Developing an Instrument for Characterizing Psychotherapy Techniques in Studies of the Psychotherapy of Borderline Patients*

Harold W. Koenigsberg, Otto F. Kernberg, Lawrence Rockland, Ann Appelbaum, Arthur Carr, and Paulina Kernberg

1. Introduction

A recent methodological advance in psychotherapy research has been the introduction of manuals to operationally define the treatment under study. Such manuals describe the prototypic form of the treatment, but can not insure that treatment actually delivered corresponds to this ideal form. Even with clearly defined treatments, deviations from the described treatment are apt to occur for several reasons. Passage of time after the initial training period may allow a drift in technique – a problem that becomes even more serious in studies of long term psychotherapy. In addition, as Mintz, Luborsky and Auerbach (1971) have shown, the actual employment of therapeutic techniques is determined by patient as well as therapist factors. Thus, a treatment defined in the abstract may take a substantially different form in the actual patient encounter. While this may be clinically advantageous in some situations or harmful in others, it complicates efforts to study a defined therapeutic approach.

An instrument to measure the therapeutic techniques actually employed is necessary to insure that the treatment under study is actually the treatment specified. In addition to its usefulness for monitoring the application of the intended techniques, an instrument for measuring techniques would also provide useful empirical data about the relative timing and mixture of techniques in actual treatment situations. In this chapter we will describe the development of such an instrument, the

*An earlier version of this manuscript was presented at the annual meeting of Society for Psychotherapy Research, Lake Louis, Canada, June 1984 and appeared in the *Journal of Nervous and Mental Diseases*. The authors thank Allison Orr-Andrawes, M.D., Charles Gardner, M.D., John Urbach, M.D., Gretchen Haas, Ph.D., John Clarkin, Ph. D., Frank Yeomans, M.D., and Andrew Lotterman, M.D. for their contributions to this work. Howard Hunt, Ph.D. provided valuable statistical consultation.

Therapist Verbal Intervention Inventory (TVII), for use in a study of the psychodynamic treatment of patients with Borderline Personality Disorder (BPD), report on its reliability, and present preliminary data on its ability to distinguish "supportive" and "expressive" psychotherapy.

Several instruments for measuring technique have been described in the psychotherapy research literature. Some are tailored to specific forms of treatment, particular patient populations, or problem areas. Strupp (1957) has described a multidimensional system for analyzing psychotherapeutic techniques. Interventions are classified along five dimensions: (1) Type of Therapeutic Activity, (2) Depth-directedness, (3) Dynamic Focus, (4) Initiative, and (5) Warmth-Coldness. Included within the first dimension are seven categories of technique: Exploratory Operations, Clarification, Interpretive Operations, Structuring, Direct Guidance, Minimal Activities, and a miscellaneous category. Strupp and colleagues (1966) introduced a number of therapist variables into a therapy rating instrument developed by Bellak and Smith (1956). Among the therapist variables are five specific technique measures: frequency of interventions, frequency of interpretations, depth of interpretation, initiative, and support. The choice of these particular variables was "intuitively derived" (p. 369). Harway et al. (1955) developed a seven point scale to measure depth of interpretations. The measurement of five specific confrontative interventions is described by Mitchell and Berenson (1970). Rounsaville, Chevron and Weissman (1984) developed scales to measure the extent to which a particular form of psychotherapy, Interpersonal Therapy (IPT), was employed in the treatment of depressed patients. Hoyt and co-workers (1981) designed a Therapist Action Scale (TAS) which measures twenty-five operationally defined therapist actions employed in the brief psychotherapy of patients with stress response syndromes.

While many of these instruments tap sets of essential psychotherapeutic techniques, some achieve poor interrater reliabilities and others are tailored to specific therapy models. In developing a scale for our study of dynamic psychotherapy of BPD patients, it was necessary to address a number of design issues.

2. General Considerations

Selection of the specific techniques to be included in the instrument should be governed by research objectives. Early technique scales selected techniques rather arbitrarily on the basis of the intuitive judg

ment of the investigators. Since a limited number of techniques may be included in any instrument, decisions must be made about which to include and which to exclude. Interventions considered essential in one model of therapy, however, may be seen merely as adventitious remarks from the the perspective of a different model. While an atheoretical approach to data collection might appear desirable, any systematic selection of categories for observation implies an underlying theoretical framework. The choice of techniques must take into account the treatment model under study and the theoretical framework underlying that treatment. All techniques necessary to define the particular therapeutic approach must be included. Since treatments may also be described in terms of interventions that are disallowed, the monitoring instrument must be designed to identify instances of such techniques as well.

A second issue in the development of a technique instrument is the choice of the conceptual level at which the interventions will be described. Two factors must be balanced – the need for a reliable rating system vs. the need for clinically meaningful constructs. Description in terms of operationally defined concrete therapist behaviors should enhance interrater reliabilities, but if carried too far, runs the risk of divorcing observation from clinical concepts.

An approach must be selected for quantification. Several strategies are available; techniques may be rated as simply present or absent, instances of use may be counted, or the use of an intervention may be measured in terms of its clinical valence within context of the session. Harway et al. (1955) have shown, for example, that "depth of interpretation" is rated differently if judged on the basis of individual therapist remarks summed over an entire session than if assessed by an overall rating the same interview. A session may be powerfully influenced by a succinct phrase uttered once.

An additional consideration is the size of the session segment to which the instrument will be applied. Short sampling units may have the advantage of increasing interrater reliabilities and of enhancing the resolving power of the instrument for identifying fine structure within the treatment. The use of sampling units that are too small, however, may break interventions into clinically meaningless fragments. Strupp and co-workers (1966) attributed the low reliabilities that they obtained with their scale to the use of too large a time unit for sampling, a full session. In a comparison of 4 minute segments and entire sessions, Mintz

and Luborsky (1971) have shown, on the other hand, that reliable ratings of certain variables, e.g. "optimal empathic relationship," could not be made on the basis of short segments.

3. Desired Features for the Therapist Verbal Intervention Inventory (TVII)

For the purposes of our study of the dynamic psychotherapy of patients with Borderline Personality Disorder, we require an instrument that addresses the full range of interventions used by psychodynamically oriented therapists in treating patients with severe character pathology. While various approaches have been proposed for the treatment of borderline patients (Kernberg 1975, 1982; Masterson 1976; Zetzel 1971) most draw from a set of basic exploratory and supportive interventions. A useful summary of these interventions has been formulated by Rosen (1974). Kernberg (1984) has proposed that purely expressive, purely supportive, or mixed supportive-expressive approaches may have differential effects. To permit testing of this hypothesis, we have attempted to develop an instrument that distinguishes these approaches.

The TVII has been designed to identify the techniques characteristic of these forms of psychotherapy. In doing so, it includes the techniques which characterize psychoanalytically-oriented psychotherapy in general. The instrument thus should have applicability for studying dynamic psychotherapy as it is used with many types of patients.

The existing instruments, described above, have been developed for use with healthier patients than those to be treated in our study. Our instrument differs in the inclusion of a number of interventions that may be called upon in work with patients of lower ego strength. Thus, limit setting, direct correction of distorted perceptions of the therapist, intervention in the patient's life situation, enlisting the aid of others, and deviating from a neutral therapeutic stance were all included in our instrument.

In addressing the question of the conceptual level at which the techniques are described, we sought to balance a need for reliability with a need for categories that relate to the clinical literature. Experimental studies of psychotherapy can be valuable in providing empirical grounding for abstract theory. To achieve this, the experimental cate

gories should be designed as bridges between raw experiential data and accepted clinical constructs. For this reason, we have chosen to define our categories at a level of abstraction somewhat above that of operationally defined therapist behaviors. This choice was made in the belief that raters with some clinical sophistication could reliably rate interventions at an intermediate level of inference. The reliability studies described in this article afforded an opportunity to test this assumption.

Since the impact of a particular technique within a session depends not only upon the number of times the technique is invoked, but also upon phrasing, timing, overall context and other factors, we have chosen to rely upon the raters' clinical judgment of the "relative emphasis" placed upon a particular technique within the sampling unit. The extent to which each technique is utilized is quantified on a scale from 0 (absent) to 5 (maximally present). Drawing from the experience of Strupp et al. (1966) and Mintz and Luborsky (1971), we have chosen 15 minute segments as sampling units. We have restricted the instrument to verbal interventions. We seek to measure interventions as they are invoked by the therapist, independent of their effect upon the patient; thus the TVII instructions specify that the techniques are to be rated without regard to the patient responses. Since an individual statement by the therapist may serve several functions at once, TVII raters are permitted to score a given intervention in more than one category.

4. Development of the Instrument

Items for inclusion in the TVII were derived from review of related instruments (Harway et al. 1955; Hoyt et al. 1981; McNair and Lorr 1964; Mitchell and Berenson 1970; Rounsaville et al. 1984; Strupp 1957; Strupp et al. 1966), a survey of interventions described in the psychodynamic literature (Rosen 1974), and the clinical experience of the authors. An initial version of the scale was designed and distributed to eleven clinicians experienced in dynamic psychotherapy. This group included two clinical psychologists and nine psychiatrists with a range of 5 to 28 years of post-training clinical experience; six group members were psychoanalysts. The group applied the instrument to videotaped psychotherapy sessions. They viewed treatment sessions of four BPD patients conducted by three therapists. Group discussions of these sessions led to refinement of the instrument. Items relating to reiteration of the basic treatment contract and to arranging the practical details of treatment were added. In addition the rating instructions were clarified.

The authors developed a manual for the instrument, which provided definitions of the items and illustrative clinical examples.

5. Assessment of the TVII in Expressive Therapy

The individual items of the instrument were assessed for interrater reliability and agreement. In addition, a study was carried out to determine whether the instrument could be taught to an independent group of raters who had not participated in its development. This work is reported in detail elsewhere (Koenigsberg et al. 1985) and will be briefly summarized here.

Study 1: Interrater Reliabilities – Experienced Clinicians

Eight members of the research team, who had participated in the design of the instrument, applied the TVII to videotaped segments of psychotherapy sessions. These raters were experienced clinicians (6 psychiatrists and 2 clinical psychologists with from 5 to 28 years of postgraduate clinical experience). A total of 11 segments were rated. The segments were 15 minute portions of 45 minute interviews, drawn from the recorded treatments of two patients by one of the authors (O.F.K.). Both patients carried diagnoses of personality disorders. Each treatment was characterized by the treating therapist as an "expressive psychotherapy" in which relatively few supportive techniques were employed.

For each intervention, the frequency of its use throughout the 11 segments was determined by averaging, over the raters, the proportion of segments in which the rater scored it as present. The frequencies of occurrence for the 35 TVII items are displayed in Table I. The distribution of frequencies indicates that 16 techniques were used less than 20 percent of the time, while the remaining techniques were rather evenly distributed in frequency of use from 30 percent to 100 percent. The 16 infrequently used techniques group into two classes. Ten are supportive techniques: advice, sympathy, encouragement, cognitive support of defenses, emotional support of defenses, correcting transference distortion, deviating from neutrality, limit setting, intervening directly in the patient's life, and enlisting the assistance of others in intervening. Four are techniques that would be expected more often in expressive (insight) psychotherapy: clarification, confrontation, interpretation of childhood events, and genetic reconstruction (linking past, present and transference).

Table 1 Interrater Reliabilities and Agreements for TVII Items in Two Studies

TVII ITEM			Research Group		Independent Raters		
			Freq r	Finn's[a]	Freq r	Finn's[b]	T[c]
1a.	Provides factual information	- concerning external reality	39%	.76	36%	.90	.68
1b.		- concerning practical arrangements of treatment	13%	.99*	21%	.96*	.68
1c.		- concerning general psychological processes	30%	.81	11%	.93*	.84
2.	Gives advice		16%	.85	0%	--	--
3a.	Seeks clarification	- of the transference	74%	.58	64%	.80	.68
3b.		- of external reality	41%	.72	93%	.80	.53
3c.		- of the patient's childhood	7%	.95	0%	--	--
3d.		- of the present defenses	53%	.58	32%	.87	.68
3e.		- of the patient's internal reality	74%	.71	71%	.64	NS
4.	Offers sympathy		7%	.97	4%	.99*	1.00
5.	Offers encouragement		9%	.89	4%	.99*	1.00
6a.	Confronts	- in the transference	94%	.71	68%	.87	.68
6b.		- in the external reality	49%	.44	46%	.64	.37
6c.		- in the patient's childhood	6%	.98	0%	--	--
6d.		- in the present defenses	85%	.55	29%	.81	.53
6e.		- in the patient's internal reality	84%	.48	39%	.42	.53
7.	Supports defense cognitively		2%	.98	14%	.95*	1.00
8.	Supports defense emotionally		3%	.99*	0%	--	--
9a.	Interprets "here-and-now" of the transference	- in the	82%	.57	32%	.83	.68
9b.		- in the external reality	34%	.56	25%	.73	.68

Table 1 (Continued)

		Research Group		Independent Raters		
		Freq	Finn's[a] r	Freq	Finn's[b] r	T[c]
9c.	- in the patient's childhood	8%	.93	0%	--	--
9d.	- in the patient's defenses	71%	.56	39%	.65	.37
9e.	- in the patient's internal reality	63%	.69	11%	.85	.84
9f.	- in linking the observations of a), b), c), d), and/or e) with assumed unconscious past (genetic reconstruction)	15%	.83	0%	--	--
10.	Stresses reality to reduce transference distortion	56%	.71	25%	.93*	.84*
11.	Reduces transference distortion by deflection onto extra-transferential objects	5%	.95	0%	--	--
12.	Deviates from technical neutrality	6%	.84	0%	--	--
13.	Intervenes to reinstate technical neutrality	0%	--	0%	--	--
14.	Sets limits within the hours	9%	.83	0%	--	--
15.	Directly intervenes in patient's life	6%	.80	0%	--	--
16.	Intervenes in patient's life by enlisting the aid of others	0%	--	0%	--	--
17.	Reinforcement of treatment contract	64%	.73	21%	.96*	1.00*
18.	Focus on external reality	47%	.75	100%	.88	.84*
19.	Focus on transference	97%	.74	75%	.82	.53*
20.	Therapist's relative verbal activity	100%	.82	100%	.78	.53*

[a] All Research Group Finn's r values are significant at $p < .01$, $df = 77$, except those marked *.

[b] All Independent Raters Finn's r values are significant at $p < .01$, $df = 21$, except those marked *.

[c] All T values are significant at $p < .05$ except those marked *

This distribution of techniques is consistent with the therapist's report that he was conducting an expressive treatment and with the fact that exploration of childhood antecedents and genetic reconstruction customarily occur in the later phase of such treatments (Kernberg 1975). Since the raters were not blind to the therapist's intended technique in this study, the agreement between the TVII ratings and the therapist's stated technique may be artifactual and should not be taken as evidence of validity. This issue is reexamined in the second study, reported below, where the raters were blind to the therapist's intention.

For each variable, Finn's r gives the agreement among the raters across the 11 segments on the presence or absence of that variable (Finn 1970). It takes account of the possibility of chance agreement. The Finn's r values are presented in Table 1.

Study 2: Interrater Reliabilities – Independent Raters

A second study was undertaken to determine if an independent group of clinicians with less extensive clinical experience could be trained to use the TVII reliability.

Three raters who had not participated in the development of the TVII were trained by one of the authors (H.W.K.) in use of the instrument. Two raters were in their first year post psychiatry residency and one was a third year psychiatric resident. Training consisted of an initial review of the TVII and its manual, followed by practice rating of videotaped segments. The raters were encouraged to make notes of the specific interventions that formed the basis of their ratings. These interventions were reviewed and the scoring discussed. The raters received about five hours of training, before the reliability testing began.

Scoring was based on seven 15-minute psychotherapy segments. The patient was a woman with a character disorder and the treatment was characterized as a brief expressive psychotherapy by the treating therapist (O.F.K.). The raters were blind to the therapist's description of his technique. Segments were drawn at random from the first eight sessions.

Occurrence frequencies for each of the 35 techniques were calculated as above and are reported in Table 1. Seventeen techniques had a frequency of occurrence of less than 20 percent. These items were identical to the low frequency items in Study 1, with two exceptions: providing factual information about general psychological processes

replaced factual information about treatment, and interpretation addressing the patient's internal reality was added. The low frequency of these items, with the exception of the last, is consistent with the early phase of an expressive psychotherapy. The TVII, then, correctly characterized the treatment, a finding supportive of the validity of the instrument. The low occurrence rating for the last item may be explained by the fact that the raters consistently reported difficulty in identifying interventions that addressed "internal reality."

Interrater reliabilities were calculated using Finn's r.[1] The results are displayed in Table 1. All but one of the values ranged from .64 to .99; the single low reliability of .42 was associated with the item "confrontation in internal reality." Since a spuriously high interrater agreement is possible for techniques rarely used, the high r value for this technique should be viewed with caution.

We have also examined the interrater agreement achieved with the TVII, using the approach of Lawlis and Lu (1972) and the value, T, defined by Tinsley and Weiss (1975). For our six point rating scale, we consider raters to agree if they come within one scale point of each other for a given rating. In calculating the extent of agreement, the probability for chance agreement has been conservatively estimated, using the values provided by Lawlis and Lu (1972) for a 5-point scale. Where chi-square indicates that the agreement is significantly different from chance, we have calculated the T values. These are displayed in Table 1.

6. The TVII Applied to Supportive Therapy

With the achievement of acceptable interrater reliabilities in our pilot studies with expressive psychotherapy sessions, we applied the instrument to a broader range of treatments conducted by a number of different therapists. The therapies examined were characterized as "supportive" or "mixed supportive-expressive" by the treating therapists. These treatments were conducted by psychoanalysts with experience and special interest in supportive psychotherapy.

On the basis of early experience applying the TVII to these treatments, we empirically identified a number of additional supportive interventions that had not been included in the original TVII. These included techniques in which the therapist explicitly offers himself as a

[1]Finns' r (Finn 1970) is an estimate of the reliability of judges assigning items to a set of categories.

model, directly fulfills the patient's stated or inferred wishes, seeks help or advice from the patient, discloses information about himself, and actively discourages certain behaviors, attitudes, or feeling states. These techniques were defined operationally and incorporated into version 2 of the TVII.

Study 3: Reliability of Version 2 of the TVII

We examined the reliability of the expanded version of the TVII by obtaining ratings of 16 segments of "supportive" and "mixed supportive-expressive" sessions conducted by three different therapists and involving five different patients. Three experienced members of the research group rated the sessions. Interrater reliabilities were calculated using Finn's r. These values are displayed in Table 2, along with the frequencies of use of each intervention. Finn's r values for techniques endorsed as present in fewer than 20 percent of the segments are enclosed in parentheses since low frequency interventions may have spuriously high interrater reliabilities. Excluding the infrequent interventions, reliabilities range from .97 to .64.

Study 4: A Comparison of "Expressive" and "Supportive" Sessions

In order to determine whether psychotherapy sessions identified as "expressive" and "supportive" by the therapists could be distinguished in terms of the interventions identified by the TVII, we compared the frequency of use of 27 specific techniques in the two types of treatment. We report here on some preliminary findings from a comparison of eight segments taken from "supportive" psychotherapy sessions with eleven segments from "expressive" psychotherapy sessions. The "supportive" segments were drawn from three psychotherapies conducted by two different therapists (A.A. and P.K.). The "expressive" segments were drawn from psychotherapy sessions of two patients conducted by the same therapist (O.F.K.). All therapists were psychoanalysts with over ten years of experience.

Table 2 Interrater Reliabilities for TVII - Version 2

TVII ITEM	Research Group Frequency (%)	Finn's r^d
1. Informs, instructs, educates:		
a) External Reality	25	.94
b) Psychological Processes	21	.90
2. Informs, instructs re treatment:		
a)General conditions	23	.95
b)Arrangements	33	.97
c)Behavior between sessions	6	(.97)*
d)Behavior within sessions	10	(.96)
3. Seeks clarification:		
a) Transference	35	.86
b) External reality	98	.72
c) Childhood	13	(.91)
d) Defenses	48	.63
e) Internal reality (affects, fantasies)	77	.71
4. Reinforces treatment contract	2	(.99)*
5. Accepts, confirms, suggests, praises:		
a) Defenses	33	.77
b) Superego	13	(.97)*
c) Therapeutic alliance	4	(.96)
d) Positive transference	4	(.99)*
e) Negative transference	4	(.96)
f) Feelings states	73	.70
e) Actions	42	.72
6. Dissuades, criticizes, rejects, prohibits:		
a) Defenses	10	(.91)
b) Superego	4	(.99)*
c) Positive transference	4	(.96)
d) Negative transference	6	(.97)*
e) Destructive attitudes behavior	8	(.92)
f) Feeling states	4	(.96)
g) Actions	13	(.89)
7. Sets limits:		
a) Within session	0	0
b) Outside session	0	0

Table 2 (Continued)

		Research Group Frequency (%)	Finn's r^d
8.	Offers sympathy	6	(.98)*
9.	Expresses hopefulness	10	(.96)*
10.	Confronts:		
	a) Transference	29	.94
	b) External reality	19	(.93)
	c) Childhood	4	(.99)*
	d) Defenses	27	.86
	e) Internal reality	35	.80
11.	Intervenes in patient's life:		
	a) Collaterals	0	--
	b) Medication	0	--
	c) Direct help	0	--
12.	Explains deviation from technical neutrality	0	--
13.	Stresses reality of relationship:		
	a) Response to negative feelings	8	(.93)
	b) Response to positive feelings	0	--
14.	Interprets:		
	a) Transference	21	.91
	b) External reality	17	(.87)
	c) Childhood	15	(.96)
	d) Defenses	33	.74
	e) Internal reality	33	.67
	f) Linking above with past	13	(.94)
15.	Directs focus away from therapist	6	(.94)
16.	Reveals information about self:		
	a) Factual data	17	(.94)
	b) Feeling states	15	(.98)*
	c) Confirms patient's perceptions	8	(.98)*
17.	Accepts help from patient	2	(.99)*
	GLOBAL ASSESSMENTS:		
18.	Focus on external reality	98	.76

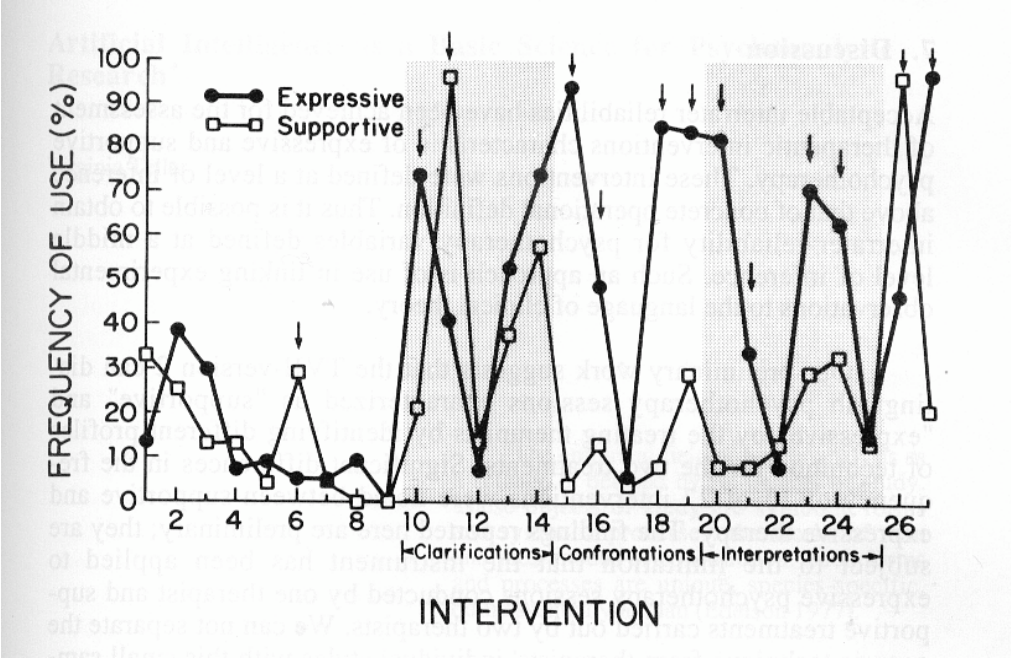
Table 2 (Continued)

		Research Group Frequency (%)	Finn's r^d
19.	Focus on transference	42	.85
20.	Therapist verbal activity	100	.88
21.	Deviation from technical neutrality	27	.81
22.	Restore technical neutrality	0	--
23.	Fulfills patient's wishes:		
	a) Stated	15	(.94)
	b) Inferred	29	.81
24.	Offers self as model	15	(.94)
25.	Supports defenses	31	.76

[d] All Finns's r values significant at $p < .05$ except those marked *.
( ) = Item endorsed <20% of the time.

The frequency of use of each intervention in each treatment was defined as the ratio of the number of segments of that treatment in which a rater scored the technique as present divided by the total number of segments of that treatment and averaged over all raters' assessments. The frequency of use of each of the 27 techniques examined is plotted for each type of therapy in Figure 1. This provides a graphic TVII profile of each form of treatment. The chi-square statistic was used to identify significant differences in the frequencies of use of each technique between the two types of treatment. Significant differences ($p < .05$) were identified for thirteen of the interventions.

The "supportive therapy" used significantly more support of defenses (technique 6), clarification in the external reality (technique 11), and overall focus on the external reality (technique 26). "Expressive therapy" used more clarification of the transference (technique 10), confrontation of the transference (technique 15), confrontation in the external reality (technique 16), confrontation of defenses (technique 18), confrontation in the internal reality (technique 19), interpretation of the transference (technique 20), interpretation in the external reality (technique 21), interpretation of the defenses (technique 23), interpretation of internal reality (technique 24), and overall focus on the transference (technique 27).



Arrows indicate significant differences ($p < .05$).

Key to Type of Intervention:

1. Provides Information about Treatment Arrangements, 2. Provides Information about other Reality Issues, 3. Provides Information about Psychological Processes, 4. Offers Sympathy, 5. Encourages/Hopefulness Expressed, 6. Supports Defenses, 7.Deflects Focus Away from Therapist, 8. Sets Limits in Sessions, 9. Intervenes in Patient's Life, 10. Clarifies Transference, 11. Clarifies External Reality, 12. Clarifies Childhood Experiences, 13. Clarifies Defenses, 14. Clarifies Internal Reality, 15. Confronts Transference, 16. Confronts External Reality, 17. Confronts Childhood Memories, 18. Confronts Defenses 19. Confronts Internal Reality, 20. Interprets Transference, 21. Interprets External Reality, 22. Interprets Childhood Experiences, 23. Interprets Defenses, 24. Interprets Internal Reality, 25. Interpretation linking any of items 20 - 24 above, 26. Overall Focus on External Reality, 27. Overall Focus on Transference

Figure 1 TVII Profiles of Supportive and Expressive Psychotherapy Sessions.

7. Discussion

Acceptable interrater reliabilities have been achieved for the assessment of therapeutic interventions characteristic of expressive and supportive psychotherapy. These interventions were defined at a level of inference above that of concrete operational definition. Thus it is possible to obtain interrater reliability for psychotherapy variables defined at a middle level of inference. Such an approach is of use in linking experimental observations to the language of clinical theory.

Our preliminary work suggests that the TVII-version 2 can distinguish psychotherapy sessions characterized as "supportive" and "expressive" by the treating therapists by identifying different profiles of technique for the two treatments. Significant differences in the frequency of 13 of 27 interventions were noted between supportive and expressive therapy. The findings reported here are preliminary; they are subject to the limitation that the instrument has been applied to expressive psychotherapy sessions conducted by one therapist and supportive treatments carried out by two therapists. We can not separate the generic technique from therapists' individual styles with this small sample of psychotherapists. Nevertheless, the differences identified – more focus upon the transference and use of confrontation and interpretation in the expressive treatment and more focus upon external reality and support of defenses in the supportive treatment – are consistent with the treating therapists' own conceptualizations of their treatments.

The TVII appears to be a potentially useful tool for monitoring pychotherapeutic techniques in ongoing studies of the treatment of patients with severe personality disorders. We have been able to train therapists at the advanced resident or immediate post-resident level to use the TVII to rate videotaped sessions reliably. Further work with a variety of treatment approaches applied by a range of therapists is necessary to confirm its construct validity.