

Advances in Measures of Psychodynamic Formulations

Jacques P. Barber and Paul Crits-Christoph

Recently published methods are reviewed to assess dynamic formulations derived from psychodynamic case material, especially those methods that emphasize central maladaptive interpersonal patterns. These methods are divided into those that yield psychometric data based on samples that are at least of moderate size and those in which the measures are derived from case studies. The psychometric properties of the first group of measures are reviewed, including interjudge reliability, internal consistency, and content, predictive, and concurrent validity. This article then examines how the measures have addressed and could address the question of construct validity. Also discussed are issues involving the comparison of the methods, particularly their importance for examining specific hypotheses in psychodynamic psychotherapy. Concrete suggestions are made for future research.

Almost every school of psychotherapy recommends that at the outset of treatment its practitioners define each patient's problems and then develop a treatment plan that addresses both the presenting and the underlying problems (Persons, 1991). This is often referred to as a case formulation, but psychodynamic psychotherapists frequently call it a dynamic formulation. Early attempts by researchers (e.g., Malan, 1976; Seitz, 1966) to extract systematic dynamic formulations from case material were unsuccessful because they did not attain the level of reliability required for research purposes. For example, Malan described an elaborate method for formulating psychodynamic themes, which included the listing of (a) all the symptoms, (b) the hypotheses that organize the patient's symptoms around an idiosyncratic conflict (this being the distinctively dynamic part of the formulation), and (c) a priori follow-up criteria for improvement (i.e., the specific behavioral and emotional changes that would represent an optimal outcome). The therapists followed and scored these factors. The dynamic formulation and outcome criteria were reached by consensus, and, as pointed out by Mintz (1981), no attempts were made to estimate their reliabilities. Trying to replicate Malan's results, DeWitt, Kaltreider, Weiss, and Horowitz (1983) found that independent judges could not agree on the core dynamic formulation (including the core conflicts) and suggested the development of "improved methods for reaching agreement on the initial dynamic formulation" (p. 1127).

During the past 15 years, researchers and clinicians have been developing systematic criteria for describing or representing dynamic and interpersonal problems. In a comparison of this new generation of measures to earlier attempts to derive

dynamic formulations, the recent measures use more specific operationalizations of the central concepts and have developed systematic guidelines for judges to make inferences. The results of some of these efforts may serve as the basis for the next generation of psychodynamic formulations; however, the quality of these measures for research use varies widely.

In this article, we review the emerging literature on efforts to evaluate the methods of dynamic¹ formulation in the hope that the review concisely identifies and clarifies what has already been learned about this topic and what issues still need to be identified and studied. We further hope that some of the most sensitive and precisely stated concepts or measures may be adopted as guidelines and heuristic devices for psychodynamically inspired clinicians to improve on the reliability and validity of their clinical formulations. Perhaps rigorous and accurate psychodynamic formulations can offer guidance for psychotherapy beyond phenomenological diagnosis in the style directed by the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.; *DSM-III-R*; American Psychiatric Association, 1987) and could thus be used as outcome measures that are both theory specific and individualized to the patient. In addition, scientifically based psychodynamic formulations can have practical effects on treatment (J. C. Perry, 1989), for example, by guiding interpretations (Crits-Christoph, Cooper, & Luborsky, 1988). Following Luborsky's (1990a) review of 15 existing measures of dynamic formulation and transference, we cover a subset (to be defined later) of these measures and focus on their psychometric characteristics, including the presentation of recent studies. Also, this review differs from previous ones by examining the construct validity of these measures and suggesting specific ways to develop them further.

We restrict ourselves to a specific aspect of dynamic formulation: the assessment of intrapsychic conflicts with or without interpersonal relationship themes as formulations and modes

Jacques P. Barber and Paul Crits-Christoph, Department of Psychiatry, University of Pennsylvania.

The writing of this article was enabled by National Institute of Mental Health Clinical Research Center Grant MH 45178 and Career Development Award MH-00756 to Paul Crits-Christoph.

Correspondence concerning this article should be addressed to Jacques P. Barber, Room 704, Center for Psychotherapy Research, Department of Psychiatry, University of Pennsylvania, 3600 Market Street, Philadelphia, Pennsylvania 19104-2648.

¹ Although traditionally the term *psychodynamic* referred to theories that explained psychic phenomena as the result of multiple (opposing) forces, we use the term more broadly here to include interpersonal theories.

of understanding the patient's major problems. Most of the methods we review here have been developed to assess patients' central maladaptive interpersonal patterns. Our focus is, therefore, narrower than that chosen by S. Perry, Cooper, and Michels (1987), who viewed dynamic formulation as including (a) a description of the current problems and situation and the developmental history of the patient; (b) a description of nondynamic factors that could have contributed to the disorder; (c) a "psychodynamic explanation of the central conflicts, describing their role in the current situation and their genetic origins in the developmental history" (p. 544); and (d) a prediction of how their aforementioned Criterion c will influence treatment. Most of the research reviewed addresses only part of their Criterion c, leaving out the genetic origins of the conflict.

Organization of the Review

The criteria for including a measure in the review are as follows: (a) The measure is described in at least one published work, (b) reliability and validity figures have been reported, and (c) the measure focuses on intrapsychic or interpersonal functioning. Because of our interest in reviewing the psychometric foundations of these measures, we divided the measures into two groups. The first group of measures, which we describe in greater detail than the second group, has already accumulated some published psychometric data and has been used on moderately sized (or larger) samples of patients. The second group has been used mostly in case studies only and still requires more empirical development (additional measures are described in Luborsky, 1990a, and in Luborsky, Barber, et al., in press).

We briefly present the theory underlying each method, and then we describe the kind of data required for applying the method—its database. The scoring method is then presented, followed by how the data is integrated into a dynamic formulation. Finally, we describe the available data in terms of existing reliability and validity. General issues considering the validation of complex constructs such as dynamic formulation or maladaptive interpersonal pattern or conflict are complex and are elucidated in part in the discussion.

Measures Used With Samples of Moderate or Larger Size

Core Conflictual Relationship Theme Method

Luborsky and Crits-Christoph (Luborsky, 1976, 1977; Luborsky & Crits-Christoph, 1990) were the first to develop a clinically based and systematic method of abstracting the interpersonal relationship patterns called the Core Conflictual Relationship Theme (CCRT), which uses narratives told during psychotherapy sessions. The CCRT describes the relationship pattern or conflict in terms of three components: (a) wishes, needs, or intentions expressed by the subject (wishes); (b) expected or actual responses from others (ROs); and (c) responses of self (RSs; i.e., the subject's responses to others' responses), which can include emotions, behaviors, symptoms, and so forth.

The CCRT has several alternative scoring systems. The original, "tailor-made" scoring method includes the judges' infer-

ences about the CCRT component expressed in their own preferred language. These tailor-made categories were later expressed in preset categories (see Luborsky's first edition in Barber, Crits-Christoph, & Luborsky, 1990), later called standard categories. In an effort to increase interrater reliability and to make the judge's task easier, Crits-Christoph, Luborsky, et al. (1988) used these standard categories. The dimensions of psychodynamic themes were examined by using cluster-analytic methods to determine the underlying grouping of the wishes, the ROs, and the RSs (Barber et al., 1990).

Database and Judges

Any subject's interview material that includes relatively concrete descriptions of the subject's interpersonal relationships can be used successfully. Transcripts from a variety of sources, such as therapy sessions or a specially designed interview (e.g., the Relationship Anecdotes Paradigm [RAP]; Luborsky, 1990b) have been used. A minimum of 10 objectively defined narratives describing interpersonal interactions (named relationship episodes) are usually used as a basis for scoring the CCRT.

In general, judges are experienced PhD- and MD-level, psychodynamically oriented psychotherapists who have been selected because of their ability to formulate CCRTs close to those formulated by experienced CCRT judges. Judges are trained by learning the CCRT guide (Luborsky & Crits-Christoph, 1990), trying several standard practice cases, and then receiving feedback from the research team about their performance after they complete each one. With this training procedure, several BA-level assistants have also been selected as judges.

Scoring and Formulation

First, a research assistant delineates the relationship episodes from within the transcribed session. Second, independent judges read each relationship episode in the transcript and identify each of the three components: wishes, ROs, and RSs. Third, for each component, the types with the highest frequency across all relationship episodes are identified, and their combination constitutes a preliminary CCRT formulation. Fourth, on the basis of this preliminary CCRT formulation, the same judge re-identifies, where needed, the types of wishes, ROs, and RSs. Fifth, the judge can change the original rating on the basis of the recount of all wishes, ROs, and RSs. In addition, judges may be asked to follow their tailor-made scoring by standard categories.

Reliability of the CCRT Method

The interjudge agreement on delimiting relationship episodes (Bond, Hansell, & Shevrin, 1987) and rating their completeness (Crits-Christoph, Luborsky, et al., 1988) has been shown to be acceptable. The agreement between judges in identifying the main person described in the narrative was high: for 80 episodes, 97% had the same other person selected (Crits-Christoph, Luborsky, et al., 1988).

Reliability of the CCRT formulation was evaluated in a sample of 35 patients drawn from the Penn Psychotherapy Project

(Crits-Christoph, Luborsky, et al., 1988). To calculate interjudge reliability, each judge's tailor-made CCRTs were translated into the CCRT standard categories by a different set of two independent judges. Each judge picked the standard category that was closest in meaning to the tailor-made wording. The interjudge agreement on the coding of tailor-made CCRTs into standard categories was 95%. The reliability of the original judges' ratings, which were translated into standard categories, was then measured by using Cohen's (1968) weighted kappa. For these calculations, the highest weight (full agreement) was given when CCRT judges listed the same wish, RO, or RS as the most frequent one across the 10 relationship episodes; a lower weight was given for the highest frequency category from one CCRT judge matching the second highest frequency from the other CCRT judge, and an even lower weight was given when the match occurred with the second highest frequency for each CCRT judge.

The weighted kappas for the wish and negative RS were both .61; for the negative RO, the weighted kappa was .70. All of the values were statistically significant ($N = 35$, $p < .001$). These kappas fell toward the upper end of the "fair-to-good" range.

Validity of the CCRT

The validity of the CCRT has been examined in a series of studies (for more detailed reviews, see Luborsky, Barber, & Crits-Christoph, 1990; Luborsky, Barber, & Diguier, 1992; Luborsky & Crits-Christoph, 1990).

The "adequacy with which a specified domain of content is sampled" (Nunnally, 1978, p. 91) is called *content validity*. The question of content validity in regard to the measures reviewed in this article can be asked at the level of the components of the measure (e.g., wishes for the CCRT) and at the level of the categories within each component. The former level addresses whether the components of the methods cover the domain proposed by the measure; the latter level addresses whether the list of wishes, for example, cover the universe of wishes, needs, and intentions. Little discussion or empirical research has addressed whether wishes, ROs, and RSs adequately represent the underlying components of core conflicts or whether other domains, such as defenses, should also be included. Content validity is, for the most part, assessed by using reasoning and common sense. This aspect of validity for the CCRT and other methods is discussed in more detail when we address the issue of the scope of dynamic formulation. The content validity of the CCRT list of categories has received some attention. Whether the CCRT lists cover the universe of wishes, ROs, and RSs, however, remains to be shown. On the one hand, many specific wishes, ROs, and RSs are not included in the Barber et al. (1990) list of standard categories. On the other hand, some of them are likely to be represented as variants of the higher level categories already developed (e.g., the wish "to merge" could be referred as a variant of the wish "to be close"), and thus existing standard categories may cover a wide range of wishes, ROs, and RSs. The problem is to maintain a level of abstraction that is neither too high nor too low when including different wishes under the same standard category (Luborsky & Crits-Christoph, 1990). Although the standard categories were initially developed inductively on several samples of patients (Edi-

tion 1), the most recent version (Edition 2) was an attempt to cover the list of Murray's (1938) presses and needs (Barber et al., 1990).

Nunnally (1978) defined predictive validity as any estimate of "some important form of behavior that is external to the measuring instrument itself" (p. 87). In terms of predictive (concurrent) validity, the pervasiveness of the RO component of the CCRT of depressed patients has been shown to correlate moderately ($r = .32$, $n = 21$) with concurrent severity of depression (Eckert, Luborsky, Barber, & Crits-Christoph, 1990). More specifically, the extent to which depressed patients expected or experienced "being rejected by others" (RO) across their different relationships was found to correlate with their level of depression. Another aspect of concurrent validity was explored in a study comparing the patients' own interpretations of their RAP narratives with the independent judges' CCRTs (Crits-Christoph & Luborsky, 1990c); in that study, patients tended to rate highly the same wishes that clinical judges abstracted from the narratives.

Indirect evidence for the predictive validity of the CCRT is provided by a study of accurate interpretations, in which Crits-Christoph, Cooper, and Luborsky (1988) demonstrated an important consequence of obtaining a concise and reliable dynamic formulation. Accurate interpretations (i.e., interpretations addressing the CCRT that had been formulated independently by other judges) are associated with better outcome in dynamic psychotherapy. For example, the therapists' accurate interpretation of the wish and the RO of their patients during the early sessions correlated .44 with residual gain in general adjustment in a group of 43 patients receiving moderate-length dynamic psychotherapy. Furthermore, in a sample of 33 patients, Crits-Christoph, Barber, and Kurcias (1993) found that the extent to which therapists accurately addressed the CCRT in their interpretations predicted the development of the therapeutic alliance (i.e., maintenance of good alliances or improvement in bad alliances, over the course of treatment). Finally, the level of patients' self-understanding of their CCRT, as manifested with their therapists in early sessions, correlated ($r = .31$, $n = 43$) with successful outcome in moderate-length dynamic psychotherapy (Crits-Christoph & Luborsky, 1990b).

Modification of the CCRT Method

For many purposes, the CCRT method has many strengths; for example, a similarity to the steps a clinician might use to formulate patients' themes. Nevertheless, for research purposes on large samples of patients, a number of disadvantages are apparent: (a) the possibility of judge bias toward describing consistency in themes across episodes, as all episodes for one patient are scored as a set; (b) the uncertain validity of the standard categories used; and (c) the method yields insufficient quantitative data. Similar disadvantages are found in most of the measures reviewed in this article. A modification of the CCRT method, labeled the Quantitative Assessment of Interpersonal Themes (QUAINT; Crits-Christoph & Baranackie, 1992; Crits-Christoph, Demorest, & Connolly, 1990), has been developed in an attempt to address the aforementioned concerns. In the QUAINT method, relationship episodes are presented to judges in random order across patients. The Struc-

tural Analysis of Social Behavior (SASB; Benjamin, 1974) is used as the system of standard categories for each CCRT component, and judges rate each SASB-based standard category on a 5-point scale for each episode on the extent to which that category is present in an episode. Whether the randomization of episodes leads to a loss of information or validity, however, is not known. Further research is needed to clarify the relative strengths and weaknesses of the QUAINT method. Luborsky, Luborsky, et al. (in press) have similarly modified the CCRT method to involve rating each of its standard categories instead of the SASB items and, therefore, provides the CCRT with more quantitative data than currently available.

Idiographic Conflict Formulation Method (ICF)

J. C. Perry, Augusto, and Cooper (1989) presented another method of dynamic assessment. The method has two stages: The first yields the ICF, and the second assesses the ICF for specific conflicts by using the Psychodynamic Conflict Rating Scales (PCRS). Descriptions of the conflicts are theoretically organized into the eight stages of human development described by Erik Erikson.

Database and Judges

In general, entire videotaped psychodynamic interviews or any psychiatric interview that covers the major areas of life have been used. For the most part, judges have been experienced, dynamically oriented clinicians. Two teams of two expert clinicians each rate an entire videotape of a dynamic assessment, and then each team derives a consensual, well-supported formulation.

Scoring Method and Formulation

Judges describe the major components of the conflicts and support their claims with evidence from the interview. The ICF components (J. C. Perry, Augusto, & S. H. Cooper, 1989) are as follows: (a) wishes (conscious and unconscious); (b) fears, which often conflict with the wishes; (c) resultants of the conflicts in terms of symptomatic and avoidant outcomes—the latter referring to the characteristic ways in which the patients avoid or reduce the experience of their conflicts, as well as the patients' constrictions in various domains such as emotional expression and interpersonal relationships; (d) vulnerability to specific stressors that activate the patients' conflicts; and (e) best level of adaptation to conflicts. The first three components represent the traditional dynamic conflicts. Using evidence found in the interview, judges may modify or refine the definitions of the components.

Using the above ICF components and the evidence collected, the teams of clinicians write their formulations. Optionally, the patient's ICF can be rated using the PCRS or a standardized list of wishes and fears. The PCRS provides raters with brief definitions of 14 predefined conflicts. Judges rate the ICF on a 4-point scale regarding their confidence in the presence and centrality of these conflicts.

Reliability of the ICF

The measure uses two teams of two expert clinicians to rate an entire videotape of a dynamic assessment and to derive a consensual formulation. In a recent study of the reliability of the ICF, J. C. Perry, Augusto, & Cooper (1989) used matched and mismatched pairs of idiographic formulations from 20 patients. They asked two teams of pairs of graduate students to rate these formulations on their degree of similarity for each of the ICF components. Each team had to reach a consensus rating. The intraclass correlations for the ICF components ranged from .54 to .75. Matched pairs of formulations were found to be more similar than mismatched pairs. More specifically, using a 7-point scale, they found that the average degree of similarity of correctly matched pairs (4.41) was significantly higher ($p < .001$) than that for mismatched formulation pairs with the same diagnoses (3.05) or with different diagnoses (2.91).

J. C. Perry (1992) reported that two independent judges re-rated 35 ICFs, using standard categories for fears and wishes. The 15 ICFs had been previously rated with the individualized categories. To compute the weighted kappa, the judges gave a weight of 1 to different wishes or fears within the same Eriksonian stage and a weight of 2 to disagreements from different stages. Among the wishes and fears mentioned at least 5% of the time, the median kappa was .64 for wishes and .46 for fears, indicating acceptable interjudge reliability for wishes but marginal interjudge reliability for fears.

Validity

Like the other methods, the ICF seems to cover many important aspects of dynamic conflicts, including the CCRT components. In contrast to the other methods, the inductively created list of ICF wishes has been organized around Erikson's theory of personality development. Thus, it seems likely that the ICF wishes cover the domain of Erikson's developmental stages.

In terms of concurrent validity, J. C. Perry, Augusto, & Cooper (1989) examined whether formulations within any diagnostic group (borderline personality disorders and two relatively close diagnoses, antisocial personality disorder and bipolar Type II affective disorder) were more similar than those for other diagnostic groups. They compared mismatched cases with the same diagnosis to mismatched cases on diagnosis. However, they found little support for differential formulations across diagnostic categories. A possible reason for this lack of support could be that they used diagnostic groups with overlapping psychopathology, thereby reducing the likelihood of finding differences between groups. An alternative hypothesis would be that specific diagnostic entities are not characterized by specific formulations. More research is clearly needed to examine these hypotheses.

The Plan Formulation (PF) Method

The way in which Weiss, Sampson, and the Mount Zion Psychotherapy Research Group (1986) conceptualized their patients shares similarities with Popper's (1979) view of how organisms acquire knowledge. According to Popper, organisms attempt to increase their knowledge by refuting previous knowledge. Similarly, Weiss et al. assumed that patients interact with

their therapists in such a way as to disconfirm their pathological beliefs. In accordance with their control mastery psychoanalytic theory, Weiss et al. initially developed the Plan Diagnosis method for formulating their patients' difficulties. The PF method evolved out of the Plan Diagnosis method (Curtis, Silberschatz, Sampson, & Weiss, 1992).

The Plan Diagnosis and the PF methods describe four components relevant to the control mastery psychoanalysis of Weiss et al. (1986): (a) the patients' conscious and unconscious goals for therapy; (b) the patients' inner obstacles (i.e., their pathogenic beliefs and related negative emotions that prevent patients from reaching their goals); (c) the plan or manner the patients use to disconfirm their pathogenic beliefs (e.g., tests) during treatment; and (d) the specific insights required for patients to benefit from treatment and to achieve the goals of psychotherapy.

Database and Judges

Transcripts from the beginning sessions of psychodynamic psychotherapy (in general, the intake interview plus the first 2 sessions) or the first 10 sessions of a psychoanalysis are used. This method has been used in a research program that emphasizes the in-depth study of individual cases.

In the past, judges have been experienced clinicians, well-trained in the Weiss et al. (1986) approach, but there are indications that less experienced mental health trainees can be used.

Scoring and Formulation

First, the judges independently develop a written PF specifying all four of the Plan Diagnosis and PF method components. They are then asked to develop lists of relevant items as well as alternative, less relevant items describing the four components. Second, the researchers combine the judges' lists separately for each component. Third, these master lists are then returned to the judges, who are asked to rate the relevancy of each item for the case on a 5-point Likert scale. Fourth, arithmetic means are calculated for each item. Items that are below the median for their component are dropped. Fifth, a second group of judges identifies and eliminates redundancies among the items. The remaining items constitute the PF.

The formulation includes the nonredundant items, their summary, and a description of the patient, including his or her complaints and current life situation (Curtis et al., 1992).

Reliability

Interjudge reliability. Initial reliability data has been presented for single cases (Curtis, Silberschatz, Sampson, Weiss, & Rosenberg, 1988) and for a small group of patients (Rosenberg, Silberschatz, Curtis, Sampson, & Weiss, 1986). Four therapists trained in cognitive-psychoanalytic theory received the transcript of the sessions and rated the relevance of the items included in the master lists. The pooled interjudge agreement for goals, obstacles, tests, and insights ranged from .78 to .97 (except for tests for Patient 1, which was .39). In addition, there was good agreement between the two teams of judges concerning the relevance of the various statements describing the PF com-

ponents. No reliability data are available for the initial formulation from which master lists are created because "as a narrative it still defies easy comparison with another narrative formulation developed by a separate set of clinicians" (Curtis et al., 1988, p. 256).

Internal consistency. Curtis et al. (1992) reported internal consistencies of the items for each of the aforementioned master lists, ranging from .84 (obstructions) to .90 (insights and goals).

Stability. The PF method is one of the few methods for which stability has been examined (Collins & Messer, 1991). Correlations between judges' mean at 3 months posttest were all above .94 for each component for the one patient on which this aspect of reliability was examined. Using a different approach to assess stability, Collins and Messer have also shown that for the previously mentioned patient and for another one, the percentage of selected items retained at 3 months' interval was at least 85%.

Validity

In terms of the different levels of analysis of content validity, at least theoretically, the PF components seem to cover the essential tenets of Weiss et al.'s (1986) theory. In regard to the list of categories generated by the clinicians, no attempt is made to cover the universe of obstacles, because the judges' goal is to provide an idiographic description of the patient. A potential question is whether providing judges with standard lists of goals, tests, obstacles, and insights would be helpful.

Indirect evidence for the predictive validity of the measure is provided in Silberschatz, Fretter, and Curtis's (1986) study of the accuracy of interpretation. They defined suitability of interpretations as the compatibility of the therapists' interventions with the patients' plan. They were able to show that the degree of suitability of interpretations with the plan correlated positively within session with the Experiencing Scale (Klein, Mathieu-Coughlan, & Kiesler, 1986)—a measure of insight, involvement in treatment, and productive talk—in their 3 patients, with correlations ranging from .25 to .54. Similar findings using outcome measures were presented on a single case using the Plan Diagnosis method (Weiss et al., 1986).

That clinicians trained successfully in the PF method but having differing theoretical perspectives could come up with a reliable but different formulation of the same patient (Collins & Messer, 1988, 1991) has implications regarding the specificity, reliability, and validity of the method. A reliable method used by trained judges should lead to the same formulation, irrespective of the judges' theoretical biases. Because clinicians from different perspectives can come up with different formulations, it might be more appropriate to view the PF as a heuristic enabling judges from similar theoretical stances to come up with an agreed-on formulation. The PF as a heuristic would then be useful in providing a structure for researchers for creating measures relevant to their own theory. The advantage of such a structure, as Collins and Messer (1991) pointed out, is that it would allow for the testing of the "differential effectiveness of differential predictions made by two different plans for the same patient" (p. 80).

Consensual Response Method (CRM)

L. M. Horowitz's CRM represents another approach at resolving the earlier reports of reliability problems of dynamic formulation (L. M. Horowitz, Rosenberg, Ureño, Kalehzan, & O'Halloran, 1989; L. M. Horowitz & Rosenberg, 1992) by averaging formulations made by a group of clinicians. The principle underlying this approach is that, in general, the ratings made by a group of judges are more stable and reliable than the ones made by one judge. This is in contrast to other investigators who have developed sophisticated systems of guidelines on how to derive dynamic formulations; the CRM averages seminatural dynamic formulations. The problem these authors faced was how to average various dynamic formulations.

Database and Judges

Hour-long clinical videotaped semistructured interviews were rated by experienced dynamic clinicians.

Scoring Method and Formulation

First, clinicians independently derive their dynamic formulation from the videotape and write it up on one page. They are instructed to include the following factors about the patient, the treatment, or both: what brought the patient into treatment, personal history, characteristic interpersonal relationships, personality, conflicts, defenses, and psychodynamic themes. Second, the clinicians discuss the case for 30 min. Third, they rewrite their dynamic formulation without reviewing the original written formulation. Fourth, each formulation is divided into thought units by a group of four students using a detailed set of instructions. Fifth, five judges, using a standardized procedure, review the thought units and identify those that have similar meanings. Those similar thought units are tabulated across the eight judges.

In addition to the biographical data and description of the symptoms, the thought units that have been mentioned by three or more judges are then integrated into a single narrative that is the consensual formulation.

Reliability and Replicability

Reliability of the number of thought units was computed and found acceptable. These scores ranged from .84 to .96 both before and after the discussion (L. M. Horowitz et al., 1989). Very few thought units were mentioned by all eight experienced, psychodynamic clinicians who rated the interviews; in fact, most items were mentioned by only three or four of the judges. Thus, the reliability of this method is limited when only a few judges are used.

One case was reformulated by a different group of eight clinicians. Two methods were used to measure the replicability (L. M. Horowitz et al. 1989):

1. The proportion of thought units that appeared in both formulations. Eighty-one percent of the thought units contained in either the pre- or postdiscussion for the first consensual formulation reappeared in the second formulation as judged by four of five graduate students.

2. Five students were able to recognize correctly the consen-

sual formulation (the idiosyncratic information having been removed) from among the 15 previous consensual formulations.

Validity

As with the other methods, content validity has not been assessed directly. Nevertheless, the fact that formulations were derived from a large group of clinicians makes it likely that most aspects of what these clinicians thought should be included in such a formulation were indeed included. Specific aspects such as genetic explanations for the presenting problem might have been omitted. Predictive validity was investigated by examination of the correspondence between the formulation and the Inventory of Interpersonal Problems (IIP; L. M. Horowitz, Rosenberg, Baer, Ureño, & Villaseñor, 1988). If the formulation is valid, then one would expect that the issues raised in the formulations would also be discussed during psychotherapy. In a test of this hypothesis, the IIP was distributed at Sessions 10 and 20 to assess which issues were discussed in therapy. The issues mentioned in the formulations were accurate in their predictions of which issues would be discussed during treatment.

The rationale for examining another aspect of the predictive validity was based on a previous study, in which L. M. Horowitz et al. (1988) found that patients who primarily had interpersonal problems were better candidates for brief dynamic psychotherapy than patients who had noninterpersonal distress. They found that the proportion of thought units referring to interpersonal issues included in the formulation predicted outcome as indicated by an overall measure of clinical improvement.

The Role-Relationship Models (RRMs) Formulation

As part of the more complex configurational analysis, M. J. Horowitz (1989) recently described the Role-Relationship Models Configuration (RRMC), which incorporates several RRM's. An RRM represents an interaction between a self-schema and another person, using up to seven of the following elements: (a) the self-schema, (b) the schema of the other person, (c) anticipated actions or emotions toward the other (wish), (d) the response of the other, (e) the response of self, (f) the self-estimation of these reactions, and (g) the other's expected self-estimation of these reactions (M. J. Horowitz, 1989, p. 261). The RRMC method differs from other methods by allowing for formulations of multiple states of mind. M. J. Horowitz and Eells (1993) referred to four specific states of mind: dreaded, desired, problematic compromise, and adaptive compromise. Conflicts can occur at three levels: (a) between the wish to act or express emotion and responses from others and responses of self, (b) between "two or more . . . RRM's as alternative schemas for organizing experience and action" (pp. 261-262), and (c) among different views of oneself (self-schemas).

Database and Judges

Intake interviews, therapists' notes, and therapy sessions are used as the database. For the purpose of reliability assessment, M. J. Horowitz and Eells (1993) used RRMCs derived from the

entire length of a 12-session brief dynamic psychotherapy. Judges have been experienced members of the research group working in teams.

Scoring and Formulation

Because of the complexity of the RRMC method, we do not attempt to summarize it; instead, we refer the interested reader to M. J. Horowitz (1991; especially chapter 5 and its appendix). The RRMC provides a detailed description of the interactions between different aspects of the self and others, the pattern of their interactions (action sequence), as well as how the consequences of these interactions result in a different state of mind, which in turn leads to other types of interpersonal interactions. Furthermore, the RRMC provides a description of these patterns for different aspects of the self. In that sense, both the CCRT and Schacht and Henry's (1992) cyclical maladaptive pattern (CMP) are included as elements of the RRMC.

Reliability

M. J. Horowitz and Eells (1993) reported initial reliability data on the RRMC method from a sample of 4 women with pathological grief reactions following the death of a parent. The consensually constructed RRMCS from two teams of experts were used. The RRMCS were derived from the close examination of the entire span of videotaped brief dynamic psychotherapy. Four groups of five judges who viewed only the first four sessions of each patient's therapy were asked to rate how well the previously formulated four RRMCS fit with the material emerging from those four sessions. Results indicated that non-expert judges were best able to match the entire RRMC, rather than components, to the patient's videotapes.

Validity

The formulations obtained by the RRMC method may be the most comprehensive of all the formulation methods and, therefore, may match what therapists learn from their patients during an entire span of psychotherapy; unfortunately no data to assess this conjecture are available. Except for case studies in which the RRMC and the CCRT were shown to arrive at similar formulations, no validity data are yet available.

Measures Used in Single-Case Studies

In this section, we briefly review other methods that have been reported and appear promising as methods for identifying interpersonal themes (these and additional methods are also reviewed in Luborsky, 1990a).

Cyclical Maladaptive Pattern

Strupp and Binder (1984) introduced the concept of dynamic focus, later renamed the CMP, as a concept related to Mann's central issue (Mann & Goldman, 1982) and to the CCRT. Strupp and Binder relied on Wachtel's (1983) concept of the "cyclical psychodynamic pattern" to characterize the chronic repetitiveness and self-perpetuating nature of problems that patients bring to therapy. This concept explains the neurotic repe-

tion that results from the creation and maintenance of vicious circles. Recently, Schacht and Henry (1992) have operationalized the CMP by using Benjamin's (1974) SASB with the intention to reliably describe the CMP interpersonal aspects. Its four components are (a) interpersonal acts, (b) introjective acts, (c) expectancies regarding interpersonal actions, and (d) introjective actions (predictions, wishes, and fears). Although SASB codings have been shown to be reliable (Benjamin, 1982; Henry, Schacht, & Strupp, 1986), there is no available data on the reliability of the CMP formulation as the integration of the four CMP components into a causal, circular, psychodynamic formulation. Because of the strong heuristic power of the CMP, we hope that researchers will begin its systematic investigation, including the development of procedures to delimit the relevant statements, reliability, and validity studies.

Frame Analysis

Teller and Dahl (1981) developed a system to examine psychotherapy sessions in terms of recurrent, structured sequences of events representing people's wishes and beliefs as they are inferred from the person's actions, thoughts, perceptions, or emotions. At least three methods have been suggested on how to find, infer, and formulate the frame (see Dahl, Kachele, & Thoma, 1988; Teller & Dahl, 1992). A fourth, more systematic and promising method has been recently presented (Dahl, 1992). One of the methods, for example, includes the examination of the narrative, the ordering of the events described within, and their summary in the form of predicates. The sequence of predicates is the frame. The frame analysis method is one of the few methods that views the sequentiality of the patient material as meaningful. In contrast to other methods, the judge is not provided with a priori categories (e.g., fears) to parse the patient's narratives. Teller and Dahl (1992) presented evidence for the preliminary validity of their method, most of which is based on the in-depth study of one patient. They also reported on an interesting unpublished dissertation in which twelve 3-year-olds were recorded during 10 interactions with their mothers and during 10 interactions with each of two other children. The results showed that similar frames were apparent in each child's interaction with their mother and the other children and that the frames were different for each child.

Comparisons of the Different Methods

Now that we have presented the different methods, we turn to examine how they are used in attempts to achieve construct validity. We also examine the few attempts at comparing the methods quantitatively and discuss their implications. Then, we discuss the main methodological and research issues raised by this line of research and delineate possible directions for future research.

The goal of much research is to develop the validity of scientific constructs such as the one discussed in this article. In case of complex variables such as dynamic formulation or one of its components (e.g., core conflict), the process of achieving construct validity is long, multiphased, and multifaceted. One of the reasons causing this difficulty is that no independent measure or criterion exists for comparison of the methods. Nun-

nally described three aspects of construct validity that need to be addressed:

- "(1) specifying the domain of observables related to the construct;
- (2) . . . determining the extent to which the observables tend to measure the same thing, several different things, or many different things; and (3) studies . . . to determine the extent to which supposed measure of the construct produce results which are predictable from highly accepted theoretical hypotheses concerning the construct." (1978, p. 98)

One aspect of construct validity that has received some attention is the degree to which these different methods yield similar data: that is, do different methods produce a similar or recognizable formulation? Investigating the similarity of dynamic formulation between patients or the degree of change within a specific patient during therapy regardless of the method used is a complex task. J. C. Perry, Luborsky, Silberschatz, and Popp (1989) used the ICF, the CCRT, and the Plan Diagnosis methods to derive a dynamic formulation from the same videotaped interview. Judges were asked to rate on a 7-point Likert scale the similarity of formulations from the specific case, or from a control case derived from the different methods. J. C. Perry, Luborsky, Silberschatz, and Popp found that the Plan Diagnosis method goals were similar to the most frequent CCRT wish but that the ICF wishes were somewhat more similar to the second most frequent CCRT wish. This indicates some convergence among these three methods.

When comparing different methods of formulation, an additional complexity arises because each method covers a different domain, and formulations can be systematically compared only on the mutual domains they cover. For example, as J. C. Perry, Luborsky, Silberschatz, and Popp (1989) remarked, "the CCRT focuses on relationship patterns," whereas the "PF focuses on dynamic features related to transference, resistance and insight in therapy" and the ICF "focuses on stress and internal conflict" (p. 302) and how one copes with these in daily life and in treatment. In addition, most formulations result in multiple ratings for each component; it is common for CCRTs, for example, to have two or three wishes.

So, although there is a growing interest in comparing these methods, it turns out that such comparison is difficult. We suppose that part of the difficulty is inherent in the comparison of two different descriptions of the same person. One solution is to follow the direction of J. C. Perry, Luborsky, Silberschatz, and Popp (1989) and to instruct judges to rate the degree of similarity between the two formulations being compared. This is not too difficult when the formulations are derived from related and similar underlying frameworks such as the ICF and the CCRT. One needs to explore, however, what the judges' ratings mean when they indicate that two formulations are similar and that they cover different domains. Thus, in many cases, the relevant question is not whether the different methods yield similar formulations but whether the assessment of a certain domain (e.g., wishes) from one method is comparable to the assessment of the same domain from a different method. To the extent that the answer to this question is affirmative, then such comparisons could be helpful in the future to assess the methods' convergent validity. Alternatively, one could translate the two different descriptions using a common and well-accepted "language," such as Benjamin's (1974) SASB.

Two additional suggestions for dealing with the difficult conceptual and measurement problems that this type of evaluation research poses can be made, neither of which are panaceas. One is to ask the formulators to write up their theoretically derived predictions, along with a specification of what data would be necessary for refuting or corroborating these predictions *in advance* of carrying out the research itself. Of course, as with any other refutation, it indicates either that the specific formulation is wrong or that the theory from which it is derived is wrong. The other suggestion is to keep in mind that the best formulation would be the one with the better explanatory power. The major shortcoming of this proposal is that these predictions are difficult to formulate, and different experts might disagree on them because the underlying theories from which they are generated are not always adequately specific. Nevertheless, one could envision a study where patients would be first interviewed and then assigned to treatment. From the interview, two groups of experts could derive separate CCRT and PF formulations, for example, on the same patients, and then predict which therapist's action will result in a better in-session and treatment outcome.

Although comparing the methods and examining their correspondence is interesting in itself and may also increase confidence in the validity of these methods by providing evidence for convergent and divergent validity, other avenues for construct validation are available. One of the options is to address Nunnally's (1978) first aspect of construct validity and to define the domain of observables of the construct. In other words, what is the phenomenon or phenomena that these methods are trying to capture? As Nunnally noted, this is rarely the first step in a research program, but it logically should be the initial step. All the developers of the methods have addressed this issue, but for the most part, the domain of human behaviors that the methods are trying to capture has only been implicitly delimited. Which human behaviors do these methods not try to explain? Do the methods focus on interpersonal behaviors only? Are maladaptive patterns as frequent in "cold" relationships as they are in more intense, "hot" relationships? For example, it is not clear whether maladaptive interpersonal patterns are also expected to be present in a nonpatient population, and, if present, whether they will be less stereotypical. Finally, it could be that various methods differ in terms of which behaviors the measures try to capture. A way to answer this question is to look at the scope of the different measures: To a large extent, the scope of the dynamic formulation is a function of the amount of material included in its database and the range of human behaviors the formulation is trying to capture.

Although most methods are flexible on the issue of what is the database, some require only a limited amount of information, such as relationship episodes in the CCRT, whereas others, such as the ICF, require the full clinical interview. In addition, some methods require their judges to make inferences about unconscious factors (e.g., the ICF), whereas others tend to maintain a more conscious focus (e.g., the CMP).

The scope of the formulation also tends to vary. Some methods, such as the ICF and the PF, include patients' information and history, and therefore derive a comprehensive formulation much like the dynamic formulations recommended by S. Perry

et al. (1987); others, like the CCRT and QUAINT, focus on the formulation of the maladaptive pattern or transference. Reliability studies reported for the ICF and PF addressed the component of the maladaptive pattern rather than the entire dynamic formulation. Also, most methods are flexible enough so that J. C. Perry, Luborsky, Silberschatz, and Popp (1989) could provide judges by using different formulation methods with the same database. Nevertheless, different methods require their judges to focus on different aspects of the patient material. The use of different information could be associated with some degree of divergence in the formulations obtained from the different methods.

Another target for evaluating the construct validity of these methods is to examine the empirical relations among the observables (Nunnally, 1978). The J. C. Perry, Augusto, and Cooper (1989) comparison of patients with the same diagnosis versus those with different diagnoses is one example of this approach to construct validity. Another potential study would be to examine whether people with similar core conflicts have similar personality profiles. Yet another study would be to derive dynamic formulations for patients with a specific *DSM-III* diagnostic category and then to examine whether patients with a similar ICF, for example, displayed other similarities (e.g., response to different forms of therapy) while being different from patients with the same diagnosis but different ICF. Similarly, if two different methods will show the same kind of change (or lack of change) following a treatment, we will tend to increase our confidence that these two measures would measure the same construct. More specifically, if the wish components of the ICF remain unchanged following psychotherapy but the resultants do change (corroborating Crits-Christoph & Luborsky's [1990a] results), we will tend to view the ICF and the CCRT as measuring similar constructs.

Such a line of investigation would also help resolve the question of whether a formulation should include more than one focus; that is, whether people, or some people, are better described by a monoschematic versus multischematic formulation. Psychodynamic formulations may be grossly categorized along a dimension of complexity into those that yield a monoschematic (relatively simple) representation, such as the CCRT, and those that yield a (complex) multischematic representation, such as M. J. Horowitz's (1989) RRCM. Although the CCRT might, at times, be formulated as multischematic (e.g., including several wishes), judges are encouraged to come up with a single formulation. It remains possible, however, for proponents of the monoschematic view to suggest that the various schematics of a patient represent one underlying main core issue. Obviously, if one uses a level of analysis that is abstract enough, any wish, for example, can be translated into a highly inferential wish, such as a wish for pleasure, survival, and so forth (Fried & Agassi, 1977). There are little data to support either a monoschematic model or a multischematic model. One exception is the study by Crits-Christoph et al. (1990), which demonstrated that data from one patient revealed the presence of multiple themes rather than one, single predominant theme.

Assessment of Reliability

Besides addressing validity, future research should continue addressing the issue of reliability. Measuring interjudge reliabil-

ity in the use of dynamic formulation methods is a complex task because judges' ratings involve many stages. Here, we begin by addressing some of the issues involved in the selection and training of rating judges.

Judges

Most methods have been successfully tied with experienced and interested PhD-level clinicians as judges. Because of the difficulty in keeping such highly educated judges interested and because of their cost, it is important to determine whether less highly educated judges could be used without sacrificing quality. Perhaps selection procedures and preparation for the judging task is the answer. Clear selection criteria for judges could be developed and evaluated. Additionally, with some of the methods, for example, the CCRT and the PF, a short training course can easily be provided, followed by a series of rating tasks that can be compared with the ones produced by experienced judges.

Consensual Teamwork Versus Independent Judges

Some of the methods (e.g., the ICF) require their judges to work in teams. This procedure is likely to increase reliability because it decreases the likelihood of judges' idiosyncrasies. On the other hand, using teamwork is more difficult and expensive, especially if one needs to use experienced clinicians. Furthermore, a team's work is also not applicable to clinical practice in which the therapist is working alone. We therefore recommend, for practical reasons, that researchers begin to use independent judges rather than teams. In addition, if teams are used, more sophisticated methods to achieve consensus may be available.

Use of Transcripts

Many of the methods described in this article require the use of transcripts from sessions or interviews. It is quite likely that the use of transcripts rather than the direct rating of audiotapes or videotapes leads to more precise and possibly more reliable formulations, but it comes at a price, financially and scientifically. The financial expense leads to a scientific bias toward selecting case studies or small samples—decisions that ideally should be made on scientific grounds alone. From the practitioner's point of view, to the extent that any of these methods require transcribed material, they are not practical and are not likely to be used in clinical settings. Also, the use of transcripts might lead to loss of potentially important nonverbal information. Our recommendation, therefore, is to conduct reliability studies examining whether there are differences in the formulations obtained from videotapes versus transcripts. Initial data regarding the comparison of videotaped versus transcript-based CCRT indicated no difference between the two modalities (Cierpka, Krannich, Ratzke, Reich, & Homburg, 1992).

Levels of Analysis of Interjudge Reliability

Reliability needs to be assessed at every step in the process of deriving a formulation. This might include reliability of identifying the basic unit of observations (e.g., in the CCRT method,

relationship episodes and thought units), of extracting information from the basic units, of transforming themes into standard categories, and of the final formulation.

Use of Standard Categories

To increase reliability, investigators, such as J. C. Perry (1992), have recently added a list of standard categories (e.g., lists of wishes and fears). Although the introduction of standard categories is likely to increase judges' reliability, few studies have specifically shown this result. J. C. Perry organized his list of standard categories around the eight stages of development described by Erikson, whereas others such as the CCRT's categories have a more empirical basis. Organizing inductively generated categories in terms of a theory may increase the chance that the content validity is adequate.

The drawback of translating idiographic formulations using standard categories is that important information might be lost. On the other hand, using standard categories decreases the probability of judges' bias (i.e., judges might tend to use a limited number of categories and not think of all potential categories). Further research into this important question is warranted because of the danger that a focus on solely increasing reliability might lead to a decrease in the meaningfulness and validity of what one is measuring.

Test-Retest Reliability and Stability

Little data relevant to either the stability or the test-retest reliability of formulations are available for any of the methods reviewed. Such data are extremely important to obtain because all these measures assume that the content of the maladaptive interpersonal pattern is pervasive across situation and time. Thus, showing that patients' formulations are stable across relatively long periods of time would strengthen the validity of the constructs of interpersonal pattern or dynamic formulation and increase the confidence in the existence of a stable, long-lasting pattern. Preliminary stability data are available for the PF method on 2 patients (Collins & Messer, 1991). On a larger sample, the only indirect evidence addressing the issue of stability is available from the study on change in the pervasiveness of the CCRT during dynamic psychotherapy: wishes do not change, but negative ROs and RSs tend to become less negative (Crits-Christoph & Luborsky, 1990a).

What Is Next

Although preliminary, many of the psychometric values obtained by the group of researchers in this article are extremely encouraging. In contrast to earlier typical dynamic formulations, the methods presented in this review showed that at least the central maladaptive interpersonal pattern (an important aspect of dynamic formulation) can be judged reliably (see J. C. Perry, Luborsky, Silberschatz, & Popp, 1989). As J. C. Perry, Augusto, & Cooper (1989) noted, reasons for this improved reliability include the structure provided by the methods, the training of the judges, the midlevel inferences requested from judges, and the lack of instructions to come up with a comprehensive, encompassing dynamic formulation that explains all

the patient's behaviors. Initial validity data are available for some of the measures, and we want to encourage researchers to address further aspects of construct validity along the lines suggested in our previous discussion.

Although formulation methods are available for theories underlying psychodynamic psychotherapies such as the control-mastery psychoanalysis of Weiss et al. (1986) and the supportive-expressive psychoanalytic psychotherapy (Luborsky, 1984), methods have not been developed for other theories of psychotherapy such as interpersonal psychotherapy (IPT: Klerman, Weissman, Rounsaville, & Chevron, 1984) or even for cognitive therapy. Because the efficacy of IPT has been studied in several clinical trials (e.g., Elkin et al., 1989), research on the nature of clinical formulations within the IPT would seem to be indicated. Of course, one could use any of the methods reviewed in this article to address, for example, whether IPT therapists make interpretations that address the patient's central conflict.

One of the reasons research in this area is difficult is that there are little data available to guide decisions regarding what should be included under the heading of dynamic formulation. For example, defenses and coping mechanisms have a central place in a variety of dynamic psychotherapies. Nevertheless, major methods (e.g. the CCRT) do not directly refer to them. This acknowledged omission (e.g., Luborsky & Crits-Christoph, 1990, pp. 271–272) may reflect the more general problem: The developers of the methods have written little on the theoretical rationale of the components included in their methods. For example, although many methods refer to wishes, there is little mention of the relationships between wishes and psychoanalytic theoretical entities, such as drive derivatives and compromise formations.

We therefore recommend that future studies address the question of which components of dynamic formulations will be most useful. One way to determine the usefulness of dynamic formulation components is through examining their predictive validity in various therapeutic contexts. For example, one could ask whether adding an assessment of defense (as in the ICF) will lead to different therapeutic interventions that, in turn, lead to different outcomes.

Despite some of the limitations noted in our review, we emphasize the major contribution coming from this line of research: the utilization of these methods to address important issues in psychotherapy research. Before these methods were available, there was no reliable way to assess, for example, whether therapists' interpretations addressed a patient's specific maladaptive pattern. The availability of reliable measures of central conflicts enabled researchers to answer positively questions such as whether interpretations accurately targeting the core conflict lead to better within-session outcome (Silberschatz et al., 1986) or to good treatment outcome (Crits-Christoph, Cooper, & Luborsky, 1988). Both Weiss et al. (1986) and Luborsky and Crits-Christoph (1990) have shown that having a reliable formulation of the patient's central issues allows investigators to examine questions related to what patients learn during psychotherapy and how they change in terms of the core conflict. Thus, researchers have begun using these methods as theory-specific measures of outcome. If some of these methods could be translated into self-report format, there is little doubt that they will be used even more often to assess outcome. We

also suggest that other future studies address, for example, the importance of focusing on the core conflict versus conducting relatively unfocused treatment.

References

- Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1990). A guide to the standard categories and their classification. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme Method* (pp. 37–50). New York: Basic Books.
- Benjamin, L. S. (1974). Structural analysis of social behavior. *Psychological Review*, 81, 392–425.
- Benjamin, L. S. (1982). Use of SASB to guide intervention in psychotherapy. In D. Kiesler & J. Anchin (Eds.), *Handbook of interpersonal psychotherapy* (pp. 190–212). New York: Pergamon Press.
- Bond, J., Hansell, J., & Shevrin, H. (1987). Locating a reference paradigm in psychotherapy transcripts: Reliability of relationship episode location in the Core Conflictual Relationship Theme (CCRT) method. *Psychotherapy*, 24, 736–749.
- Cierpka, M., Krannich, S., Ratzke, K., Reich, G., & Homburg, H. (1992, June). *Sex differences in RAP narratives*. Paper presented at the annual international meeting of the Society for Psychotherapy Research, Berkeley, CA.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement on partial credit. *Psychological Bulletin*, 70, 213–220.
- Collins, W. D., & Messer, S. (1988, June). *Transporting the Plan Diagnosis method to a different setting: Reliability, stability, and adaptability*. Paper presented at the annual international conference of the Society for Psychotherapy Research, Santa Fe, NM.
- Collins, W. D., & Messer, S. (1991). Extending the plan formulation method to an object relations perspective: Reliability, stability, and adaptability. *Psychological Assessment*, 3, 75–81.
- Crits-Christoph, P., & Baranackie, K. (1992, June). *The quantitative assessment of relationship themes method*. Paper presented at the annual international meeting of the Society for Psychotherapy Research, Berkeley, CA.
- Crits-Christoph, P., Barber, J. P., & Kurcias, J. S. (1993). The accuracy of therapists' interpretations and the development of the therapeutic alliance. *Psychotherapy Research*, 3, 25–35.
- Crits-Christoph, P., Cooper, A., & Luborsky, L. (1988). The accuracy of therapists' interpretations and the outcome of dynamic psychotherapy. *Journal of Consulting and Clinical Psychology*, 56, 490–495.
- Crits-Christoph, P., Demorest, A., & Connolly, M. B. (1990). Quantitative assessment of interpersonal themes over the course of psychotherapy. *Psychotherapy*, 27, 513–521.
- Crits-Christoph, P., & Luborsky, L. (1990a). Change in CCRT pervasiveness during psychotherapy. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 133–146). New York: Basic Books.
- Crits-Christoph, P., & Luborsky, L. (1990b). The measurement of self-understanding. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 189–196). New York: Basic Books.
- Crits-Christoph, P., & Luborsky, L. (1990c). The perspective of patients versus clinicians in the assessment of the central relationship themes. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 197–208). New York: Basic Books.
- Crits-Christoph, P., Luborsky, L., Dahl, L., Popp, C., Mellon, J., & Mark, D. (1988). Clinicians can agree in assessing relationship patterns in psychotherapy: The Core Conflictual Relationship Theme Method. *Archives of General Psychiatry*, 45, 1001–1004.
- Curtis, J. T., Silberschatz, G., Sampson, H., & Weiss, J. (1992). *The plan formulation method*. Manuscript submitted for publication.
- Curtis, J. T., Silberschatz, G., Sampson, H., Weiss, J., & Rosenberg, S. E. (1988). Developing reliable psychodynamic case formulation: An illustration of the Plan Diagnosis method. *Psychotherapy*, 25, 256–265.
- Dahl, H. (1992, June). *Frames: How to find them and what they're good for*. Panel presented at the annual international meeting of the Society for Psychotherapy Research, Berkeley, CA.
- Dahl, H., Kachele, H., & Thoma, H. (Eds.). (1988). *Psychoanalytic process research strategies*. Berlin, Federal Republic of Germany: Springer-Verlag.
- DeWitt, K. N., Kaltreider, N., Weiss, D. S., & Horowitz, M. J. (1983). Judging change in psychotherapy: Reliability of clinical formulation. *Archives of General Psychiatry*, 40, 1121–1128.
- Eckert, R., Luborsky, L., Barber, J. P., & Crits-Christoph, P. (1990). The CCRT of patients with a DSM III diagnosis of major depression. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 222–234). New York: Basic Books.
- Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J., & Parloff, M. B. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971–982.
- Fried, Y., & Agassi, J. (1977). *Paranoia: A study in diagnosis*. Boston studies in the philosophy of science (Vol. L). Dordrecht, The Netherlands: D. Reidel.
- Henry, W. P., Schacht, T. E., & Strupp, H. H. (1986). Structural analysis of social behavior: Application to a study of interpersonal process in differential therapeutic outcome. *Journal of Consulting and Clinical Psychology*, 54, 27–31.
- Horowitz, L. M., & Rosenberg, S. E. (1992). *The consensual response psychodynamic formulation*. Manuscript submitted for publication.
- Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureño, G., & Villaseñor, V. S. (1988). Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology*, 56, 885–892.
- Horowitz, L. M., Rosenberg, S. E., Ureño, G., Kalebzan, B. M., & O'Halloran, P. (1989). Psychodynamic formulation, consensual response method, and interpersonal problems. *Journal of Consulting and Clinical Psychology*, 57, 599–606.
- Horowitz, M. J. (1989). Relationship schema formulation: Role-relationship models and intrapsychic conflict. *Psychiatry*, 52, 260–274.
- Horowitz, M. J. (Ed.). (1991). *Person schemas and maladaptive interpersonal patterns*. Chicago: University of Chicago Press.
- Horowitz, M. J., & Eells, T. D. (1993). Case formulations using role-relationship model configurations: A reliability study. *Psychotherapy Research*, 3, 57–68.
- Klein, M. H., Mathieu-Coughlan, P., & Kiesler, D. (1986). The Experiencing Scale. In L. S. Greenberg & W. M. Pinsof (Eds.), *The psychotherapeutic process: A research handbook* (pp. 21–71). New York: Guilford Press.
- Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). *Interpersonal psychotherapy of depression*. New York: Basic Books.
- Luborsky, L. (1976). Helping alliance in psychotherapy: The groundwork for a study of their relationship to its outcome. In J. L. Claghorn (Ed.), *Successful psychotherapy* (pp. 92–116). New York: Brunner/Mazel.
- Luborsky, L. (1977). Measuring a pervasive psychic structure in psychotherapy: The core conflictual relationship theme. In N. Freedman & S. Grand (Eds.), *Communicative structures and psychic structures* (pp. 367–395). New York: Plenum Press.

- Luborsky, L. (1984). *Principles of psychoanalytic psychotherapy: A manual for supportive-expressive (SE) treatment*. New York: Basic Books.
- Luborsky, L. (1990a). Alternative measures of the central relationship pattern. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 235–250). New York: Basic Books.
- Luborsky, L. (1990b). The relationship anecdotes paradigm (RAP) interview as a versatile source of narratives. In L. Luborsky & P. Crits-Christoph, *Understanding transference: The Core Conflictual Relationship Theme method* (pp. 102–116). New York: Basic Books.
- Luborsky, L., Barber, J. P., Binder, J., Curtis, J., Dahl, H., Horowitz, L. M., Horowitz, M. J., Perry, J. C., Schacht, T., Silberschatz, G., & Teller, V. (in press). Transference-related measures: A new therapy-based class. In N. Miller, L. Luborsky, J. P. Barber, & J. Docherty (Eds.), *Handbook of dynamic psychotherapy research and practice: A how-to-do-them guide*. New York: Basic Books.
- Luborsky, L., Barber, J. P., & Crits-Christoph, P. (1990). Theory-based research for understanding the process of dynamic psychotherapy. *Journal of Consulting and Clinical Psychology*, 58, 281–287.
- Luborsky, L., Barber, J. P., & Diguier, L. (1992). The meanings of the narratives told during psychotherapy—The fruits of a new observational unit. *Psychotherapy Research*, 2, 277–290.
- Luborsky, L., & Crits-Christoph, P. (1990). *Understanding transference: The Core Conflictual Relationship Theme method*. New York: Basic Books.
- Luborsky, L., Luborsky, E., Diguier, L., Schaffler, P., Schmidt, K., Dengler, D., Faude, J., Morris, M., Buchsbaum, H., & Emde, R. (in press). Is there a core relationship pattern at age 3 and does it remain at age 5? In G. Noam & K. Fisher (Eds.), *Development and vulnerabilities in relationships*. New York: Erlbaum.
- Malan, D. M. (1976). *Toward the validation of dynamic psychotherapy*. New York: Plenum Press.
- Mann, J., & Goldman, R. (1982). *A casebook in time-limited psychotherapy*. New York: McGraw-Hill.
- Mintz, J. (1981). Measuring outcome in psychodynamic psychotherapy. *Archives of General Psychiatry*, 38, 503–506.
- Murray, H. (1938). *Explorations in personality: A clinical and experimental study of fifty men*. Oxford: Oxford University Press.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Perry, J. C. (1989). Scientific progress in psychodynamic formulation. *Psychiatry*, 52, 245–249.
- Perry, J. C. (1992). *The idiographic conflict formulation method*. Manuscript submitted for publication.
- Perry, J. C., Augusto, F., & Cooper, S. H. (1989). Assessing psychodynamic conflicts: I. Reliability of the Idiographic Conflict Formulation Method. *Psychiatry*, 52, 289–301.
- Perry, J. C., Luborsky, L., Silberschatz, G., & Popp, C. (1989). An examination of three methods of psychodynamic formulation based on the same videotaped interview. *Psychotherapy*, 52, 302–323.
- Perry, S., Cooper, A. M., & Michels, R. (1987). The psychodynamic formulation: Its purpose, structure, and clinical application. *American Journal of Psychiatry*, 144, 543–550.
- Persons, J. B. (1991). Psychotherapy outcome studies do not accurately represent current models of psychotherapy. *American Psychologist*, 46, 99–106.
- Popper, K. R. (1979). *Objective knowledge: An evolutionary perspective* (rev. ed.). Oxford, England: Oxford University Press.
- Rosenberg, S. E., Silberschatz, G., Curtis, J. T., Sampson, H., & Weiss, J. (1986). A method for establishing the reliability of statements from psychodynamic case formulations. *American Journal of Psychiatry*, 143, 1454–1456.
- Schacht, T. E., & Henry, W. P. (1992). *Modeling recurrent relationship patterns with Structural Analysis of Social Behavior: The SASB-CMP*. Manuscript submitted for publication.
- Seitz, P. F. D. (1966). The consensus problem in psychoanalytic research. In L. Gottschalk & A. Auerbach (Eds.), *Methods of research in psychotherapy* (pp. 209–225). New York: Appleton-Century-Crofts.
- Silberschatz, G., Fretter, P. B., & Curtis, J. T. (1986). How do interpretations influence the process of psychotherapy? *Journal of Consulting and Clinical Psychology*, 54, 646–652.
- Strupp, H. S., & Binder, J. L. (1984). *Psychotherapy in a new key: A guide to time-limited dynamic psychotherapy*. New York: Basic Books.
- Teller, V., & Dahl, H. (1981). The framework for a model of psychoanalytic inference. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, 1, 394–400.
- Teller, V., & Dahl, H. (1992). *The characteristics, identification, and application of frames*. Manuscript submitted for publication.
- Wachtel, P. L. (1983). Vicious circles: The self and the rhetoric of emerging and unfolding. *Contemporary Psychoanalysis*, 18, 259–273.
- Weiss, J., Sampson, H., & the Mount Zion Psychotherapy Research Group. (1986). *The psychoanalytic process: Theory, clinical observation & empirical research*. New York: Guilford Press.

Received May 13, 1992

Revision received December 10, 1992

Accepted January 21, 1993 ■