Measuring Therapist Technique in Psychodynamic Psychotherapies

Development and Use of a New Scale

John S. Ogrodniczuk, Ph.D.
William E. Piper, Ph.D.

Treatment manuals are becoming a requirement for conducting quality psychotherapy research. What the field lacks, however, are reliable, valid, and cost-efficient instruments that can be used to measure a wide variety of prescribed therapeutic techniques. This article describes the development and use of a new instrument, the Interpretive and Supportive Technique Scale (ISTS). It is designed to measure interpretive and supportive features of technique for a broad range of dynamically oriented psychotherapies. Data concerning the psychometric properties of the ISTS are presented from two studies. The findings suggest that the ISTS is a potentially useful tool for measuring interventions for different forms of dynamically oriented psychotherapy.

(The Journal of Psychotherapy Practice and Research 1999; 8:142–154)

Therapist technique refers to the technical procedures that are used to facilitate therapeutic change. Instruments that measure technique have the potential to identify active ingredients of psychotherapy, permit the replication of research, and allow the analysis of commonalities and differences between treatments. The advancement of such work requires instruments with strong psychometric properties.

Treatment manuals provide guidelines for therapists to follow in conducting therapy. They focus on the specific techniques and strategies that are viewed as desirable.[1] Manuals have emerged as an important development in psychotherapy research. Although treatment manuals describe the prototypic form of therapy, they cannot ensure that the treatment that was actually provided conformed to the protocol. *Adherence* is defined as the degree to which the therapist uses the techniques specified in a treatment manual.[2,3]

Adherence, which is sometimes referred to as *treatment fidelity* or *treatment integrity*, has been heralded as a primary variable to consider when determining the therapist's contribution to the therapeutic process. However, despite the increasing use of treatment manuals and

the growing awareness of the need for adherence checks in psychotherapy research, the inclusion of adherence measurements is still not standard practice.[4]

The challenge is to develop reliable measures of technique that can be used to empirically verify the treatments that are provided in psychotherapy studies. An important related question concerns the relationship between different levels of therapist adherence and both the process (e.g., the therapeutic alliance) and outcome of therapy.

Many instruments for measuring therapist adherence have been developed,[5–9] but no widely accepted methodology exists. Thus, it is likely that different adherence measures assess different constructs. In addition, a common practice has been to use measures of technique *frequency* as measures of adherence. Although equating the two may be accurate in some circumstances, it is questionable in others. Treatment manuals generally do not instruct therapists to provide prescribed techniques as frequently as possible. Doing so may have undesirable effects. Clinical experience suggests that a balance between primary and secondary techniques is more likely to be beneficial. Thus, measures of the two constructs (frequency, adherence) need to be differentiated.

Lack of common methodology makes it difficult to compare and integrate the results of different studies. Consequently, it is hard to draw general conclusions from the few studies[10–14] that have addressed the question of whether adherence is significantly related to the process and outcome of psychotherapy.

Although a number of important measures of therapy technique exist, the practical utility of some scales is limited by their excessive length, weak psychometric properties, or representation of very specific therapy models. Examples include scales that consist of nearly 100 items,[2] scales that require many (e.g., eight) raters to achieve reliable ratings,[8] and scales that are designed for a very specific treatment approach, such as supportive-expressive dynamic therapy for cocaine dependence.[9]

In this article, psychometric data are reported from two studies for a new rating scale, the Interpretive and Supportive Technique Scale (ISTS), which is intended to be brief, reliable, and easily applicable to different forms of dynamic therapy. It is not our intention to present the ISTS as a replacement for existing measures of technique, but to offer it as a reliable alternative that has general applicability. The scale is designed to measure amount of technique and adherence for a range of interpretive and supportive forms of psychodynamic psycho-

therapy. Thus, it has the potential to facilitate comparisons among different therapies.

The first study determined the psychometric properties of rater reliability and internal consistency of the ISTS. The second study provided an opportunity to cross-validate the findings of the first study and examine the factor structure of the scale. In addition, the second study investigated the scale's validity.

Examination of construct validity included assessing the associations between ratings from the ISTS and two independent measures of therapist technique. We also examined the scale's capacity to differentiate two forms of therapy.

With regard to predictive validity, we investigated the associations that adherence and amount each had with the therapeutic alliance and treatment outcome. We tested the hypothesis that the more the therapists adhered to the guidelines of the treatment protocol, the stronger the alliance would be and the more favorable the outcome of therapy. Additionally, we investigated whether amount of technique had a curvilinear relationship with alliance and outcome. Specifically, we wished to know whether low to moderate amounts were positively related, and high amounts negatively related, to alliance and to outcome. Previous findings[15] had indicated that high levels of certain types of therapist interventions were negatively associated with alliance and outcome for some types of patients.

STUDY 1: METHODS

Sources of Data

Study 1 used preexisting data from two clinical investigations of psychotherapy outcome. The first was a controlled clinical trial in which patients treated with short-term interpretive therapy were compared with patients in a wait-list control group.[16] The second was a comparative clinical trial in which patients treated with short-term interpretive therapy were compared with patients treated with short-term supportive therapy.[17]

Subjects

For Study 1, we chose a sample size of 50 cases. Thirty-six cases were available from the comparative trial. They were independent of the sample of 144 therapy completers used in Study 2. Six completed therapy as pilot cases and 13 completed therapy with additional

nonproject therapists. Four completed therapy but had missed more than 6 sessions, and 13 were dropouts whose average attendance was 6.5 sessions (range 1–13). To achieve our objective of 50 cases, we randomly selected 14 cases from the 86 patients who completed interpretive therapy in the controlled trial.

Eighty-four percent of the 50 patients received Axis I diagnoses according to DSM-III[18] or DSM-III-R.[19] Most of these were mood disorders (69%) and adjustment disorders (12%). Of the patients diagnosed with mood disorders, 76% presented with major depression. Thirty-two percent of the patients received Axis II diagnoses, most of which were dependent (31%) and borderline (25%) personality disorders. With regard to comorbidity, 28% of the patients received Axis I and Axis II diagnoses. Thirty-eight percent of the patients were on medication, the majority of these (91%) receiving antidepressants. The average age of the patients was 31.6 years (SD = 9.7, range 18–60). Sixty percent were women. Thirty-eight percent were married or living with a partner, 42% had never been married, and 20% were separated or divorced. Fifty-four percent were educated beyond high school; 70% were employed. The racial composition was white, 94%; East Indian, 4%; and Native American, 2%.

Therapists

Eighteen staff therapists who had considerable experience practicing both interpretive and supportive forms of dynamic therapy provided treatment for the 50 cases of Study 1. They came from the disciplines of psychiatry, psychology, social work, and nursing. Seventeen were white and one was East Indian. Ten were female. The therapists' average age was 42.6 years (SD = 7.7, range 34–65), and their average experience practicing individual psychotherapy was 11.6 years (SD = 7.01, range 3–35). Therapists were trained to carry out treatment according to a manual. To facilitate their understanding and use of the manual, they participated in 6-month training programs prior to taking cases in the clinical trials. The training program included weekly 1-hour supervision sessions and weekly 1-hour group seminars where session material was presented and technical principles were discussed. These seminars continued throughout the trials. In addition, each therapist treated 2 pilot cases of 20 weekly, 50-minute sessions. In total, therapists received approximately 4 hours of training per week for the particular forms of treatment involved in the current studies.

Therapies

Each patient received a form of short-term, time-limited, individual psychodynamic psychotherapy that emphasized interpretive or supportive features (W. E. Piper, H.F.A. Azim, A. S. Joyce, et al., 1993, unpublished). These were labeled *interpretive therapy* and *supportive therapy*, respectively. Twenty once-weekly, 50-minute sessions were contracted. The primary objective of interpretive therapy (STI) was to enhance the patient's insight about repetitive intrapsychic and interpersonal conflicts that serve to underlie and sustain the patient's difficulties. Interpretation and clarification were emphasized relative to support and direction. The primary objective of supportive therapy (SUP) was to improve the patient's immediate adaptation to his or her life situation. Support and guidance were emphasized relative to interpretation. The treatment manual instructed therapists to provide a relative emphasis on the prescribed features but to avoid overemphasis on them. All therapy sessions were audiotaped.

Measures

The 14-item Interpretive and Supportive Technique Scale (Appendix A) was used to quantify the degree of therapist adherence to the intended strategies of STI and SUP psychotherapies. It also indicated the amount of interpretive and supportive technique provided. The 14 items cover a set of interpretive and supportive technical features that characterize differences among dynamic psychotherapies. Initial item generation was based on the treatment features described in the interpretive and supportive treatment manual. Additional items were generated from a review of relevant theoretical and clinical literature[20] and the clinical experience of the authors. It was our intention to develop a scale that could be used by bachelor's-degree-level raters. Thus, efforts were made to make all items as clear and concrete as possible, requiring a relatively low level of inference and allowing judgments that could be provided solely on the basis of observations of audiotaped sessions. An initial version of the scale was applied to audiotaped sessions of interpretive therapy and supportive therapy. Redundant, unclear, or ambiguous items were then revised or eliminated. As well, the rating instructions were clarified. A manual for the scale was developed, providing definitions of the items and illustrative clinical examples. The final version of the ISTS consists of the 14 items presented

in Appendix A. The scale does not focus on theoretical concepts of any one particular author. Rather, the scale focuses on general techniques that are common to a number of forms of psychodynamic therapy. The seven odd-numbered items of the scale describe supportive features, and the seven even-numbered items describe interpretive features. Conceptually, each item represents a continuum. Nonparticipant observers rate each item on a 5-point Likert-type scale (0 = no emphasis, 4 = major emphasis).

Two subscale scores (Interpretive, Supportive) are derived from the ISTS. Each subscale score is calculated by adding the scores of the seven items that correspond to each form of therapy. The subscale scores represent the quantified amount of interpretive and supportive technique. The range of scores is 0–28, with higher scores representing greater amounts.

For the comparative clinical trial, the two subscale scores were used to calculate a full-scale score, which was the measure of adherence. This operationalization of adherence reflected two important features: 1) the characterization of the two psychotherapies by the use of a set of continua, and 2) the instruction given in the therapy manual that the therapist is to provide a *relative* emphasis on the prescribed technique. In order to create a meaningful index of adherence that reflected these two features—that is, the use of a continuum representing relative emphasis—we devised a formula to calculate the full-scale score. This formula is

(Interpretive subscore – Supportive subscore) + 28.

Thus, full-scale scores, indicating relative emphasis, fall on a continuum that ranges from 0 to 56. Zero to 27 represents the supportive range of the continuum, with lower scores indicating greater supportive emphasis. Twenty-nine to 56 represents the interpretive range of the continuum, with higher scores indicating greater interpretive emphasis. A score of 28 represents a therapy with equal supportive and interpretive emphases. Completing the ISTS requires approximately one hour per 50-minute session.

Procedures

Study 1 examined the rater reliability and internal consistency of the ISTS. We wished to demonstrate that the ISTS can be used effectively by a variety of raters. For this case, the appropriate rater reliability index is Shrout and Fleiss's[21] intraclass correlation coefficient (ICC) model 2, a random effects model. ICC (2,1) provides an estimate of the reliability of a rating that might be obtained by an independent rater and represents the generalizability of the rating. To determine the ICC (2,1), a random sample of raters is selected from a larger population, and each rater independently rates each target. The reliability coefficient indicates the degree to which any single rater can be used to represent the score.

In Study 1, two male independent raters provided ratings for one session from each of the 50 cases. The raters were randomly chosen from a larger pool of 10 trained, bachelor's-degree-level raters. As part of the training process, all raters were provided with a didactic overview of psychodynamic theory and a copy of the treatment manual. The raters were then introduced to the rating scale. Group discussions of the conceptual background of the ISTS were held so that each rater could gain an adequate understanding of the material. Training also involved rating a number of practice sessions prior to applying the scale to therapy cases in the research project.

Ratings were made of audio recordings of whole, 50-minute therapy sessions. The treatments were equally represented (25 interpretive cases, 25 supportive cases). Except for this constraint, cases were chosen randomly from the larger population (the database described above).

STUDY 1: RESULTS

Rater Reliability

Rater reliability was assessed for the full scale, the subscales, and the individual items of the ISTS. The ICC (2,1) coefficient of 0.95 for the full scale was high, as were the coefficients for the two subscales (0.93 for the supportive subscale and 0.88 for the interpretive subscale). As shown in Table 1, the ICC (2,1) coefficients for the individual supportive and interpretive items were in the moderate to high range with the exception of one interpretive item that was low (item 14, impression of others). The average ICC (2,1) coefficient for all 14 items was 0.74.

Internal Consistency

Generally speaking, *internal consistency* refers to how highly the items of a scale are related to each other. One

of the most commonly used indexes of internal consistency is Cronbach's alpha,[22] which is based on the average correlation of items within a scale. Cronbach alphas were calculated to determine the internal consistency of the full scale and the two subscales (Supportive, Interpretive). The full scale refers to the entire set of 14 items. The seven supportive items were reverse-scored for this part of the analysis so that all 14 items were keyed in the same direction. The ratings of each of the two raters were examined separately. The alpha coefficients were 0.92 and 0.95 for the full scale, 0.92 and 0.94 for the supportive subscale, and 0.86 and 0.88 for the interpretive subscale, indicating high internal consistency.

STUDY 2: METHODS

Patients

Study 2 used data from 144 cases (all of the therapy completers) from the comparative trial described above under Study 1.

All patients received diagnoses according to DSM-III-R.[19] Axis I diagnoses were identified by the computer-administered Mini-SCID (Structured Clinical Interview for DSM-III-R[23]) and validated by an independent clinical diagnosis assigned jointly by the intake assessor and a psychiatrist, both of whom saw the patient on the day of intake. Axis II diagnoses were determined by the computer-administered SCID-II PQ and Auto-SCID II (Structured Clinical Interview for DSM-III-R[24]).

Seventy-three percent of the patients received an Axis I diagnosis. The most frequent were mood disorders (61%) and adjustment disorders (7%). Of those diagnosed with mood disorders, 80% presented with major depression. Sixty percent of the patients received an Axis II diagnosis, the majority of which were avoidant (29%), obsessive-compulsive (24%), borderline (22%), and paranoid (22%) disorders. A total of 46.5% of the patients received both Axis I and Axis II diagnoses. Forty-two percent of the patients were prescribed psychotropic medication. In nearly all of these cases (93%), the medication was an antidepressant.

The average age of the patients was 34.3 years (SD = 9.6, range 18–62). Sixty-one percent were women. Forty-two percent were married or living with a partner, 37% had never been married, and 21% were separated or divorced. Sixty-seven percent were educated beyond high school, and 71% were employed. The racial composition was white, 92%; East Indian, 4%; native American, 2%; and Asian, 2%.

Therapists

Eight therapists (three male, five female) treated the patients in the comparative clinical trial. They came from the disciplines of psychiatry, psychology, social work, and occupational therapy. Seven were white and one was East Indian. The therapists' average age was 42.63 years (SD = 6.86, range 35–52). The average number of years practicing individual psychotherapy was 10.88 (SD = 4.82, range 3–19). Although the therapists were experienced in providing a variety of interpretive and supportive forms of therapy, they participated in a 6-month training program prior to taking cases in the comparative trial. This training program included following a technical manual, treating 2 pilot cases, and attending a weekly 1-hour supervision session and a weekly 1-hour group seminar that continued throughout the trial. Overall, therapists received approximately four hours of training per week for the two forms of therapy involved in the current study.

TABLE 1. Rater reliability of 2 raters for the Interpretive and Supportive Technique Scale (ISTS)

Item or Scale	Study 1	Study 2
ISTS Item		
1. Gratify	0.81	0.58
2. Pressure	0.75	0.41
3. Noninterpretive interventions	0.80	0.46
4. Uncomfortable emotions	0.71	0.47
5. Guidance	0.84	0.71
6. Interpretations	0.90	0.77
7. Problem solving	0.81	0.43
8. Impression of therapist	0.87	0.87
9. Explanations	0.76	0.44
10. Linking	0.74	0.53
11. Praise	0.87	0.43
12. Patient-therapist relationship	0.70	0.72
13. Personal information	0.61	0.28
14. Impression of others	0.25	0.50
Supportive subscale	0.93	0.69
Interpretive subscale	0.88	0.84
Full scale	0.95	0.95

➥ *Note:* Rater reliability was estimated by using intraclass correlation coefficients (2,1).

Therapies

Each patient received a form of short-term, time-limited, individual psychotherapy that emphasized interpretive or supportive features. A description of these therapies is provided above under Study 1.

Measures

In addition to the ISTS, two other measures of therapist technique were used. One was the Therapist Intervention Rating System (TIRS).[25] The TIRS, which is a very different type of rating system from the ISTS, provides a content analysis of individual therapist interventions. Each therapist statement from each session is assigned to one of nine categories that range from simple utterances (e.g., "mm-hmm") to complex interpretations. The five lower categories include brief expressions, reflections, clarifications, questions, and directives that do not make reference to patient dynamic components such as wishes, anxiety, or defenses. Thus, they are defined as interventions but not interpretations. The upper four categories make reference to patient dynamic components and are defined as interpretations. They differ only in the number of dynamic components referred to (one, two, three, or four). The TIRS was designed to be sensitive to variations in therapist technique across different forms of dynamic psychotherapy. Trained listeners assign each therapist intervention to one of nine categories. This method of rating requires 2 to 6 hours to complete a 50-minute session. The TIRS provides information on the *frequency and duration of techniques used* in a therapy session. Rater reliability of category agreement has been consistent across three studies.[16,17,25] The kappa coefficients and sample sizes in these studies were 0.71 (15 patients), 0.66 (40 patients), and 0.71 (12 patients), respectively. With regard to validity, the TIRS has been used in studies that have demonstrated significant relationships between the concentration and correspondence (accuracy) of transference interpretations and both the therapeutic alliance and treatment outcome.[26] The TIRS has also been used to identify significant relationships between characteristics of therapist interventions and patient responses.[27]

The other measure of technique was the Perception of Technique Scale (W.E. Piper, A.S. Joyce, M. McCallum, et al., 1993, unpublished). Both the therapist-rated version (PTS-T) and the patient-rated version (PTS-P) were used. This is a brief eight-item rating scale that measures the therapist's and the patient's perceptions of the technical approach the therapist used during the therapy session. The scale is based on the structure of the ISTS; that is, the eight items reflect key features of interpretive and supportive forms of therapy. Five items represent supportive features and three items represent interpretive features. Two subscales, representing each form of treatment, and one full scale keyed in the interpretive direction are derived from the PTS. After each therapy session, the therapist and patient independently rated each of the eight items on the *degree to which the therapist attempted to provide these key features.* The ratings were made on a 5-point, Likert-type scale ranging from $0 = $ no emphasis to $4 = $ major emphasis. The internal consistency for each version of the scale was examined by using ratings from the sample of 144 cases from the comparative trial. The alpha coefficients for the full scale, interpretive subscale, and supportive subscale of the therapist-rated version were 0.95, 0.96, and 0.92, respectively. For the full scale, interpretive subscale, and supportive subscale of the patient-rated version, they were 0.87, 0.80, and 0.90, respectively.

Although the ISTS and PTS are similar in structure and have some similarity in content addressed, the two scales also differ. Specifically, the PTS addresses the participants' subjective impression of the therapist's behavior in eight key areas during the session. The ISTS, on the other hand, is an externally rated quantitative measure of 14 specific therapist techniques.

The battery of outcome measures employed in the comparative trial provided the post-therapy outcome data used in Study 2. This battery included nine measures (questionnaire or interview) that covered 13 variables (listed below) in the areas of interpersonal distress and functioning, psychiatric symptomatology, self-esteem, life satisfaction, and use of defenses. The patient, therapist, and external assessors were used as sources for outcome ratings. The variables were measured at pre-therapy and post-therapy, permitting the calculation of residual gain scores.

The residual gain scores for the 13 outcome variables were subjected to a factor analysis. The primary goal of the analysis was to reduce the large set of variables to a small set of factors. Each factor contains highly correlated outcome variables. This goal of data reduction was achieved by using a principal components analysis with varimax rotation. The number of factors retained was determined by a screeplot of the eigenvalues. A three-factor varimax solution was selected. The eigenvalues

for the three factors were 5.5, 1.3, and 1.1. Collectively, they accounted for 60.4% of the variance.

Factor I accounted for 42.1% of the variance and included the ratings of target objective severity provided by the patient and therapist, the Beck Depression Inventory,[28] the global severity index of the SCL-90,[29] the Trait Anxiety Inventory,[30] the Rosenberg Self-Esteem Scale,[31] a life satisfaction rating, the overall rating of the Inventory of Interpersonal Problems,[32] and ratings of maladaptive defensive style.[33] All variables, except for life satisfaction, were scored in the pathological direction. The scores for life satisfaction were reversed in order to be consistent with the other variables. The factor was interpreted as representing *General symptomatology and dysfunction.* Factor II accounted for 9.6% of the variance. It included ratings from the Social and Sexual subscales of the Social Adjustment Scale (SAS; M.M. Weissman, E.S. Paykel, R. Siegel, et al., 1972, unpublished). This factor is understood as representing *Social-sexual maladjustment.* Factor III accounted for 8.7% of the variance and included ratings of the nonuse of mature defenses[33] and family pathology according to the Family subscale of the SAS. As the outcome variables suggest, it represented *Nonuse of mature defenses and family pathology.* For all three factors, higher scores represent greater pathology.

Therapeutic alliance was defined as the working relationship between the patient and therapist. It was assessed by soliciting brief ratings by the patient and by the therapist after each session. The patient and therapist each rated six 7-point, Likert-type items that ranged from "very little" to "very much." The items focus on whether the patient 1) had talked about private, important material, 2) felt understood by the therapist, 3) understood and worked with what the therapist said, and 4) felt that the session enhanced understanding. The remaining two items focused on 5) whether the therapist was helpful and 6) whether the therapist and patient worked well together.[34] The six items were averaged across their respective assessments. Principal components analyses of each set of items (patient-rated, therapist-rated) resulted in one patient-rated factor and one therapist-rated factor. Examination of the internal consistency of the two sets of items revealed high coefficient alphas for each (0.97 for patient-rated, 0.96 for therapist-rated). An overall alliance score was devised by calculating the average of the six items. Thus, two scores (Patient, Therapist) served as summary measures of the therapeutic alliance over the entire course of therapy. The correlation between the patient-rated alliance factor and the therapist-rated

alliance factor was significant but not high ($r = 0.32$, df = 140, $P < 0.001$).

Procedures

In Study 2, the procedure varied according to the research question being addressed (i.e., confirmation of the rater reliability and internal consistency findings from Study 1 or examination of the scale's factor structure and validity). To check the replicability of the findings of Study 1, two independent raters provided ratings for one session from each of 50 cases. The treatments for these cases were equally represented (25 interpretive cases, 25 supportive cases). Cases were randomly selected from the first 96 completers in the comparative trial. The raters (one male, one female) were chosen from the larger pool of trained, bachelor's-level raters. Neither rater participated in Study 1. Rater reliability estimates were based on the collection of ratings from the two new raters. Also, as in Study 1, internal consistency was determined for the ratings of each of the two raters.

In regard to investigating the factor structure and validity of the ISTS, all 10 trained raters (6 females, 4 males) were used to assess therapist adherence for all 144 completer cases in the comparative trial. Each rater was randomly assigned to cases as the investigation progressed. The rating of the 144 cases took place over a period of 3 years. Nine sessions of each therapy case were rated in a fixed order starting with the third session and proceeding with every other subsequent session (sessions 3, 5, 7, 9, 11, 13, 15, 17, 19). Whenever a tape was unavailable or inaudible, the audiotape from the previous or following session was used. A total of 1,296 sessions were rated by the 10 raters: 648 interpretive therapy sessions and 648 supportive therapy sessions.

STUDY 2: RESULTS

Rater Reliability

In Study 2, the ICC (2,1) coefficient of 0.95 for the full scale was high and similar to the coefficient in Study 1. The coefficient for the supportive subscale (0.69) was moderate and the coefficient for the interpretive subscale (0.84) was high. The ICC (2,1) coefficients (see Table 1) for the individual supportive items were in the moderate range with the exception of one item that was low (item 13, therapist disclosure). The coefficients for the individual interpretive items were in the moderate to

high range. The average ICC (2,1) coefficient for all 14 items was 0.54.

In addition to the examination of rater reliability in Study 2, rater reliabilities were calculated on three occasions in the comparative trial. For each of the three reliability checks, each available rater independently rated one session from each of eight cases. ICC (2,1) coefficients were calculated for the full scale, the two subscales, and each of the 14 items of the ISTS. Rater reliabilities for the full scale and two subscales were consistently high. The average ICC (2,1) coefficients across the three periods were full-scale = 0.92, supportive subscale = 0.87, and interpretive subscale = 0.88. The rater reliabilities for the individual items of the ISTS varied over the three periods, yet most remained in the moderate to high range. The mean ICC (2,1) coefficients and the range of coefficients for the three occasions were 0.63 (range 0.11–0.84), 0.68 (range 0.07–0.94), and 0.63 (range 0.51–0.83), respectively.

Internal Consistency

For each of the two raters, the alpha coefficients for the full scale were 0.92 and 0.86; for the supportive subscale, 0.81 and 0.87; and for the interpretive subscale, 0.92 and 0.81.

Factor Structure

The 14 items of the ISTS were subjected to a factor analysis in order to identify the underlying dimensions of the scale; that is, to identify a set of more general factors that explain the correlations among the ISTS items. On the basis of this goal and the recommendations of Floyd and Widaman,[35] a principal-axis (common factors) factor extraction method was chosen. The analysis was performed on a data set of the 144 therapy completers. The item ratings averaged across the nine rated sessions for each case were used.

One factor emerged from the analysis, accounting for 64.9% of the variance. The eigenvalue for this factor is 9.1. The factor loadings for this factor are shown in Table 2. It can be seen from this table that all of the supportive items have considerable positive loadings (0.69 or greater) on the factor. The interpretive items have high negative loadings (–0.51 or greater) on the same factor. It seems clear, therefore, that this bipolar factor represents the set of interpretive-supportive continua on which the ISTS was based.

Convergent Validity

Convergent validity refers to how highly a scale correlates with other measures of the same (or a very similar) construct. To demonstrate the convergent validity of the ISTS, we examined the correlation between ratings from the ISTS and ratings from two other measures of psychodynamic technique (TIRS, PTS). Pearson correlation coefficients were calculated between ISTS and TIRS frequency ratings of noninterpretive interventions, interpretations, and transference interpretations. The ISTS ratings for these three variables were derived from item 3, noninterpretive interventions; item 6, interpretations; and item 8, subjective impression of therapist (transference), respectively. Because of the considerable time required to provide ratings, the TIRS was used for only six sessions per case (approximately every third session beginning with session 3) for the first 80 therapy completers. ISTS and TIRS ratings were averaged across the six sessions. TIRS raters were independent of the ISTS raters. The results indicated that the ISTS and TIRS ratings of noninterpretive interventions were significantly correlated ($r = 0.73$, df = 78, $P < 0.000$). In the case of interpretations, there was also a significant association ($r = 0.84$, df=78, $P < 0.000$) between ISTS ratings and TIRS ratings. As well, ISTS ratings of transference interpretations were significantly related ($r = 0.77$, df=78, $P < 0.000$) to comparable TIRS ratings.

TABLE 2. Factor solution for the Interpretive and Supportive Technique Scale

Item	Factor Loading
Supportive	
1. Gratify	0.95
3. Noninterpretive interventions	0.83
5. Guidance	0.87
7. Problem solving	0.77
9. Explanations	0.79
11. Praise	0.85
13. Personal information	0.69
Interpretive	
2. Pressure	–0.73
4. Uncomfortable emotions	–0.79
6. Interpretations	–0.93
8. Impression of therapist	–0.80
10. Linking	–0.79
12. Patient-therapist relationship	–0.89
14. Impression of others	–0.51

The full-scale and subscale ratings of the ISTS were compared with the comparable ratings of the PTS-T and PTS-P. These analyses were based on all cases ($N= 144$). The relationships between comparable scores were calculated by using Pearson correlation coefficients. The findings revealed that the ISTS ratings for the full scale, supportive subscale, and interpretive subscale were highly associated with the therapist's perception of his or her own technique ($r= 0.96$, df $= 142$, $P< 0.000$; $r= 0.86$, df $= 142$, $P< 0.000$; and $r= 0.94$, df $= 142$, $P< 0.000$, respectively). Although the correlations were lower, ISTS ratings for the full scale, supportive subscale, and interpretive subscale were significantly related to the patient's perception of the therapist's technique ($r= 0.70$, df $= 140$, $P< 0.000$; $r= 0.42$, df $= 140$, $P< 0.000$; and $r = 0.28$, df $= 140$, $P< 0.001$, respectively).

One aspect of the validity of the ISTS is its ability to differentiate interpretive and supportive treatments. Independent-samples t-tests were used to compare the means for the full scale, subscales, and individual items of the ISTS from each form of treatment (72 interpretive cases, 72 supportive cases). As shown in Table 3, the full scale and subscales significantly distinguished the two forms of treatment. Furthermore, each of the 14 items of the scale was able to significantly differentiate the two treatments.

Predictive Validity

Predictive validity refers to how well a test can predict future criteria. In Study 2, we wished to determine whether ISTS ratings of adherence and amount of technique could be used to predict the therapeutic alliance and treatment outcome. Pearson correlation coefficients were used to assess the relationship that adherence and amount (interpretive, supportive) each had with alliance and outcome. In addition, curvilinear relationships between amount and alliance and outcome were assessed by a series of hierarchical regression analyses. All analyses were conducted within each of the treatment subsamples (STI, SUP) and across all cases.

Adherence was significantly associated with therapist-rated alliance in STI therapy ($r= 0.23$, df $= 70$, $P< 0.05$), as well as across all cases ($r= 0.21$, df $= 142$, $P< 0.05$). Adherence was not significantly related to therapy outcome.

Amount of interpretive technique was also significantly associated with therapist-rated alliance in STI therapy ($r= 0.36$, df $= 70$, $P< 0.01$). Amount of interpretive

technique was not significantly associated with therapy outcome. No significant curvilinear relationships were found.

Amount of supportive technique was significantly associated with therapist-rated alliance across all cases (SUP and STI; $r= 0.18$, df $= 140$, $P< 0.05$). Amount of supportive technique was not significantly related to therapy outcome. No significant curvilinear relationships were found.

DISCUSSION

This article has presented the development and validation of a measure of therapist technique for different forms of dynamically oriented psychotherapy. Despite considerable interest in the role of technique in psychotherapy, only a handful of reliable, valid, and cost-efficient measures have been developed.

TABLE 3. Comparison of mean emphasis ratings for the items, subscales, and full scale of the ISTS from each treatment modality (STI, SUP)

| | Rating | | |
Variable	STI	SUP	t^a
Supportive items			
1. Gratify	0.36	2.91	27.07
3. Noninterpretive interventions	1.96	3.73	15.76
5. Guidance	0.19	2.38	18.68
7. Problem solving	0.07	1.55	14.48
9. Explanations	0.19	1.58	13.37
11. Praise	0.07	1.78	17.85
13. Personal information	0.15	1.56	11.39
Interpretive items			
2. Pressure	1.92	0.27	12.99
4. Uncomfortable emotions	2.32	0.52	13.67
6. Interpretations	3.31	0.57	32.75
8. Impression of therapist	1.34	0.02	12.63
10. Linking	1.17	0.02	12.65
12. Patient-therapist relationship	2.08	0.25	19.89
14. Impression of others	1.91	0.74	7.54
Supportive subscale	3.00	15.46	29.18
Interpretive subscale	14.04	2.40	28.29
Full scale	39.05	14.93	39.81

➥ *Note:* ISTS=Interpretive and Supportive Technique Scale; STI= interpretive; SUP= supportive.
^{a}For all comparisons, df=142, $P<0.000$.

Rater Reliability, Internal Consistency, and Factor Structure of ISTS

Rater reliability for the full scale and subscales was high across the two studies—among the highest reported in the literature. This result is particularly promising given that bachelor's-degree-level raters were used. It is possible that their naïveté—that is, their freedom from clinical or theoretical biases—assisted them in achieving reliable ratings of psychodynamic constructs. Experienced clinicians often encounter difficulty. It is also possible that assessing observable features of technique is straightforward and relatively devoid of inference compared with assessing more latent psychodynamic constructs. This might enable both clinically experienced and inexperienced raters to achieve high reliability.

Rater reliability for the individual items was moderate to high. Only item 13 and item 14 suffered from low rater reliability. For each, this occurred in only one of the two studies. Should further research using the ISTS reveal consistently low reliability estimates for either item, consideration will be given to either modifying the item or omitting it from the scale. Current inclusion of these items increases the comprehensiveness of the domain of technical features assessed by the scale. The high levels of internal consistency found in both studies compare favorably to those of other measures of technique reported in the literature.[8]

Factor analysis of the ISTS yielded a meaningful factor structure underlying observed differences between SUP and STI therapy sessions. One factor, representing both supportive features and interpretive features, emerged. This bipolar factor supports our conceptualization of the full scale as a supportive–interpretive continuum. As well, each set of interpretive items and supportive items loaded highly together, thus providing support for the rationally developed subscales. One may question, however, whether a unidimensional, bipolar factor can adequately represent a therapy session that emphasizes both interpretive and supportive features. We believe that the answer is yes. Our interpretation of this factor is that it represents the continuum on which our full scale is based. The more a session gravitates toward one end of the continuum, the more features of one form of therapy and the fewer of the other are present. Likewise, the more a session gravitates toward the middle of the continuum, the more equal are the technical emphases from each form of therapy. Technical emphases within a session may be both low or both high. Although the full-scale score does not specify how much of each therapy is present, the two subscales provide this information.

Overall, the findings provided evidence that the ISTS is a reliable measure of therapist technique and addressed the technical features of the two forms of dynamic psychotherapy as intended.

ISTS Compared With PTS and TIRS

Technical features assessed by the ISTS were significantly related to the therapist's and the patient's perceptions of the therapist's technical emphasis during the session. The very high correlations between the PTS-T and the ISTS ratings suggest that the ISTS may provide redundant information and therefore may not be needed. If further research with these two scales consistently reveals very high correlations, then the suggestion to use the simpler and more time-efficient PTS rather than the ISTS would have to be entertained. However, there may be other reasons for using the ISTS in a particular study. First, an externally rated scale is usually considered to be more objective than a self-report scale. Second, the ISTS provides information about specific therapist techniques. The PTS provides only an impression of general therapist behavior. Third, the ISTS has broader application. It can be applied to any audiotaped therapy session, regardless of whether the scale was included as part of the study. The PTS, on the other hand, must be originally included in a study for ratings to be provided.

Scores from the ISTS were also significantly related to ratings of similar constructs provided by the TIRS, an objective measure of therapist technique with established reliability and validity. As an externally rated measure, however, the ISTS has the decided advantage of requiring substantially less time than the TIRS (1 hour versus 2–6 hours).

The ISTS was designed for multiple uses. One is to measure therapist adherence to interpretive and supportive forms of dynamic psychotherapy. In Study 2, the ISTS demonstrated that the two forms of treatment had been carried out as intended. Additional analyses indicated that the ISTS was able to differentiate the two forms of treatment. From a methodological standpoint, treatment differentiation serves a critical purpose in comparative outcome research, where it is essential to ensure that the various forms of treatment differed as intended.

Relationships Between Therapist Technique and Other Variables

Another use of the ISTS is to study the relationship between therapist technique and other important variables. Study 2 examined the relationship that adherence, as well as amount of technique, had with the therapeutic alliance and therapy outcome.

Therapist adherence was significantly associated with therapist-rated alliance within STI therapy and across all cases. Amount of interpretive technique was also significantly associated with therapist-rated alliance within STI therapy. One explanation for these findings is that the alliance is strengthened when the therapist provides a greater emphasis on interpretive features, such as providing thoughtful interpretations and exploring uncomfortable emotions. An alternative explanation is that a stronger alliance allowed the therapist to be more interpretive. The therapist may have believed that once a strong working relationship with the patient had been established, the patient could tolerate a more intense interpretive emphasis in therapy.

Amount of supportive technique was also significantly related to therapist-rated alliance across all cases. It is possible that therapists in both forms of treatment perceived supportive interventions as valuable tools for building a strong alliance. Outcome had no significant relationship to adherence or to amount for either of the two forms of therapy.

Overall, it must be concluded that adherence had a weak, but statistically significant, association with the alliance and a minimal relationship with therapy outcome. Several explanations are possible.

One explanation for these findings is that therapist adherence to a technical manual is not that important to the development and maintenance of a strong working relationship and the achievement of patient change. In general, the literature indicates that technical adherence has been a weak and inconsistent predictor of the alliance and treatment outcome.[36]

A related possibility is that common therapeutic factors are more influential and that once they are provided, variation in technical features may have little impact. This explanation attributes successful therapy outcome to common factors such as the presence of a helping relationship, a convincing rationale, and feedback concerning progress, which are believed to increase morale, a sense of mastery, and positive expectations. These factors may be both necessary and sufficient to induce therapeutic change, thereby minimizing the contribution of specific technical interventions.

Conversely, adherence to therapist technique may be influential, but high levels of adherence may create problems that cancel the beneficial effects of the technique. In an effort to follow the manual as closely as possible, therapists may take too many opportunities to implement prescribed interventions without due consideration of the appropriateness or consequences of the interventions. In other words, a rigid or mechanical application of technique can be counterproductive. Support for this hypothesis is offered by Henry et al.,[37] who found deterioration in certain interpersonal and interactional aspects of therapy with increased technical adherence.

An alternative explanation is that once a minimal, but sufficient, level of technique is provided, further adherence is not important. Perhaps the levels of adherence present in studies that have monitored it have exceeded this minimal level and thus have lacked the variability necessary to test its effect.

Yet another possibility is that there may be limitations to the particular measure of adherence that was used in the present studies. Some of the items of the ISTS may represent unimportant features. Only certain technical interventions, such as transference interpretations or problem solving, may operate to bring about favorable patient change. Inclusion of additional, unimportant features would prevent the sensitive measure of these important interventions.

A final point of possible relevance is that the manual used in our studies has certain limitations. Because it was intended to be brief, it did not describe specific occasions for flexibly implementing one technique over another. That is, it may be necessary for the manual to provide more detailed guidelines about which techniques to use (or avoid) for particular patients at certain times in therapy. As an example, for certain patients, it may be unwise to persist with transference interpretations in an effort to diminish resistance early in therapy. This type of information could be communicated in a more detailed or comprehensive manual.

Of the correlational relationships that reached significance, many were low in magnitude, accounting for a small amount of variance. However, this does not mean that the relationships were statistical anomalies, nor that further exploration of such associations should be inhibited. The low magnitude of the correlations does, however, limit the practical relevance of the findings.

CONCLUSION

Despite methodological limitations in the current studies, the ISTS appears to be a potentially useful tool for measuring therapeutic interventions for different forms of dynamically oriented psychotherapy. Future work to further substantiate the reliability of the ISTS and establish its validity should involve a wider range of modalities and different applications of the scale in research focused on therapeutic technique.

REFERENCES

1. Luborsky L, DeRubeis RJ: The use of psychotherapy manuals: a small revolution in psychotherapy research style. Clinical Psychology Review 1984; 4:5–14
2. Shapiro DA, Startup M: Measuring therapist adherence in exploratory psychotherapy. Psychotherapy Research 1992; 2:193–203
3. Strupp HH, Butler SF, Rosser CL: Training in psychodynamic therapy. J Consult Clin Psychol 1988; 56:689–695
4. Waltz J, Addis ME, Koerner K, et al: Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. J Consult Clin Psychol 1993; 61:620–630
5. Barber JP, Crits-Christoph P: Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: a preliminary report. Psychotherapy Research 1996; 6:81–94
6. Hollon SD, Evans MD, Elkin I, et al: System for rating therapies for depression. Paper presented at the American Psychiatric Association annual meeting, Los Angeles, CA, 1984
7. Koenigsberg HW, Kernberg OF, Haas G, et al: Development of a scale for measuring techniques in the psychotherapy of borderline patients. J Nerv Ment Dis 1985; 173:424–431
8. Hill CE, O'Grady KE, Elkin I: Applying the Collaborative Study Psychotherapy Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. J Consult Clin Psychol 1992; 60:73–79
9. Barber JP, Krakauer I, Calvo N, et al: Measuring adherence and competence of dynamic therapists in the treatment of cocaine dependence. J Psychother Pract Res 1997; 6:12–24
10. Barber JP, Crits-Christoph P, Luborsky L: Therapist competence and treatment outcome in dynamic therapy. J Consult Clin Psychol 1996; 64:619–622
11. DeRubeis RJ, Feely M: Determinants of change in cognitive therapy for depression. Cognitive Therapy and Research 1990; 14:469–482
12. Luborsky L, McLellan AT, Woody GE, et al: Therapist success and its determinants. Arch Gen Psychiatry 1985; 42:602–611
13. Winston A, Pollack J, Flegenheimer W, et al: A brief psychotherapy fidelity scale: reliability, validity and relation to outcome. Paper presented at the Society for Psychotherapy Research, Ulm, Germany, 1987
14. Winston A, Pinsker H, Muran JC, et al: Brief supportive psychotherapy: a preliminary study of adherence and outcome. Paper presented at the American Psychiatric Association annual meeting, Washington, DC, 1992
15. Piper WE, Azim HFA, Joyce AS, et al: Transference interpretations, therapeutic alliance and outcome in short-term individual psychotherapy. Arch Gen Psychiatry 1991; 48:946–953
16. Piper WE, Azim HFA, McCallum M, et al: Patient suitability and outcome in short-term individual psychotherapy. J Consult Clin Psychol 1990; 58:475–481
17. Piper WE, Joyce AS, McCallum M, et al: Interpretive and supportive forms of psychotherapy and patient personality variables. J Consul Clin Psychol 1998; 66:558–567
18. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 3rd edition. Washington, DC, American Psychiatric Association, 1980
19. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 3rd edition, revised. Washington, DC, American Psychiatric Association, 1987
20. Piper WE: Psychodynamic psychotherapy, in American Psychiatric Press Review of Psychiatry, vol 15, edited by Dickstein LJ, Riba MB, Oldham JM (volume), MacKenzie KR (section). Washington, DC, American Psychiatric Press, 1996, pp 109–128
21. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86:420–428
22. Cronbach LJ: Coefficient alpha and the internal structure of tests. Psychometrika 1951; 16:297–334
23. First MB, Gibbon M, Williams JBW, et al: Mini-SCID. Toronto, Ontario, Multi-Health Systems Inc., 1990
24. First MB, Gibbon M, Williams JBW, et al: SCID-II PQ and Auto-SCID II. Toronto, Ontario, Multi-Health Systems Inc., 1991
25. Piper WE, Debbane EG, de Carufel FL, et al: A system for differentiating therapist interpretations and other interventions. Bull Menninger Clin 1987; 51:532–550
26. Piper WE, Joyce AS, McCallum M, et al: Concentration and correspondence of transference interpretations in short-term psychotherapy. J Consult Clin Psychol 1993; 61:586–595
27. Joyce AS, Duncan SC, Piper WE: Task analysis of "working" responses to dynamic interpretation in short-term individual psychotherapy. Psychotherapy Research 1995; 5:49–62
28. Beck AT, Steer RA: Beck Depression Inventory manual. New York, Harcourt Brace Jovanovich, 1987
29. Derogatis LR: SCL-90 administration, scoring, and procedures manual. Baltimore, John Hopkins University Press, 1977
30. Spielberger CD: Manual for the State-Trait Anxiety Inventory. Palo Alto, CA, Consulting Psychologists Press, 1983
31. Rosenberg M: Conceiving the Self. New York, Basic Books, 1979
32. Horowitz L, Rosenberg SE, Baer BA, et al: Inventory of Interpersonal Problems: psychometric properties and clinical applications. J Consult Clin Psychol 1988; 56:885–892
33. Andrews G, Singh M, Bond M: The Defense Style Questionnaire.

J Nerv Ment Dis 1993; 181:246–256

34. Luborsky L: Principles of Psychoanalytic Psychotherapy: A Manual for Supportive-Expressive Treatment. New York, Basic Books, 1984

35. Floyd FJ, Widaman KF: Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment 1995; 7:286–299

36. Binder JL: Is it time to improve psychotherapy training? Clin Psychol Rev 1993; 13:301–308

37. Henry WP, Strupp HH, Butler SF, et al: Effects of training in time-limited dynamic psychotherapy: changes in therapist behavior. J Consult Clin Psychol 1993; 61:434–440

APPENDIX A.   The Interpretive and Supportive Technique Scale

0 No Emphasis	1 Minor Emphasis	2 Moderate Emphasis	3 Considerable Emphasis	4 Major Emphasis

The therapist attempted to:

1. Gratify the patient, i.e., make the patient feel good rather than anxious in the session.

2. Maintain pressure on the patient to talk, e.g., by at times remaining passive, by not breaking pauses, by not answering questions.

3. Make noninterpretive interventions, e.g., reflections, questions, provisions of information, clarifications, and confrontations.

4. Encourage the patient to explore uncomfortable emotions.

5. Provide guidance similar to the role of family doctor, e.g., advise a course of action more appropriate to healthy functioning regarding self-care, life skills, or interpersonal behavior.

6. Make interpretations.

7. Engage in problem-solving strategies with the patient, i.e., generating and evaluating alternative solutions to external life problems.

8. Direct attention to the patient's subjective impression of the therapist.

9. Offer explanations that locate the responsibility for the patient's difficulties outside him- or herself, e.g., in the patient's environment, as a function of interpersonal transactions, or in the patient's body chemistry or physiology.

10. Make links between the patient's relationship with the therapist and the patient's relationships with others.

11. Praise the patient.

12. Focus on the patient and therapist in the treatment situation rather than the patient and significant others outside the treatment situation.

13. Display personal information, opinions, and/or values.

14. Direct attention to the patient's subjective impression of others outside the treatment situation.