses to be empathic. His data no doubt contributed strongly to the finding that the therapists belonging to the Low Accurate Empathy Group rated themselves as displaying a high level of empathic behavior.

Rogers (in Rogers, Gendlin, Kiesler, Truax, 1967), has pointed out in his work with patients diagnosed as schizophrenic, that the therapeutic conditions appear to stabilize by the eighth interview and to remain at a fairly consistent level after stabilizing. Prior to the eighth interview, it appears that the conditions of the therapist, as judged by raters, do fluctuate considerably. Therefore, the use of ratings of therapist behavior taken from the early interviews may decrease the probability of correctly predicting successful versus an unsuccessful outcome from therapy.

The methodology use in this study involved an averaging of empathy ratings for the second, fourth, and eighth interviews. This means that two out of the three ratings used to derive the average empathy score were taken from the time when the therapist conditions were likely unstable. This then, may partly account for the lack of confirmation of the hypotheses.

Pare (1970) hypothesized that a number of factors such as personality factor, educational level, socioeconomic status, and age, influence ratings of accurate empathy. The results of his study disconfirmed his hypotheses and indicated that different people can agree closely on their ratings of accurate empathy. In Pare's study inter-rater reliability was found to be between r=.77 and r=.91. His raters were given only a one-half hour to become familiar with the scale. In our study, the raters were not able to reach such a high degree of reliability even after several instructional sessions and practice ratings. It seems to be that certain therapist-patient dyads are easier to rate than others - perhaps due to the small range of the samples being rated, etc. This probably has a greater influence on the reliability of the ratings than do the rater characteristics.

The scale used for rating in this study was a five level scale published by Carkhuff (1969). It is important to note that all of the studies cited in the introductory notes used a previous version of the scale for the rating of accurate empathy - Truax's nine level scale of accurate empathy - The Accurate Empathy Scale (1967). The five level scale was chosen for the rating because it was constructed as an "attempt to reduce the ambiguity and increase the reliability" of the nine level scale. (See Appendix A). The use of the more compact scale has given rise to a very relevant question, which, if answered in the future, would provide worthwhile information. The question which deserves investigation is whether the use of the five level scale in fact makes the rating process

less complex without significantly reducing the accuracy of the ratings and the meaningfulness of them. As pointed out in the section on methodology, the raters found it necessary to decide upon a number of rating conventions, in addition to those provided by Carkhuff's scale itself. From what can be gathered from published literature, this had not been necessary when the nine level scale was used for rating. Was it that the dyads found in our sample were more difficult to rate because they were so homogeneous, or because the five level scale used for rating was too simplified?

Because of the methodological limitations of our study, as discussed above, we can not really be sure if the theory underlying the hypotheses tested is in fact valid or invalid. The question of whether or not the level of accurate empathy by itself can be used to predict success of therapeutic outcome, has not yet been answered.

Rogers (in Rogers et al, 1967), points out the importance of patient qualities in the outcome of the therapeutic relationship. The patients' characteristics will serve to elicit the therapist's attitudes. "High therapeutic conditions seem to be a product of interaction between the person of the therapist and the person of his client (p. 90)." Rogers concludes that "the best therapeutic relationship develops between a therapist who is understanding and real, and a client or patient who is able to be somewhat expressive, who is not too remote from his own experiencing (p. 72)."

Thus, patient qualities may need to be taken into account in order to be able to make a more accurate prediction of the therapeutic outcome. If this is so, our study and its underlying theory may have been incomplete. Further research in this area will be necessary, in order to test the validity of the hypotheses and their underlying theory. It is our hope that the methodological limitations found in this study will be taken into consideration in future research.

REFERENCES

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

Carkhuff, R. R. Helping and human relationships II. New York: Holt, Rinehart and Winston, 1969.

Chinsky & Rappaport Brief critique of the meaning and reliability of "accurate empathy" ratings. Psych. Bull., 1970, 73, 5, 379-382.

Fitts, W. H. Tennessee Self Concept Scale, Counselor Recordings and Tests, Nashville, 1965.

Halkides, G. in Barrett-Lennard, G. T. Dimensions of therapist response as causal factors in therapeutic change. Psychol. Monogr. 76, No. 43 (whole No. 562), 1962.

Hart, J. T. in Truax, C. B. Effective ingredients in psychotherapy: An approach to unravelling the patient-therapist interaction. J. counsel Psychol., 1963, 10, 256-263.

Mullen, J., & Abeles, N. Relationship of liking, empathy, and therapist's experience to outcome of therapy, J. of counsel Psych., 1971, 18, No. 1, 39-43.

Pare, D. D. Accurate empathy: relative or absolute. Doctoral Dissertation, University of Florida, 1970.

Rogers, C. R. The necessary and sufficient conditions of therapeutic personality change. J. Consott. Psych., 1957, 21, 95-103.

Rogers, Gendlin, Kiesler, & Truax the therapeutic relationship and its impact Univ. of Wisconsin Press. Madison, 1967.

Truax, C. B. The process of group psychotherapy: Relationships between hypothesized therapeutic conditions and intra-personal exploration. Psychol. Monogr., 1961, 75, 7 (whole No. 511) (a).

Truax, C. B. A scale for the measurement of accurate ampathy. Psychiatric Institute Bull., Wisconsin Psychiatric Institute, Univer. of Wisconsin, 1961, 1, 12 (b).

Truax, C. B. Effective ingredients in psychotherapy: An approach to unravelling the patient-therapist interaction. J. counsel Psychol. 1963, 10, 256-263.

Truax, C. B. Effects of client centered psychotherapy with schizophrenic patients: nine years pre-therapy and nine years post-therapy hospitalization, J. of consult & clin Psychol., 1970, 35, 417-421.

Truax & Carkhuff, For better or for worse: the process of psychotherapeutic personality change. Chapter 116:Recent advances in the study of behavior change. McGill Univ. Press, Montreal, 1963, 118-163.

Truax, Carkhuff, and Kodman, F., Jr. Relationships between therapist-offered conditions and patient change in group psychotherapy: J. clin. Psychol., 1965, 21, 327-329.

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

Truax, C. B., and Wargo, D. G. Antecedents to outcome in group psychotherapy with juvenile delinquents: effects of therapeutic conditions, alternate sessions, vacarious therapy pretraining, and patient self-exploration. Unpublished manuscript, University of Arkansas, 1966 (b).

Truax, C. B., Wargo, D. G., and Carkhuff, R. R. Antecedents to outcome in group psychotherapy with out-patients: Effects of therapeutic conditions, alternate sessions, vacarious therapy pretraining and patient self-exploration. Unpublished manuscript, University of Arkansas, 1966.

Truax, Wargo, Frank, Imber, Battle, Hoehn-Saric, Nash, & Stone, Therapist empathy, genuineness, and warmth and patient therapeutic outcome. J. Consult. Psychol., 1966, 30, 395-401

Truax, C. B., Wargo, D. G., and Silber, L. D. Effects of high accurate empathy and non-possessive warmth during group psychotherapy upon female institutionalized delinquents. J. abnorm. Psychol., 1966, 71, 267-274.

Van der Veen, F. Basic elements in the process of psychotherapy: A research study. J. of Consult. Psych. 1967, 31, 295-303.

APPENDIX A

Empathic Understanding in Interpersonal Processes. II

A Scale for Measurement[1]

Robert R. Carkhuff

State University of New York at Buffalo

Level 1

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.

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 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 affect and meaning.

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 a 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 wavelength. 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.

[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 counselling 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 counselling and psychotherapy (summarized in Carkhuff and Berenson, 1967). In addition, similar measures of similar constructs have received extensive support in the literature of counselling 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 co-respondent: Level 2 and Stages 2 and 3 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.

AGREED UPON TRAINING CONVENTIONS

The agreed upon definition of accurate empathy was that of a "communicated understanding of the client's intended message." The word "intended" in the definition was further explained as referring to what the client is trying to say right now; if the therapist brings in the past the rating is lowered. The issue is not whether the therapist is getting at what is "really true"; what is "really true" is not to be inferred by the rater and used as a basis for rating.

It was found to be necessary to specify just how one four minute sample was to be rated. It became agreed upon that only a "meaningful unit" was to be rated. The rater must hear what the client says, and then the therapist's following comment. If the sample begins in the middle of the client's statement or therapist's statement, this part of the sample cannot be rated. The statement must be complete before a rating can be made. Also, there may be more than one meaningful unit in a four minute sample. When this occurs, the rater must weigh the significance of the unit, assign a rating to the unit, and average the ratings taking the weighting into account, in order to assign the final rating to the sample

A number of more specific questions arose during the training sessions, and resulted in the following agreed upon conventions:

1. If the therapist adds meaning to what the client has said but detracts feeling, rate as level 4.

2. If the therapist finishes the client's sentence, the therapist is in essence saying "this is what you are trying to say." The rating will be at least as high as Level 2 if the client did mean what the therapist said. A rating of 1 would be given if the therapist was not attending to the client and interpreted him incorrectly.

3. If the therapist says "I don't understand," a rating at Level 2 would be given if the therapist is asking for clarification and he is truly attempting to understand the client, rather than to redirect him.

APPENDIX B

Progress of Counselling Rating Scale

We would like to have you rate the progress, or lack of progress, in your counselling up to the present point. Regardless of whether you are planning to terminate in the near future or are planning to continue for some time in counselling, we would like your opinion of the progress of the first ten sessions. Your ratings will be used only for research and will in no way affect the course of your counselling. First circle the number which best represents your overall evaluation of progress so far:

1	2	3	4	5	6	7	8	9
much negative or undesireable result			no results				much positive or desireable results	

Now, more specifically, below are listed a number of areas which people sometimes mention as goals in counselling. Each goal is followed by numbers 1-9. After each possible goal you are to place two ratings. First, place the letter "G" at the point which indicates how much this particular area has, at any time in counselling, been one of your goals. For example, if the particular goal has nothing to do with you, place a "G" on number 1. If it is a very important goal for you, place the "G" on number 8 or 9. If it is moderately important, place the "G" somewhere in the middle.

Second, place the letter "A" on the number which indicates the degree to which you feel you have achieved this goal in counselling.

Thus, for example, if you have made a great deal of progress toward a very important goal, both the "G" and "A" would be placed on high numbers on the line describing that goal. An "A" on number 1 indicates no progress toward that goal.

1.	Improving my ability to have close relationships with the opposite sex	1	2	3	4	5	6	7	8	9
2.	Dealing with unhappiness and depression	1	2	3	4	5	6	7	8	9
3.	Becoming more aware of the true nature of my feelings	1	2	3	4	5	6	7	8	9
4.	Relieving tension and anxiety	1	2	3	4	5	6	7	8	9
5.	Discovering "who I am" -- my identity	1	2	3	4	5	6	7	8	9
6.	Dealing with panic reactions to such things as tests	1	2	3	4	5	6	7	8	9
7.	Improving my relationships with people in general	1	2	3	4	5	6	7	8	9
8.	Changing specific behavior (what behavior?) _____	1	2	3	4	5	6	7	8	9
9.	Resolving problems with my parents	1	2	3	4	5	6	7	8	9
10.	Dealing with sexual problems	1	2	3	4	5	6	7	8	9

11. Improving my ability to control my emotions 1 2 3 4 5 6 7 8 9

12. Other (specify)

_____

_____ 1 2 3 4 5 6 7 8 9

13. Other (specify)

_____

_____ 1 2 3 4 5 6 7 8 9

14. Other (specify)

_____

_____ 1 2 3 4 5 6 7 8 9