

Journal of Social and Clinical Psychology, Vol. 24, No. 2, 2005, pp. 156-171

## INCREASING EMPATHIC ACCURACY THROUGH PRACTICE AND FEEDBACK IN A CLINICAL INTERVIEWING COURSE

DAVID F. BARONE  
*Illinois State University*

PHILINDA S. HUTCHINGS  
*Argosy University/Phoenix*

HEATHER J. KIMMEL, HOWARD L. TRAUB, JOAN T. COOPER,  
AND CHRISTINE M. MARSHALL  
*Nova Southeastern University*

Accurate empathy, long argued to be important in psychotherapy, now is an object of social-cognitive research. Graduate-level psychology students viewed brief portions of a therapy session and inferred the thoughts and feelings of the client. Accuracy scores were the rated similarity of their inferences to the client's reported thoughts and feelings. Throughout the semester course in interviewing, experimental participants practiced such judgments with feedback, while controls did not. Both groups' accuracy increased from pre- to post-test on inferred feelings, in part because the post-test was easier. Nonetheless, experimental participants on the post-test had greater accuracy of inferred feelings than controls. Women were more accurate than men in inferences for thoughts at post-test. Implications for training and future research are discussed.

Empathy is a factor common to most theoretical orientations in clinical/counseling psychology as a means of increasing practitioners' effectiveness (Truax & Carkhuff, 1967). Empathy usually is conceptualized as a skill that can be developed and learned if properly instructed. The

---

The authors wish to acknowledge the statistical consultation provided by Edward Simco, Nova Southeastern University, and Matthew Hesson-McInnis, Illinois State University. @FtnText = The authors would also like to thank two anonymous reviewers for suggesting the methodological follow-up study.

Correspondence should be addressed to David F. Barone, Department of Psychology, Campus Box 4620, Illinois State University, Normal, IL 61790; E-mail: [dbaron@ilstu.edu](mailto:dbaron@ilstu.edu).

construct of empathy has been discussed in psychology for over 125 years, the English term having been coined by experimental psychologist Edward Titchener (Wispe, 1987). Sigmund Freud recognized the importance of empathy in establishing a strong working alliance and conducting successful analysis (Pigman, 1995). Carl Rogers (1959) defined empathy as the ability “to perceive the internal frame of reference of another *with accuracy*, and with the emotional components and meanings which pertain thereto” (p. 210, emphasis added). Rogers (1957, 1975) argued for empathy being “one of the most potent factors in bringing about change and learning” (1975, p. 3).

Many psychologists have explicated empathy as a process involving cognitive and social processes (Davis, 1983; Marks & Tolsma, 1986; Riggio, Tucker, & Coffaro, 1989). Reik (1949; as cited in Marcia, 1987) conceptualized empathy as having four distinct stages: identification, incorporation, reverberation, and detachment. Later, Barrett-Lennard (1981) proposed that there are three cyclical phases in an empathic interaction: inner understanding, communication or expressed understanding, and received empathy. Dymond (1949) asserted that one can more accurately understand another’s thoughts and feelings through cognitive means and imaginative role-taking. She developed a scale of individual differences in the ability to do so. Stotland (1969) argued that empathy not only involves the recognition of another’s feelings on a cognitive level, but also includes an emotional responsiveness and a sharing of one’s feelings. Gladstein (1977) added a third aspect that combined the cognitive and the affective.

Researchers began to develop and evaluate individual-differences measures of empathy. The Truax Relationship Inventory (Truax & Carkhuff, 1967) and the Barrett-Lennard Relationship Inventory (Barrett-Lennard, 1962) measure the counselor’s or the client’s judgment of empathy in a session. Truax and Carkhuff’s (1967) Accurate Empathy Scale and Carkhuff’s (1969) Empathic Understanding in Interpersonal Process Scale require observers to judge empathy in counseling sessions. A review of these instruments revealed that none serves to assess all aspects of empathic ability (Feldstein & Gladstein, 1980). Other measures of empathy include the Lister Empathy Scale (Hargrove, 1974) and the Response Empathy Rating Scale (Elliot et al., 1982). The Hogan Empathy Scale (HES; Hogan, 1969) and the Questionnaire Measure of Emotional Empathy (QMEE; Mehrabian & Epstein, 1972) are two other widely used self-report measures of empathy. Davis (1980) developed the Interpersonal Reactivity Index (IRI), a more complex measure of four aspects of empathy: perspective-taking, fantasy, empathic concern, and personal distress. His results support the multidimensionality of the empathy construct and its relation to social functioning, self-esteem,

emotionality, and sensitivity to others (Davis, 1983). A later study of the HES, QMEE, IRI, and empathy indices from the Social Skills Inventory showed that empathy and social skills overlap, which suggests that social skills may be another facet of empathy (Riggio et al., 1989).

Recent attention has been focused on teaching empathy as a learnable skill that can aid in more accurately understanding an individual's experience and that is essential in counseling relationships (Barak, Engle, Katzir, & Fisher, 1987). Despite challenges concerning whether empathy can be taught (Carkhuff & Berenson, 1967; Davis, 1980; Hatcher et al., 1994; Mehrabian & Epstein, 1972), empathy training typically attempts to teach two skills: understanding another's thoughts and feelings and communicating this understanding to a client (Truax & Carkhuff, 1967; Carkhuff, 1969). Researchers have studied a variety of techniques for teaching empathy, including verbal instruction (Shaffer & Hummel, 1979), role-playing (Guzzetta, 1976; Stone & Vance, 1976), videotape (Dalton & Sunblad, 1976; Eisenberg & Delaney, 1970), audiotape (Perry, 1975), written and taped patient session transcripts (Finke, 1990), communication skills training (Evans, Stanley, & Burrows, 1993), a Rogerian-based curriculum in peer facilitation training (Hatcher et al., 1994), peer mediation (Lane-Garon, 1998), and an empathy training game that focuses on competition and enjoyment to increase trainee motivation and interest (Barak et al., 1987). The lack of agreement on the definition of empathy and a reliable way to measure it has hindered the adoption of conclusions from this research about how best to train empathy (Evans et al., 1993).

Recent advances in social-personality psychology provide a promising new approach to the study of empathy. Ickes, Robertson, Tooke, and Teng (1986) created a research paradigm to study empathic accuracy. Dyads briefly interact; immediately afterwards and in separate rooms, they watch a videotape of the interaction and stop it to record thoughts and feelings they experienced during the interaction. These stop-points and the target-supplied thoughts and feelings provide a series of items and correct answers for a test of judgments of accurate empathy. After observers watch the tape and infer the target's thoughts and feelings at the stop-points, raters score the similarity between the observers' and target's responses. Ickes, Stinson, Bissonnette, and Garcia (1990) found that this performance test of empathic accuracy in mixed-sex dyads did not correlate with participants' self-reports of empathic accuracy. This finding may help to explain the elusiveness of the empathy construct, which typically has been measured by self-report (Ickes, 1993, 1997). In addition, Ickes, Stinson, et al. (1990) found no gender differences, another finding later replicated in multiple studies (Hancock & Ickes, 1996; Marangoni, Garcia, Ickes, & Teng, 1995; Riggio et al., 1989).

Ickes' methodology has been used to study empathic accuracy for therapy clients (Marangoni et al., 1995). Using videotapes of simulated therapy sessions, the authors were able to show that individual differences in empathic accuracy remain consistent across targets, that empathic accuracy improves with continued exposure to individual targets, and that feedback increases the rate at which it improves. That study also revealed that some targets are less readable than others, which suggests that tests of accurate empathy can include items that vary in difficulty. Another study with the same stimuli showed that order of presentation of the targets and of individual segments within targets had no effect (Gesn & Ickes, 1999). Furthermore, Hancock and Ickes (1996) found that readability of the target produced differences in empathic accuracy while the relationship of the perceiver to target did not.

In the current study, we used Ickes' paradigm to evaluate a method of teaching empathic accuracy in a graduate psychology program. We have modified the technique to allow for group testing and training. Taking a step beyond previous research on empathic accuracy in a clinical setting (Marangoni et al., 1995), we used as the test stimulus a videotape of an actual therapy session conducted by a doctoral-level psychology intern. Although the distinction between thoughts and feelings was found to be unimportant in previous research, we included the distinction because of its potential relevance in a clinical situation. Clinicians of several theoretical orientations draw a distinction between thought and emotion and attempt to influence one in an effort to change the other, such as in cognitive-behavioral therapies.

The experimental group practiced inferring thoughts and feelings during in-class role-plays throughout the semester. In some cases, students observed two other students role-playing therapy situations in front of the class. At other times, students were participants in dyad role-plays. Following the recording of thoughts and feelings, class members received criterion information from the targets and were encouraged to compare their inferred responses to the actual responses provided by each role-played "client." In the control group, standard role-playing of therapy situations took place; that is, without explicit recording of inferences and feedback on their accuracy. Based on the finding that feedback increases the rate of improvement of accuracy (Marangoni et al., 1995), we hypothesized that, at the end of the course, students in the experimental group would show significantly more improvement in empathic accuracy than those in the control group. We also were interested in evaluating gender differences but made no hypothesis because of conflicting findings in previous research on empathic accuracy.

## METHOD

### PARTICIPANTS

Participants in the study were 54 first-year graduate students in five interviewing course sections. All were enrolled in a Ph.D. or Psy.D. program in clinical psychology. Thirty-five were female and 18 were male (one did not indicate gender). The mean age for the sample was 25.2 years, with a range in age from 21 to 41 years. Although 72 students agreed to participate—31 in the experimental group and 41 in the control group—18 students failed to complete the post-test because they left the class session early—eight in the experimental group and 10 in the control group. Of the 54 participants who completed the study, 23 participants who were in the experimental group and 31 in the control group. There were no gender or age differences between the two groups, nor between those who dropped out of the two groups.

### STIMULUS TAPE

A client in a community mental health center being seen by an intern for individual therapy gave her informed consent to have one of her therapy sessions videotaped for use in this study. After the taped session, the intern therapist viewed the tape with the client and instructed her to stop the tape in places when she could remember what she was thinking and feeling during the session. Each time the tape was stopped, the client wrote down these thoughts and feelings and the place on the tape time-counter was recorded. Two 15-minute segments of the tape were shown to participants. In the first, used as the pre-test, there were nine places in which the client had recorded thoughts and feelings. In the second segment, used in the post-test, the tape was stopped and thoughts and feelings recorded in eight places.

### PROCEDURE

Participants took the pre-test during the third week of a 15-week semester and the post-test in the 14th week. They signed informed consent forms before doing the pre-test. A research assistant read instructions to the class using the following script for the pre-test:

We are asking you to participate in a study on empathic accuracy. We will be comparing your inferences regarding clients' and/or students' thoughts and feelings with the actual thoughts and feelings as recorded by clients and/or students who have been interviewed. Please read the consent form. If you agree to participate, we will play a segment of a

video of a client therapy session. We will stop the tape at several places and ask you to write on the assessment form what you think the client is thinking and feeling at that moment in a few words or phrases, with enough information to be complete and as specific as possible, but not lengthy or elaborate. For example, rather than write that the client was thinking about work, write that the client was thinking about how much work she had to do today. Rather than write that the client was upset, write that the client was angry-upset or sad-upset. We'll do this now and again at the end of the semester.

The participants, who had completed an ethics course the previous semester and begun practicum activities, agreed to keep client information anonymous and confidential. Participants were instructed to attend to the client, not to the therapist's interventions or therapy. Pre-test and post-test, the participants recorded an inference of thought and an inference of feeling each time the videotape was stopped. The form that participants were given had a space for recording a thought and a feeling for each tape segment.

### SCORING

Two research assistants, who were advanced graduate students in clinical psychology and blind to participants' condition, scored for accuracy. They made a qualitative judgment of the accuracy of each inference written down by participants and entered scores of 0, 1, or 2. These scores indicate that, compared to the thought or feeling reported by the client, the inference reported by the participant was judged to be essentially different (0), somewhat similar but not the same (1), or essentially the same content (2). To aid in reliability, qualitative decision rules were developed to guide which of the three judgments of accuracy were made (see Appendix). The raters divided up the responses so that the score entered into the analysis of accuracy for each participant response was that entered by one of the raters.

In order to establish inter-rater reliability, research assistants first rated the accuracy of the thoughts and feelings from videotapes of practice role-plays by other students not involved in the study. Because the numerical ratings are at best ordinal, rather than a psychometric equal-interval scale, the appropriate index of reliability is percent of agreement: that is, the percent of rating instances that both raters made the same judgment from the three possibilities. Note that this is a more conservative index of inter-rater reliability than a correlation coefficient, which gives "partial credit" for closer ratings (e.g., 1 and 2) than for more distant ones (e.g., 0 and 2). Participant responses for this study began be-

ing scored by one rater when inter-rater agreement for these two trained scorers reached 90%—a very stringent criterion. Midway through scoring, however, the raters checked their ratings on some common responses and found that their agreement had fallen to below 70% for thoughts. They revised the decision rules, reestablished their agreement at 90%, and then each went back and re-rated the responses he or she had rated. Because the raters also were serving as the experimenters and knew from the content of the responses whether they were pre-tests or post-tests, other raters who were blind to this information and untrained were brought in and briefly trained on the decision rules. Their levels of agreement for thoughts and feelings were similar, so the results of the trained raters were deemed unbiased.

#### TRAINING EXPERIENCE

Two of the five class sections of participants served as the experimental group and the remaining three sections formed the control group for the study. Each section was taught by a different professor but had the same course goals and included role-playing of interviewing skills. Typically involved was the assignment of some simulated therapy situation in which one student interviewed another playing the role of the patient. In addition, the experimental group practiced inferring thoughts and feelings during in-class role-plays and received immediate feedback on the accuracy of their inferences. The control group did not receive feedback. The number of practice times inferring thoughts and feelings varied in each class session for each participant.

#### RESULTS

Data were subjected to a multivariate analysis of variance for two within-subject factors—measure (feelings/thoughts) and test (pre/post)—and one between-subjects factor—group (experimental/control). Alpha was set at .05 and partial eta-squared was used for effect size.  $F(1, 52)$  values were significant for measure, (309.74,  $\eta^2 = .86$ ), test (8.62,  $\eta^2 = .14$ ), measure by test (10.91,  $\eta^2 = .17$ ), and condition, (5.28,  $\eta^2 = .09$ ), but not for the other interactions. Given the interaction, group results for pre-test and post-test are shown by measure in Figure 1 and Figure 2. Table 1 presents descriptive statistics. Feelings were more accurately inferred than thoughts ( $M = 1.00$  vs.  $.41$ ), post-test scores were higher than pre-test scores ( $M = .76$  vs.  $M = .66$ ), and experimental scores were higher than controls ( $M = .76$  vs.  $M = .67$ ). Pre- and post-test scores were uncorrelated except modestly for thoughts in the control condition. Targeted comparisons related to the hypothesis of the study were

TABLE 1. Empathic Accuracy by Group Before and After an Interviewing Course

Group	Thoughts					Feelings				
	Pre-test		Post-test			Pre-test		Post-test		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>r</i>
Experimental ( <i>n</i> = 24)	.42	.19	.43	.22	-.36*	.93	.28	1.24	.40	.02
Control ( <i>n</i> = 31)	.41	.18	.38	.18	.20	.88	.29	1.01	.24	.16

Note. Empathic accuracy was rated 0 (different), 1 (similar), or 2 (the same). Correlations are between pre-test and post-test. \* $p < .05$ .

made with *t*-tests ( $\alpha = .05$ ). Paired-samples two-tailed *t*-tests found a significant increase in accuracy from pre-test to post-test for inferred feelings, in the experimental group ( $t [22] = -2.93$ ) and a marginal one in the control group ( $t [30] = -2.10$ ), but no increase in the accuracy of inferred thoughts in either the experimental group ( $t [22] = -.28$ ) or the control group ( $t [30] = .48$ ). Independent-samples two-tailed *t*-tests found that, at post-test, accuracy in inferring feelings was significantly greater for the experimental than control group ( $t [53] = 2.65$ ), but there were no differences for thoughts ( $t [53] = .92$ ). Thus, there was improvement in empathic accuracy, but just for feelings, in both groups. The trained group showed more accuracy in inferring feelings than the control group.

Secondary analyses were conducted to check for differences related to class section, gender, and practice ( $\alpha = .05$ ). Using pre-test levels as a covariate, a one-way analysis of covariance showed no effect of class section at post-test on the accuracy of inferred thoughts or feelings ( $F_s[4, 48] = .88, 1.73$ ). Thus, the differences found cannot be explained by the differential effects of class structure or instructor, irrespective of the training experience introduced. There were twice as many women participants as men, and homogeneity of variance was lacking for ANOVAs. Gender differences were evaluated with independent-samples two-tailed *t*-tests (equal variances not assumed). Prior to training (pre-test), there were no differences in the accuracy of inferred feelings or thoughts. Post-test, women scored higher than men ( $M = .45$  vs.  $M = .32$ ) on accuracy of inferred thoughts ( $t [48.2] = -2.71$ ), but there were no differences on accuracy of inferred feelings. Regression analysis of the mean accuracy ratings of experimental participants across class practice sessions revealed a positive relationship between session and the accuracy of inferred feelings ( $F = 3.90, R^2 = .023$ ), but not for accuracy of inferred thoughts ( $F = 1.20, R^2 = .007$ ). However, total times practiced dur-



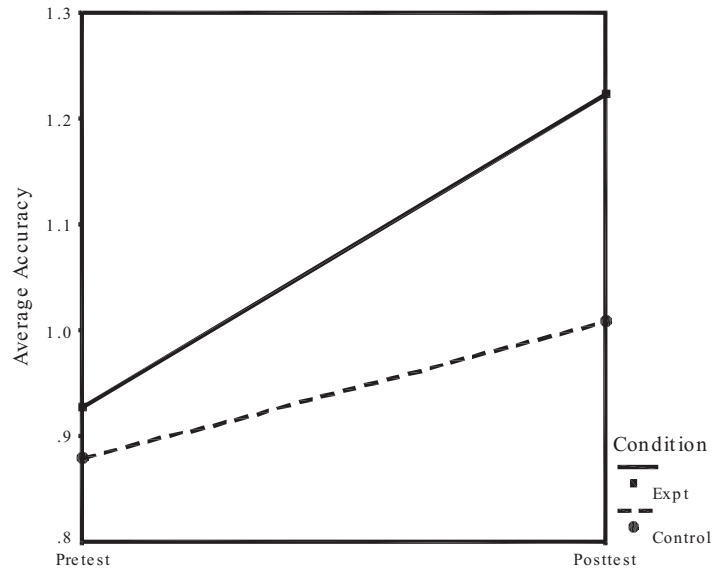


FIGURE 1. Accuracy of inferred feelings before and after an interviewing course.

ing class ( $M = 19.3$ ,  $SD = 5.31$ ) did not predict the accuracy of inferred feelings or thoughts on the post-test.

### FOLLOW-UP STUDY

The above study includes results on change from pre-test to post-test, in addition to comparisons of post-test scores. Although the tests were excerpted from the same psychotherapy session to minimize differences, the study cannot discount the possibility that changes reflect differences in the difficulty of the two tests rather than improvements from training. Thus, a follow-up study was conducted to check the comparability of the two tests. A separate group of 21 students in their first year of graduate study in clinical psychology were recruited from a class in adult psychopathology. As in the procedure described above, these participants watched the videotaped therapy session, the tape was stopped at the appropriate times, and the participants recorded what they inferred the client was thinking and feeling at those times. Ten of these participants watched the pre-test items first and the post-test items second, while 11 of them watched the post-test items first and the pre-test items second. Then, an additional measure of difficulty was obtained. The test

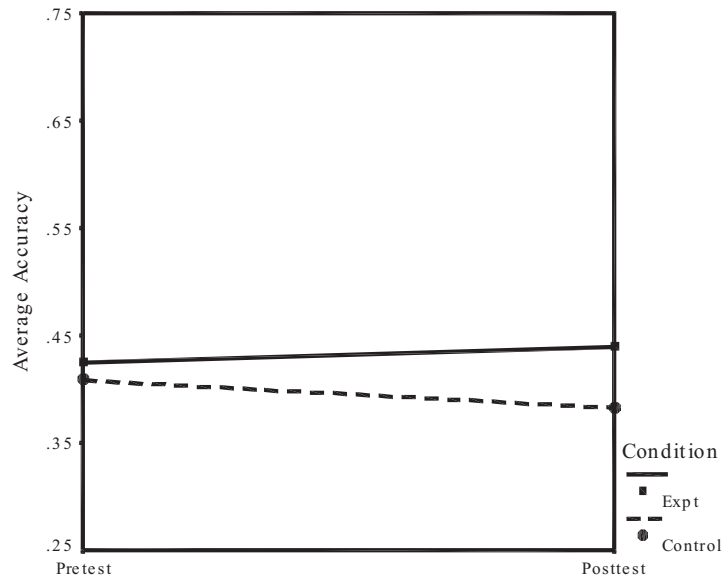


FIGURE 2. Accuracy of inferred thoughts before and after an interviewing course.

administrator announced what the client said she was thinking and feeling at each of the stops in the tapes. As each correct answer was given, the participants rated their own subjective difficulty in inferring that thought and feeling on a 7-point Likert scale (1 = very easy, 4 = neither difficult nor easy, and 7 = very difficult).

Two independent raters, unaware of the purpose of the study, were trained in the empathic accuracy scoring system. Over five training sessions, using training materials and the decision rules from the previous study, their inter-rater reliability of accuracy judgments increased from 75% agreement on thoughts and 76% agreement on feelings ( $N = 85$ ) to 85% agreement on thoughts and 88% agreement on feelings ( $N = 128$ ). Given that these raters had much less experience than those in the main study, it was decided to accept this level of agreement and have both raters independently judge the accuracy of the participants' inferences on every pre-test and post-test item. The raters then met to compare their ratings, identified any differences, and agreed upon each final accuracy judgment.

These data were subjected to two multivariate analyses of variance, one for empathic accuracy and one for subject difficulty. Each analysis included two within-subject factors—test (pre/post) and measure (feel-

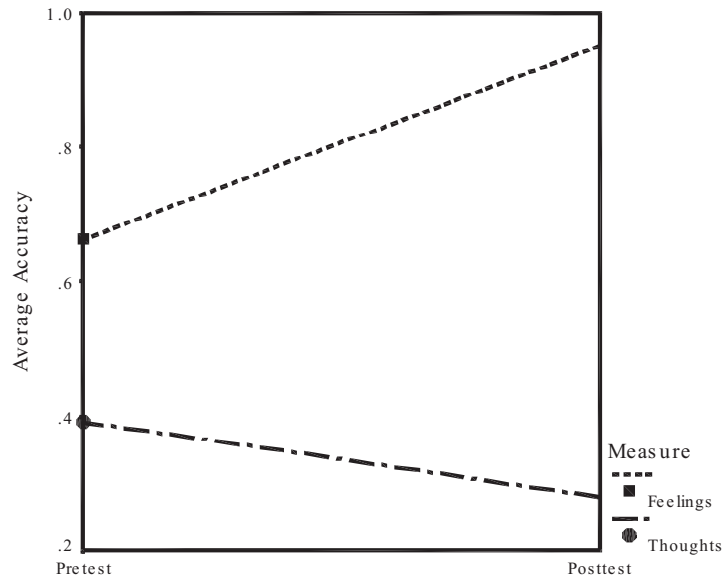


FIGURE 3. Accuracy of inferred feelings and thoughts during a single session.

ings/thoughts)—and one between-subjects factor—test order (pre/post or post/pre). Alpha was set at .05 and partial eta-squared was used for effect size. For empathic accuracy,  $F(1, 19)$  values were significant for measure ( $97.33, \eta^2 = .84$ ) and time by measure ( $27.52, \eta^2 = .59$ ), but not for test (2.76) or order (1.01) or other interactions. Consistent with the primary study, empathic accuracy for feelings was higher than for thoughts ( $M = .80$  vs.  $M = .34$ ). To further explore the interaction, shown in Figure 3, paired-sample two-tailed  $t$ -tests were run. The difference between the two tests was significant for feelings ( $t [20] = -3.60$ ) and marginal for thoughts ( $t [20] = 2.16$ ). The broken line for feelings measured during a single session in Figure 3 shows a larger difference than the dotted line for feelings in the control condition in Figure 1, suggesting that the latter's change results from an easier post-test rather than from improvement.

For subjective difficulty,  $F(1, 19)$  values were significant for measure ( $14.37, \eta^2 = .43$ ) and order ( $6.63, \eta^2 = .26$ ), but not for test (1.94) or any interactions. Thus, despite performance differences between pre-test and post-test, participants reported no differences in subjective difficulty. Also, despite a lack of performance differences for order, participants who took a test second, regardless of which one, reported it to be more

TABLE 2. Empathic Accuracy and Subjective Difficulty Scores During a Single Session

Measure	Pre-test				Post-test			
	Order 1		Order 2		Order 1		Order 2	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Feelings								
EA	.81	.27	.51	.23	.94	.26	.97	.29
SDS	3.80	.94	4.63	.87	3.74	.85	4.30	1.0
Thoughts								
EA	.40	.22	.38	.21	.25	.12	.31	.16
SDS	4.43	.85	5.06	.80	3.95	.96	4.92	.78

*Note.* For Order 1,  $n = 10$ , and for Order 2,  $n = 11$ . Empathic accuracy (EA) was rated 0 (different), 1 (similar), or 2 (the same). Subjective Difficulty Scores (SDS) were rated on a 7-point scale from easy to difficult. Order 1 was pre-test followed by post-test; Order 2 was post-test followed by pre-test.

difficult than those who took it first ( $M = 4.72$  vs.  $M = 3.98$ ). The one consistent finding across measures was that the participants reported greater difficulty in judging thoughts than feelings ( $M = 4.61$  vs.  $M = 4.12$ ) and had lower accuracy scores in inferring thoughts compared to feelings.

## DISCUSSION

Using Ickes' model, this study assessed empathic accuracy by clinicians-in-training for an actual client during psychotherapy. Of participants taking a semester-long course for developing interviewing skills, some practiced making empathic judgments about a variety of targets and received immediate feedback on accuracy from the target. In confirmation of the study's hypothesis, these participants demonstrated greater accuracy of inferring feelings, but not thoughts, than those who did not receive such training during the course. Additional support for the hypothesis was provided by experimental participants' improved accuracy in inferring various targets' feelings, but not thoughts, during practice sessions throughout the course. A follow-up study showed that increases in the post-test accuracy of inferred feelings in the control group could not be attributed to the interviewing course because there were higher scores on the post-test taken in a single setting, whether taken before or after the pre-test. Thus, the overall results of this study found no increase in empathic accuracy, as measured in the Ickes' paradigm, as a result of a standard interviewing course, but found that

targeted practice with feedback for this skill resulted in significantly greater accuracy for inferring feelings.

These findings are consistent with those of Marangoni et al. (1995) and Gesn and Ickes (1999), which were based on training college students on non-clinical targets. Their study and other research using Ickes' method dropped separate measures of thoughts and feelings after failing to find differences between them. It is possible that the greater difficulty of inferring thoughts found in the present study is a consequence of having a clinical target; such targets' experiences and interpretations are likely to be less normative and therefore less readable or predictable. In addition, while emotion may be deduced from displayed affect or facial expressions, thoughts must be inferred from the context of the interaction, which may be a more difficult task. Furthermore, number of emotions is limited (despite a large vocabulary for them) across persons, while the range of thoughts and different levels of description are unlimited. The decision rules for determining the similarity level of thoughts and feelings used by the raters, presented in the Appendix, indicate the different strategies used in scoring these two dimensions. It also is possible, however, that the small number of thoughts and feelings sampled differed in difficulty. Further research is needed to test more adequately (with many more items) the thought/feeling distinction in clinical settings. If supported, the training of cognitive empathic accuracy, as opposed to affective empathic accuracy, in clinical settings needs further study.

Unlike previous nonclinical studies (Ickes et al., 1990; Hancock & Ickes, 1996; Marangoni et al., 1995; Riggio et al., 1989) that found no gender differences, this study found a gender difference in the accuracy of thought inferences at post-test. Both males and females averaged less than "similar" in inferring thoughts, however, at least according to the decision rules for accuracy ratings used in this study. Furthermore, there was only one test target, who was female. Further research is needed to investigate possible gender differences in clinical empathic accuracy. This research should include conditions in which participants know that they are being evaluated on empathy (Ickes, Gesn, & Graham, 2000).

This study demonstrates that research on empathic accuracy in social psychology offers a fruitful method of researching the training of empathic accuracy in clinicians and counselors. This method provides a performance assessment for the important clinical skill of empathic accuracy, and this study demonstrates that emotional empathic accuracy can be improved as part of clinical training. Undoubtedly, the process of empathy, especially in a long-term relationship like psychotherapy, is much more complex than this measurement of empathic accuracy. But, like any good test, this one provides an efficient measure, and as a performance test with a criterion, it is an advance over judgments of empa-

thy by self or others. This preliminary study demonstrated, however, that more research is needed on the development of such a test in order to have different forms of it with equal difficulty.

It remains for future research to test whether improvement in the therapeutic process of empathy is correlated with positive outcomes of psychotherapy, as has long been claimed. Are there critical levels of empathic accuracy associated with effective therapy? Are there entry levels necessary to benefit from training in clinical and counseling programs? Can training be developed to improve the accuracy of inferred thoughts as well as of feelings? Finally, it would be of value to develop a set of test tapes of psychotherapy clients (or, if that is not ethically permissible, then of actors posing as clients), varying in types of disorder and difficulty in readability, so that we can test the crucial aspects of social intelligence needed by counselors and clinicians.

## REFERENCES

- Barak, A., Engle, C., Katzir, L., & Fisher, W. A. (1987). Increasing the level of empathic understanding by means of a game. *Simulation and Games, 18*, 458–470.
- Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs, 76*, 1–33.
- Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology, 28*, 91–100.
- Carkhuff, R. R. (1969). *Helping and human relations* (Vols. 1 & 2). New York: Holt, Rinehart & Winston.
- Carkhuff, R. R., & Berenson, G. (1967). *Beyond counseling and therapy*. New York: Holt, Rinehart, & Winston.
- Dalton, R. F., & Sunblad, L. M. (1976). Using principles of social learning in training for communication of empathy. *Journal of Counseling Psychology, 23*, 454–457.
- Davis, M. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology, 10*, 85.
- Davis, M. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.
- Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology, 13*, 127–133.
- Eisenberg, S., & Delaney, D. J. (1970). Using video simulation of counseling for training counselors. *Journal of Counseling Psychology, 17*, 15–19.
- Elliot, R., Filipovich, H., Harrigan, L., Gaynor, J., Reimschuessel, C., & Zapadka, J. K. (1982). Measuring response empathy: The development of a multicomponent rating scale. *Journal of Counseling Psychology, 29*, 379–387.
- Evans, B. J., Stanley, R. O., & Burrows, G. D. (1993). Measuring medical students' empathy skills. *British Journal of Medical Psychology, 66*, 121–133.
- Feldstein, J., & Gladstein, G. A. (1980). A comparison of the construct validities of four measures of empathy. *Measurement and Evaluation in Guidance, 13*, 49–57.
- Finke, J. (1990). Can psychotherapeutic competence be taught? *Psychotherapy and Psychosomatics, 53*, 64–67.

- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology, 77*, 746–761.
- Gladstein, G. A. (1977). Empathy and counseling outcome: An empirical and conceptual review. *Counseling Psychologist, 6*, 70–79.
- Guzzetta, R. (1976). Acquisition and transfer of empathy by the parents of early adolescents through structured learning training. *Journal of Counseling Psychology, 23*, 449–453.
- Hancock, M., & Ickes, W. (1996). Empathic accuracy: When does the perceiver–target relationship make a difference? *Journal of Social and Personal Relationships, 13*, 179–199.
- Hargrove, D. S. (1974). Verbal interaction analysis of empathic and nonempathic responses of therapists. *Journal of Consulting and Clinical Psychology, 42*, 305.
- Hatcher, S. L., Nadeau, M. S., Walsh, L. K., Reynolds, M., Galea, J., & Marz, K. (1994). The teaching of empathy for high school and college students: Testing Rogerian methods with the interpersonal reactivity index. *Adolescence, 29*, 961–974.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*, 307–316.
- Ickes, W. (1993). Empathic accuracy. *Journal of Personality, 61*, 587–610.
- Ickes, W. (Ed.). (1997). *Empathic accuracy*. New York: Guilford.
- Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathy accuracy: Differential ability or differential motivation? *Personal Relationships, 7*, 95–109.
- Ickes, W., Robertson, E., Tooke, W., & Teng, G. (1986). Naturalistic social cognition: Methodology, assessment, and validation. *Journal of Personality and Social Psychology, 51*, 66–82.
- Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed–sex dyads. *Journal of Personality and Social Psychology, 59*, 730–742.
- Lane–Garon, P. S. (1998). Developmental considerations: Encouraging perspective taking in student mediators. *Mediation Quarterly, 16*, 201–217.
- Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology, 68*, 854–869.
- Marcia, J. (1987). Empathy and psychotherapy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 81–102). New York: Cambridge University Press.
- Marks, T., & Tolsma, R. (1986). Empathy research: Some methodological considerations. *Psychotherapy, 23*, 4–20.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525–543.
- Perry, M. A. (1975). Modeling and instructions in training for counselor empathy. *Journal of Counseling Psychology, 22*, 173–179.
- Pigman, G. (1995). Freud and the history of empathy. *International Journal of Psychoanalysis, 76*, 237–256.
- Riggio, R. E., Tucker, J., & Coffaro, D. (1989). Social skills and empathy. *Personality and Individual Differences, 10*, 93–99.
- Rogers, C. R. (1957). Necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95–103.
- Rogers, C. R. (1959). A theory of therapy, personality, and interpersonal relationships as developed in the client–centered framework. In S. Koch (Ed.), *Psychology: The study of a science. Volume 3. Formulations of the person and the social context* (pp. 184–256). New York: McGraw–Hill.
- Rogers, C. R. (1975). Empathic: An unappreciated way of being. *The Counseling Psychologist, 2*, 2–10.

- Shaffer, W. F., & Hummel, T. J. (1979). Three experiments using an algorithm for empathic responses. *Journal of Counseling Psychology, 26*, 279–284.
- Stone, G. L., & Vance, A. (1976). Instruction, modeling, and rehearsal: Implications for training. *Journal of Counseling Psychology, 23*, 272–279.
- Stotland, E. (1969). Exploratory investigations of empathy. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press.
- Truax, C. B. & Carkhuff, R. R. (1967). *Towards effective counseling and psychotherapy*. Chicago, IL: Aldine.
- Wispe, L. (1987). History of the concept of empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 17–37). New York: Cambridge University Press.