Testing Hypotheses of Psychotherapeutic Change Processes*

George Silberschatz, John T. Curtis, Polly B. Fretter, and Thomas J. Kelly

1. Introduction

How the therapist influences the process and outcome of psychotherapy is a widely debated topic in the psychotherapy research literature. Despite a large body of research (see Parloff, Waskow, and Wolfe 1978; Schaffer 1982 for reviews) surprisingly little progress has been made in understanding how the therapist contributes to the success or failure of psychotherapy. In this paper we will present the argument that the lack of progress is due to (1) inadequate conceptualization of how therapist interventions affect particular patients, and (2) imprecise, overly global methods of evaluating therapist behaviors. In particular, the relevance or suitability of the therapist's behavior to the particular problems and needs of a given patient has not been adequately assessed.

We will briefly review some of the literature on how the therapist influences psychotherapy and discuss some of the methodological problems in these studies. We will then describe a new conceptual approach developed by Weiss (1986), for understanding how therapist behaviors affect patients and show how this approach leads to more precise methods of assessing the therapist's contribution to psychotherapy. Research studies of psychoanalysis and of brief dynamic psychotherapy using this approach will be presented and a procedure for testing alternative psychoanalytic hypotheses about how the therapist's behavior affects the patient's progress will be described.

*Preparation of this chapter was supported by a grant from the National Institute of Mental Health (MH35230) and by the Research Support Program of Mount Zion Medical Center.

2. Background

Most theories of psychotherapy assume that the therapist plays a significant role in the change process and that specific therapist behaviors and techniques constitute effective ingredients of the treatment. Psychoanalytic theories place a strong emphasis on the therapist's facilitating patient insight through interpretation: "Insight through interpretation is the supreme agent in the hierarchy of therapeutic principles characteristic of analysis" (Bibring 1954, p. 747). The interpretation of transference is thought to be a particularly potent intervention by many psychoanalytic writers (e.g. Gill 1982; Greenson 1967; Loewald 1960, 1971; Macalpine 1950; Malan 1963, 1976a, 1976b; Stone 1967). Nonetheless, empirical evidence for the efficacy of specific therapist techniques (including interpretation) is lacking or inconsistent (Clairborn 1982; Gomes-Schwartz 1978; Luborsky et al. 1980; Parloff et al. 1978). For example, Garduk and Haggard (1972) showed that in psychoanalytic therapy interpretations increase patient productivity more than non-interpretations. However, in another study (Sloane, Staples, Cristol, Yorkston, and Whipple 1975), the opposite was found to be true: patients who received more interpretations showed *less* improvement at outcome than did patients who received fewer interpretations.

The failure to demonstrate the differential effectiveness of one therapy technique over another has led some investigators to focus on so-called "non-specific factors" that are thought to operate in all helping relationships (e.g. Frank 1982; Strupp 1978b). These non-specific variables include therapist behaviors that facilitate an accepting, therapeutic atmosphere (e.g. therapist warmth, friendliness, empathy, encouragement). While early studies of these variables seemed promising (Rogers 1957; Truax 1963; Truax and Mitchell 1971), more recent studies have failed to demonstrate a consistent relationship between these non-specific factors and treatment outcome (e.g. Gomes-Schwartz 1978; Luborsky et al. 1980; Mitchell, Bozarth, and Krauft 1977; Orlinsky and Howard 1978; Parloff et al. 1978).

Some reviewers have suggested that the difficulty measuring the therapist's influence is due to the inadequate methodology employed in most studies of the psychotherapeutic process (Elliott 1983, 1984; Fiske 1977; Gottman and Markman 1978; Greenberg 1986; Parloff et al. 1978; Schaffer 1982). According to these reviewers, most psychotherapy research is based on the erroneous assumption that there is little variability between or within patients, therapists, and treatments. Kiesler

(1966) referred to this assumption as the homogeneity myth and showed how this false assumption has weakened most studies of psychotherapy (see also Gottman and Markman 1978). Just as patients and therapists differ substantially, so too does the meaning of various process events. A prevailing assumption in much of the previous literature is that:

> . . . given client or therapist behavior is either 'good' or 'bad' without regard to the context in which it appears. This fits poorly with the observation of experienced therapists that a given kind of therapist response or client performance seems to be crucial at one point and irrelevant or even detrimental at another. (Rice and Greenberg 1984, p. 10).

A strategy which avoids the uniformity assumption involves identifying clinically significant critical incidents or key events within therapy sessions. This method, initially applied to psychotherapy research by Luborsky (1967, 1970; Luborsky and Auerbach 1969; Luborsky, Singer, Hartke, Crits-Christoph, and Cohen 1984) in his "symptom-context" studies, has been successfully applied in a number of psychotherapy process studies (e.g. Elliott 1983, 1984; Fretter 1984; Gassner, Sampson, Weiss, and Brumer 1982; Greenberg 1982, 1986; Horowitz, Sampson, Siegelman, Wolfson, and Weiss 1975; Mahrer 1985; Rice and Greenberg 1984; Silberschatz 1978, 1986; Silberschatz, Fretter, and Curtis 1986a). With this type of approach – recently designated the "events paradigm" in process research (Stiles, Shapiro, and Elliott 1986) – the investigator identifies significant episodes within the therapy session (e.g. the patient's describing a particular problem or conflict) and evaluates the extent to which the therapist's intervention facilitates problem resolution. The events model of research is:

> . . . geared to the goal of understanding. Rather than assuming a given kind of process to have equal significance at any point in therapy, this new approach relies on the segmentation of therapy into different episodes or events in order to understand process in the context of these clinically meaningful units. This approach of breaking therapy down into classes of recurring events prevents one from being swamped by the data through a selective focus on those episodes in the therapeutic interaction that hold promise of illuminating the change process. (Rice and Greenberg 1984, p. 11)

The events approach asks, "Which specific therapist interventions, introduced in which momentary therapeutic contexts, will lead to which immediate and subsequent impacts (outcomes) for the client?" (Stiles et al. 1986, p. 174).

The focus on key change episodes in therapy is a clear advance over previous homogenized approaches to psychotherapy research. However, the events approach does not address another methodological problem evident in most studies of how therapist's behaviors influence psychotherapy – namely, the quality or suitability of the therapist's behavior. The events paradigm may lead an investigator to focus on critical incidents in therapy – for instance, significant transference re-enactments – but it does not provide a framework for determining whether the therapist's interventions in response to these key episodes are well suited to the patient's particular needs. For instance, if a patient fails to respond to a therapist's interpretation it could be because interpretation as a technique is ineffective or because the particular interpretation (or line of interpretation) is not pertinent to that patient. Thus, what is needed is a procedure for evaluating the "goodness-of-fit" between the therapist's behavior and the patient's particular problems and treatment goals.

3. A Framework for Determining the Suitability of Therapist Behaviors

How does one develop a measure of suitability? Some investigators have proposed that variables such as therapist "skillfulness" (Schaffer 1982) or "generic helpfulness" (Elliott 1984) can be rated using broadly defined, "generic" rating scales. However, such a broad approach does not recognize that interventions that are helpful to one patient may not be helpful to another. The helpfulness of an intervention can not be meaningfully determined without first knowing the specific needs of a particular patient. Thus, assessing the quality or suitability of therapist behaviors requires:

(1) identifying what the patient's problem(s), need(s), and therapy goal(s) are; and

(2) determining whether any given intervention appropriately addresses the patient's problem(s) or facilitates the attainment of the goal(s).

The concept of suitability used by the Mount Zion Psychotherapy Research Group (Weiss, Sampson, and the Mount Zion Psychotherapy Research Group, 1986; see also, Curtis and Silberschatz 1986; Silberschatz and Curtis 1986; Silberschatz et al. 1986b) is based on a theory developed by Weiss (1986). Weiss has proposed that psychopathology stems from unconscious pathogenic ideas or false beliefs that are typically based on traumatic childhood experience. According to Weiss, patients enter psychotherapy with an unconscious plan for solving problems and disconfirming pathogenic beliefs. The patient's plan may be thought of as an unconscious strategy for disconfirming certain pathogenic beliefs by developing greater understanding of them in the therapy and by testing them in the relationship with the therapist.

In testing a pathogenic belief, the patient carries out a trial action which is intended to provide information about the belief. For example, a patient whose parents were bothered by his autonomous strivings might develop the belief that his autonomy is harmful or upsetting to others and thus might stifle certain desires and needs. This patient might test the belief that his autonomous behaviors are harmful by behaving independently in the therapy (e.g. by coming up with his own insights, being late to sessions, ignoring the therapist's comments) to see if the therapist can comfortably tolerate these behaviors. If the therapist does not become defensive or act critically toward the patient, he passes the test; that is, the therapist's behavior disconfirms the patient's false idea that his autonomy would hurt the therapist. On the other hand, the therapist might fail the test by acting in a way that the patient would experience as being hurt by or critical of his independence. Such a response would tend to confirm the patient's pathological expectation that it is dangerous to act autonomously and to disagree. This example of testing illustrates one prominent way in which patients may work to disconfirm pathogenic beliefs. Tests may vary according to their relevance to central pathogenic beliefs; *key tests* are those tests that are most critical to the patient because they are central to the pathogenic beliefs which the patient is working to disconfirm (Silberschatz and Curtis 1986).

According to Weiss's theory, the therapist's attitudes, overall style of response, and particular interventions can help the patient relinquish pathogenic beliefs. Therapeutic styles or interventions that are consistent with the patient's unconscious plan are considered to be "plan compatible," while interventions that are counter to the patient's plan are "plan incompatible." A therapist can help a patient disconfirm a pathogenic belief by passing the patient's test or by intervening in a plan-compatible fashion; therapeutic progress tends to follow such plan-com-

patible interventions. Interventions which are incompatible with the patient's plan or fail a patient's test tend to be followed by therapeutic retreat. Studies by the Mount Zion Psychotherapy Research Group have shown that patients' plans can be reliably inferred (Bush and Gassner 1986; Caston 1986; Curtis, Silberschatz, Sampson, Weiss, and Rosenberg, 1988; Rosenberg, Silberschatz, Curtis, Sampson, and Weiss 1986), and that formulations of patients' plans can be used to assess the suitability of therapist behaviors (Fretter 1984; Silberschatz 1986; Silberschatz et al. 1986a).

The work of the Mount Zion Psychotherapy Research Group (Weiss and Sampson 1986) thus provides a framework for assessing how the therapist's behavior influences the patient's therapeutic progress. It suggests two different types of significant events in therapy – patient initiated events (key tests) and therapist initiated episodes (interpretations) – when therapist interventions are likely to be instrumental in effecting change. Weiss's theory provides a conceptual framework for assessing the goodness-of-fit between therapist behaviors (passing/failing tests, plan-compatible/plan-incompatible interventions) and the patient's particular problems and goals. We will now report several studies that illustrate how these concepts have been applied to studies of psychoanalysis and brief dynamic psychotherapy.

4. An Empirical Study of Significant Events in Psychoanalysis

The Patient's Tests of the Analyst

As noted, Weiss (1986) has identified the testing of the therapist by the patient as one type of critical incident or significant event in psychoanalytic therapy. The therapist's response to the patient's tests is thought to play a decisive role in the process and outcome of therapy: if the therapist's response to the patient's test is perceived by the patient as disconfirming a pathogenic belief (passing the test), the patient will feel relieved, reassured, and is likely to feel less anxious and more productive in the therapy. If the therapist's response confirms the pathogenic beliefs (fails the tests), the patient is likely to feel distressed and may show signs of therapeutic retreat.

A study (Silberschatz 1978, 1986) designed to test this hypothesis was carried out on the verbatim transcripts of the first 100 hours of a tape-recorded psychoanalysis. The patient, Mrs. C, a 28-year-old professional woman with an obsessive-compulsive personality structure,

sought treatment because of her inability to enjoy sexual relations with her husband. The analysis, carried out by an experienced (Freudian) psychoanalyst, was completed long before this study was planned. The method for the study involved three steps: (1) identification of the patient's key tests (significant events); (2) ratings of the degree to which the analyst's response to these tests would be perceived by the patient as disconfirming the belief which the patient was testing (passing/failing the test); and (3) assessment of the patient's behavior and affects immediately before and after the test in order to determine whether the patient changed in the predicted direction.

The first step entailed selecting a pool of all possible tests. Nine judges read verbatim transcripts of the first 100 hours of Mrs. C's psychoanalysis and selected all instances in which the patient attempted to elicit a response from the therapist. It was assumed that many tests would have this form and that such instructions would help judges select a large pool of potentially relevant test episodes. In all, 87 such instances were identified. Typescripts of the patient's attempts to elicit a response as well as the therapist's interventions (which included silences) were then prepared. Three psychoanalytically trained judges (who were familiar with the concept of testing) read a formulation of Mrs. C's plan (which had been reliably developed as part of a separate study; see Caston 1986) and identified which of the pool of incidents represented the patient's key tests of the analyst. A sample of 46 episodes was selected by all three judges as instances of key tests.

The second step used four experienced psychoanalysts who were familiar with the clinical application of the testing concept. These judges read the plan formulation and then independently rated (on a 7-point scale) the extent to which the analyst had passed or failed each key test.

The immediate effects on the patient of the therapist's passing or failing a test were assessed using several patient measures: ratings of the patient's level of experiencing (i.e., degree of involvement and productivity; see Klein, Mathieu, Gendlin, and Kiesler 1970), boldness, relaxation, and an affect classification system which measured the patient's level of fear, anxiety, love, and satisfaction (Dahl 1978, 1979b; Dahl and Stengel 1978). A segment of speech preceding the test sequence (pre-segment) and a segment of speech following the test sequence (post-segment) were rated on each of the measures by different groups of judges. The segments (approximately six minutes of patient speech) were presented in random order without any context and with the judges unaware whether the segment was a "pre-test" segment or a "post-test"

segment. Reliabilities for all of the above ratings were adequate, ranging from .65 to .94.

Correlations between the ratings of the therapist's intervention (the degree to which he passed or failed a key test) and changes (residualized gain scores – Cohen and Cohen 1975) in each of the patient measures indicated that the patient became significantly more involved, more productive, and more relaxed when the therapist passed a key test (Table 1). These results support the hypothesis that the patient was reassured by the therapist's passing tests and that the patient's satisfaction was demonstrated by her becoming less anxious, more involved, and more productive in the analytic work.

Table 1 Correlation between the Degree to which the Analyst Passed the Patient's Tests and Changes in the Patient Measures for the Key Tests

Measure [a]	r
Experiencing	.33*
Boldness	.32*
Relaxation	.35*
Love	.37*
Fear	-.34*
Satisfaction	.15*
Anxiety	-.29*

[a] N = 46.

* $p < .05$, two-tailed test

(Data for this table are taken from Silberschatz, G. et al. (1986b). Testing pathogenic beliefs. In J. Weiss, H. Sampson, and the Mount Zion Psychotherapy Research Group, *The Psychoanalytic Process: Theory, Clinical Observation, and Empirical Research*. Guilford: New York.)

The results of this study on patient tests also have implications for identifying key events. Correlations between therapist behaviors and the patient's immediate responses were not significant for the larger sample

of 87 instances (i.e., those episodes broadly defined as the patient's attempts to elicit a response). However, correlations were significant for the subsample of episodes which were *directly pertinent* to the patient's plan – i.e., the sample of 46 *key* tests. This pattern of findings suggests that significant events in psychotherapy must be identified in a case-specific fashion and cannot be adequately identified using broad criteria (e.g. patient demands). A clinical formulation of the patient's particular problems, needs, and treatment goals is needed to identify events that are most significant for a given patient.

5. An Empirical Study of Alternative Hypotheses of the Psychoanalytic Process

Selection of key events in psychotherapy sessions is inevitably based on theories of psychotherapeutic change. Just as the practicing clinician bases his interventions on a conceptual framework, psychotherapy researchers have argued that empirical studies of psychotherapy must be guided by theory (e.g. Bergin and Lambert 1978; Gendlin 1986; Strupp 1986; Yeaton and Sechrest 1981). In a review paper on dimensions of successful treatments, Yeaton and Sechrest (1981) pointed to the importance of powerful theories in practice and in research studies: "What we need is good theory, in the sense of an understanding of the mechanisms relating the causes and the problems as well as the presumed manner by which the treatment alleviates the problem" (p. 157). A well articulated theory both delineates significant events and specifies the kinds of therapist interventions during those key events that are likely to be helpful. This degree of specificity is necessary to test a theory.

Within psychoanalysis, Weiss (1986) has identified two distinct models that meet the above criteria for testability. Because the two models differ in their predictions about how a patient responds in a particular situation, it is possible to test empirically which theory better fits observation. We will briefly describe an empirical study which tested these competing hypotheses in a psychoanalytic case (Silberschatz 1978; Silberschatz, Sampson, and Weiss 1986b).

The study focused on an event frequently observed in psychoanalytic treatment: the patient's transference demands – those instances in which the patient, either overtly or covertly, makes a demand on the analyst to respond to him in some particular way. The patient may, for example, demand affection, special attention, advice, criticism, punishment, rejection, or humiliation. A central aspect of the psychoanalytic

theory of therapy is that the analyst should maintain an analytic or neutral stance and should not accede to the patient's demands. There are, however, two fundamentally different theories regarding the nature of the patient's transference demands. These theories contain different explanations for the therapeutic value of the analyst's not acceding to the patient's demand and make different (opposite) predictions about how the patient is likely to behave if the analyst does or does not accede to the demands.

The first explanation is based on what Weiss (1986) has termed an Automatic Functioning (AF) model. According to this model, the patient makes a transference demand in order to gratify an unconscious wish. When the analyst does not accede to the patient's demand, the patient's unconscious wish (transference longing) is frustrated. As a result, the wish is intensified and is pushed into consciousness. The other explanation is referred to by Weiss (1986) as a Higher Mental Functioning (HMF) model. According to this model, when a patient makes a demand of the analyst he does so primarily to test a pathogenic belief. For example, the patient may demand advice from the therapist to test the distressing pathogenic belief that the therapist, like a parent in childhood, wishes to run his life. If the therapist does not accede to this demand, the patient will feel reassured, more relaxed, and more productive in the therapy.

Both the AF and HMF models agree that the analyst should maintain a neutral stance in response to the patient's demands. However, they differ sharply in explaining how the analyst's neutrality is helpful to the patient. In fact, these two models make opposite predictions about the patient's affective response. According to the AF model, a patient would be likely to feel unhappy, distressed, upset (frustrated) by the analyst's neutrality. This hypothesis was explicitly stated as one of the formal predictions made in the Menninger Foundation Psychotherapy Research Project:

> . . . patients whose neurotic needs are not gratified within the transference respond to this frustration with regressive and/or resistive reactions, and/or painful affects . . . (Sargent, Horowitz, Wallerstein, and Appelbaum 1968, p. 85)

By contrast the HMF model predicts that the patient is generally reassured by the analyst's not acceding to the demand (because doing so disconfirms a pathogenic belief), and that the patient's satisfaction is often demonstrated by his becoming more relaxed and productive in the session.

Because the two models differ in their predictions about the patient's response to the analyst's neutral stance, it was possible to test empirically which model better fits observation. Is the patient frustrated and distressed as the AF model predicts, or is the patient generally satisfied and relaxed as the HMF model predicts?

In order to compare the AF and HMF hypotheses, it was necessary to identify instances of the patient making transference demands which fit the criteria of *both* models – that is, instances which psychoanalysts who utilize AF concepts would identify as the patient seeking to gratify a key unconscious wish, and which psychoanalysts applying HMF concepts would identify as the patient posing a key test of the analyst. The analyst's responses to the patient's transference demands were rated by AF psychoanalyst judges for the degree to which they were neutral, in the sense of frustrating the patient's wish, and by HMF psychoanalyst judges for the degree to which they "passed or failed" the patient's tests. A passed test is one in which the analyst's response is likely to disconfirm the pathogenic belief which the patient is testing; a failed test is one in which the analyst's response is likely to confirm the pathogenic belief. Finally, the patient's behavior immediately before and after each response was compared (using several patient measures) in order to test the predictions of each model.

The verbatim transcripts of the first 100 hours of a tape-recorded psychoanalysis were the primary data for this study (the same data utilized in the testing study described above). Nine clinical raters read through these transcripts and identified all instances of the patient's transference demands. Eighty-seven transference demands were identified (these included attempts by the patient to elicit approval, affection, guidance, punishment, etc.). Typescripts of each of the segments were prepared; they included the patient's transference demand (or control segment) and the analyst's response (which in some instances was silence).

Five psychoanalyst judges accustomed to applying the AF model in their clinical work and four judges accustomed to applying the HMF model independently rated the analyst's interventions. The AF judges rated each analyst's intervention (on a 7-point scale) for its degree of neutrality from the AF perspective – that is, the degree to which the analyst frustrated the patient's wish. Similarly, the HMF judges rated the extent to which the analyst passed or failed the patient's test. Interrater reliabilities were satisfactory for both ratings.

The next step in this study entailed identifying those transference demands which were pertinent to *both* the AF and HMF models. Three AF judges identified those instances in which the patient was attempting to gratify a key unconscious wish, and three HMF judges identified all instances of key tests. Each judge made his selection on the basis of a case formulation written from his perspective (i.e., AF or HMF). A total of 34 transference demands were identified by both groups of judges as significant events or key transference episodes to which their theories applied. Data analyses were based only on the 34 overlapping instances that were identified as both key wishes and key tests.

To test the predictions of the AF and HMF models, the patient's behavior before and after each of the 34 key incidents was rated on several process scales. Segments of patient speech (averaging about six minutes in length) immediately before (pre-segment) and immediately after (post-segment) the transference demand were rated independently by different teams of judges on the Experiencing Scale, the Boldness Scale (the patient's capacity to boldly confront new material), the Relaxation Scale (a measure of associative freedom and relaxation), and an affect classification system (Dahl 1978, 1979b; Dahl and Stengel 1978) which measured the patient's level of fear, anxiety, love, and satisfaction. All of these ratings were made with satisfactory levels of interjudge reliability.

The results of this study are summarized in Table 2. Predictions derived from the HMF model were supported while predictions of the AF model were not. All seven correlations in Table 2 are in the direction predicted by the HMF model and are opposite to the direction predicted by the AF model. Four of the seven correlations are statistically significant. These findings indicate that when the analyst did not accede to the patient's key transference demands the patient did not feel frustrated or upset; rather, the patient became more relaxed and spontaneous, more bold in tackling issues, and more positive in her attitude toward others. These results support the view that when the patient expressed a transference demand, she was testing a pathogenic belief. By not acceding to these demands, the analyst's behavior provided reassurance against the danger associated with the patient's pathogenic belief.

6. Studies of Brief Psychodynamic Psychotherapy

In both of the studies described above, the significant events studied were patient initiated episodes (the patient's transference demands or key

Table 2 Correlation between Ratings of the Therapist's Behavior and Changes in the Patient Measures for Segments Identified as both Key Frustrations

Measure [a]	r	Predicted by AF	Predicted by HMF
Experiencing	+.23	-	+
Boldness	+.41*	-	+
Relaxation	+.35*	-	+
Love	+.36*	-	+
Satisfaction	+.15	-	+
Fear	-.31	+	-
Anxiety	-.34*	+	-

Note.

AF = Automatic Functioning Paradigm.

HMF = Higher Mental Functioning Paradigm.

+ or – = sign of the correlation predicted by the theory.

[a] N = 34.

* $p < .05$, two-tailed test

(Data for this table are taken from Silberschatz, G.et al. (1986b). Testing pathogenic beliefs vs. seeking transference gratifications. In J. Weiss, H. Sampson, and the Mount Zion Psychotherapy Research Group, *The Psychoanalytic Process: Theory, Clinical Observation, and Empirical Research*. Guilford: New York.)

tests). A recent study carried out by Fretter (1984; Silberschatz et al. 1986a) focused on therapist initiated events – namely, therapist interpretations. The study was designed to show that suitability of interpretations would be a better predictor of immediate (in-session) patient

progress than type of interventions. The suitability of the therapist's intervention was defined as the compatibility of the intervention with the patient's plan – plan compatibility. The type or category of intervention studied was the transference interpretation.

The verbatim transcripts of the brief (16 weekly sessions) psychodynamically-oriented psychotherapies of three cases were the primary data for this study. The research design involved 6 steps:

(1) locating all therapist interpretations;
(2) identifying all transference and non-transference interpretations;
(3) rating the plan compatibility of interpretations;
(4) measuring the patient's behavior (in-session productivity) immediately before and after interpretations;
(5) assessing changes in patient behavior (from pre- to post-interpretation); and
(6) comparing the extent to which the category of the interpretation (transference vs. non-transference) and the plan compatibility of the interpretation predicted these change scores.

The data were analyzed separately for each case in a repeated single-case design.

Malan's intervention typology (Malan 1963, 1976b; Marziali 1984; Marziali and Sullivan 1980) was used to categorize all therapist interpretations as either transference or non-transference interpretations. To assess the plan compatibility of the interpretation, previously developed plan formulations of each case were employed (Curtis et al., in press; Rosenberg et al. 1986). A group of clinical judges (experienced psychologists and psychiatrists who were familiar with the plan concept) read the plan formulation and then rated each interpretation for the degree to which it was compatible with the patient's plan (Plan Compatibility of Intervention Scale – PCIS). Excellent inter-judge reliabilities were obtained.

Immediate patient progress was evaluated by applying the Experiencing Scale (Klein et al. 1970) to pre- and post-interpretation segments of patient speech. Six judges independently rated segments of patient speech that immediately preceded (pre-segment) and immediately followed (post-segment) each selected interpretation. These segments – approximately 3 to 5 minutes of patient speech – were isolated from the transcripts and presented to the judges in random order. Rater bias was

controlled by keeping judges blind to the status of the segment (pre- or post-segment) and to therapy outcome.

The results of this study showed that transference interpretations did not further patient progress more than non-transference interpretations. That is, none of the patients showed significantly greater levels of experiencing following transference interpretations than following non-transference interpretations. By contrast, Plan Compatibility scores were significantly correlated with the EXP (residualized gain) scores for each case. Interpretations judged to be plan-compatible tended to be followed by an increase in the patient's level of experiencing, whereas interpretations judged to be incompatible with the patient's plan tended to be followed by a decrease in the patient's level of experiencing.

These results are consistent with the findings obtained in studies of psychoanalysis. These findings, together with results obtained in the patient testing studies described above, suggest that simple assessment of events in psychotherapy – be they patient initiated or therapist initiated – is unlikely to yield consistent results unless the meaning of such events for a particular patient is taken into account. For instance, a patient who grew up with an overly involved and intrusive parent could be hindered by frequent transference interpretations if they were experienced by the patient as intrusive and thus closely parallel to the way in which the patient had been traumatized as a child. For this kind of patient, a heavy focus on transference would clearly be unsuitable and could possibly be detrimental. A case-specific method for assessing the suitability of the therapist's behavior is needed to assess the effectiveness of any particular interpretation.

7. Methodological and Theoretical Implications

The studies reported here – as well as others carried out by the Mount Zion Psychotherapy Research Group (see Weiss and Sampson 1986) – support the value and feasibility of studying significant events in psychotherapy and of testing theories about the meaning of these events. The identification of significant events provides a useful strategy for studying psychotherapy and has the potential to bridge the wide gap between the practice of psychotherapy and research on psychotherapy (Stiles et al. 1986). However, the results described above suggest that key events need to be identified in a case-specific fashion so that the meaning of the event can be assessed. Strupp (1986) noted that analyses of psychotherapy process must be "relatively specific for the individual patient-therapist

dyad" (p. 126). He concluded that for psychotherapy research to advance, research methods must be geared to the specific dynamics of particular patient-therapist interactions. Although others in the field have drawn similar conclusions, empirical studies using such case-specific methods are extremely rare. The studies reported here illustrate how such methods can be applied to study psychoanalysis and brief psychotherapies.

The studies summarized in this paper illustrate the importance of testing theories of psychotherapeutic change. As noted, most theories of psychotherapy are too abstractly stated and hence can not be easily tested. The theory proposed by Weiss (1986) stipulates how the patient's problem develop (pathogenic beliefs based on childhood trauma), how the patient works in therapy to master problems (efforts to disconfirm pathogenic beliefs), and how the therapist helps or hinders the patient's therapeutic progress (passing or failing tests and plan compatibility of interventions). The theory is based on what Weiss has termed the Higher Mental Functioning Paradigm and is consistent with recent developments in cognitive psychology and cognitive science (e.g. Abelson 1981; Gardner 1985; Mandler 1984; Neisser 1976; Schank and Abelson 1977; Simon and Newell 1970). The studies described here and elsewhere (Weiss and Sampson 1986) provide empirical support for the theory and show how the predictions based on this theory can be tested against other theories.

The accuracy and usefulness of a theory may be argued from many different vantage points. Freud noted that:

> As a rule, however, theoretical controversy is unfruitful. No sooner has one begun to depart from the material upon which one ought to be relying, than one runs the risk of becoming intoxicated with one's own assertions and, in the end, of representing opinions which any observation would have contradicted. For this reason it seems to me to be *incomparably more useful to combat dissentient interpretations by testing them upon particular cases or problems.* (Freud 1918b, p. 48; emphasis added)

Despite Freud's recognition of the importance of systematically studying case material, it is only recently that psychoanalysis has begun to go beyond the informal case study method. As Wallerstein and Sampson (1971) noted, it has generally been quite difficult for psychoanalysis to combine rigorous scientific methods with the complexity of

its explanatory concepts, perhaps because objective research tends to sacrifice clinical relevance for rigor. This paper illustrates that controlled research methods can be added to the more traditional methods of clinical observation without sacrificing clinical relevance.