



eCOMMONS

Loyola University Chicago
Loyola eCommons

Dissertations

Theses and Dissertations

1977

Hostility, Gender and the Learning of Interpersonal Communication Skills

John E. Dalton
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

 Part of the [Psychology Commons](#)

Recommended Citation

Dalton, John E., "Hostility, Gender and the Learning of Interpersonal Communication Skills" (1977).
Dissertations. 1696.
https://ecommons.luc.edu/luc_diss/1696

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](#).
Copyright © 1977 John E. Dalton

HOSTILITY, GENDER, AND THE LEARNING OF INTERPERSONAL COMMUNICATION SKILLS

by

John E. Dalton

A Dissertation Submitted to the Faculty of the Graduate School
of Loyola University of Chicago in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

January

1977

ACKNOWLEDGMENTS

The author is grateful to Drs. Frank J. Kobler, Gerard Egan, and James E. Johnson for their encouragement, concern, and valuable assistance as members of the dissertation committee. The author also wishes to thank his wife, Barbara, for her generous and much-needed emotional support.

VITA

The author, John E. Dalton, is the son of Harold Joseph Dalton and Celestine (Adam) Dalton. He was born December 4, 1950, in Oak Park, Illinois.

His elementary education was obtained at St. Peter Canisius School in Chicago, Illinois, and secondary education at Archbishop Weber High School in Chicago, where he was graduated in 1968.

In September, 1968, he entered Loyola University of Chicago, and in June, 1972, received the degree of Bachelor of Science, Honors, Magna Cum Laude, with a major in psychology. While attending Loyola University of Chicago he was elected in 1971 for membership in Alpha Sigma Nu and Phi Sigma Tau.

In September 1972, he entered the graduate program in Clinical Psychology at Loyola University of Chicago. He fulfilled his clerkship requirement in clinical training at West Side Veterans Administration Hospital in Chicago from June to August of 1973. From September, 1973, to June, 1974, he was the teaching assistant for the Director of the Clinical Division of the Psychology Department. In June, 1975, he was awarded the Master of Arts in psychology. From September, 1975, to January, 1976, and from September, 1976, to January, 1977, he was a lecturer for the Loyola University Psychology Department. His internship requirement in clinical training was fulfilled at Hines Veterans Administration Hospital in Hines, Illinois, and at the Loyola Child Guidance Center in Chicago. He is currently a psychologist at the Chicago Alcoholic Treatment Center.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
VITA	iii
LIST OF TABLES	vi
 Chapter	
I. INTRODUCTION	1
II. REVIEW OF RELATED LITERATURE	10
Hostility and Learning in an Academic Setting	10
The Scrambled Sentence Test	10
Additional Research on the Reliability and Validity of the Scrambled Sentence Test	13
The Scrambled Sentence Test and Academic Learning	17
Hostility and Learning in Psychotherapy	20
Methodological Issues in Psychotherapy Research	24
The Skills Approach to Psychotherapy and the Training of Therapists	27
Assessment of Helping and Human Relating Skills	32
Convergent and Discriminant Validity	32
Rating of Tapes	36
Clients' Perceptions of Therapeutic Conditions	40
Other Measures of Therapeutic Conditions	47
A Comparison of Clients' Perceptions and Ratings of Audiotapes	50
Gender Differences in Perceptions of Empathy	56
Overview	58
III. METHOD	59
Subjects	59
Materials	60
Procedure	61
Dependent Measures and Statistical Design	62
IV. RESULTS	67

	Page
Hostility and Gender	67
Agreement Between the Scrambled Sentence Test and Trainer-Perceived Hostility	69
Hostility, Gender, and the Learning of Skills	70
The Scrambled Sentence Test	70
Trainer-Perceived Hostility	72
Gender Differences in Levels of Skills	74
Hostility of Trainers and Perception of Skills by Trainers	76
Supplementary Data on the Relationship Inventory	79
Agreement Between Co-Trainers	79
Pre- and Post-Treatment Variance in RI Scores Given by Trainers	79
Pre- Versus Post-Treatment Overall Skills Level	81
Scale Intercorrelations	81
V. DISCUSSION	83
Gender Differences in the Measurement of Hostility	83
The Scrambled Sentence Test as a Measure of Hostility.	86
Hostility and Interpersonal Communication Skills	89
Gender Differences in Skills Level	91
The Relationship Inventory as a Measure of Skills	92
The Effect of Hostility on Reported Perceptions.	92
Inter-Rater Agreement	93
Depth of the Relationship	94
Scale Intercorrelations	95
Comments by Trainers	96
Efficacy of the Group Training Experience	97
VI. SUMMARY	98
REFERENCES	106
APPENDIX A	117
APPENDIX B	121
APPENDIX C	125

LIST OF TABLES

Table	Page
1. Overall Post-Treatment Levels of Skills According to Gender and Trainer-Perceived Hostility	73
2. Mean Scores on the Relationship Inventory by Gender	75
3. Relationship Inventory Scores Given to Male Non-Leader Group Members by Male Versus Female Co-Trainers	77
4. Relationship Inventory Scores Given to Female Non-Leader Group Members by Female Versus Male Co-Trainers.	78
5. Agreement Between Co-Trainers on the Relationship Inventory	80
6. Pearson Correlations Between Scales on the Relationship Inventory	82
7. Male and Female Scores on the Scrambled Sentence Test in Several Studies	84

CHAPTER I

INTRODUCTION

Despite the considerable attention that psychologists have traditionally given to the study of the learning process, there has been curiously little exploration of possible relationships between learning and various personality traits. A few studies have examined the relationship between learning and the personality trait of hostility, with the results suggesting that there is a negative relation between the two (Goldman, 1955; Latane & Arrowood, 1963; Lieberman, 1966; Phillips, 1960).

Perhaps the most formidable obstacle to pursuing this line of research is the need for a reliable and valid measure of hostility. In an attempt to develop such a measure of hostility, Costin (1969) devised the Scrambled Sentence Test (SST), a semi-disguised measure of hostility, by modifying a previous measure of hostility (Watson, Pritzker, & Madison, 1955). In its final form (Form C), the SST consists of 70 sets of four words each, with the subject's task being to underline any three of the four words so as to make a sentence. Forty of the items are buffer items, and the other 30 are scored as either "hostile" or "neutral," so that the total hostility score is simply the sum of the items scored as hostile. Several studies indicated that the SST has reasonably good test-retest reliability and internal consistency (Costin, 1969, 1970, 1975). Evidence of validity for the SST has been more problemati-

cal, but this is also characteristic of hostility measures in general (Rabinowitz, 1975). The research that has been done on the validity of the SST has been supportive. Scores on the SST have been found to be significantly and positively related to ratings of hostility by psychologists (Costin, 1969). In each study conducted by Costin (1969, 1970, 1975), males have always scored significantly higher on the SST than females, a finding consistent with cultural expectations and with the pattern found on most of the hostility scales examined by Sarason (1961). Further research (Costin, 1975) also supported the validity of the SST, and suggested that it is at least as valid a measure of hostility as the more commonly-used inventory instruments.

Costin (1970, 1971) was satisfied with the reliability and validity of the SST, and therefore used this test to study the relationship between hostility and learning. The results of these studies indicated that, at least for males, hostility was negatively related to learning. The obtained correlations for females were also negative, but were non-significant. The data led Costin to the conclusion that hostility somehow interferes with learning, but he did not speculate as to the source of the interference.

The above conclusion by Costin is certainly tenable, but further examination is clearly needed. First, Costin (1970, 1971) studied learning in classroom settings, with grades used as the measure of learning. It would be informative to test whether the negative relationship found by Costin holds for other forms of learning. Second, if this negative relation does exist, exactly how does hostility interfere with learning? Finally, the whole concept of hostility needs clarification (Tedeschi, Smith, & Brown, 1974).

An initial step toward explaining Costin's data would be an attempt to replicate Costin's findings with a type of learning that is quite different from the academic learning studied by Costin. The learning of "interpersonal communication" or "helping" skills (Carkhuff, 1969a, 1969b; Egan, 1970, 1971, 1973, 1975, 1976) is indeed different from academic learning. A large body of research has emphasized the importance of various communication skills: empathy, positive regard, respect, congruence, genuineness, nonpossessive warmth, etc. These skills have been grouped under different generic labels depending on the circumstances in which they are used. When used by a therapist or helper, they have been termed "helping skills," "therapeutic conditions," and "facilitative conditions." When used in the context of a peer relationship where there is mutual helping and relating, these skills have been called "human relating skills," "interpersonal skills," "communication skills," and "interpersonal communication skills." Despite the variation in generic description, it is important to note that the skills themselves remain the same; the behavioral definitions do not vary depending on the nature of the relationship.

An exploration of the relationship between hostility and the learning of interpersonal communication or helping skills could have noteworthy theoretical value as well as practical ramifications. The primary theoretical value would be an extension of Costin's work, while practical benefits might arise in several areas. Research (Truax & Carkhuff, 1967) has indicated that levels of therapist-offered facilitative conditions are positively related to client outcome in psychotherapy. Accordingly, there are many psychologists, including Truax, Carkhuff, and Egan, who advocate the position that a crucial component of the

training of would-be helpers or therapists is direct training in the helping or interpersonal communication skills. A negative relation between hostility and the learning of such skills would suggest implications for both the selection and method of training of potential helpers. Furthermore, systematic and direct training in interpersonal communication skills represents one school of psychotherapy. Carkhuff (1969a, 1969b, 1976) proposed that skills training is the treatment of choice for patients in psychotherapy, and Egan (1975) also takes this position. Skills training has been used with psychiatric patients, with the available research indicating that this approach is at least as effective as more traditional approaches (Cohen, Johnson, & Hanson, 1971; Johnson, Hanson, Rothaus, Morton, Lyle, & Moyer, 1965; Morton, 1965; Pierce & Drasgow, 1969; Rothaus, Morton, Johnson, Cleveland, & Lyle, 1963). The proposed study would then have implications for the treatment of psychiatric patients. For example, it might be advisable to focus on the reduction of excessive hostility at the beginning of therapy, so that the potential for subsequent learning might be enhanced. Positive findings would also lead to hypotheses regarding the selection of patients most likely to benefit from psychotherapy, or least likely to be harmed by the experience. While the proposed study would not provide unambiguous answers to all these questions, it would at least point out worthwhile directions for future research.

To conduct the proposed study, accurate measures of both hostility and interpersonal skills are needed. Costin's research supports the use of the SST to assess hostility, but the measurement of human relating skills is more problematical. The source of the difficulty is the complexity and subtlety of the behaviors defining the various skills. Of

the various approaches that have been used to assess levels of helping or human relating skills, the two that have been most extensively used are (a) ratings of audio tapes by objective judges, and (b) inventories on which individuals report their perceptions regarding the skills displayed by another. The first approach, ratings by objective judges, is the one favored by Truax and Carkhuff (1967). The typical procedure is to provide judges with excerpts from counseling sessions, and the judges then rate levels of therapist-offered facilitative conditions according to a specific rating scale. Barrett-Lennard's (1962) Relationship Inventory follows the second approach to the measurement of helping or human relating skills--perceptions of skills displayed by another, with the perceptions provided by the person who directly related to the other rather than by an outside observer. The Relationship Inventory can be used to assess perceptions of skills in any relationship, including but not limited to the therapist-client relationship. This inventory consists of a series of statements (such as, "He respects me."), and for each the subjective judge records his level of agreement or disagreement. The Inventory yields a total score and scores on four scales: empathy, congruence, regard, and unconditionality of regard. Both of these approaches to measurement of interpersonal communication skills, ratings by objective judges and perceptions from subjective judges, have been validated in studies showing a significant, positive relationship between the measure of skills and outcome in psychotherapy (Barrett-Lennard, 1962; Hansen, Moore, & Carkhuff, 1968; Mullen & Abeles, 1971; Rogers, 1967; Truax & Carkhuff, 1967). However, the two approaches have not been found to correlate highly with each other (Bozarth & Grace, 1970; Caracena & Vicory, 1969; Carkhuff & Berenson, 1967; Fish, 1970;

Hansen, Moore, & Carkhuff, 1968; Hill & King, 1976; Kiesler, 1966; Kurtz & Grummon, 1972; McWhirter, 1973; Welkowitz & Kuc, 1973). In light of the lack of strong agreement between the two approaches, four studies have attempted to directly compare the two approaches to see which is the more valid. Using outcome in therapy as the criterion for validity, two studies (Hansen, Moore, & Carkhuff, 1968; Truax, 1966a) found that ratings by objective judges have greater validity than client perceptions on the Relationship Inventory. Two other studies (Caracena & Vicory, 1969; Kurtz & Grummon, 1972) indicated that client perceptions provide the more valid measure. There is at present no strong evidence to conclude that either approach is more valid.

One feasible solution to the problem of choosing a specific approach to the measurement of skills is to combine elements of the two approaches. Fortunately, a course directed by Gerard Egan at Loyola University of Chicago offers the opportunity for such a combination. The major component of this course is experiential groups which focus on training in the various skills of helping and human relating. The trainers (also known as leaders or facilitators) for these groups are individuals who have demonstrated behavioral proficiency in the skills in one or more previous group experiences, and the trainers are similar to objective judges in terms of both level of psychological functioning and ability to make accurate discriminations regarding levels of skills displayed by others. Yet the group facilitators function as both leaders and members, pursuing the same contractual goals (Egan, 1970, 1971) of interpersonal growth as the non-trainer group members. Since the groups involve mutual helping relative to the written, contractual goals toward which all group members agree to strive,

the trainers are often in the role of helpees being helped by the non-trainer members. Thus the trainers, like any helpees, can report their perceptions of the levels of helping skills displayed by the non-trainer members. The Relationship Inventory is a good instrument for this purpose, and amount of learning of skills can be evaluated by having the trainers-helpees report their perceptions early in the group experience and again at the end. Moreover, validation of the SST as a measure of hostility can be given further scrutiny by having the trainers provide ratings of hostility on the group members.

Some of the research on helping and human relating skills has indicated that gender may be an important variable relative to the levels of displayed skills. Abramowitz, Abramowitz, and Weitz (1976) found that female graduate students were rated as more empathic therapists than were their male counterparts. The authors interpreted this finding as reflecting a cultural difference in sex roles, with females being raised to be more attuned to emotional experiencing. However, other studies (Breisinger, 1976; Olesker & Balter, 1972) suggested the possibility of an interaction effect, with people being more empathic when relating to individuals of the same rather than opposite sex. The literature on gender differences in levels of skills is scant, and the present study attempted to determine whether females display higher levels of interpersonal communication skills and whether there is an interaction effect for same-gender versus opposite-gender dyads. Thus there was an examination of gender differences regarding not only empathy, but also the other three skills (congruence, regard, and unconditionality of regard) for which the Relationship Inventory has scales.

The following hypotheses and sets of hypotheses were therefore

tested:

- 1) Males have higher hostility scores on the Scrambled Sentence Test than do females.
- 2) Males are rated as being more hostile than are females.
- 3) Hostility scores on the Scrambled Sentence Test are positively related to ratings of hostility by group trainers.
- 4) There is a negative relation between hostility, as measured by the Scrambled Sentence Test, and the learning of helping or interpersonal communication skills, with the relation being stronger for males than for females.
- 5) There is a negative relation between hostility, as measured by ratings by group trainers, and the learning of helping or interpersonal communication skills, with the relation being stronger for males than for females.
- 6) Females are perceived as displaying higher levels of helping or interpersonal communication skills than are males.
- 7) Both males and females are perceived as displaying higher levels of helping or interpersonal communication skills when the perceptions are provided by a person of the same rather than opposite sex.

In addition to testing of formal hypotheses, other statistical analyses were done to provide supportive information. Mean changes in levels of skills were calculated both within and across gender, in order to obtain an indication of the efficacy of the experiential training groups. Trainer hostility, as measured by the SST, was correlated with perceptions by trainers on the Relationship Inventory, to explore the possibility that more hostile trainers perceive lower levels of skills in others. The hostility ratings by trainers were correlated

with the trainers' perceptions of skills as measured by the Relationship Inventory; this was done to determine whether or not there is a relationship between perceived hostility and perceived skills. Agreement between co-trainers on perceptions of skills was analyzed. Data were also analyzed to provide the following information regarding the Relationship Inventory: internal consistency of the scales and of the instrument as a whole, scale intercorrelations, and correlations between each scale and the sum of the other three scales. The possibility of a relationship between age and hostility as measured by the SST was explored. Since perceptions of skills have been found to be related to the variable of same-sex versus opposite-sex dyad (Olesker & Balter, 1972), there was an analysis of the possible effects of this variable on the trainers' ratings of the hostility of group members. Finally, variance in the trainers' perceptions of skills was analyzed according to gender of trainer.

CHAPTER II

REVIEW OF RELATED LITERATURE

Hostility and Learning in an Academic Setting

The Scrambled Sentence Test. Costin (1969) described his development of the Scrambled Sentence Test (SST), a semi-disguised measure of hostility, stating that the SST "was derived from Watson, Pritzker, and Madison's (1955) individual test, which was administered by projecting on a screen a set of four words arranged in a scrambled order" (p. 461). The subject's task was to select three of the four words and order them so as to make a sentence. The sixty sets of four words were designed to elicit the formation of sentences that could be scored as either "hostile" or "neutral." A subject's responses were audio recorded, and his total score was obtained by simply summing the number of hostile sentences he had constructed. Watson, Pritzker, and Madison assumed that their test was a measure of repressed hostile impulses. They therefore hypothesized that neurotics would score higher than normals, and their research confirmed this.

Costin was impressed by the potential of this assessment technique, but realized that the form of administration was rather cumbersome. He therefore set out to develop a paper-and-pencil version that would allow for quick administration to groups. However, Costin (1975) did not assume that the scrambled sentences were necessarily measuring repressed hostility:

As a beginning point it seemed sufficient to assume that because of its semi-disguised format (or ambiguity, if you will), the SST would tap a sensitivity to 'hostile' (sic) stimuli in the environment; and that this sensitivity, or readiness to respond to hostile cues, could be considered as a basic personality trait. This assumption was not unlike the major one made by Watson and his colleagues, but avoided committing the SST to their particular psychoanalytic concept of repression. Thus, hostility, as measured by the SST, was conceived to be a general predisposition--an habitual propensity for disliking others, for wishing them harm, or behaving aggressively towards them (Kaufmann, 1970). (p. 101)

Costin assumed that the scrambled sentences would be a measure of hostility as a personality trait, but he did not expect that the test would necessarily have a direct relationship to overt aggression, as the latter reflects both personality characteristics and environmental factors.

From the original (Watson, Pritzker, & Madison, 1955) individual test consisting of sixty sets of four words, Costin constructed two parallel forms, A and B, of thirty sets each. The subject's task was to "underline any three words which made a complete sentence; they were requested to do this according to their first impression, and to work rapidly" (Costin, 1969, p. 462). Following the completion of this task, the subjects were asked to briefly describe what they thought the test was measuring, and approximately 70 percent of these undergraduate students correctly discerned the "disguised" purpose of the SST. On the basis of this finding, several changes were made in the two forms, including the extension of both forms to 50 items, 20 of which were non-scorable buffer items.

The revised forms were then given to undergraduate students taking a variety of courses. Each student completed both forms, either in immediate succession or six weeks apart, and correlations between the two forms were calculated. For immediate succession, the correlation

was .82 for males ($N = 103$) and .79 for females ($N = 118$). For a six-week interval, the correlation was .65 for males ($N = 35$) and .73 for females ($N = 58$). Costin (1969) concluded:

The correlation coefficients seem to be reasonably good evidence of equivalence reliability and stability; one may also infer construct validity from the fact that on each form the mean hostility score of men was significantly higher than that of women--a finding consistent with what one would expect in the expression of hostility in our culture. (p. 464)

To assess concurrent validity, students being seen at the Student Counseling Service at the University of Illinois took both forms of the SST. Each of these students was also given a rating on hostility by the counseling psychologist who interviewed the student. The results showed that students rated as hostile had significantly higher SST scores than students rated as non-hostile. Construct validity was also indicated in that the mean SST scores were again higher for males than for females; this pattern was consistent with that found on most of the hostility scales examined by Sarason (1961).

In order to obtain a single instrument with maximum efficiency, Form C was developed by selecting those items from Forms A and B which were more highly related to the ratings by the psychologists and which discriminated equally well for males and females. Thirty items were selected, and the 40 buffer items from Forms A and B were added to make a total of 70 items. Form C was then administered to students enrolled in various undergraduate courses. Coefficients of internal consistency (KR 21) were .75 for males ($N = 140$) and .76 for females ($N = 177$); test-retest reliability coefficients over a six-week interval were .67 for males ($N = 52$) and .64 for females ($N = 75$). Consistent with previous data, it was again found that the mean SST scores for males was

higher than that for females, and that students rated as hostile by counseling psychologists had significantly higher SST scores than those rated as non-hostile (biserial $r = .65$ for males and $.66$ for females).

Additional research on the reliability and validity of the Scrambled Sentence Test. Costin (1975) described two subsequent investigations of the reliability of the SST. In the first, the SST was given twice over a six-week interim to 77 undergraduates at the University of Edinburgh, and, in the second, to 1,201 British students of ages 14 to 16 (using a six-month interval). Coefficients of internal consistency (KR 20) ranged from $.75$ to $.86$ for males and from $.69$ to $.80$ for females. Coefficients of test-retest reliability over a six-week period were $.79$ for males ($N = 33$) and $.69$ for females ($N = 44$). For reliability over a six-month interval, the coefficients ranged from $.69$ to $.77$ for males and from $.72$ to $.78$ for females. On the basis of these and previous studies, it can be said that the SST has at least satisfactory reliability, especially when considering the fact that samples were drawn from two different age groups and from two different countries.

The above studies also supported construct validity in that males again scored significantly higher than females in both studies. Further evaluation of construct validity came from correlations of SST scores with "scores on the dominance and conflict avoidance scales of the Kuder Preference Record--Personal, Form A, and with scores on the verbal parts of the School and College Ability Test, Form U" (Costin, 1969, p. 467). SST scores were not significantly related to verbal ability or to dominance, but were significantly and negatively related to conflict avoidance. These findings were judged to be consistent with the stated purpose

of the test--the measurement of hostility. The greater preponderance of buffer items was more effective in disguising the purpose of the test, in that only about one-third of the students who took Form C only once discerned the purpose of the test. Moreover, the correlations between SST scores and correct/incorrect evaluations of test purpose were low, ranging from $-.09$ to $.02$. Nevertheless, Costin (1975) felt it was important to give further study to the question of whether or not understanding the purpose of the test affects test scores. Costin selected 30 items from the Green-Stacey Questionnaire (Green & Stacey, 1967), which measures hostility and aggression, and administered this test with the SST in counterbalanced order to 46 undergraduate males and 48 undergraduate females. Upon completion of each instrument, the students were asked what they thought the test was measuring. For the Green-Stacey, 68% of the males and 62% of the females correctly guessed the purpose of the test; the corresponding data for the SST were 36% for males and 34% for females. Moreover, the correlations between correct estimates and scores were much lower for the SST than for the Green-Stacey. For the SST, the correlations were $-.08$ for males and $-.06$ for females, both nonsignificant. The correlations on the Green-Stacey were $-.33$ for males and $-.30$ for females, both significant at the $.05$ level. The data also indicated a moderately high positive relationship ($.65$ for males and $.57$ for females) between the Green-Stacey and the SST. The results of this study appear quite favorable for the SST. The correlation with the Green-Stacey demonstrated concurrent validity, and yet the SST was a less obvious measure of hostility with lower correlations between test scores and correct estimates of test purpose.

In order to further evaluate discriminant validity, Costin (1975) sought to determine whether the SST was less subject to a social desirability effect than a more obvious test of hostility. He interspersed 27 items from the Manifest Hostility Content Scale (Wiggins, 1966; Wiggins, Goldberg, & Applebaum, 1971) with the 33 items of the social desirability scale by Crowne and Marlowe (1960) to make a single measure of 60 items. This measure and the SST were administered in counter-balanced order to 56 male and 96 female undergraduate students. Although the correlations between the SST and the Manifest Hostility Content Scale were positive and significant (.31 for men and .32 for women), the correlations with the social desirability scale were lower for the SST (-.29 for men and -.06 for women) than for the Manifest Hostility Content Scale (-.54 for men and -.19 for women). Thus, the SST appears to be less prone to social desirability effects than the more obvious inventories of hostility.

In reviewing the available research on the SST, the instrument appears to be a promising measure of hostility. The test can be quickly and easily administered to either groups or individuals, and scoring is simple. Research samples have included American and British students of various ages, but there is a need to sample from other segments of the population, particularly clinical groups. Test-retest reliability and internal consistency have been shown to be at least satisfactory. Concurrent validity has been supported by correlations between the SST and three other measures of hostility: the Green-Stacey Questionnaire, the Manifest Hostility Content Scale, and ratings by counseling psychologists. Moreover, the semi-disguished format appears to be superior to the more traditional and obvious inventory approach; research suggests

that the SST is indeed more disguised, is less subject to a social desirability effect, and has lower correlations between correct estimate of test purpose and test score. Self-report measures of hostility have been found to have generally little relationship to aggressive or hostile behavior (Wolff & Merrens, 1974), whereas the relationship between the semi-disguished SST and hostile behavior remains largely untested. There has not been extensive research on the construct validity of the SST, but that which has been done supports the convergent and discriminant validity of the instrument. Scores on the SST have been shown to be negatively related to conflict avoidance, and essentially unrelated to dominance, verbal ability, intelligence, and reading comprehension (Costin, 1969, 1975). In all studies, males have shown a higher mean score than females, a finding consistent with research on other measures of hostility (Sarason, 1961) and with cultural norms regarding the expression of hostility. Still, it is not clear whether the observed sex differences reflect differences in innate hostility or in willingness to express or feel hostility.

Considering the available research on the SST and other measures, the SST seems to be at least as good a measure of hostility as the others. Nevertheless, the SST is similar to other measures of hostility in that validity remains the most significant problem area (Rabinowitz, 1975). There is considerable ambiguity about exactly what is meant by "hostility." Chaplin's (1968) Dictionary of Psychology defines hostility as "the tendency to inflict harm on others; the tendency to feel anger toward others" (p. 222). This definition, as well as the research literature, lacks specificity and shows confusion and circularity between such concepts as hostility, aggression, and anger. Accordingly, the

lack of construct and criterion-related validity is not surprising. Not only is there often a lack of clear differentiation between hostility and aggression, but it may be that both of these concepts are in themselves too heterogenous (Tedeschi, Smith, & Brown, 1974). What is sorely needed is concept clarification and definitive criteria against which to evaluate instruments such as the SST. Unfortunately, hostility can be behaviorally expressed in extremely subtle ways, and thus there are no simple solutions to the problem of validation.

The Scrambled Sentence Test and academic learning. Costin (1970) was impressed by previous research (Goldman, 1955; Latane & Arrowood, 1963; Lieberman, 1966; Phillips, 1960) suggesting that hostility is negatively related to learning, but observed that more stringent methodology was needed, particularly regarding the possibility of sex differences. He decided to explore the relationship between hostility, as measured by the SST, and learning in an introductory psychology course. Costin's (1970) hypothesis was that "in general, student hostility would be negatively correlated with acquisition of knowledge but that this relationship would be more characteristic of men than of women" (p. 370).

The subjects for Costin's experiment were 50 male and 51 female undergraduate students, all of whom were given a 60-item pretest at the start of the course. Thirty of these items were later included in the midterm examination, while the other 30 were included in the final examination. The SST was administered to the students during the first and last week of the course. The test-retest reliability of the SST over a 7-week interval was satisfactory: .70 for males, .71 for females, and .71 overall. The internal consistency coefficients (KR 21) for the

SST were as follows: .63 and .76 for males, .71 and .80 for females, and .72 and .78 overall. Mean scores for males were again significantly higher than those for females: 11.7 to 9.1, and 11.8 to 9.4. A series of zero-order, first-order, and second-order correlations generally supported the hypothesis of a negative relationship between hostility and academic performance when both "college ability" and pretest knowledge were held constant. The partial correlations between hostility and scores on the two posttests, with ability and pretest knowledge held constant, were: $-.29$ and $-.32$ overall, $-.40$ and $-.44$ for males, and $-.18$ and $-.21$ for females, with the overall correlations and those for males being significant at the .05 level.

Costin concluded that, while no definitive conclusions regarding a causal relationship can be made on the basis of correlations, "it does seem reasonable to interpret the role of hostility, in the present context, as an 'interference' with learning the subject matter of the course" (p. 373). The data also showed the obtained negative relationship to be consistently stronger for men than for women, although the differences were not significant.

Costin (1971) thought that further research was needed to support his conclusion that hostility somehow interferes with learning. Again using students at the University of Illinois as his subjects, he found that "end-of-semester grade point averages of male students enrolled in the Special Educational Opportunity Program at the University ($N = 129$) were found to be negatively correlated with presemester scores on the Scrambled Sentence Test" (p. 1015). This data, however, did not control for either ability or pre-course knowledge.

Costin carried out another study to see if this relationship be-

tween hostility and learning holds with a different subject population and a more technical course. The subjects were 60 enlisted men at an Air Force Technical Training Center, all of whom were enrolled in a meteorology course. Costin (1971) was not able to obtain a pretest measure of pre-course knowledge of meteorology, but he did have scores from the Air Force Qualification Test, "a group measure of general mental ability" (p. 1016). The author found a zero-order correlation between hostility and achievement of $-.41$, significant at the $.01$ level; the partial correlation, with general mental ability held constant, was $-.39$, again significant at the $.01$ level. Not only were these results consistent with those of the previous study (Costin, 1970), but the correlations themselves were quite similar. Costin saw this study as being further support for his position that, at least for males, hostility interferes with learning.

The results of Costin's (1969, 1970, 1971, 1975) studies are both interesting and consistent, but, as usual, more questions are raised than are answered. Most importantly, learning is a complex concept, and it would be worthwhile to determine whether the observed negative relationship between hostility and learning holds for other types of learning. Costin concluded that hostility "interfered" with learning, but what was the source of this interference? Did the more hostile students study less? Does a student's hostility reduce positive identification with the instructor and thereby reduce imitative learning? Since achievement seems to please concerned authority figures, can lack of achievement sometimes represent an acting-out of hostility? Does hostility require a focus of energy that inhibits attention to other pursuits? Does hostility reduce the possibility of obtaining the poten-

tial benefits of active interaction with instructors and successful students? Why do females have generally lower hostility scores on the SST, and why is the negative relationship between hostility and learning apparently stronger for males than for females? There is a clear need for further research to analyze these and other possibilities.

Hostility and Learning in Psychotherapy

There are many in the field of psychology who conceptualize psychotherapy as a learning experience. Behavioral therapists commonly speak of learning, while followers of Harry Stack Sullivan's (1953) theory consider psychotherapy to be interpersonal learning. Truax and Carkhuff (1967) stated that "the typical patient's inability to relate well to other human beings can be thought of as deriving at least in part from a deficit in learning or experience" (p. 152). Most schools of psychotherapy emphasize some form of "insight," which can also be described as learning. Accordingly, the question arises as to whether Costin's (1970, 1971) observed negative relationship between hostility and academic learning also holds for learning in psychotherapy. If so, there might be important practical ramifications. It might be possible to improve the efficacy of therapy by focusing on the reduction of patient hostility at the onset of therapy. Given the limited availability of psychotherapists and the fact that therapy can be for better or worse, a good measure of hostility might serve as a device to screen patients regarding the statistical prognosis for beneficial change. Three studies relevant to the relationship between hostility and change in psychotherapy will be discussed.

Schoenberg and Carr (1963) examined the efficacy of treating neuro-

dermatitis patients with a brief psychotherapy that focuses on fostering the expression of hostility regarding current life conflicts. This particular form of psychotherapy for neurodermatitis had been found to be effective, except that there had been a high rate of drop-outs. The authors wanted to determine the reason for the many drop-outs and search for variables which would predict successful outcome. One of the experimental hypotheses was that patients who drop out of therapy or who fail to improve would show a greater degree of both overt and covert hostility at the start of therapy. The 26 experimental subjects were either clinic patients or private patients referred by the dermatology department of an urban hospital. The subjects, 10 males and 16 females, ranged in age from 15 to 62.

Before therapy began, each patient was interviewed by a psychiatrist, who made 3-point ratings (slight, moderate, and marked) on both overt (directly expressed) and covert (unexpressed) anger. Each patient was also given a test battery consisting of WAIS, MMPI, TAT, Buss-Durkee Hostility-Guilt Inventory, and Rorschach. Following completion of the therapeutic program, ratings of change in neurodermatitis (worse, no change, slight improvement, moderate improvement, or marked improvement) were made by the therapists, patients, and members of the department of dermatology. All patients were then classified as either improved (moderate or marked improvement) or unimproved (worse, no change, or slight improvement). It is not clear why statistical analysis involved only these dichotomous ratings when the original ratings were made on a 5-point scale. Furthermore, the authors do not describe the process by which a single rating was obtained for each patient from the three sources of ratings. As such, no judgments regarding the reliability

or validity of the ratings can be made.

Of the 26 patients, 11 were rated as markedly improved, 5 as moderately improved, and 10 as slightly improved, not improved, or worse. Of the 16 patients classified as improved, 4 had unusual circumstances. One dropped out after 3 of the 12 weekly sessions, and another after 8 sessions. Two others inadvertently received steroids during the treatment program, which was supposed to consist of psychotherapy in the absence of chemotherapy for neurodermatitis. Of the 10 patients classified as unimproved, 3 discontinued treatment--2 after the third session and one after the eighth. Although the five drop-outs constituted over 19% of the subjects, the authors included them in the statistical analysis. The same is true of those patients who received steroids.

The results revealed a significant positive relationship between the psychiatric ratings of overt hostility and successful outcome. A non-directional test of significance was used, as the hypothesized relationship was directional and in the opposite direction of the obtained results. Ratings of covert hostility were not significantly related to outcome. Neither hostility scores from the TAT nor Buss-Durkee scores were significantly related to outcome, but the improved group had significantly more hostile content responses on the Rorschach. The authors concluded that there appears to be a positive relationship between initial hostility and successful outcome in this brief psychotherapy for neurodermatitis, but noted that some inconsistencies existed. For example, they failed to find the expected positive relationships between ratings of overt hostility and Buss-Durkee scores, and between ratings of covert hostility and hostility on the Rorschach.

The results of this experiment provide some indication that hostil-

ity is not negatively related to learning in psychotherapy. However, it must be emphasized that the nature of the patients' symptomatology and the form of therapy should be carefully considered. Psychotherapy was based on the assumption that the cause of neurodermatitis lie in the lack of overt expression of hostility. Overt hostility could not be a barrier to learning in therapy, in that the goal of therapy was to increase the expression of hostility. That is, it is not surprising that patient hostility was not negatively related to the learning of behavioral expressions of hostility. Moreover, the experiment was so methodologically weak that little confidence can be placed in the results. The outcome criteria, ratings of change in neurodermatitis, were particularly questionable. There was no evaluation of the validity and reliability of the ratings, and no explanation of how a single rating per patient was obtained from the three sources of ratings. The change ratings, like all change scores, may have obscured a relationship between outcome and initial status.

The second study relevant to the relationship between hostility and learning in psychotherapy is that by Leary and Harvey (1956). The authors used Leary's Interpersonal System to measure personality changes in psychotherapy. This system has sets of variables to classify behavior at each of the five defined levels of personality: public communications, conscious descriptions, private preconscious descriptions, the unexpressed unconscious, and values. One of the authors' conclusions was that "men who are hostile or weak (at the symptomatic level) are more likely to change than women with the same pre-therapy diagnosis" (p. 131). Leary and Harvey did not state whether the observed changes were for better or for worse, and thus the results have no clear indications

for the hostility-learning relationship.

Cohen, Johnson, and Hanson (1971) cited previous research (Rothaus, Morton, Johnson, Cleveland, & Lyle, 1963) which found that patients increased in feelings of hostility at the fantasy level following human relations training. Cohen and his associates used Leary's Interpersonal Check List and the MMPI to measure behavior, and found that "this change in the nature of fantasy material also occurs at the behavioral level where patients are seen as becoming more assertive while remaining oppositional" (p. 477). On the basis of an initial impression it might seem inconsistent that human relations training seems to result in an increase of hostile feelings and behaviors. However, such a conclusion may not be warranted. The measurement of hostility in any research becomes increasingly more difficult as the hostility, whether at the fantasy or behavioral level, becomes more subtle. When an individual brutally beats another person, it is relatively simple for an observer or the actor to label this behavior as hostile. However, hostility can also be expressed by silence and withdrawal, condescending help, humor, forgetting, tardiness, illness, etc. The hostile intent of these more subtle behaviors is far more likely to go unnoticed in any research, whether the criteria is observed hostility or self-reported hostility. It may be that the patients receiving human relations training did not become more hostile at either the fantasy or behavioral level, but rather became more accepting of their hostility and more willing to express it in a more direct and readily identifiable manner.

Methodological Issues in Psychotherapy Research

In the previous section on the relationship between hostility and

learning in psychotherapy, the three studies reviewed were characterized by an obvious lack of definitive findings. The confusion regarding the conclusions to be drawn from these studies can be at least partially attributed to the methodological problems involved in psychotherapy research. It is far easier to develop reliable and valid outcome criteria for learning in a college course than it is for learning in psychotherapy. Meltzoff and Kornreich (1970) addressed this difficulty:

It is difficult to attempt to make a value judgment about the success of psychotherapy. Success is rarely an absolute even when it can be measured in concrete quantities. It is necessary to talk of perceived success, for one man's success is another man's failure. (p. 172)

Psychotherapy is a complex interpersonal experience, with myriad factors and interactions involved. Meltzoff and Kornreich (1970) and Cartwright (1966) mention the following as variables that have been considered relevant in psychotherapy research: (a) patient variables --age, IQ, sex, education, marital status, problem-solving style, ego strength, expectations for therapy, biosocial characteristics, motivation, type and degree of disturbance, etc., (b) therapist variables --professional discipline, the therapist's personal therapy, experience, amount and type of training, sex, personality type (A-B, etc.), similarity to the patient, liking of the client, values, empathy, etc., (c) time variables--frequency, duration, etc., (d) technique variables, and (e) environmental and situational variables. The number of variables to be controlled, manipulated, or randomized presents a formidable problem. The wide array of possible outcome measures exacerbates the complexity of the research. Outcome criteria that have been used include behavioral observations, status data, test data, therapist judgments, peer judgments, and patient judgments. Referring to the research

on outcome predictors, Meltzoff and Kornreich (1970) said:

The lack of consistency in this conglomeration of studies is probably due to the fragmentation that comes from examining one personality variable at a time without controlling for a host of others. The matter of who will profit from psychotherapy is undoubtedly complexly determined, and examination of any single patient variable in its relationship to outcome is apt to account for only a small portion of the total variance. (p. 229)

It is therefore understandable that so little is known about the relationship between hostility and learning in psychotherapy.

Considering the developmental state of the research on the relationship between hostility and learning, it is probably premature to examine learning in psychotherapy. While promising results have been obtained in studies relating hostility to the learning of concrete tasks and academic material, even here many questions remain unanswered. Exploratory research is needed, and the many variables involved in psychotherapy research would tend to obfuscate any results regarding the relationship between hostility and learning. It might be feasible, however, to measure some form of specific learning in a psychotherapeutic experience, and to relate this learning to initial hostility. A basic problem in psychotherapy research is the elusiveness of concepts such as change and improvement. It is impossible to accurately measure psychotherapeutic learning if one is not clear on what the clients are supposed to learn. One form of psychotherapeutic experience in which learning is relatively well-defined is the interpersonal communication skills approach. The skills approach generally assumes that deficits in psychological functioning can be attributed to lack of fulfillment in interpersonal relationships, and that the key to more adaptive functioning lies in the learning of skills which improve interpersonal re-

lationships.

An examination of the relationship between hostility and the learning of interpersonal skills would serve three purposes. First, it would help to clarify Costin's (1971, 1971) findings by determining whether hostility is negatively related to a type of learning that is quite different from the academic learning studied by Costin. Second, significant findings might have practical implications for the selection and training of potential helpers or therapists in the learning of helping skills. Finally, given that the skills training method represents one approach to psychotherapy, there might be ramifications for the treatment and selection of clients in other schools of therapy.

The Skills Approach to Psychotherapy and the Training of Therapists

In reviewing the research on psychotherapy, Truax and Carkhuff (1967) concluded that psychotherapy is, on the average, ineffective-- that is, no better than mere passage of time in the absence of therapy. Noting that treatment groups showed more variability than control groups, they found that the reason for the average ineffectiveness was that some patients improved in therapy while others got worse. Thus, psychotherapy changed people more than the mere passage of time, but these changes were both positive and negative. In numerous studies conducted over several years, the authors and various colleagues identified three therapist qualities that were crucial to the direction of change: accurate empathy, nonpossessive warmth, and genuineness. This research indicated that clients improve if the therapist is high on these qualities and get worse if the therapist is low on these qualities. The authors provided the following definitions of accurate empathy, non-

possessive warmth, and genuineness:

Accurate empathy involves more than just the ability of the therapist to sense the client or patient's "private world" as if it were his own. It also involves more than just his ability to know what the patient means. Accurate empathy involves both the therapist's sensitivity to current feelings and his verbal facility to communicate this understanding in a language attuned to the client's current feelings. (p. 46)

The dimension of nonpossessive warmth or unconditional positive regard, ranges from a high level where the therapist warmly accepts the patient's experience as part of that person, without imposing conditions; to a low level where the therapist evaluates a patient or his feelings, expresses dislike or disapproval, or expresses warmth in a selective and evaluative way. (p. 58)

This scale is an attempt to define five degrees of therapist genuineness, beginning at a very low level where the therapist presents a facade or defends and denies feelings; and continuing to a high level of self-congruence where the therapist is freely and deeply himself. . . . He is being himself in the moment rather than presenting a professional facade. Thus the therapist's response must be sincere rather than phony; it must express his real feelings or being rather than defensiveness. (pp. 68-69)

In this book, Truax and Carkhuff suggest that the aforementioned qualities are not only important for being an effective therapist or helper, but also for interpersonal functioning of all types. They state that "the growing body of converging evidence has important implications for our own personal conduct in human encounters whether we are functioning as a therapist, an educator, a parent, or more generally, as a person" (p. 142). Although the bulk of the research by these authors focuses on the training and characteristics of effective helpers, a skills training orientation to psychotherapy is implied by the authors' belief that "fear or avoidance of interpersonal relationships is a potent symptom in almost all patients" (p. 151). Accordingly, later works by Carkhuff (1969a, 1969b) specifically propose that the most effective form of psychotherapy is the training of clients in these same

skills ("training as treatment"). Skill training is thus conceptualized as an optimal approach to both training of helpers and treatment of patients with the exact same skills involved in both processes.

Egan's (1975) theoretical position is akin to that of Carkhuff in that Egan also emphasizes the importance of the "helping skills" in a broad context that explicitly advocates a communication skills approach to psychotherapy. Egan stated that the helping skills are "primarily the skills of effective interpersonal relating"; the skills "belong first in everyday life and are not merely the inventions or tools of something apart from real life which is termed 'helping'" (p. 17). The so-called helping skills are actually skills needed by everyone for effective interpersonal functioning of all types. Naturally, this includes therapists and clients.

Although clients seek therapy because of problems or dissatisfactions in life, it is not necessary to assume that maladaptive behaviors occur because the client is doing something wrong. Instead, Carkhuff and Egan work on the assumption that the client's psychological distress is more a function of what he is not doing. Symptomatic behavior occurs when an individual faces stresses without the skills needed to handle them. Since it is impossible to attain effective interpersonal functioning in the absence of basic skills in relating to others, psychotherapy should focus on training patients in these skills. Rather than dealing with specific problems presented by the client, this approach attempts to provide the client with the skills needed to cope with whatever life problems may arise. That is, the client is given tools rather than a "cure." Although there has not been extensive research on the use of human relations training as an alternate to more

traditional forms of psychiatric treatment, that which has been done generally supports the skills training approach as being at least as effective (Cohen, Johnson, & Hanson, 1971; Johnson, Hanson, Rothaus, Morton, Lyle, & Moyer, 1965; Morton, 1965; Pierce & Drasgow, 1969; Rothaus, Morton, Johnson, Cleveland, & Lyle, 1963).

The skills cited by Egan (1970, 1971, 1973, 1975, 1976) as needed for effective helping and effective interpersonal relating include those cited by Truax and Carkhuff (1967) but go beyond these: accurate empathy (primary and advanced levels), respect, genuineness, concreteness, self-disclosure, immediacy ("you-me" talk), confrontation, providing an alternate frame of reference, elaboration of action programs, and support. The helper uses these skills to pursue the ultimate goal of training the helpee in the communication skills so that the helpee not only attains effective interpersonal relating, but, in so far as this is possible, also learns the skills so well that he becomes a helper to others. Helping is seen as an important, mutual aspect of interpersonal living.

If individuals within the human relations movement are to share a common body of research, there must be a consensus as to the behaviors which define the helping skills. Egan's (1975) descriptions of skills show good agreement with those provided by Truax and Carkhuff (1967).

Egan defines accurate empathy in the following manner:

A person is accurately empathic if he can do two things: (a) get inside the other person, look at the world through the perspective or frame of reference of this other person, get a feeling for what the world of the other is like (this is discrimination), and (b) communicate to the other this understanding in such a way that the other knows that the helper has picked up both his feelings and the behavior and experience underlying these feelings. (pp. 67-68)

What Egan (and Truax and Carkhuff) call genuineness essentially refers to what Rogers (1957) terms "congruence." The genuine person, one who is basically himself in all his interactions, exhibits the following behaviors: spontaneity, nondefensiveness, lack of discrepancies, and avoidance of professional role-taking. Respect "means prizing another person simply because he is a human being" (Egan, 1975, p. 90), and is characterized by: being "for" the other, willingness to work with the other, regard for the client as unique and self-determining, assuming good will on the part of the other, attending to the other, refusal to exploit the other, suspending critical judgment (analogous to Rogers' "unconditional positive regard"), and warmth. By referring to the previously-quoted definitions by Truax and Carkhuff (1967), the reader can see that Egan's accurate empathy, genuineness, and respect closely resemble accurate empathy, genuineness, and nonpossessive warmth as given by Truax and Carkhuff. While some variations may exist in nomenclature and classification of helping behaviors, Egan's model would certainly include the three helping skills emphasized by Truax and Carkhuff. At no point do the two models seem inconsistent or contradictory.

Egan's model for helping skills presents a particularly desirable opportunity for conducting research on the relationship between hostility and learning. Egan directs experiential group courses on skills training at Loyola University of Chicago, and the students may be thought of as "normals." Yet the skills learned by these "normals" are the same skills in which psychotherapy patients would be trained. It would then be possible to conduct research on the hostility-learning relationship using "normals" being trained in relatively well-defined skills, and yet hypotheses regarding learning in psychotherapy would be

suggested.

In order to find whatever relationship might exist between hostility and the learning of helping or interpersonal communication skills, it is important to employ an accurate method of measuring the various skills. The skills being assessed are behavioral, and so it might seem that it would be easy to find such measures. However, the behaviors are indeed complex, dealing with emotions and requiring phenomenological knowledge of both helper and helpee. Evaluation of levels of helping skills is a more concrete task than measuring "improvement" in psychotherapy, but it is also far more difficult than evaluating learning in a college course. Given the importance of employing valid and reliable measures of helping skills, research on the available measures will be discussed in considerable detail. Unfortunately, the research does not point to any one measure as being clearly and decidedly superior to the others.

Assessment of Helping and Human Relating Skills

Convergent and discriminant validity. Most of the available research has focused on three general types of skills: (a) accurate empathy, (b) genuineness or congruence, and (c) nonpossessive warmth, respect, or regard. However, there is some controversy as to the convergent and discriminant validity of these concepts. Muehlberg, Pierce, and Drasgow (1969) factor analyzed 5-point scales of empathy, positive regard, genuineness, self-disclosure, and concreteness. Three experienced therapists each saw the same client for a single psychotherapeutic interview. Two raters were used, from which an average rating was obtained. The intercorrelations among the five scales ranged from .78

to .91. The authors concluded that, for both high and low levels of therapist and client functioning, "a single factor was found to account for practically all the conditions" (p. 95). This major factor may have been the therapist's friendliness, helpfulness, and likeability--being a "good guy". However, another study (Zimmer & Anderson, 1968) using factor analysis obtained contradictory results. The experimenters trained 149 junior and senior undergraduate students to rate positive regard and 123 freshman undergraduate students to rate empathy. Ratings were done on 100 counselor responses. Factor analysis indicated that positive regard and empathy do indeed consist of orthogonally related factors. While these two studies were informative, factor analytic studies can be confusing in that different methods of factor analysis can produce different factors on the same data, and it is also difficult to attach theoretical meaning to the mathematically-derived factors.

In their review of the considerable research by Truax and others on the importance of helping skills, Rappaport and Chinsky (1972) flatly state that the accurate empathy scale "lacks discriminant validity" (p. 400). The authors were specifically referring to the measurement of accurate empathy by ratings of audio tapes. Accurate empathy is the skill which has received the most attention, and Rappaport and Chinsky concluded that, on the basis of available research, accurate empathy has a questionable relationship to outcome. However, Truax and Carkhuff (1967) cite research by themselves and their colleagues showing that higher levels of the helping skills are related to positive outcome. Lanning (1971) cited several studies (Barrett-Lennard, 1962; Gross & DeRidder, 1966; Kamin & Caughlin, 1963; Hountras & Anderson,



1969) which indicate that clients showing greater change perceive more accurate empathy, congruence, and unconditional positive regard in their therapists than do clients who show less change.

A comprehensive study by Kurtz and Grummon (1972) supports the contention that convergent and divergent validity have not been clearly demonstrated for the accurate empathy construct. The authors used several measures of the therapist's empathy, and correlated these with each other, with a measure of therapeutic process (depth of self-exploration), and with several outcome measures. The first measure of empathy was the Affective Sensitivity Scale (Campbell, Kagan, & Krathwohl, 1971). The Affective Sensitivity Scale was termed a situational measure of empathy by Kurtz and Grummon, since the scale consists of a standardized test situation. The scale does not measure accurate empathy per se, but the ability to perceive and accurately identify the affect of others. This ability is considered to be an essential component of the ability to communicate accurate empathy. Two "predictive" measures of accurate empathy were used: the Interpersonal Checklist and a version of the Kelly Role Concept Repertory Test. These were considered measures of empathy in that the therapist had to predict how the client would respond to a self-description instrument. "Perceived" empathy was measured by the Empathy Scale of the Barrett-Lennard Relationship Inventory (1962), an inventory in which measures of therapeutic conditions are derived from the responses of clients to a series of items describing the therapist's perceived behavior. The clients in the study filled out the inventory twice--after the third and the last sessions. This empathy scale was also given to the therapists after the last session so that they could rate their own level of empathy. A final measure of

empathy was a scale on which judges would rate empathy by listening to audio tapes. The process measure was a scale of client self-exploration. Five measures of therapeutic outcome were employed: (a) the Tennessee Self-Concept Scale, (b) pre-therapy and post-therapy MMPIs, (c) therapists' ratings on clients' change, (d) clients' ratings on self-change, and (e) a composite score based on all the outcome measures. Of all the correlations between the various empathy measures, only one was significant: the correlation between the Empathy Scale of the Barrett-Lennard Relationship Inventory (RI) completed after the third session and that completed after the last session ($\underline{r} = .66, \underline{p} < .05$). Only one other correlation approached significance: RI after the last session and the judges' tape ratings ($\underline{r} = .31, \underline{p} < .10$). Other than this one possible exception, the various measures of empathy appeared to be unrelated to each other. Only tape-rated empathy was positively and significantly related to the process measure (depth of self-exploration), but the process measure itself was not significantly related to any outcome measure. The best correlation between composite outcome and a measure of empathy was the RI completed after the third session ($\underline{r} = .55, \underline{p} < .01$). The second best predictor of composite outcome was judges' tape ratings, with a correlation of .30. Of the remaining correlations between empathy measures and composite outcome, one was .01 and the other three were negative and nonsignificant. Kurtz and Grummon concluded that the low and nonsignificant correlations between measures of empathy suggest that they are measuring different qualities, except for some possible overlap between client-perceived empathy (RI) and tape ratings. The amount of overlap may depend on the client's level of functioning and corresponding capacity to make accurate assess-

ments. One essential finding was that "client-perceived empathy was the best predictor" (p. 111) of outcome. RI scores correlated significantly with 4 of the 6 outcome measures (The Tennessee Self-Concept Scale yielded two scores), and almost significantly with the other two. By contrast, the next best predictor, tape ratings, was significantly related to 1 of the 6. There were no other significant correlations between empathy and outcome.

Because of the complexity of the Kurtz and Grummon study, it is difficult to reach coherent understanding of the results. However, it may be safely concluded that there is a lack of consensus as to what accurate empathy is. The ambiguity in construct validity can be attributed to the theoretical constructs and/or the instruments used for assessment. Careful consideration of the various instruments and their rationales is necessary.

Rating of tapes. Truax and Carkhuff (1967) favor the use of judges' ratings of audio tapes as the instrument to measure the therapeutic conditions. In this method judges are trained to rate the therapist's statements according to defined levels of each of the therapeutic conditions. The reliability of the rating scales is usually reported in terms of interrater agreement. Evaluation of the validity of the rating scales generally relates to two areas: (a) face validity, and (b) a determination that whatever is being measured does positively relate to outcome. Clearly, the validity of the rating scales has been more problematical than the reliability. Fischer, Paveza, Kickerty, Hubbard, and Grayston (1975) reported reliabilities of .78 on empathy, .65 on nonpossessive warmth, and .75 on genuineness, but noted that these correlations appear to be slightly higher than those generally found in

some of the previous research. Most of the research has reported adequate but not excellent inter-rater reliability.

One natural question in this line of research is the extent to which certain characteristics of the judges affect the value and accuracy of the ratings. In regard to the training of the judges, Shapiro (1968) obtained correlations indicating that "untrained raters are able to differentiate high and low levels of psychotherapeutic behavior in a manner which is similar to that of trained raters" (p. 88). However, Cannon and Carkhuff (1969) found that both the judge's experience and level of functioning had a significant effect on his ability to discriminate therapeutic conditions, with the latter more strongly related to this ability. In the belief that judge characteristics can affect the reliability of tape ratings, Carkhuff (1969a) developed a scale (the Carkhuff Discrimination Scale) that can be used to select the most reliable and accurate raters.

Another question regarding the use of tape ratings is the degree to which the ratings are influenced by the client's responses to the therapist's interventions. Investigating this question, Truax (1966b) conducted a study in which the judges heard either the entire tape recording or heard only the therapist's statements, with the client's verbalizations deleted. Truax concluded that, overall, knowing the client's responses did not significantly contaminate the ratings. Although Truax interpreted the results of this study as supportive of the validity of tape ratings, Chinsky and Rappaport (1970) interpreted the results as just the opposite. They proposed that ratings of accurate empathy should depend on the client's responses to the therapist's interventions; otherwise, how can one judge the therapist's ability to

accurately empathize with the client? Chinsky and Rappaport thereby suspected that some quality or qualities other than empathy were being measured. They further questioned the statistical method of determining reliability coefficients in this and other studies. Truax (1972) published a rejoinder to this critique of his findings. Truax defended his statistical techniques, citing recent research which used different techniques yet obtained similar results. He also defended the construct validity of the empathy scale, citing research that a global "good guy" quality was not being assessed. He proposed that the empathic quality of a therapist's statements was not contingent upon the client's responses. Furthermore, the raters had a series of each therapist's statements, and could discern whether an intervention was "on target" by looking at the following statement by the therapist.

Rappaport and Chinsky (1972) replied to the rejoinder by Truax (1972), stating that the empathy scale has a questionable relation to outcome and that Truax failed to cite entire studies or portions of studies that contradicted his theories. The authors criticized Truax's methodology, and proposed that nothing can be clearly known about the relation between empathy and outcome unless the researchers know exactly what the accurate empathy scale is measuring. In partial defense of Truax, the helping models by Carkhuff (1969a, 1969b), Carkhuff and Berenson (1976), and Egan (1970, 1971, 1973, 1975, 1976) are developmental models in which the primary purpose of the helper's empathy is to encourage the client's self-exploration and self-understanding. The models are developmental in that client self-exploration and self-understanding are seen as phases along the path to constructive change. Accurate empathy need not be related to outcome in that other phases,

such as a behavioral action program, must be successfully completed if there is to be positive outcome. Thus, one would not expect a perfect relation between empathy and outcome, although the two should be positively related.

A more cogent facet of Rappaport's and Chinsky's critique was the high correlation between raters hearing the client's responses and the raters not hearing these. Since the helping skills approach is rooted in client-centered theory, it is odd that the raters are generally told to ignore the client's statements. According to Rogers' (1957) pioneering work in this field, empathy exists only as it is perceived by the client. The scales developed by Truax and Carkhuff (1967) are defined specifically and in relation to the client, so why is it that raters can seemingly rate in a comparable manner whether they hear the client or not? The following quotes from Rappaport and Chinsky (1972) clearly state this position:

The implication is that the client does not matter and that, in fact, if one sets up a tape recorder on which are recorded highly rated accurate empathy statements, and plays these to a client, regardless of what the client says, the tape recorder would foster positive behavior change. (p. 402)

More importantly, it may be possible to rank order the responses presented by Truax (1972) in terms of what sounds like a "good" empathic statement or series of statements, but the distinctions made by the accurate empathy scale are far more specific and client related. (p. 402)

This criticism appears to be perhaps the strongest point against the use of tape ratings. It may well be that raters are rating on the "form" of the therapist's verbalizations. For example, assume that a client is worried about his low grades in college courses. Judges would probably agree that therapist responses like "If you weren't so lazy you would do better," or "Why don't you try bribing your teachers?" should be

rated low on the accurate empathy scale. But what is the client's predominant emotion? It could be frustration, depression, anger, fear, etc. Without hearing the client's response, how could judges make differential evaluations of the empathy shown in the following interventions: "You feel depressed because it seems that you can't accomplish what you thought you could" or "You are fearful that you may never get your degree" or "You are frustrated because you don't know what to do to improve your situation?" The preceding interventions would probably be rated as at least moderate in empathy, because they are in the proper "form". Yet the accuracy of these statements cannot be fairly evaluated in lieu of the client's verbalizations. Ratings of empathy should depend on the client and not just the therapist.

Clients' perceptions of therapeutic conditions. An alternate method of assessing the therapeutic conditions, one seemingly more in line with a client-centered framework, is the use of the client's perceptions. Citing research by Truax (1966a), Truax and Carkhuff (1967) claimed that using questionnaires completed by clients is a less valid technique for measuring therapeutic conditions than the use of tape ratings. Yet they also said that this approach is appealingly economical and may be of value with more intact clients. Thus, they developed the Relationship Questionnaire, a series of 141 items describing the client's relationship with his instructor, therapist, or other significant person (Truax & Carkhuff, 1967). The client merely responds "true" or "false" to each item, and responses are then scored on six scales: accurate empathy, nonpossessive warmth, genuineness, overall therapeutic relationship, intensity and intimacy of interpersonal contact, and concreteness. In theory and design, the Relationship Questionnaire

resembles the Relationship Inventory (RI), an earlier inventory developed by Barrett-Lennard (1962). Truax and Carkhuff report correlations of .53 to .56 between scores on the Relationship Questionnaire and tape ratings when non-psychotic clients are used. Using subjects who were meeting each other for the first time in a peer rather than therapeutic relationship, Welkowitz and Kuc (1973) found alpha reliability coefficients of .79 for the empathy scale, .87 for nonpossessive warmth, and .62 for the genuineness scale of a modified Relationship Questionnaire. These coefficients indicate relatively good reliability. Two disadvantages of the Relationship Questionnaire are its length and the wording of several items specific to the instructor-student or therapist-client relationship. Accordingly, some studies have used modified versions of the Relationship Questionnaire items (Frankel, 1971; Ivey, Normington, Miller, Morrill, & Haase, 1968; Welkowitz & Kuc, 1973). Two lesser-used instruments to measure a client's perceptions of therapeutic conditions are the Counselor Effectiveness Scale (Frankel, 1971; Ivey, Normington, Miller, Morrill, & Haase, 1968) and the Wisconsin Relationship Orientation Survey (Archer & Kagan, 1973; Danish, 1971; Frankel, 1971).

The pioneer instrument to measure clients' perceptions, the RI (Barrett-Lennard, 1962), seems to also be the most widely used instrument. In developing the RI, Barrett-Lennard proposed five areas of the therapist's functioning which are central to positive client change: regard for the client, unconditionality of regard, empathic understanding, congruence, and willingness to be known by the client. In agreement with Carl Rogers, the focus of attention is the therapeutic conditions as experienced by the client:

It follows from this that the relationship as experienced by the client (rather than by the therapist) will be most crucially related to the outcome of therapy. Moreover, although it is not supposed that a client's conscious perceptions would represent with complete accuracy the way he experiences his therapist, it would seem that his own report, given under suitable conditions, would be the most direct and reliable evidence we could get of his actual experience. (p. 2)

Barrett-Lennard defined empathic understanding as "the extent to which one person is conscious of the immediate awareness of another" (p. 3). Empathy involves recognition of both directly and indirectly expressed emotions within the appropriate context. Level of regard represents the affective component of the therapist's behavior--a continuum ranging from intense negative feelings to intense positive feelings. Unconditionality of regard refers to the extent to which the therapist's level of regard (affect) depends upon the behavior of the client. Congruence can be described as honesty, directness, and sincerity, the "absence of conflict or inconsistency between his total experience, his awareness, and his overt communication" (p. 4). Congruence does not require that a therapist or other person say everything he is aware of, but that there are no inconsistencies and nothing is withheld for personal, non-therapeutic reasons. Willingness to be known is akin to reasonable self-disclosure, i.e., self-disclosure to the extent that the other desires it. As with the therapeutic conditions previously described (Egan, 1970; Truax & Carkhuff, 1967), the construct validity of these conditions is a question requiring exploration. Although Barrett-Lennard proposed that these five conditions represent separate constructs, his ideas regarding construct validation are questionable:

The theoretical relatedness of the relationship measures is sufficient to expect a moderate degree of positive association between valid measures of them. However, each one is considered

to have significant contributing influence, in its own right, on therapeutic change, so that each one should (over a moderately large and diverse sample of therapy relationships) be associated with change. (pp. 5-6)

This position on construct validity is similar to the previously-mentioned arguments by Truax and Carkhuff (1967) that whatever is being measured is related to positive change. This validation is not acceptable, since defining the "whatever" is important in interpretation of results. This issue will be given further attention following a description of the RI and Barrett-Lennard's findings.

The revised RI consists of 85 statements describing the manner in which one person "could feel or behave in relation to another person" (p. 34). One example of these statements is: "He usually understands all of what I say to him." Responses to each statement are scored on one and only one of the scales for the five therapeutic conditions, with 16 to 18 statements representing each scale. The items for the scales are dispersed throughout the inventory rather than clustered together. For each statement the subject indicates his degree of agreement (+1 to +3) or disagreement (-1 to -3). The range of possible scores on a given scale is $-3n$ to $+3n$, where n is the number of scale items. A total score is obtained by summing the scores on the five scales.

The participants for Barrett-Lennard's study were 42 clients and 21 therapists from the University of Chicago Counseling Center. The mean split-half reliability coefficient for the five scales was .86, while the mean intercorrelation coefficient was .45. These correlations provided some support for the construct validity of the scales. The obtained data were generally supportive of the experimental hypotheses. There was little linear relationship between the therapist's perception

of offered conditions (as measured by RI scores) and the client's perceptions (also measured by RI scores). There was an overall tendency for the therapists' perceptions of offered conditions to be slightly more positive than the clients' perceptions. With the exception of "Willingness to be Known", it was found that: (a) levels of the conditions positively predicted outcome, with the client's RI scores predicting better than the therapist's perceptions (RI) of conditions offered, and (b) clients of "expert" therapists produced more positive RI scores and greater change than clients of "nonexpert" therapists. Clients' perceptions were more predictive of outcome than therapists' perceptions, but the best predictions came from a combination of the two. In light of the results, Barrett-Lennard proposed that "Willingness to be Known" is a component of congruence rather than a separate variable. Nevertheless, the results supported the use of clients' perceptions in the measurement of therapeutic conditions, with the RI being a promising instrument for that purpose.

Subsequent research analyzed the strength of the RI as an assessment technique. Mills and Zytowski (1967) used a 64-item version of the RI, with 16 items for each of the conditions of empathy, congruence, regard, and unconditionality of regard. The RI was given to 79 female undergraduates in two forms, one to reflect the student's relationship with her mother and one to evaluate the student's perceptions of her mother's feelings toward her. This did not require any modification in the wording of items, since the RI can be used to describe any dyadic relationship. Test-retest reliability was done over a 3-week interval, with coefficients ranging from .74 to .90. Overall, these were slightly lower than those reported by Barrett-Lennard (1962) for a 4-week inter-

val, but still satisfactory. Mills and Zytowski computed four separate component analyses, using their own data and that reported by Barrett-Lennard. Three components were identified: (a) a general factor accounting for about 67% of the variance in each analysis, (b) a component indicating a reciprocal relationship between level of regard and unconditionality of regard, which accounted for about 15% of the variance, and (c) a relationship-distorting component accounting for about 10% of the variance. These results are supportive of the reliability of the RI but not supportive of the convergent and discriminant validity of the four scales. The data do not provide any clear conclusions about the validity of the test as a whole.

Walker and Little (1969) also suspected that the correlations among the four scales of the RI might indicate general factors accounting for most of the score variance. (In accordance with Barrett-Lennard's conclusions, most researchers have ignored the "willingness to be known" variable.) Dissatisfied with the methodology in the Mills and Zytowski study, Walker and Little conducted their own factor analysis. Using the same 64-item RI employed in the Mills and Zytowski study, the authors asked 150 students in an introductory psychology course to complete the inventory in relation to one significant person in their lives. Factor analysis identified three basic factors: (a) nonevaluative acceptance, seemingly most related to unconditionality of regard, (b) psychological insight, most related to empathy and congruence, and (c) likeability, most related to positive regard. The results of this study are generally supportive of the validity of the four scales of the RI, and therefore are not in agreement with the data obtained by Mills and Zytowski. This is not sur-

prising, in that the two studies used different statistical techniques and slightly different experimental instructions, and thus found different factors. Moreover, factors may differ depending on whether the relationship is one of therapist-client, mother-daughter, etc. One might conclude that these two studies did as much to obfuscate as to clarify the validation of the RI.

An experiment by Lin (1973) provided further support that the RI is a reliable inventory. Lin found the following internal consistency (alpha) coefficients for the 64-item RI: .88 for empathy, .91 for regard, .76 for unconditionality, and .92 for congruence. These were judged to be high, except for unconditionality. While good reliability was demonstrated, doubts were raised regarding construct validity. Using three different questionnaires, Lin found that the levels of therapeutic conditions perceived by the clients "was linearly related to the level of counselor's self-confidence" (p. 293). Counselor's self-confidence accounted for a small (11% to 28%) but significant amount of variance. There were three inexperienced counselors who saw clients for five weekly, 30-minute sessions. No significant effect was found for the unconditionality scale, possibly because the variable is poorly defined. Although the counselors allegedly used equivalent techniques, it might have been that more confident counselors were higher in self-confidence because they were also higher in helping skills. Yet it might also be that the clients' perceptions of therapeutic conditions were colored by the self-confidence of the counselors; such an effect might be associated with social influence theory (Dell, 1973; Frank, 1973; Strong, 1968; Strong & Dixon, 1971), according to which the helper, at least in the beginning, who exudes an air of con-

fidence would be perceived as more competent. If this effect did occur, this would be one contraindication for the use of client perceptions.

Wiebe and Pearce (1973) noted that some of the intercorrelations between RI scales were high, and conducted an item analysis to ferret out unreliable or non-discriminatory items. The subjects, 57 freshmen at the University of North Dakota, completed the RI in relation to a friend. Wiebe and Pearce found alpha correlations that were "slightly but consistently lower" (p. 496) than those reported by Barrett-Lennard: .83 for regard, .64 for empathy, .80 for congruence, .73 for unconditionality, .76 for willingness to be known, and .93 for the summed inventory. The results indicated that each scale was significantly and positively related to each other scale, with correlations ranging from .49 to .81. Item analysis suggested that:

a shorter and more robust inventory may be achieved by including only those items which are most strongly correlated with the scale on which they appear, which have an item-scale correlation greater than .50, and which discriminate significantly ($p < .05$) between high and low scorers. Such a revision would produce an RI of 4 scales and 32 items. (p. 496)

The authors concurred with Barrett-Lennard's recommendation that the W (willingness to be known) scale be dropped because it is actually a component of congruence. The suggested revision contained 10 items for Regard, 7 for Empathy, 10 for Congruence, and 5 for Unconditionality. Although this revision would seem to produce a briefer and more refined instrument, it creates a problem in that differences in the number of items for the scales effectively "weights" the contribution of the four conditions in the calculation of the total score.

Other measures of therapeutic conditions. The Affective Sensitivity

scale uses a very different approach to the measurement of a therapeutic condition. The scale, which is available in several forms, utilizes videotapes of excerpts from counseling sessions. Following the presentation of an excerpt or sequence, the subjects are asked to make selections among several multiple choice items which describe the affect that the client is actually experiencing. The quality being assessed is "affective sensitivity," described by Campbell, Kagan, and Krathwohl (1971) as the "ability to detect and describe the immediate affective state of another (affective sensitivity or empathy)" (p. 407). Affective sensitivity is viewed as a necessary component in the ability to effectively empathize with others. One limitation of this scale is that it measures an individual's ability to recognize the immediate emotional state of another, but not necessarily to communicate that recognition. Studies (Campbell, Kagan, & Krathwohl, 1971; Danish & Kagan, 1971; Kagan, Krathwohl, Goldberg, Campbell, Schauble, Greenberg, Danish, Resnikoff, Bowes, & Bondy, 1967) have found small but significant increases in affective sensitivity following counselor training programs. Yet the Affective Sensitivity Scale has not been shown to have strong relationships to outcome criteria or to other measures of therapeutic conditions. According to Gormally and Hill (1974), written responses "lack generalization to real helping situations" (p. 541). Kurtz and Grummon (1972) had considered the scale to be a promising instrument but also found that the ability being measured was not necessarily used in actual counseling sessions, leading them to conclude "that the Affective Sensitivity Scale is not a useful instrument for studying counseling and psychotherapy, even though it may be useful in training situations" (p. 113). A review of the research suggests that

the scale is a less reliable and valid tool for the assessment of therapeutic conditions than either the RI or tape ratings.

The Hogan Empathy Scale (Hogan, 1969) conceptualizes empathy as primarily a personality characteristic. Hogan defined empathy as "the intellectual or imaginative apprehension of another's condition or state of mind" (p. 307). Hogan's definition of empathy goes beyond mere affective sensitivity; empathy is likened to social sensitivity or role-taking ability. An empathic person takes a moral viewpoint--considering the effect of one's actions on others. Hogan developed his 64-item scale by comparing the responses of two groups, one rated high on empathy and one rated low, on a combined MMPI and CPI item pool. Hogan found correlations in the .50's between scores on his empathy scale and measures of sociability and extroversion. Since Hogan's theory assumes that empathic persons have good ability to make appropriate interpersonal differentiations, Martin and Toomey (1973) hypothesized that persons with high scores on the Hogan Empathy Scale would tend to be field-independent on the Embedded Figures Test. The hypothesis and Hogan's theory were supported. Hekmat, Khajavi, and Mehryar (1975) found that persons scoring high on the Hogan Empathy Scale tended to show fewer neurotic and psychotic signs than low-scoring persons. Hogan's scale has not been used extensively, but the available research is generally promising.

The final assessment approach to be discussed is the use of peer ratings. On the basis of a literature review, Jansen, Robb, and Bonk (1972) stated that:

Accumulated research data suggest that peers choose certain fellow counselors as being most or least competent rather

consistently, and that peer ratings correlate positively and significantly with other assessments of counselor competency, such as supervisor ratings. (p. 333)

The experimenters found a significant relationship between competency, knowledge, and likeability. A weakness in this study was that peer rankings of competency were obtained by asking the participants the order in which they would wish to seek help from among their peers. It is then not surprising that likeability was related to competency. The same weakness occurred in a study by McWhirter and Marks (1972), who found practically no relationship between peer ratings and tape ratings of accurate empathy, nonpossessive warmth, and genuineness.

Peer ratings would appear to be an accurate and economical avenue for the assessment of offered therapeutic conditions. Yet they have not been used extensively in the research on helping skills. It may be that peer ratings are often plagued by a crucial problem, the lack of objective criteria on which ratings are to be made. The ratings are only worthwhile to the extent to which they are made on uniform, relevant, and valid criteria. The search for appropriate criteria has generally been elusive.

A comparison of clients' perceptions and ratings of audiotapes. A review of the research on therapeutic conditions indicates that objective tape ratings and clients' perceptions are the two most frequently used methods of assessment, and also appear to be the most reliable and valid. In regard to the present research, a specific measurement approach must be selected as preferable. The choice of approach has important consequences, since the two approaches do not correlate highly and seem to represent, in part, the measurement of different phenomena. The relative efficacy of these two vantage points has been touched

upon in earlier sections of this paper, and a closer scrutiny will now be presented.

The available research shows great discrepancies in the correlations of agreement between judges' tape ratings and client perceptions of therapeutic conditions. The absence of any clear patterns may be at least partially attributable to methodological differences in the following areas: (a) client variables, such as education and level of psychological functioning, (b) therapist variables, such as experience and level of functioning, (c) relationship variables--whether the therapy relationship is real or is simulated for the purpose of conducting research, and (d) measurement variables, including the instruments used to rate audio tapes and client perceptions and the times at which measurements are obtained.

Fish (1970) and Hansen, Moore, and Carkhuff (1968) found no significant correlations between RI scores and tape ratings. Welkowitz and Kuc (1973) also found client perceptions to be generally unrelated to tape ratings, while Burstein and Carkhuff (1968) obtained comparable findings using moderate to low functioning therapists. Bozarth and Grace (1970) found that the therapeutic relationship was viewed differently by judges rating tapes than by clients at a university counseling center, with the total scores of the two methods having a nonsignificant correlation of .48. Both Caracena and Vicory (1969) and Kurtz and Grummon (1972) found a nonsignificant correlation of .31 between empathy scores measured by the RI and by tape ratings. However, Carkhuff and Berenson (1967) found that client ratings were relatively consistent with tape ratings when graduate student trainees were used as clients. Subsequent research (Carkhuff & Burstein, 1970) also implied that the

therapist's level of functioning might be a moderating variable in the agreement between client perceptions and tape ratings. The importance of another moderating variable--the client's level of functioning--was emphasized by Truax and Carkhuff (1967):

In summary, the evidence with respect to perceived therapeutic conditions seems to suggest that measures such as the relationship questionnaire presented in Chapter 2 are indeed useful when used with patients who are not seriously disturbed in their ability to accurately perceive and report. Such positive findings have been obtained with juvenile delinquents, outpatients, and a heterogenous population of vocational rehabilitation clients. By contrast, in schizophrenic or psychotic patients who have severe distortions in perception, such measures as the relationship questionnaire appear to be less useful as measures of the therapist-offered therapeutic conditions. (pp. 137-139)

However, other studies seemingly contradict the above statement. Rogers (1967) found relatively strong agreement between tape ratings and the perceptions of schizophrenic clients. Caracena and Vicory (1969) used "normals," undergraduates taking an introductory psychology course--and yet found no significant relationship between empathy scores on the RI and on tape ratings. McWhirter (1973) used the Guilford-Zimmerman Temperament Survey to select emotionally stable undergraduates to be used as paid, coached clients, and found no significant relationship between RI scores and tape ratings. Yet Kiesler (1966) found moderately good agreement between the two measures using counseling clients, and Hill and King (1976) found a significant positive relationship between the two measures when clients had been trained in what they should be observing about the therapists. Considering the wide variations in the results of these studies, it is indeed difficult to summarize the degree of agreement between client perceptions and tape ratings. The two appear to be positively related, but at this point there is no way to unequivocally describe the strength of the relationship.

It is not clear why there is not greater and more consistent agreement between these two approaches to the assessment of therapeutic conditions. Carkhuff and Burstein (1970) inferred that the lack of agreement is due to a deficiency commonly found in the use of client perceptions; they stated that "inherent in the usual client's condition is an inability to discriminate interpersonally" (p. 395). A similar view was expressed by Truax (1966a), who proposed that the effect of the therapeutic conditions "is relatively independent of the patient's reported perceptions of them" (p. 228). However, the lack of agreement may not be the fault of deficiencies in using client perceptions. While client perceptions may lose accuracy because of emotional distortions and lack of training in making judgments of therapeutic conditions, the opportunity to make use of both verbal and nonverbal messages should be an asset. Haase and Tepper (1972) examined the relative contributions of verbal and nonverbal behaviors to the variance in judged levels of accurate empathy. They found that "with respect to the main effects the nonverbal components in the model accounted for slightly more than twice as much variance in the judged level of empathy as did the verbal message" (p. 421). This finding led the authors to suggest that the accuracy of judgments may be reduced by 66% if only the verbal component is used for data. Whereas objective judges must make their ratings on the basis of brief, audiotape excerpts, clients can base their perceptions on the entirety of the therapeutic sessions, including both verbal and nonverbal components. The lack of greater agreement between client perceptions and tape ratings may therefore be related to both differences in who is making the ratings and in what is being rated. Support for this contention comes from research

by Blaas and Heck (1975), who found that tape ratings can attain enhanced accuracy if the judges receive a description of the client's attitudes and perspectives.

Lack of agreement between the two main approaches to the assessment of therapeutic conditions would not be so important an issue if one approach had been shown to be clearly more valid. Unfortunately, such superiority has not been demonstrated. Validation for the use of tape ratings has been shown in studies finding a significant positive relationship between ratings of therapeutic conditions and process and outcome measures of improvement in therapy (Hansen, Moore, & Carkhuff, 1968; Mullen & Abeles, 1971; Truax & Carkhuff, 1967). Tape-rated empathy has been found to be related to the level of training of the therapist (Fish, 1970), but also to the therapist's verbal dominance or wordiness (Caracena & Vicory, 1969). Client perceptions of therapeutic conditions have also been found to be positively related to therapeutic outcome (Barrett-Lennard, 1962; Rogers, 1967; Truax & Carkhuff, 1967). Further validation comes from studies indicating that perceptions of empathy, genuineness, and warmth in relationships with parents and friends are positively related to self-disclosure in these relationships (Shapiro, Krauss, & Truax, 1969), and that client perceptions of empathy are inversely related to an interviewer's anxiety (Pierce & Mosher, 1967). On the negative side, perceptions of empathy have been found to be positively related to the age of the therapist (Fish, 1970).

In addition to the above studies supporting the validity of each of the two main approaches to the assessment of therapeutic conditions, four studies have attempted to compare the two approaches to see which

is the more valid. Two studies (Hansen, Moore, & Carkhuff, 1968; Truax, 1966a) found that client perceptions on the RI were less predictive of therapeutic outcome than were tape ratings, but two other studies (Carcena & Vicory, 1969; Kurtz & Grummon, 1972) indicated that client perceptions provide the more valid measure of therapeutic conditions.

Considering the many discrepancies in the research literature, there is no sound scientific basis to conclude that either approach is more valid. As Gormally and Hill (1974) have said, "we do not know who is the most objective judge" (p. 543). One reasonable solution to the problem of choosing the most accurate measure of therapeutic conditions would be combination of sources of judgment. The training groups under the direction of Gerard Egan at Loyola University of Chicago present a unique opportunity for such a combination. Egan (1970, 1971, 1973, 1975, 1976) believes that the facilitator of a training group should function as both leader and member. The group facilitator is a leader because of the special resources and human relating skills he makes available to the group; he is a member because he pursues the same contractual goals of interpersonal growth that all group members do. An integral aspect of these groups is mutuality, according to which all group members, including the trainers, function as both helpers and helpees. In working toward interpersonal growth, the leader is sometimes in the role of helpee being helped by other group members. He can then be conceptualized as a client or helpee who generally functions at a relatively high level. Having the leaders-helpees complete Relationship Inventories regarding the therapeutic conditions offered by the group members combines the advantages of both client perceptions and ratings by trained judges. The leaders are selected on

the basis of demonstrated proficiency in helping skills, and thus should be comparable to judges trained to rate audiotapes in their ability to make accurate discriminations of helping skills. The leaders' function of training others in the skills of helping and human relating make them ideal candidates to provide client perceptions. Moreover, the leaders, just as any clients completing Relationship Inventories, have access to both verbal and nonverbal information, and their perceptions are not based on only brief vignettes of the total relationship. This arrangement seems to present an optimal setting for the use of client perceptions as measured by the Relationship Inventory.

Gender Differences in Perceptions of Empathy

The literature on therapist-offered levels of facilitative conditions indicates that the gender of both therapist-helper and client-helpee may be an important factor. Olesker and Balter (1972) found that, with undergraduates, students were more empathic when relating to individuals of the same rather than opposite sex. However, a re-examination of this effect by Breisinger (1976) produced discrepant findings. Breisinger gave the Affective Sensitivity Scale to 21 male and 21 female graduate students in counselor education, and found no differences in empathy for same-sex versus opposite-sex dyads. The literature on this effect is scant, and further study is needed before definitive conclusions can be made. Since responses to the Affective Sensitivity Scale seem to "lack generalization to real helping situations" (Gormally & Hill, 1974, p. 541), there is reason to doubt whether or not Breisinger's findings actually contradict the earlier results obtained by Olesker and Balter. In American culture, males and females

do experience different, sex-related norms and expectations, and, given that empathy is the ability to phenomenologically understand the world of the other and express that understanding, there is a common-sense appeal to the possibility that empathy may be more readily developed in same-sex rather than opposite-sex dyads.

The possibility of another gender difference was suggested by Abramowitz, Abramowitz, and Weitz (1976). In using audiotapes to rate empathy, the authors found that female graduate students were rated as more empathic therapists than were male graduate students. This effect may also reflect cultural differences in the upbringing of male and female children. Men are supposed to be action-oriented, practical, and "tough", not well attuned to emotional experiencing. Conversely, women are trained to be more sensitive to both their own feelings and the emotions of others. A review of the items included on the masculinity-femininity scales of the MMPI and Guilford-Zimmerman Temperament Survey reflects these gender differences in our society. While individual differences certainly exist and training in empathy may cancel gender differences, the possibility of gender differences in empathy should be given further exploration. The training groups directed by Gerard Egan at Loyola University of Chicago include trainers and non-leader members of both sexes, and thus provide an opportunity to test for both a main effect of female superiority in empathy and an interaction effect depending on whether the dyad is same-sex or opposite-sex. In addition, there is an opportunity to explore the possibility of comparable gender differences in congruence, regard (respect), and unconditionality of regard.

Overview

A review of the literature suggests the possibility of interrelations between hostility, gender, and the learning of interpersonal communication skills. Research by Costin (1970, 1971) indicated that hostility is, at least for males, negatively related to learning, but Costin studied only learning in a classroom setting. Moreover, the test used by Costin to measure hostility, the SST, is in need of further validation. The present study affords an opportunity to determine whether hostility is negatively related to a different form of learning experience--the learning of interpersonal communication skills within the context of a training group--and to obtain further validity data on the SST and the RI.

CHAPTER III

METHOD

Subjects

The group members were 20 male and 41 female students taking an upper-level undergraduate psychology course entitled, "Interpersonal Relations: An Experiential Approach." The mean age of these Loyola University students was 25.7 (standard deviation of 7.0), with a range from 20 to 56. Of the 61 non-leader group members, 16 were graduate students and 5 were religious professionals.

The group trainers (also known as facilitators or leaders) were 11 males and 10 females, most of whom were facilitating in connection with a graduate psychology course entitled, "Practicum in Group Approaches." The mean age of the trainers was 30.0 (standard deviation of 11.6), with a range from 21 to 66. Of the 21 trainers, 7 were undergraduate students and 6 were religious professionals. All trainers were selected to be leaders on the basis of demonstrated proficiency in helping and human relating skills. All trainers had previously been enrolled in at least one experiential course in interpersonal relations in a non-leader capacity. The group trainers were supervised by more experienced trainers, reflecting a pyramid approach to training.

Data were collected on 12 groups, each consisting of 2 co-trainers and from 4 to 7 non-leader members. Each pair of co-trainers consisted of one male and one female, with the exception of one group in which

both trainers were male. Of the 21 trainers, three co-trained in each of two separate groups, but there was no repetition in pairs of co-trainers for the twelve groups.

Materials

The Scrambled Sentence Test (SST), Form C, was used to assess hostility as a personality trait (see Appendix A). The test consists of 70 sets of four words, and for each set the subject is asked to underline three words to form a sentence. Forty of the sets are buffer items, and the remaining 30 are scored as either hostile or neutral. The test provides only one score, the total number of sentences that are scored as hostile. Research on the reliability and validity of the SST was cited in Chapter II.

A modification of the Wiebe and Pearce (1973) revision of the Relationship Inventory (RI) was used to measure levels of helping skills (see Appendix B). The Wiebe and Pearce revision consists of 10 items for each of the scales of Regard and Congruence, 7 for Empathy, and 5 for Unconditionality of Regard. This revision is advantageous in that it has 32 items, and thus takes less time to complete than the original 85-item RI. Since the group trainers had to complete from 4 to 12 of these inventories within one week, it was decided that a shorter inventory would be needed to ensure good cooperation. However, the Wiebe and Pearce revision gives more weight to Regard and to Congruence than to Empathy in the total score. This is a disadvantage in that Empathy has been given more emphasis in the literature on helping skills than any of the other therapeutic conditions. Accordingly, three items (numbers 27, 62, and 92) from the Empathy scale of the original RI

(Barrett-Lennard, 1962) were added to the Wiebe and Pearce revision. The revised RI used in this study thus consisted of 35 items: 10 each for Regard, Empathy, and Congruence, and 5 for Unconditionality of Regard.

Procedure

Immediately prior to the 3rd of the 14 group sessions, trainers and other group members were given a maximum of 15 minutes to complete the SST in a classroom setting. In order to minimize the possibility of an interaction between the training program and pre-test effects, neither trainers nor other group members were informed of the purpose of any aspect of the research.

Following the third group session, Relationship Inventories were distributed to the co-trainers, who were asked to independently complete these inventories for each of the non-leader members in their group and to return them prior to the start of the fourth group session. In completing the inventories, trainers were instructed to respond in terms of themselves as helpees and the other group members as helpers. The inventories were distributed at this stage of the group experience to allow the trainers sufficient opportunity (approximately four hours of total group time) to become familiar with the members' levels of helping skills, and yet it was early enough that these completed inventories could be considered pre-treatment levels of helping skills.

Between the 13th and final (14th) of the approximately two-hour group sessions, the trainers again completed Relationship Inventories following the same procedure. Following completion of the final group

session, all trainers were asked to rank the non-leader members of their groups from most hostile (1) to least hostile (n) (see Appendix C). Upon completion of all data collection, one-third of the trainers were briefly interviewed to explore their experiences in completing Relationship Inventories and hostility rankings.

Dependent Measures and Statistical Design

For each of the 61 non-leader group members, four RI's were collected: one from each of the co-trainers at pre-treatment and at post-treatment. Prior to calculation of many of the dependent measures and subsequent testing of experimental hypotheses, there was a need for reliability data on the version of the RI used in this study. Since three of the Empathy items from the original (Barrett-Lennard, 1962) RI had been added to the modification of the RI by Wiebe and Pearce (1973), it was first necessary to determine the degree of correspondence between the three added Empathy items and the seven Empathy items in the Wiebe and Pearce revision. For the four groups of 61 RI's, the Pearson correlations between the two sets of Empathy items were as follows: .57, .55, .51, and .74 (mean = .59). Each of these correlations was significant at the .00002 level (one-tailed), and were thought to be sufficiently high to warrant inclusion of the three added Empathy items in all subsequent statistical analysis.

The Wiebe and Pearce revision of the RI used item analysis of the original, 85-item RI to form a shorter and more robust inventory of 32 items. However, Wiebe and Pearce did not collect reliability and validity data on their shortened instrument. Although the failure to collect such data represents a departure from rigorous methodological standards,

the research literature includes numerous studies in which a shortened version of the RI or Relationship Questionnaire (Truax & Carkhuff, 1967) was used as a dependent measure without first obtaining reliability and validity data on the modified instrument (Archer & Kagan, 1973; Frankel, 1971; Hansen, Moore, & Carkhuff, 1968; Ivey, Normington, Miller, Morrill, & Haase, 1968; Shapiro, Krauss, & Truax, 1969; Welkowitz & Kuc, 1973). Reliability and validity may be, to a degree, inferred from the fact that such data had been collected on the original, lengthier instruments. Nevertheless, the absence of such data represents a methodological weakness. Accordingly, correlations of internal consistency were calculated for each of the four scales and for the total instrument in the 35-item RI used in this study. The alpha coefficient was used to calculate internal consistency, as this statistic is frequently used to describe the reliability of the RI. Four RI's were collected for each subject in the study, and thus four correlations were computed for each scale. For the Regard Scale, the alpha coefficients of reliability were .76, .91, .89, and .93. The mean of these correlations is .87, which compares favorably to the .91 found by Lin (1973) when using a 64-item RI, and to the .83 found by Wiebe and Pearce (1973) when using the full 85-item RI. For the Empathy scale, the four alpha correlations were .79, .86, .81, and .91, with a mean of .84. This mean is slightly lower than the .88 reported by Lin for the Empathy scale, but considerably higher than the .64 found by Wiebe and Pearce. For the Congruence scale, the correlations were .91, .91, .90, and .91. The mean of .91 again compares favorably to the .92 found by Lin and the .80 found by Wiebe and Pearce. The correlations for Unconditionality of Regard were .79, .83, .86, and .85. The mean of .83 is superior to both the

.76 reported by Lin and the .73 obtained by Wiebe and Pearce. Finally, the coefficients for the total scale were .93, .96, .94, and .97, with a mean of .95 being slightly higher than the .93 cited by Wiebe and Pearce. The alpha correlations for the 35-item RI used in this study are generally comparable to those found by Lin, and uniformly higher than those reported by Wiebe and Pearce. Moreover, the RI used in this study is markedly shorter than either the 64 items used by Lin or the 85 items used by Wiebe and Pearce. Considering the economy in both subject and experimenter time, these data provide some support for the preferability of the 35-item instrument. It was beyond the practical scope of the present study to collect further data on test-retest reliability or validity. Again, these may be somewhat inferred from the data on the original RI, yet there is a need for future researchers to be aware of the desirability of collecting such data.

For each of the four RI's collected on each of the 61 subjects, five measures were obtained--a total score and a score for each of the four scales (Regard, Empathy, Congruence, and Unconditionality of Regard). Since RI's were completed by two co-trainers at both pre- and post-treatment, scores from the two trainers were added to provide overall "pre" and "post" measures on the four scales and on the total score. Since the RI requires subjective evaluations, a combined score based on the responses of both trainers was used wherever possible to hopefully obtain a more accurate and objective measure of the levels of helping skills of the individual group members. Change in overall helping skills (total RI score) was calculated by subtracting the sum of the two "pre" RI totals from the sum of the two "post" RI totals.

Total scores on the SST represented one hostility measure for each subject. In order to more directly compare SST scores with the hostility rankings by the trainers, subjects were also ranked within groups based on the SST scores of the group members. The number of members varied across groups, and thus trainers' hostility rankings were converted to percentages (rank divided by number of members in the group) to allow for some comparison across groups. Ranking was used instead of a rating scale in order to reduce the effects of individual differences in making ratings. The rankings given by the two trainers were summed to provide an overall ranking and an overall ranking percentage (sum of the two ranks divided by 2 times the number of persons in the groups). Just as with RI scores, these measures of combined ranking were used whenever possible to provide a more accurate indication of trainer-perceived hostility.

Analysis of covariance was used to test the hypothesis that hostility is negatively related to learning of helping skills, with the effect being stronger for males than for females. The covariate was overall, pre-treatment RI total (summed over both co-trainers), the independent variables were gender and hostility level, and the dependent variable was overall, post-treatment RI total score. This hypothesis was tested using first the SST as a measure of hostility and secondly using the hostility rankings. For both statistical analyses, subjects were rank ordered by gender and separated into high, medium, and low hostility groups according to thirds of the distribution. For males, low hostility was operationally defined as an SST score (first analysis of covariance) from 1 to 6 ($n = 6$) or as a hostility rank percentage (second

analysis of covariance) from 62% to 100% ($\underline{n} = 7$), medium hostility was defined as an SST score from 8 to 12 ($\underline{n} = 7$) or a hostility rank percentage from 49% to 61% ($\underline{n} = 6$), and high hostility as an SST score from 13 to 16 ($\underline{n} = 7$) or as a hostility rank percentage from 20% to 43% ($\underline{n} = 7$). For females, low hostility was defined as an SST score from 0 to 5 ($\underline{n} = 13$) or as a hostility rank percentage from 80% to 100% ($\underline{n} = 13$), medium hostility as an SST score from 6 to 10 ($\underline{n} = 15$) or a hostility rank percentage from 56% to 79% ($\underline{n} = 14$), and high hostility as an SST score from 12 to 18 ($\underline{n} = 13$) or a hostility rank percentage from 20% to 50% ($\underline{n} = 14$). The cut-off points for levels of hostility had to vary depending on gender because males tended to show more hostility on both measures (see Chapter 4).

Since the trainers in this study also completed SST's, it was possible to examine the relationship between trainer hostility and the RI scores given by the trainer. To do this, a mean RI total score was calculated for each trainer (sum of RI total score across all members being rated, divided by the number of members). Separate means were also calculated for "pre" and "post" RI's given by the trainers. These means and the corresponding standard deviations were also used to explore the relationship between gender of trainer and the RI scores assigned to male and female subjects.

CHAPTER IV

RESULTS

Hostility and Gender

The first hypothesis stated that males show higher hostility scores on the SST than do females. For the 20 male and 41 female non-leader group members, the mean SST score for males was 9.85 (standard deviation = 4.33), while the mean for females was 8.44 (standard deviation = 5.20). The data were in the predicted direction, but the difference was not significant, $t(59) = 1.05$, $p > .05$. For the 11 male and 10 female co-trainers, males again had a higher mean SST score than did females, 10.09 (standard deviation = 5.70) to 8.60 (standard deviation = 4.62). The difference was again nonsignificant, $t(19) = .65$, $p > .05$. Thus the first hypothesis was not supported. SST scores had a nonsignificant negative relation to age, $r(59) = -.15$, $p > .05$.

The second hypothesis stated that males are rated as being more hostile than are females. To test this hypothesis, the hostility rankings provided by the pairs of co-trainers were first summed for each of the non-leader group members; this provided a single measure of trainer-perceived hostility. These overall rankings were then changed to percentages by dividing the sum of the two rankings by twice the number of persons in the individual subject's group. The use of percentages allowed for comparisons across groups, with the assumption that the 12 groups would be approximately equivalent in the distribution of hostility

among group members, or that at least there would be no systematic differences. The co-trainers were asked to rank the members of their groups from 1, most hostile, to n , least hostile, and thus higher hostility ranking percentages indicated lower perceived hostility. Males were perceived as more hostile than females, .55 (standard deviation = .23) to .62 (standard deviation = .24). The second hypothesis was not supported in that the difference again failed to reach significance, $t(59) = 1.24, p > .05$.

There was further analysis to determine whether or not the hostility rankings by co-trainers on non-leader group members depended on the variable of same-sex versus opposite-sex dyad, that is, whether the co-trainer and subject being ranked were of the same or opposite sex. To examine possible effects of this variable, all hostility ranks were first converted to hostility ranking percentages by dividing the hostility ranks by the number of persons being ranked. There were 57 non-leader group members, 19 males and 38 females, who received hostility rankings from one male co-trainer and one female co-trainer. Male non-leader group members were perceived as being slightly more hostile by female co-trainers (mean hostility ranking percentage = .54, standard deviation = .29) than by male co-trainers (mean = .56, standard deviation = .29), but the difference was not significant using a two-tailed test, $t(18) = 0.31, p > .05$. Female non-leader group members were ranked as being slightly more hostile by male co-trainers (mean hostility ranking percentage = .61, standard deviation = .29) than by female co-trainers (mean = .63, standard deviation = .28), with the difference again being nonsignificant using a two-tailed test, $t(37) = 0.24, p > .05$. Thus, there was a tendency for perceptions of hostility to be higher for

opposite-sex than for same-sex dyads, but the differences were small and nonsignificant.

Agreement Between the Scrambled Sentence Test and Trainer-Perceived Hostility

The third hypothesis predicted a positive relation between hostility as measured by the SST and the hostility rankings by the co-trainers. To obtain a single measure of trainer-perceived hostility, the rankings by the two co-trainers for each non-leader group member were again summed and converted to an overall hostility ranking percentage. The Pearson correlation between SST scores and overall hostility ranking percentages was $-.05$, in the opposite direction from that which was predicted. The Pearson correlations computed separately by gender were $-.17$ for males and $-.04$ for females. Thus all correlations were opposite the direction predicted by the third hypothesis, and clearly nonsignificant.

In order to facilitate interpretation of the observed lack of agreement between SST scores and trainer-perceived hostility, the degree of agreement in the perceptions by co-trainers was examined. Strong agreement between co-trainers would raise more doubt regarding the validity of the SST than would a lack of agreement. Each of the rankings by the co-trainers was converted to a percentage by dividing the ranking from a given co-trainer by the number of persons in the group, that is, the number of persons being ranked. The Pearson correlation between the hostility ranking percentages by co-trainers was $.35$, significant at the $.005$ level with 59 degrees of freedom. Correlations by gender were $.32$ ($df = 18$, $p > .05$) when males were being ranked and $.36$ ($df = 39$, $p < .02$)

when females were being ranked. As a check on the procedure of using hostility rank percentages, the nonparametric Spearman correlation was calculated separately for the raw hostility rankings by each of the 12 pairs of co-trainers. The unweighted mean of the 12 Spearman correlations was .36, which is very comparable to the Pearson correlation of .35 between the hostility ranking percentages by co-trainers.

Hostility, Gender, and the Learning of Skills

The Scrambled Sentence Test. The fourth hypothesis predicted a negative relationship between hostility as measured by the SST and the learning of interpersonal communication skills, with the relationship being stronger for males than for females. To examine this hypothesis it was first necessary to obtain single measures of pre-group and post-group levels of skills. For the measure of pre-group or pre-treatment level of skills, the RI's completed by the co-trainers after the 3rd of the 14 group sessions were used. For each of the 61 non-leader subjects, there were two pre-treatment RI's--one from each of the subject's co-trainers. The total scores on the two RI's for each subject were summed to provide a measure of overall pre-treatment level of skills. The same procedure was followed for the post-treatment measure, using the RI's completed after the 13th session.

Analysis of covariance was then used to test the fourth hypothesis. The independent variables were gender of non-leader group member and level of hostility on the SST, the dependent variable was overall post-treatment level of skills, and the covariate was overall pre-treatment level of skills. There were three levels of hostility (low, medium, and high) as defined by the SST; the SST scores defining the various

levels depended on the gender of the subject and were cited in Chapter III (Method). Analysis of covariance was advantageous in that it allowed for examination of learning without having to use some form of change score.

Analysis of covariance indicated no significant main effect for either gender ($F(1,54) = 0.79, p > .05$) or for level of hostility ($F(2,54) = 0.63, p > .05$). The prediction of a significant interaction effect, in which the negative relationship between hostility and learning of skills would be stronger for males than for females, was also not supported, $F(2,54) = 1.36, p > .05$. The only significant source of variation was the covariate, overall pre-treatment level of skills, $F(1,54) = 25.23, p < .001$. The Pearson correlation between overall pre-treatment level of skills and overall post-treatment level of skills was $.55$ ($df = 59, p < .00001$).

While there was thus no support for the hypothesis that hostility as measured by the SST was negatively related to the learning of interpersonal communication skills, it was possible that hostility might yet be related to levels of skills. For males, the Pearson correlation between SST scores and overall pre-treatment levels of skills was $-.31$ ($df = 18, p > .05$), with the correlation between SST scores and overall post-treatment levels of skills being $-.16$ ($df = 18, p > .05$). Thus, for males there was a negative but nonsignificant relationship between hostility as measured by the SST and levels of skills at both pre-treatment and post-treatment. For females, the Pearson correlation between SST scores and overall pre-treatment levels of skills was $.13$ ($df = 39, p > .05$), with the correlation between SST scores and overall post-treatment levels of skills being $.17$ ($df = 39, p > .05$). For females there was

then a positive but nonsignificant relationship between hostility as measured by the SST and levels of skills at both pre-treatment and post-treatment.

Trainer-perceived hostility. The fifth hypothesis predicted a negative relationship between the hostility rankings by trainers and the learning of interpersonal communication skills, with the relationship being stronger for males than for females. The measure of trainer-perceived hostility was the overall hostility ranking percentage, which combines the rankings by the two co-trainers on each non-leader group member. Analysis of covariance was again used to test this hypothesis. The independent variables were gender of subject and overall hostility ranking percentage, the dependent variable was the overall post-treatment level of skills, and the covariate was overall pre-treatment level of skills. Each subject was placed into a high, medium, or low hostility group based on overall hostility ranking percentage; the levels defining these groups varied with gender and were cited in Chapter III (Method).

Analysis of covariance did not indicate a significant effect for gender of subject on overall post-treatment level of skills, $F(1,54) = 1.12, p >.05$. There was a significant main effect for level of hostility, $F(2,54) = 9.00, p <.001$. Thus, part of the fifth hypothesis was supported in that there was a negative relationship between trainer-perceived hostility and the learning of interpersonal communication skills. However, the predicted interaction effect between gender and hostility was not supported, $F(2,54) = 1.26, p >.05$. Cell means unadjusted for the covariate are presented in Table 1.

While analysis of covariance supported the hypothesis of a negative

Table 1
Overall Post-Treatment Levels of Skills^a According
to Gender and Trainer-Perceived Hostility

Overall Hostility Ranking Percentage	Males (20)	Females (41)
<u>Low hostility:</u>		
For males, 62% to 100%	73.29	97.69
For females, 80% to 100%		
<u>Medium hostility:</u>		
For males, 49% to 61%	40.00	88.36
For females, 56% to 79%		
<u>High hostility:</u>		
For males, 20% to 43%	36.71	32.14
For females, 20% to 50%		

^aOverall post-treatment level of skills represents, for each of the 61 non-leader group members, the sum of the two total scores of the RI's completed by the subject's two co-trainers at the end of the group sessions.

relationship between trainer-perceived hostility and learning of interpersonal communication skills, it was also decided to explore the relation between trainer-perceived hostility (hostility ranking percentages) and trainer-perceived level of skills (total scores on the RI's). The Pearson correlation between overall hostility ranking percentage and overall pre-treatment level of skills was $-.33$ for males ($df = 18, p > .05$) and $-.10$ for females ($df = 39, p > .05$). The Pearson correlation between overall hostility ranking percentage and overall post-treatment level of skills was $-.48$ for males ($df = 18, p > .05$) and $-.59$ for females ($df = 39, p < .002$). Thus, for both male and female non-leader group members, there was a nonsignificant negative relationship between trainer-perceived hostility and trainer-perceived pre-treatment skills level, and a significant negative relation between trainer-perceived hostility and post-treatment skills level.

Gender Differences in Levels of Skills

The sixth hypothesis predicted that females would be perceived by co-trainers as displaying higher levels of empathy, regard, congruence, and unconditionality of regard than males. This hypothesis was tested separately for levels of skills at pre-treatment and at post-treatment. A series of t tests revealed no significant differences between males and females, although all differences were in the predicted direction. Means and t values are depicted in Table 2.

The seventh hypothesis states that both males and females are perceived as displaying higher levels of empathy, congruence, regard, and unconditionality of regard when the perceptions are provided by a person of the same rather than opposite sex. Since the co-trainers for 1 of

Table 2
Mean Scores on the Relationship Inventory by Gender

	Males (N = 20)	Females (N = 41)	All (N = 61)	<u>t</u> ^b
Pre-Treatment ^a				
Regard	9.2	11.1	10.5	1.01
Empathy	2.0	3.6	3.1	0.83
Congruence	1.8	6.3	4.9	1.48
Unconditionality of Regard	2.4	3.5	3.2	1.00
Total Score	15.5	24.6	21.6	1.24
Post-Treatment ^a				
Regard	12.9	15.4	14.6	1.44
Empathy	3.3	6.3	5.3	1.31
Congruence	5.4	9.6	8.2	1.64
Unconditionality of Regard	3.7	4.7	4.4	0.72
Total Score	25.2	36.1	32.5	1.49

Note. The mean scores above represent the means of the scores given on the Relationship Inventories by the two co-trainers.

^aPre-treatment scores are those based on the inventories completed by the two co-trainers after the 3rd of the 14 group sessions. Post treatment scores are based on the inventories completed by the co-trainers after the 13th of the 14 group sessions.

^bNo values of t were significant at the .05 level, using a one-tailed test with df = 59.

the 12 groups were both male, the non-leader group members of this group could not be included in the testing of this hypothesis. Thus there were 57 subjects, 19 males and 38 females. Table 3 depicts the results of a series of t tests on the RI scores given to male non-leader group members by male versus female co-trainers. Of the 10 t tests done for male non-leader group members, 6 were significant at the .05 level and all but 1 were in the predicted direction. Table 4 depicts the corresponding t tests on the RI scores given to female non-leader group members. Of these 10 t tests, only 1 was significant at the .05 level while 7 were in the predicted direction. Thus the seventh hypothesis was generally supported for males being rated on skills but not for females.

Hostility of Trainers and Perception of Skills by Trainers

As one check on the validity of using RI's completed by trainers to measure the levels of helping skills displayed by non-leader group members, the hostility of the trainers was correlated with the perceptions of the trainers on the RI. First, the mean total score on the RI was calculated for each of the 21 trainers at pre-treatment and at post-treatment. This mean represented the mean of the total scores on all of the RI's completed by the given co-trainer. The Pearson correlation between trainer hostility, as measured by the SST, and the mean total score on the pre-treatment RI's was $-.13$, $df = 19$, p (two-tailed) $>.05$. The correlation between trainer hostility and mean total score on post-treatment RI's was $-.62$, $df = 19$, p (two-tailed) $<.002$. The correlation between trainer hostility and mean improvement in total RI score was $-.42$, $df = 19$, p (two-tailed) $>.05$. Moreover, trainer hostility was

Table 3

Relationship Inventory Scores Given to Male

Non-Leader Group Members by Male Versus Female Co-Trainers

Pre-Treatment	Male Co-Trainers	Female Co-Trainers	$\frac{t}{(df = 18)}$	One-Tailed Probability
Regard	9.26	9.05	0.10	.47
Empathy	1.84	2.37	-0.28	---
Congruence	3.58	0.11	1.87	.04
Unconditionality of Regard	4.00	0.84	2.33	.02
Total Score	18.68	12.37	1.21	.13
Post-Treatment				
Regard	16.68	9.05	2.40	.02
Empathy	5.32	1.68	1.51	.08
Congruence	9.26	1.53	3.09	.01
Unconditionality of Regard	5.00	2.47	1.88	.04
Total Score	36.26	14.74	2.76	.01

Table 4

Relationship Inventory Scores Given to Female

Non-Leader Group Members by Female Versus Male Co-Trainers

Pre-Treatment	Female Co-Trainers	Male Co-Trainers	$\frac{t}{(df = 37)}$	One-Tailed Probability
Regard	10.82	11.03	-0.16	---
Empathy	5.21	2.47	1.60	.06
Congruence	7.32	5.53	1.06	.15
Unconditionality of Regard	3.37	3.63	-0.24	---
Total Score	26.71	22.66	0.78	.23
Post-Treatment				
Regard	15.08	15.68	-0.36	---
Empathy	8.68	4.63	2.18	.02
Congruence	11.21	8.47	1.55	.07
Unconditionality of Regard	4.92	4.63	0.29	.39
Total Score	39.89	33.42	1.30	.10

was negatively related to the variance in both pre-treatment (-.14) and post-treatment (-.01) total RI scores, although neither correlation was significant at the .05 level using a two-tailed test. There was thus a consistent tendency for trainer hostility to be negatively related to trainer perceptions of levels of skills.

Supplementary Data on the Relationship Inventory

Agreement between co-trainers. A series of Pearson correlations were used to determine the degree of agreement between co-trainers regarding the levels of skills displayed by non-leader group members. Although each pair of co-trainers had a different group of subjects to rate on the RI, the data were analyzed across groups so that all correlations were based on 61 cases (the total number of non-leader group members). Table 5 depicts the agreement between co-trainers on RI scores given to non-leader group members. The level of agreement was moderate, with correlations ranging from .06 to .62. Agreement on total RI score at pre-treatment was .48, and the corresponding agreement at post-treatment was .45.

Pre- and post-treatment variance in RI scores given by trainers. A t test was used to determine whether there was a difference between the variances of pre-treatment versus post-treatment total RI scores. The variances in pre-treatment and post-treatment total RI scores were calculated for each of the 21 co-trainers. The variance in total RI scores was significantly greater at post-treatment than at pre-treatment, $t(20) = 2.13$, p (two-tailed) $< .05$. The unweighted mean variance (disregarding differences in the number of non-leader group members being rated by the co-trainers) for pre-treatment total RI scores was 647.8,

Table 5

Agreement Between Co-Trainers on the Relationship Inventory

Pre-Treatment	\underline{r} ($\underline{df} = 59$)	One-Tailed Probability
Regard	.35	.003
Empathy	.39	.001
Congruence	.62	.001
Unconditionality of Regard	.20	.059
Total Score	.48	.001
Post-Treatment		
Regard	.06	.313
Empathy	.40	.001
Congruence	.45	.001
Unconditionality of Regard	.52	.001
Total Score	.45	.001

while that for post-treatment scores was 1042.0.

Pre- versus post-treatment overall skills level. Since one stated purpose of the group sessions (treatment) was to train students in interpersonal communication skills, a t test was used to determine whether co-trainers perceived higher levels of skills at post-treatment than at pre-treatment. Overall pre-treatment and post-treatment skills level was calculated for each non-leader group member by summing the total scores on the RI's completed by the subject's two co-trainers. Overall post-treatment skills level was significantly greater (mean = 65.0, standard deviation = 53.6) than overall pre-treatment skills level (mean = 43.2, standard deviation = 47.9), t (60) = 3.52, two-tailed p < .001. The mean change in overall skills level was +19.6 for male non-leader group members and +23.0 for females.

Scale intercorrelations. The Pearson correlations between RI scales are shown in Table 6. The highest interscale correlations were between Empathy and Congruence, with a mean correlation of .82. Unconditionality of Regard showed the lowest correlations with the other scales, with correlations ranging from .26 to .72. With the exception of Unconditionality of Regard, the scale intercorrelations tended to be relatively high.

Table 6

Pearson Correlations Between Scales on the Relationship Inventory

Scale	Regard	Empathy	Congruence	Unconditionality of Regard
Regard				
Empathy	.64, .77, .58, .73, Mean = .68			
Congruence	.59, .77, .54, .72, Mean = .66	.85, .82, .77, .85 Mean = .82		
Uncondition- ality of Regard	.26, .60, .42, .65, Mean = .48	.34, .71, .35, .62, Mean = .51	.41, .69, .60, .72, Mean = .61	
Total Score Minus Scale	.62, .80, .61, .77, Mean = .70	.83, .86, .73, .84, Mean = .82	.83, .86, .80, .87, Mean = .84	.39, .72, .53, .72, Mean = .59

Note. Four inventories were completed for each of the 61 non-trainer group members, one from each of the co-trainers at the beginning and again at the end of the group. Thus four sets of intercorrelations were computed, with each based on 61 subjects. The means reported above represent the mean of the four correlations. All of the above correlations are significant at the .025 level, using a one-tailed test with df = 59.

CHAPTER V

DISCUSSION

Gender Differences in the Measurement of Hostility

Sarason (1961) found that males score higher than females on most of the hostility scales that he examined, and this pattern occurred without exception in all of Costin's (1969, 1970, 1975) research on the SST. In the present research, higher hostility scores were found for males on both hostility rankings and the SST and for both co-trainers and non-leader group members, but none of the differences were significant. The lack of significant results was surprising given the consistency of the findings in other studies. Table 7 depicts the mean SST scores according to gender in the various samples studied by Costin (1969, 1970, 1975) and in the present research. It can be seen that, although the standard deviations are comparable, the mean scores found in the present study are lower than any of those reported by Costin. In addition, the differences obtained in the present study between the means for males and for females were less than any found by Costin. The reasons for the lower scores and the lower differences remain uncertain, but may lie in variation in the samples used. First, the present study utilized a pre-selected sample, students who had registered for a course in interpersonal relations. It may be that these students are less hostile or show less gender differences regarding hostility than university students in general. Second, the mean age of both the non-leader

Table 7
 Male and Female Scores on the Scrambled
 Sentence Test in Several Studies

	Males			Females			Difference in Means
	Mean	SD	<u>n</u>	Mean	SD	<u>n</u>	
Costin (1969)	11.7	5.2	140	9.5	5.1	177	2.2
	11.9	5.0	52	9.8	4.7	75	2.1
	11.9	5.7	52	9.0	5.0	75	2.9
Costin (1970)	11.7	3.8	50	9.1	4.3	51	2.6
	11.8	4.2	50	9.4	4.3	51	2.4
Costin (1975)	12.6	5.7	33	10.1	4.8	44	2.5
	13.0	6.3	33	9.9	4.1	44	3.1
	12.9	3.9	240	10.5	3.8	226	2.4
	11.4	4.2	406	8.8	3.2	369	2.6
	12.1	3.6	144	10.0	3.4	100	2.1
	11.9	3.5	128	9.4	3.5	122	2.5
	9.8	4.3	20	8.4	5.2	41	1.4
Present Study	10.1	5.7	11	8.6	4.6	10	1.5

group members (25.7) and the co-trainers (30.0) is probably older than that in Costin's studies. Age may be a variable that deserves further consideration, in that the present study found a negative but nonsignificant correlation (-.15) between age and scores on the SST. A negative relation between age and hostility might be suggested by crime statistics and other indices of hostile behaviors. The lack of significant differences between the SST scores of males and females in the present study might also partially reflect the smallness of the sample used in this study as opposed to Costin's research, since the standard deviations were similar.

The interpretation of previous findings that males have higher scores on hostility scales is problematical. It may be that males are innately more hostile or become more hostile as a result of cultural forces. It may also be that males are more willing, for whatever reasons, to let their hostility be known to others, or that the manifestations of hostility that are most commonly tapped by hostility scales are more characteristic of males than females. American culture appears to be more tolerant of displays of hostility from males than from females, and females may therefore have to utilize more subtle expressions of hostile impulses. The higher hostility scores of males may then reflect the greater difficulty in assessing more disguised expressions of hostility. The development of a hostility scale that would be equally effective in measuring gross and subtle manifestations of hostility would enhance the understanding of gender differences in hostility.

An interesting hypothesis is generated by the results showing a nonsignificant tendency for non-leader group members of both sexes to be perceived as more hostile by trainers of the opposite rather than same

sex. The variable of same-sex versus opposite-sex dyad in the perception of hostility merits further exploration. Assuming the existence of gender-related cultural norms regarding the manifestation of hostility, there may be an anxiety-arousing unfamiliarity with the modes of hostile expression used by persons of the opposite sex. As a result of the greater difficulty in inferring hostile intent, there may be a tendency to increase self-protective mechanisms and err on the safe side by overestimating the hostility of the other person.

The Scrambled Sentence Test as a Measure of Hostility

In an attempt to further validate the SST as a measure of hostility, the hostility rankings by co-trainers were correlated with SST scores. Although Costin (1969) found correlations of .65 and .66 between SST scores and hostility ratings by counseling psychologists, the present study found a nonsignificant correlation of $-.05$ between SST scores and overall hostility ranking percentages. Thus there was essentially no linear agreement between the two measures of hostility used in this study. While strong agreement between the two measures could be considered to be supportive of the validity of the SST, interpretation of the observed lack of agreement is more complex.

One important aspect of interpreting the lack of agreement between the SST and the hostility rankings is the agreement between co-trainers in their hostility rankings. The Pearson correlation between the hostility ranking percentages by co-trainers was $.35$. The data indicate that co-trainers agreed with each other in the perception of hostility more than with SST scores, but that there were considerable differences between co-trainers in their hostility rankings. This situation is com-

parable to that commonly found with hostility instruments: there is a confusing lack of agreement between measures and little if any evidence for relative superiority. It would appear that different phenomena are being measured.

The source of the confusion may lie in the basic concept of hostility. It would indeed be difficult to find an operational or theoretical definition of hostility that would be agreeable to most psychologists. The concept of hostility is closely tied to that of aggression, with the latter being perhaps more behaviorally-oriented than the former. Nevertheless, the measurement of both concepts must necessarily depend on some sort of observable behavior, and appropriate behavioral criteria are lacking. A crucial problem is that the description of an act or a person as hostile must depend on not only the behavior itself, but also the intent of the actor. Spanking a baby is a hostile act if done to make the baby suffer but a caring act if done to make the baby start breathing for the first time. Prosecution of criminals may arise from a genuine concern for the public welfare and/or from a desire to harm others in a socially sanctioned manner. The hostile intent of the actor can be safely inferred from gross acts such as senseless torture, but it becomes increasingly difficult to accurately infer intent as the expression of hostility becomes more subtle. Not only is it difficult to know the intent of another, but the actor himself may not be conscious of the purpose behind his behavior. It is then understandable that measures of hostility tend to focus on physical aggression. Unfortunately, persons who express hostility in obvious, physical ways can be recognized as hostile without the use of tests. Test instruments need to focus more on the more subtle indications of hostility, for it is here

that the recognition of hostility is difficult. In the present study, the description of hostility given to co-trainers for use in making hostility rankings was purposely broad, referring to hostility as "a desire or tendency to make other persons undergo negative experiences." This description is admittedly vague, but the author contends that it must be so. The labelling of hostile behaviors must include those that inflict psychological as well as physical distress. It is more risky to infer the intent of another to inflict psychological harm, but the study of hostility will be severely limited if only the gross manifestations are examined.

The SST is a promising measure of hostility in that the sentences scored as hostile represent a variety of behavioral manifestations. There are items reflecting physical harm to persons ("Take her life.") and to objects ("Destroy the book."), items reflecting verbal hostility ("Go to hell."), and items tapping more passive forms of hostility ("Let him cry."). The semi-disguised format of the SST is probably advantageous in that a hostile person may express his hostility by obstructing the examiner's attempt to accurately assess hostility. Nevertheless, the value of the SST could be increased by further test refinement. There is a need for more items aimed at assessing more passive, covert, and socially acceptable forms of hostility. In addition, there are two items with questionnaire scoring. The four words for item 9 are "out, him, let, knock," and the item is scored as hostile if the subject makes the sentence, "Knock him out." However, another sentence, "Let him knock", is scored as neutral even though it seems to reflect the same type of passive hostility that is tapped by item 21, "Let him cry." The four words for item 65 are "life, pay, take, her," and the item is scored as

hostile for the sentence "Take her life," yet is scored as neutral for the sentence "Take her pay."

Hostility and Interpersonal Communication Skills

The results did not support the hypotheses that hostility as measured by the SST is negatively related to the learning of skills, with the relation being stronger for men than for women. There are several factors which might account for the discrepancy between the results of the present study and Costin's (1970, 1971) data indicating a negative relation between hostility and learning in academic courses. Although Costin found a significant negative relation, the strength of the relation was moderate or low (correlations of approximately .4 for males and .2 for females) and accounted for a relatively small amount of the variance in learning. The sample size in the present study may have been too small to separate variance related to hostility from error variance. Accurate measurement of interpersonal communication skills is probably more difficult to obtain than accurate measurement of classroom learning, and thus the relative contribution of error variance in the present study may have been greater than that in Costin's research.

Nevertheless, it seems more likely that the discrepant results reflect differences between the two types of learning experiences. Costin theorized that hostility somehow interferes with learning, but did not speculate as to the source of the interference. This is an essential consideration given that such an interference apparently did not occur in the learning of interpersonal communication skills within a training group. The results of the present study thus suggest that hostility does not necessarily interfere with all forms of learning, and insights as to

possible sources of interference may be gained by comparing the academic learning examined by Costin with the present learning situation. In traditional classroom learning, time spent on homework is generally an important factor in the amount of learning that takes place. Homework time was probably less of a factor in the present study, in that most of the learning of interpersonal communication skills must occur within the context of the experiential training group. Assuming that hostility may be expressed by failure to do homework assignments, this type of interference with learning may have been less of a factor in the present study than in Costin's research. Absences were a negligible factor in the present study, since students were not allowed to miss more than one group session. Hostility would interfere with learning if it were expressed through absenteeism, and students in the present study probably had less opportunity to express hostility in this manner than did students in Costin's research. In addition, most students find the group experience to be more interesting than classroom lectures, and this may have altered the usual reactions of hostile students.

Perhaps the most significant difference between classroom learning and learning in a training group is that the former is basically independent learning while the latter requires cooperative learning. Since the training group is a cooperative endeavor, there is group pressure for all group members to maintain comparable paces of learning. If a group member expresses hostility in such a way as to interfere with his own learning or the learning of other group members, the group member receives group pressure and group support to express hostility in more appropriate ways. For example, a member exhibiting hostile withdrawal would be encouraged to verbalize his hostility. The hypothesized negative

relationship between hostility and the learning of interpersonal communication skills would be inimical to the goals of the group experience.

The results did indicate that the learning of skills was negatively related to hostility rankings by trainers, and also that levels of skills at both pre-treatment and post-treatment were negatively related to hostility rankings by trainers. Nevertheless, these results cannot be interpreted as clear support for the hypothesized negative relationship between hostility and learning, since co-trainers provided subjective measures of both level of skills and of hostility. The task of co-trainers is to train group members in interpersonal communication skills, and there is ego-investment in how well the group members learn the skills. There may have been a tendency to perceive as hostile those group members who showed less learning or lower levels of skills.

Gender Differences in Skills Level

The results indicated that female non-leader group members were rated higher than males on all of the skills at both pre-treatment and at post-treatment. Although none of the differences were significant, the consistency of the findings lend some support to the contention that females display higher levels of interpersonal communication skills. An interesting finding was that this higher level was maintained even after all students had completed the group training experience. It would be worthwhile to attempt to replicate these findings using a larger sample than that used in the present study, as gender differences are probably small relative to individual differences. Female superiority in interpersonal communication skills might be expected on the basis of American culture. The two genders experience variant cultural expectations and

experiences, with females being raised to be more attuned to non-task-oriented interpersonal functioning.

The data also indicated that, while females tend to be perceived as displaying higher levels of skills, the gender of the person making the perceptions may make a difference. Males were generally perceived as showing higher levels of skills when the perceptions were provided by male rather than female co-trainers. The results for females were less consistent, and it may be that females are also seen as being more skilled by other females than by males, but that the effect here is not as strong. Further study is needed, and again it would be wise to employ a larger sample than that used in this study. An overview of this data on gender and skills level suggests an interpretive hypothesis that there are broad interpersonal communication skills in which females tend to be more facile, but also that there are sex-specific interpersonal communication patterns. These sex-specific patterns may reflect cultural differences in the way in which males and females are trained to relate to persons of the same and opposite sex. Males may be less oriented to the styles of interpersonal communication that are common to training groups, and may have a different style for relating to other males than for relating to females. The different patterns for same-sex and opposite-sex relations may also appear for females, but here there may be greater overlap between the two styles.

The Relationship Inventory as a Measure of Skills

The effect of hostility on reported perceptions. The results indicated a negative relation between the hostility (as measured by the SST) of co-trainers and the perceptions of skills by co-trainers on the

RI. This relation was moderately strong and significant with RI scores at post-treatment (-.62) and modest and nonsignificant at pre-treatment (-.13). There was also a nonsignificant tendency for trainer hostility to be negatively related to the variance in RI perceptions given by the trainer. These findings indicate a weakness in the divergent validity of the RI, in that the measurement of helping skills should not be correlated with the hostility of the person (helpee) providing the measurement. While it is best to cautiously interpret the nonsignificant relation between trainer hostility and variance in scores given to subjects by trainers, it would be interesting to give further study to the possibility that trainers may express their hostility by putting little effort into making accurate discriminations. These data on the relation between hostility and RI perceptions can be seen as a contra-indication for the use of the RI, but this negative relation may also exist between hostility and other forms of ratings. This is another area that merits further research, and lends support to the efficacy of using multiple raters, as was done in the present study. Giving low ratings may represent one way to express hostility.

Inter-rater agreement. The RI was designed primarily for use by clients or helpees, and purposely requests subjective evaluations. High inter-rater agreement would therefore not be expected. Given this consideration, the agreement between co-trainers (shown in Table 5) was quite satisfactory. The agreement for total score at pre-treatment was .48 and that for post-treatment was .45; agreement on individual scales ranged from .06 to .62. The common practice when using objective judges to make ratings on the basis of audiotapes is to train the raters to a minimum inter-rater agreement of .50. The inter-rater agreement in the

present study was then slightly lower than the common criteria for objective judges, but the co-trainers had to make ratings on both verbal and non-verbal information and so less agreement might be expected. Because the RI requires subjective evaluations, future researchers should give consideration to using more than one rater, as was done in the present study. The RI ratings in this study were provided by trainers, individuals who can be assumed to be functioning at higher levels than most helpees and who can probably provide more accurate discriminations than most clients. Thus, researchers should be prepared to statistically handle a relatively large amount of error variance when using the RI with clients or persons not trained in interpersonal communication skills.

Depth of the relationship. The mean variance in post-treatment total RI scores was significantly greater than that in pre-treatment scores. If variance is conceptualized as an indication of the ability of trainers-helpees to make accurate discriminations, the trainers were more able to make accurate discriminations after spending more time with the subjects and forming deeper relationships with them. While this conclusion might seem obvious and simple, it is a consideration that is often ignored in the research on the RI. As was pointed out in Chapter II, many of the studies on the RI had subjects completing the inventories after spending relatively little time with the person being rated. It is not surprising that such a procedure might be related to lack of agreement between the RI and ratings by objective judges. An important source of confusion on the RI may be the amount of contact between the two persons before the inventory is completed. The practice of having individuals complete RI's after knowing the person being rated for only one or two one-hour sessions is questionable. Most of the co-trainers

who were informally interviewed after completing pre-treatment RI's stated that they did not feel they had enough time (approximately three to four hours to total group time) to provide very accurate ratings. Contact between any two persons within a group may be less than it would be in an individual session, but this should be more than offset by the greater ability of trainers to make accurate assessments of therapeutic conditions. If the co-trainers found the relatively brief amount of time to be a hindrance, one must wonder what effect time might have on clients or persons not trained in assessing skills. Moreover, clients probably tend to focus on themselves and not on the expertise of the helper-therapist. The efficacy of the RI may increase with the rater's greater exposure to the person being rated.

Scale intercorrelations. The scale intercorrelations for the RI obtained in the present study (Table 6) tend to be slightly higher than those reported by Barrett-Lennard (1962) and Wiebe and Pearce (1973) for the full 85-item RI. Empathy and Congruence were the pair showing the highest intercorrelations, with Unconditionality of Regard having the lowest correlations with the other scales. Empathy and Congruence also showed the highest correlations with the sums of the other three scales, but this may be due in part to the correlation between Empathy and Congruence. The coefficients of internal stability for the modification of the RI used in this study were superior to those reported by Barrett-Lennard and Wiebe and Pearce despite the fact that the RI used in this study contained only 35 of the original 85 items. The RI used in the present study was then briefer than the original, with comparable or superior internal stability but higher interscale correlations. The reason for the higher interscale correlations is not clear, but may lie

in differences in both individuals being rated on the RI and the persons providing the ratings.

The interscale correlations for the RI tend to be high, but an evaluation of these correlations must center on the question of what the optimal correlations should be. In looking at the mean correlation of .84 between Congruence and the sum of the other three scales, one might think it would be more simple to merely administer the 10 items on the Congruence scale. Also, the high intercorrelations raise questions about whether these scales really represent separate constructs. It might appear that one could simply administer 10 items and simply label the scale "interpersonal skills." Yet it is reasonable to expect high correlations between these skills, since the person who is very high on one skill and very low on another should be a relative rarity. While the intercorrelations are high, they are low enough to warrant the use of separate scales. This does not mean that further construct validation is not needed.

Comments by trainers. Of the 21 trainers, 7 were briefly and informally interviewed to explore their experiences in completing the RI. Four co-trainers commented that the items on the RI were generally easy to understand but that they found a few items (e.g., "He behaves just the way that he is, in our relationship.") to be ambiguous and unclear. The trainers are persons who have completed at least two years of college and who are familiar with the terminology of skills training, and, if some of them found a few items to be ambiguous, it is likely that the average person would be more confused. Three trainers stated that it was harder to complete RI's for individuals whom they did not know as well, and that these persons tended to be the less active group members.

These trainers stated that they tended to give moderate ratings to such persons simply because they were less sure of how to rate their feelings toward them. It may then be generally more difficult to rate less active helpers. Three of the trainers interviewed commented that they were not sure whether or not they were able to limit their responses on the RI to themselves as helpees and the others as helpers. They thought that their responses reflected their total relationship and not just this aspect, and this may have been a methodological weakness of the present study. None of the trainers interviewed thought that the use of only masculine pronouns in the RI presented any problems in completing RI's for female subjects. The potential for such problems was of interest to the author, since Barrett-Lennard (1962) used a different form of the RI for males and females, with the only difference being the gender of the pronouns.

Efficacy of the Group Training Experience

The results indicated a significant improvement in overall skills level over the course of the group training experience. An improvement in total score would be expected as people get to know each other on a deeper and more intimate level, and this is seemingly what happened in the 10 weekly group sessions between the completion of pre-treatment and post-treatment RI's. The efficacy of the skills training groups was supported in that, within a relatively brief amount of time, the group members did fulfill their common goal by coming to know each other in a deeper and more meaningful way.

CHAPTER VI

SUMMARY

The purpose of the present research was to explore possible interrelations between hostility, gender, and the learning of helping or interpersonal communication skills. The potential for such interrelations has received relatively little attention in the psychological literature, which may be partly attributable to the difficulty in obtaining reliable and valid measures of hostility and interpersonal communication skills.

Costin (1969) described his development of the Scrambled Sentence Test (SST), a semi-disguised measure of hostility. The SST consists of 70 sets of four words each, with the subject's task being to underline three of the four words so as to make a sentence. Forty of the items are buffer items, and the other 30 are scored as either hostile or neutral. A subject's hostility score is computed by merely summing the number of items scored as hostile. In regard to test standardization, several studies (Costin, 1969, 1970, 1975) indicated moderately good test-retest reliability and internal consistency. Although further research on test validity is needed, that which has been done has been supportive. SST scores have been found to have a moderately strong correlation (.65) with ratings of hostility by psychologists (Costin, 1969). Research (Costin, 1975) comparing the SST with other tests of hostility have supported the semi-disguised format of the SST, indicating

that the SST is indeed more disguised, is less subject to a social desirability effect, and has lower correlations between awareness of test purpose and test score. Convergent and discriminant validity were supported by data showing a negative relation between SST scores and a measure of conflict avoidance, and the lack of significant correlations between the SST and measures of dominance, verbal ability, intelligence, and reading comprehension (Costin, 1969, 1975). However, the SST, as well as hostility tests in general, has not been satisfactorily validated relative to behavioral criteria, and suffers from a lack of clarity in the basic concept of hostility. Nevertheless, the available research suggests that the SST compares favorably with the more traditional hostility inventories.

The results of two studies by Costin (1970, 1971) indicated that, at least for males, hostility is negatively related to learning in a classroom setting. Costin concluded that hostility somehow interferes with learning, but did not speculate as to the dynamics behind this apparent interference. An understanding of exactly how hostility acted to interfere with classroom learning would greatly enhance both the theoretical and practical value of Costin's research. The primary purpose of the present study was to extend Costin's research by determining whether hostility is negatively related to a learning experience quite different from that examined by Costin--the learning of interpersonal communication skills within the context of a skills training group.

There is a considerable body of research emphasizing the importance of "interpersonal communication" or "helping" skills (Carkhuff, 1969a, 1969b; Egan, 1970, 1971, 1973, 1975, 1976). These skills, such as empathy, respect, congruence, genuineness, and nonpossessive warmth, have

been conceptualized as essential components of all forms of interpersonal relationships. Accordingly, therapists need to be trained in these skills, and need to be further trained to train their patients in the same skills. Since these skills are basic to human relating, they pertain not only to the therapeutic relationship but to all human interaction.

Unfortunately, accurate measurement of human relating skills is a formidable task. The behaviors defining the various skills are both subtle and complex; in order to assess an individual's skills in relating to another, one must have a phenomenological knowledge of the other. The two approaches to the assessment of interpersonal communication skills that have been most commonly used are (a) rating scales for various skills completed by objective judges on the basis of audiotapes, and (b) inventories on which individuals report their perceptions regarding the skills displayed by another. The most frequently used representative of the latter approach is the Barrett-Lennard Relationship Inventory (Barrett-Lennard, 1962), which consists of 85 statements (for example, "He respects me.") regarding which a subjective judge records his level of agreement or disagreement. The inventory can be used to measure perceptions of skills displayed by another in any dyadic relationship--therapist and patient, father and son, two friends, etc. Both approaches to the measurement of skills, ratings by objective judges and perceptions from subjective judges, have been validated in studies showing a positive relationship between the level of skills displayed by a therapist and outcome of psychotherapy (Barrett-Lennard, 1962; Hansen, Moore, & Carkhuff, 1968; Mullen & Abeles, 1971; Rogers, 1967; Truax & Carkhuff, 1967). However, the two approaches have not been

found to correlate highly, and studies attempting to compare the relative validities of the two approaches yielded discrepant findings (Bozarth & Grace, 1970; Caracena & Vicory, 1969; Carkhuff & Berenson, 1967; Fish, 1970; Hansen, Moore, & Carkhuff, 1968; Hill & King, 1976; Kiesler, 1966; Kurtz & Grummon, 1972; McWhirter, 1973; Truax, 1966a; Welkowitz & Kuc, 1973). At present there is no conclusive evidence that either approach is more valid.

Each approach offers certain advantages as well as disadvantages. When objective judges complete rating scales on the basis of audiotapes, the data are provided by persons trained in the discrimination of levels of interpersonal communication skills. Since these judges are in no way involved in the dyadic relationship, their assessments should indeed be relatively objective. However, the basis for ratings, brief excerpts from audiotaped sessions, precludes the opportunity to use nonverbal information and limits observations to a relatively small time sampling of the entire relationship. These disadvantages are overcome when the data are provided by subjective judges, such as clients or helpees, who record their perceptions of skills displayed by another on instruments such as the Relationship Inventory (RI). However, the advantages are also lost in that the data are now provided by persons who are subjectively involved and who are not trained to discriminate levels of interpersonal communication skills.

A course directed by Gerard Egan at Loyola University of Chicago provided an opportunity to combine elements of the two approaches in an advantageous manner. The major component of this course, which is titled "Interpersonal Relations: An Experiential Approach," is a small group experience focusing on training in the various skills of helping and

human relating. The trainers or facilitators for these groups are persons who have demonstrated proficiency in helping and human relating skills in previous group experiences, and thus should be similar to objective judges in their ability to discriminate levels of skills displayed by others. Yet the trainers function not only as leaders but as members, pursuing the same contractual goals of interpersonal growth as the non-leader group members. Accordingly, the trainers are often in the role of helpee being helped by the non-leader group members. It was thus possible to have the trainers evaluate the helping skills displayed by the non-leader group members by responding as helpees on the RI. Since both males and females were represented among trainers and other group members, the possibility of gender differences in levels of communication skills was explored. Previous research has suggested that females are more empathic than males (Abramowitz, Abramowitz, & Weitz, 1976), but that there may also be an interaction effect, with individuals being more empathic when relating to persons of the same rather than opposite sex (Breisinger, 1976; Olesker & Balter, 1972).

The subjects for this study were 20 male and 41 female students taking the aforementioned undergraduate psychology course in interpersonal relations. The 61 students were divided into 12 groups, each consisting of from 4 to 7 students who were led by two co-trainers. Since three persons co-trained in two separate groups, there were 11 male and 10 female co-trainers for the 12 groups. Each pair of co-trainers consisted of one male and one female, with the exception of one group in which both trainers were male.

Immediately prior to the 3rd of the 14 group sessions, co-trainers and students (non-leader group members) were given a maximum of 15

minutes to complete the SST in a classroom setting. Following the 3rd session, co-trainers independently completed a 35-item modification of the Barrett-Lennard Relationship Inventory (RI) for each of the students in their groups, responding in terms of themselves as helpees and the students as helpers. The RI provides a total skills score as well as scores on Regard, Empathy, Congruence, and Unconditionality of Regard. For a post-treatment measure, co-trainers again completed RI's between the 13th and final group sessions. Following the last group session, co-trainers made hostility rankings of the students in their groups. Upon completion of formal data collection, 7 of the 21 co-trainers were briefly interviewed to obtain feedback regarding their experiences in completing RI's.

Alpha coefficients of internal stability for the 35-item RI used in this study ranged from .76 to .97, which compares favorably to previous findings (Lin, 1973; Wiebe & Pearce, 1973) on the original, 85-item inventory (Barrett-Lennard, 1962). It was hypothesized that males would show greater hostility than females on both the SST and on hostility rankings by co-trainers. The results were in the predicted direction but the differences were not significant. It was also hypothesized that students would be perceived as more hostile by co-trainers of the opposite rather than same sex. The differences were again in the predicted direction but were small and nonsignificant. Co-trainers showed only moderate agreement ($r = .35$) in their hostility rankings, and the hostility rankings were essentially unrelated ($r = -.05$) to hostility as measured by the SST. The hypothesis that hostility is negatively related to the learning of interpersonal communication skills (regard, empathy, congruence, and unconditionality of regard) was not

supported. Females were consistently perceived as displaying higher levels of skills at both pre-treatment and post-treatment, but none of the differences were significant. There was some support, particularly for males, for the hypothesis that individuals are rated higher on communication skills when the ratings are given by a person of the same rather than opposite sex. The responses given by co-trainers on the RI showed moderate agreement but were found to be negatively related to the hostility scores of the co-trainers on the SST.

In each of the 11 samples examined by Costin (1969, 1970, 1975), males showed significantly greater hostility on the SST than did females. The direction of the results in the present study were consistent with Costin's findings, but the differences were nonsignificant and smaller than any of those reported by Costin. One possible explanation for the discrepancy between the present findings and those cited by Costin is sampling differences. The subjects for the present study were persons who registered for a course in interpersonal relations, and these individuals may be less affected by cultural, gender-related stereotypes regarding the expression of hostility than the students in more traditional courses studied by Costin. Also, the mean age of both the student group members (25.7) and the co-trainers (30.0) is probably older than that in Costin's (1969, 1970, 1975) studies. In the present study, there was a nonsignificant correlation of $-.15$ between age and hostility as measured by the SST. Given the assumption that American culture allows males to express hostility in more direct ways than females are allowed, the tendency for males to score higher on measures of hostility may be an artifact of the greater difficulty in assessing more subtle manifestations of hostility. The validation of hostility

instruments in general has been problematical (Rabinowitz, 1975), and in the present study co-trainers showed only moderate agreement ($r = .35$) in their perceptions of hostility, with essentially no agreement ($r = -.05$) between hostility scores on the SST and hostility rankings by co-trainers. A crucial problem in any measurement of hostility is that the description of an act or a person as hostile must depend on not only the behavior itself but also the intent of the actor.

The lack of a negative relation between learning and hostility in the present study suggests possible explanations for the dynamics underlying Costin's (1970, 1971) observation that hostility interferes with learning. Costin studied academic learning in traditional lecture course, while the present study examined the learning of interpersonal communication skills within the context of a training group. In a traditional lecture course, a student may act out his hostility by absenteeism, lack of studying, generalized withdrawal, etc. That is, the failure to learn may reflect not only interference due to hostility but a volitional acting out of hostile impulses. Within an experiential skills training group, group pressure and support may inhibit these forms of hostile behaviors. It may be that hostility did not interfere with learning because the group setting fostered the expression of hostility in more appropriate and constructive ways.

REFERENCES

- Abramowitz, C. V., Abramowitz, S. I., & Weitz, L. J. Are men therapists soft on empathy? Two studies in feminine understanding. Journal of Clinical Psychology, 1976, 32, 434-437.
- Arbes, B. H., & Hubbell, R. N. Packaged impact: A structured communication skills workshop. Journal of Counseling Psychology, 1973, 20, 332-337.
- Archer, J., & Kagan, N. Teaching interpersonal relationship skills on campus: A pyramid approach. Journal of Counseling Psychology, 1973, 20, 535-540.
- Barrett-Lennard, G. T. Dimensions of therapists' responses as causal factors in therapeutic change. Psychological Monographs, 1962, 76, (43, Whole No. 562).
- Blaas, C. D., & Heck, E. J. Accuracy of accurate empathy ratings. Journal of Counseling Psychology, 1975, 22, 243-246.
- Bozarth, J. D., & Grace, D. P. Objective ratings and client perceptions of therapeutic conditions with university counseling center clients. Journal of Clinical Psychology, 1970, 26, 117-118.
- Breisinger, G. D. Sex and empathy, reexamined. Journal of Counseling Psychology, 1976, 23, 289-290.
- 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, 240-241.

- Campbell, R. J., Kagan, N., & Krathwohl, D. R. The development and validation of a scale to measure affective sensitivity (empathy). Journal of Counseling Psychology, 1971, 18, 407-412.
- Cannon, J. R., & Carkhuff, R. R. Effects of rater level of functioning and experience upon the discrimination of facilitative conditions. Journal of Consulting and Clinical Psychology, 1969, 33, 189-194.
- Caracena, P. F., & Vicory, J. R. Correlates of phenomenological and judged empathy. Journal of Counseling Psychology, 1969, 16, 510-515.
- Carkhuff, R. R. Helping and human relations. Vol. 1: Selection and training. New York: Holt, Rinehart, and Winston, 1969a.
- Carkhuff, R. R. Helping and human relations. Vol. 2: Practice and research. New York: Holt, Rinehart, and Winston, 1969b.
- Carkhuff, R. R., & Berenson, B. G. Beyond counseling and therapy. New York: Holt, Rinehart, and Winston, 1967.
- Carkhuff, R. R., & Berenson, B. G. Teaching as treatment: An introduction to counseling and psychotherapy. Amherst, Massachusetts: Human Resource Development Press, 1976.
- Carkhuff, R. R., & Burstein, J. W. Objective therapist and client ratings of therapist-offered facilitative conditions of moderate to low functioning therapists. Journal of Clinical Psychology, 1970, 26, 394-395.
- Carkhuff, R. R., & Truax, C. B. Training in counseling and psychotherapy: An evaluation of an integrated didactic and experiential approach. Journal of Consulting Psychology, 1965, 29, 333-336.
- Cartwright, R. D. A comparison of the response to psychoanalytic and client-centered psychotherapy. In L. A. Gottschalk & A. H. Auerbach (Eds.), Methods of research in psychotherapy. New York: Appleton-

Century-Crofts, 1966.

Chaplin, J. P. Dictionary of psychology. New York: Dell, 1968.

Chinsky, J. M., & Rappaport, J. Brief critique of the meaning and reliability of "accurate empathy" ratings. Psychological Bulletin, 1970, 73, 379-382.

Cohen, C. P., Johnson, D. L., & Hanson, P. G. Interpersonal changes among psychiatric patients in human relations training. Journal of Personality Assessment, 1971, 35, 472-479.

Costin, F. The Scrambled Sentence Test: A group measure of hostility. Educational and Psychological Measurement, 1969, 29, 461-468.

Costin, F. Hostility and learning in an introductory psychology course. Psychology in the Schools, 1970, 7, 370-374.

Costin, F. Hostility and learning: A follow-up note. Educational and Psychological Measurement, 1971, 31, 1015-1017.

Costin, F. Measuring hostility with the Scrambled Sentence Test. In C. K. Verma & C. Bagley (Eds.), Race and education across cultures. London: Heinemann, 1975.

Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. Journal of Consulting Psychology, 1960, 24, 349-354.

Danish, S. J. Factors influencing changes in empathy following a group experience. Journal of Counseling Psychology, 1971, 18, 262-267.

Danish, S. J., & Kagan, N. Measurement of affective sensitivity: Toward a valid measure of interpersonal perception. Journal of Counseling Psychology, 1971, 18, 51-54.

D'Augelli, A. R., Deyes, C. S., Guerney, B. G., Hershenberg, B., &

- Sborofsky, S. L. Interpersonal skill training for dating couples: An evaluation of an educational mental health service. Journal of Counseling Psychology, 1974, 21, 395-389.
- Dell, D. M. Counselor power base, influence attempt, and behavior change in counseling. Journal of Counseling Psychology, 1973, 20, 339-405.
- Egan, G. Encounter: Group processes for interpersonal growth. Belmont, California: Brooks/Cole, 1970.
- Egan, G. Encounter groups: Basic readings. Monterey, California: Brooks/Cole, 1971.
- Egan, G. Face to face: The small-group experience and interpersonal growth. Monterey, California: Brooks/Cole, 1973.
- Egan, G. The skilled helper: A model for systematic helping and interpersonal relating. Monterey, California: Brooks/Cole, 1975.
- Egan, G. Interpersonal living: A skills/contract approach to human relations training in groups. Monterey, California: Brooks/Cole, 1976.
- Fischer, J., Paveza, G. J., Kickerty, N. S., Hubbard, L. J., & Grayston, S. B. The relationship between theoretical orientation and therapists' empathy, warmth, and genuineness. Journal of Counseling Psychology, 1975, 22, 399-403.
- Fish, J. M. Empathy and the reported emotional experiences of beginning psychotherapists. Journal of Consulting and Clinical Psychology, 1970, 35, 64-69.
- Frank, J. Persuasion and healing (Rev. ed.). Baltimore: The Johns Hopkins University Press, 1973.
- Frankel, M. Effects of videotape modeling and self-confrontation technique on microcounseling behavior. Journal of Counseling Psychology, 1971, 18, 465-471.

- Goldman, M. Alternative classroom standards concerning management of hostility and effects on student learning. Dissertation Abstracts, 1955, 15, 2098-2099.
- Gormally, J., & Hill, C. E. Guidelines for research on Carkhuff's training model. Journal of Counseling Psychology, 1974, 21, 539-547.
- Gottschalk, L. A., & Auerbach, A. H. (Eds.). Methods of research in psychotherapy. New York: Appleton-Century-Crofts, 1966.
- Green, R. T., & Stacey, B. G. The development of a questionnaire measure of hostility and aggression. Acta Psychologica, 1967, 26, 265-285.
- Gross, W. F., & DeRidder, L. M. Significant movement in comparatively short-term counseling. Journal of Counseling Psychology, 1966, 13, 98-99.
- Haase, R. F., & Tepper, D. T. Nonverbal components of empathic communication. Journal of Counseling Psychology, 1972, 19, 417-424.
- Hansen, J. C., Moore, C. D., & Carkhuff, R. R. The differential relationships of objective and client perceptions of counseling. Journal of Clinical Psychology, 1968, 24, 244-246.
- Hekmat, H., Khajavi, F., & Mehryar, A. H. Some personality correlates of empathy. Journal of Consulting and Clinical Psychology, 1975, 43, 89.
- Hill, C. E., & King, J. Perceptions of empathy as a function of the measurement instrument. Journal of Counseling Psychology, 1976, 23, 155-157.
- Hogan, R. Development of an empathy scale. Journal of Consulting and Clinical Psychology, 1969, 33, 307-316.
- Houtras, P. T., & Anderson, D. L. Counselor conditions for self-explora-

- tion of college students. Personnel and Guidance Journal, 1969, 48, 45-48.
- Ivey, A. E., Normington, C. J., Miller, C. D., Morrill, W. H., & Haase, R. F. Microcounseling and attending behavior: An approach to pre-practicum counselor training. Journal of Counseling Psychology Monograph, 1968, 15 (5, Pt. 2).
- Jansen, D. C., Robb, G. P., & Bonk, E. C. Peer ratings of practicum counselors. Journal of Counseling Psychology, 1972, 19, 333-339.
- Johnson, D. L., Hanson, P. G., Rothaus, P., Morton, R. B., Lyle, F. A., & Moyer, R. Follow-up evaluation of human relations training for psychiatric patients. In E. H. Schein & W. G. Bennis (Eds.). Personal and organizational change through group methods: The laboratory approach. New York: Wiley, 1965.
- Kagan, N., Krathwohl, D. R., Goldberg, A. D., Campbell, R. J., Schauble, P. G., Greenberg, B. S., Danish, S. J., Resnikoff, A., Bowes, J., & Bondy, S. B. Studies in human interaction: Interpersonal process recall stimulated by videotape. Ann Arbor: Michigan State University, Educational Publication Services, 1967.
- Kamin, I., & Caughlin, J. Subjective experience of out-patient psychotherapy. American Journal of Psychotherapy, 1963, 17, 660-668.
- Kaufmann, H. Aggression and altruism. New York: Holt, Rinehart, & Winston, 1970.
- Kiesler, D. J. Basic methodologic issues implicit in psychotherapy research. American Journal of Psychotherapy, 1966, 20, 135-155.
- 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.

- Lanning, W. L. A study of the relation between group and individual counseling supervision and three relationship measures. Journal of Counseling Psychology, 1971, 18, 401-406.
- Latane, B., & Arrowood, A. J. Emotional arousal and task performance. Journal of Applied Psychology, 1963, 47, 324-327.
- Leary, T., & Harvey, J. S. A methodology for measuring personality changes in psychotherapy. Journal of Clinical Psychology, 1956, 12, 123-132.
- Lieberman, L. R. Effects upon anxiety, depression, and hostility of postponement of an examination. Perceptual and Motor Skills, 1966, 23, 1051-1054.
- Lin, Tien-Teh. Counseling relationship as a function of counselor's self-confidence. Journal of Counseling Psychology, 1973, 20, 293-297.
- Martin, P. L., & Toomey, T. C. Perceptual orientation and empathy. Journal of Consulting and Clinical Psychology, 1973, 41, 313.
- McWhirter, J. J. Two measures of the facilitative conditions: A correlation study. Journal of Counseling Psychology, 1973, 20, 317-320.
- McWhirter, J. J., & Marks, S. E. An investigation of the relationship between the facilitative conditions and peer and group leader ratings of perceived counseling effectiveness. Journal of Clinical Psychology, 1972, 28, 116-117.
- Meltzoff, J., & Kornreich, M. Research in psychotherapy. New York: Atherton Press, 1970.
- Mills, D. H., & Zytowski, D. G. Helping relationship: A structural analysis. Journal of Counseling Psychology, 1967, 14, 193-197.
- Morton, R. B. The patient training laboratory: An adaptation of the

- instrumented training laboratory. In E. H. Schein & W. C. Bennis (Eds.), Personal and organizational change through group methods: The laboratory approach. New York: Wiley, 1965.
- Muehlberg, N., Pierce, R., & Drasgow, J. A factor analysis of therapeutically facilitative conditions. Journal of Clinical Psychology, 1969, 25, 93-95.
- Mullen, J., & Abeles, N. Relationship of liking, empathy, and therapist's experience to outcome of therapy. Journal of Counseling Psychology, 1971, 18, 39-45.
- Olesker, W., & Balter, L. Sex and empathy. Journal of Counseling Psychology, 1972, 19, 599-562.
- Phillips, B. N. Authoritarian, hostile, and anxious students' ratings of an instructor. California Journal of Educational Research, 1960, 11, 19-23.
- Pierce, R., & Drasgow, J. Teaching facilitative interpersonal functioning to psychiatric inpatients. Journal of Counseling Psychology, 1969, 16, 295-299.
- Pierce, R. M., & Schauble, P. G. Graduate training of facilitative counselors: The effects of individual supervision. Journal of Counseling Psychology, 1970, 17, 210-215.
- Pierce, W. D., & Mosher, D. C. Perceived empathy, interviewer behavior, and interviewee anxiety. Journal of Consulting Psychology, 1967, 31, 101.
- Rabinowitz, A. Hostility measurement and its relationship to fantasy capacity. Journal of Personality Assessment, 1975, 39, 50-54.
- Rappaport, J., & Chinsky, J. M. Accurate empathy: Confusion of a construct. Psychological Bulletin, 1972, 77, 400-404.

- Rogers, C. R. The necessary and sufficient conditions of therapeutic personality change. Journal of Consulting Psychology, 1957, 21, 95-103.
- Rogers, C. R. (Ed.). The therapeutic relationship and its impact: A study of psychotherapy with schizophrenics. Madison: University of Wisconsin Press, 1967.
- Rosenthal, R. Estimating effective reliabilities in studies that employ judges' ratings. Journal of Clinical Psychology, 1973, 29, 342-345.
- Rothaus, P., Morton, R. B., Johnson, D. L., Cleveland, S. E., & Lyle, F. A. Human relations training for psychiatric patients. Archives of General Psychiatry, 1963, 8, 572-581.
- Sarason, I. G. Intercorrelations among measures of hostility. Journal of Clinical Psychology, 1961, 17, 192-195.
- Schoenberg, B., & Carr, A. C. An investigation of criteria for brief psychotherapy of neurodermatitis. Psychosomatic Medicine, 1963, 25, 253-263.
- Shapiro, J. G. Relationship between expert and neophyte ratings of therapeutic conditions. Journal of Consulting and Clinical Psychology, 1968, 32, 87-89.
- Shapiro, J. G., Krauss, H. H., & Truax, C. B. Therapeutic conditions and disclosure beyond the therapeutic encounter. Journal of Counseling Psychology, 1969, 16, 290-294.
- Strong, S. R. Counseling: An interpersonal influence process. Journal of Counseling Psychology, 1968, 15, 215-224.
- Strong, S. R., & Dixon, D. N. Expertness, attractiveness, and influence in counseling. Journal of Counseling Psychology, 1971, 18, 562-570.

- Sullivan, H. S. The interpersonal theory of psychiatry. New York: Norton, 1953.
- Tedeschi, J. T., Smith, R. B., & Brown, R. C. A reinterpretation of research on aggression. Psychological Bulletin, 1974, 81, 540-562.
- Truax, C. B. Therapist empathy, warmth, and genuineness and patient personality change in group psychotherapy: A comparison between interaction unit measures, time sample measures, and patient perception measures. Journal of Clinical Psychology, 1966, 22, 225-229. (a)
- Truax, C. B. Influence of patient statements on judgments of therapist statements during psychotherapy. Journal of Clinical Psychology, 1966, 22, 335-337. (b)
- Truax, C. B. The meaning and reliability of accurate empathy ratings: A rejoinder. Psychological Bulletin, 1972, 77, 397-399.
- Truax, C. B., & Carkhuff, R. R. Toward effective counseling and psychotherapy: Training and practice. Chicago: Aldine, 1967.
- Walker, B. S., & Little, D. F. Factor analysis of the Barrett-Lennard Relationship Inventory. Journal of Counseling Psychology, 1969, 16, 516-521.
- Watson, R. E., Pritzker, L., & Madison, P. Hostility in neurotics and normals. Journal of Abnormal and Social Psychology, 1955, 50, 36-40.
- Welkowitz, J., & Kuc, M. Interrelationships among warmth, genuineness, empathy, and temporal speech patterns in interpersonal interaction. Journal of Consulting and Clinical Psychology, 1973, 41, 472-473.
- Wiebe, B., & Pearce, W. B. An item-analysis and revision of the Barrett-Lennard Relationship Inventory. Journal of Clinical Psychology,

1973, 29, 495-497.

Wiggins, J. S. Substantive dimensions of self-report in the MMPI item pool. Psychological Monographs, 1966, 80 (22, Whole No. 630).

Wiggins, J. S., Goldberg, L. R., & Applebaum, M. MMPI content scales: Interpretative norms and correlations with other scales. Journal of Consulting and Clinical Psychology, 1971, 37, 403-410.

Wolff, W. T., & Merrens, M. R. Behavioral assessment: A review of clinical methods. Journal of Personality Assessment, 1974, 38, 3-16.

Zimmer, J. M., & Anderson, S. Dimensions of positive regard and empathy. Journal of Counseling Psychology, 1968, 15, 417-426.

APPENDIX A

SST

This is a test of how people perceive word relationships according to their first immediate impression. It consists of sets of words which are in a "scrambled" order. By underlining three words in a set, you can make a complete sentence. Here is an example:

your hand head raise

You can make a complete sentence from these scrambled words by underlining three words as follows:

your hand head raise (or) your hand head raise

Simply underline three words which make a complete sentence. DO THIS ACCORDING TO YOUR FIRST IMPRESSION. YOU MAY CHOOSE ANY COMBINATION OF 3 WORDS YOU WISH, AS LONG AS YOU MAKE A COMPLETE SENTENCE.

Here are two more sets of scrambled words for practice. Underline the three words in each set which make a complete sentence:

close swing door the

you know see I

Now turn the page and begin. WORK RAPIDLY. UNDERLINE WORDS ACCORDING TO YOUR FIRST IMPRESSION. Please print your name below and also on top of the following page. If you wish, you may print your first name and the first four letters of your last name. Names will only be used for statistical reasons, and all data will be kept confidential. We are not interested in individual performances, but in overall patterns.

NAME:

SEX:

NAME:

- | | |
|------------------------------|------------------------------|
| 1. see the hear bee | 24. her found knew I |
| 2. me knows him he | 25. fall him in let |
| 3. cloth sell tear the | 26. the book close read |
| 4. out sort bawl them | 27. teacher call the accuse |
| 5. present he away is | 28. eggs buy some fry |
| 6. mend the sort clothes | 29. go hell sleep to |
| 7. me you disgust convince | 30. iron the pack clothes |
| 8. the taste make tea | 31. ruin page read the |
| 9. out him let knock | 32. songs again sing it |
| 10. winter has he gone | 33. missed I her most |
| 11. see now it hear | 34. grapes the grow crush |
| 12. grow the choose apples | 35. the rake repair leaves |
| 13. clever happy she is | 36. spring is he here |
| 14. saves time she money | 37. thief a woman she's |
| 15. destroy book examine the | 38. find money keep the |
| 16. meet mother his believe | 39. a write read story |
| 17. the work begin finish | 40. seam find rip the |
| 18. now hire fire them | 41. you me know I |
| 19. him go me let | 42. earns wages she respect |
| 20. orange the taste smash | 43. his trick meet father |
| 21. cry him in let | 44. her go him let |
| 22. boy a man he's | 45. his deceive greet father |
| 23. the swing fix axe | 46. give up don't I'll |

(turn to next page)

47. hear I you hate
48. some enjoy buy coffee
49. find key use the
50. all damn find them
51. cards again play it
52. down him carry push
53. time wastes takes studying
54. cake the bake take
55. a read recite poem
56. teacher greet the blame
57. son a father he's
58. the plunge sharpen knife
59. hear I'll you hit
60. the door open fix
61. me you sicken tell
62. watch grow eat it
63. liar a woman she's
64. feed the train dog
65. life pay take her
66. pick the eat peaches
67. time waste take exams
68. find mother his leave
69. hope don't I'll lose
70. girl suit hang the

APPENDIX B

RELATIONSHIP INVENTORY (HELPEE FORM)

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 an individual group member. 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.

Please write your name along with the name of the group member regarding whom you are responding. If you wish, you may use first names along with the first four letters of last names. Names are needed only for statistical evaluation, and we are not interested in individual relationships but rather in overall results. All results will be confidential. If you are completing more than one inventory (responding in relation to more than one group member), please be sure to also provide names on other inventories.

* Your name:

* Name of other group member:

His/her sex:

NOTE: Please fill out this inventory in terms of you as helpee and the other as helper. That is, try to confine your responses to this aspect of your relationship. Try to be as honest and accurate as possible. Individual ratings will only be known to the experimenter, who is not connected in any way with those directing the course. Thus, individual ratings can have no effect on your grades or on the grades of others. Thus, accurate and honest discriminations should not produce harmful consequences.

RI

- ___ 1. He respects me.
- ___ 2. He pretends that he likes me or understands me more than he really does.
- ___ 7. He understands my words but not the way I feel.
- ___ 12. He is interested in knowing what my experiences mean to me.
- ___ 13. He is disturbed whenever I talk about or ask about certain things.
- ___ 16. He likes seeing me.
- ___ 23. He behaves just the way that he is, in our relationship.
- ___ 26. He appreciates me.
- ___ 28. I do not think that he hides anything from himself that he feels with me.
- ___ 34. If I feel negatively toward him he responds negatively to me.
- ___ 36. He cares about me.
- ___ 37. His own attitudes toward some of the things I say, or do, stop him from really understanding me.
- ___ 43. I feel that I can trust him to be honest with me.
- ___ 44. Sometimes he is warmly responsive to me, at other times cold or disapproving.
- ___ 46. He is interested in me.
- ___ 47. He appreciates what my experiences feel like to me.
- ___ 49. Depending on his mood, he sometimes responds to me with quite a lot more warmth and interest than he does at other times.
- ___ 56. He does not really care what happens to me.
- ___ 57. He does not realize how strongly I feel about some of the things we discuss.
- ___ 58. There are times when I feel that his outward response is quite different from his inner reaction to me.

(continued)

RI

- _____ 59. His general feeling toward me varies considerably.
- _____ 61. He seems to really value me.
- _____ 63. I don't think that he is being honest with himself about the way he feels toward me.
- _____ 68. I feel that he is being genuine with me.
- _____ 69. Sometimes he responds quite positively to me, at other times he seems indifferent.
- _____ 73. Sometimes he is not at all comfortable but we go on, outwardly ignoring it.
- _____ 76. He feels deep affection for me.
- _____ 77. He usually understands all of what I say to him.
- _____ 78. He does not try to mislead me about his own thoughts or feelings.
- _____ 81. He regards me as a disagreeable person.
- _____ 86. At times he feels contempt for me.
- _____ 87. When I do not say what I mean at all clearly he still understands me.
- _____ 88. Sometimes he thinks that I feel a certain way, because he feels that way.
- _____ 89. He responds to me mechanically.
- _____ 92. He can be deeply and fully aware of my most painful feelings without being distressed or burdened by them himself.

APPENDIX C

HOSTILITY RANKINGS

Hostility can be described as a desire or a tendency to make other persons undergo negative experiences. It would be impossible to list all the ways that hostility can be expressed in training groups, but some possible examples are: negative evaluations, non-caring confrontation, condescending support, silence and withdrawal, exclusively adopting either the helper or helpee role, and attempting to obstruct the group from meeting its contractual goals. Again, these are only examples. The described behaviors do not necessarily reflect hostility, nor are these the only ways in which hostility can be expressed.

On the basis of your observations in your role as group trainer, would you please rank the members in your group on hostility. Do NOT include yourself or your co-trainer. Depending on the number of persons you have to rank, you may not need all of the spaces provided.

Your name: _____

<u>Most</u> hostile	↑	1. _____
		2. _____
		3. _____
		4. _____
		5. _____
		6. _____
<u>Least</u> hostile	↓	7. _____

APPROVAL SHEET

The dissertation submitted by John Dalton has been read and approved by the following Committee:

Dr. Frank Kobler, Chairman
Professor, Psychology, Loyola

Dr. James Johnson
Associate Professor, Psychology, Loyola

Rev. Gerard Egan
Associate Professor, Psychology, Loyola

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval by the Committee with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

January 10, 1977
Date

Frank Kobler
Director's Signature