

# A Method for Establishing Reliability of Statements From Psychodynamic Case Formulations

Saul E. Rosenberg, Ph.D., George Silberschatz, Ph.D.,  
John T. Curtis, Ph.D., Harold Sampson, Ph.D., and Joseph Weiss, M.D.

---

*The authors describe a method for assessing the reliability of statements derived from psychodynamic case formulations. By reducing a narrative formulation into a series of distinct statements and rating their relevance to a particular patient, they obtained good interjudge reliabilities.*

(Am J Psychiatry 143:1454-1456, 1986)

---

Psychodynamic case formulations are widely used in clinical practice to understand the meaning of a patient's problems, to guide therapeutic interventions, and to evaluate response to treatment. Nonetheless, dynamic formulations have rarely been used in psychotherapy research because of difficulties in establishing satisfactory interjudge reliability for complex narratives (1). A few investigators have achieved satisfactory interrater agreement by reducing complex formulations into smaller units, for example, a description of a patient's goal or wish (2, 3). In this paper, we describe a method for determining the reliability of statements from dynamic formulations of patients in brief psychodynamic psychotherapies.

In this study, the dynamic case formulations were based on a cognitive psychoanalytic theory developed by Weiss and Sampson (4). According to this theory, many inhibitions and psychological symptoms stem from pathogenic beliefs and the fears, guilt, and anxiety associated with them. Because pathogenic beliefs are often frightening and disturbing and may lead a patient to renounce certain goals, the patient is powerfully motivated to overcome these beliefs in therapy. One of the patient's primary efforts in therapy is to

disconfirm pathogenic beliefs by testing them in the relationship with the therapist.

We refer to a psychodynamic formulation based on this model as a "plan formulation." Caston (2) developed a method, which he called "plan diagnosis," for assessing the reliability of statements derived from clinical formulations. We have applied this method to formulations of brief psychodynamic therapy cases.

## METHOD

Five patients were selected from a project on the process and outcome of brief (16-session) psychodynamic psychotherapy. All patients in the project gave signed informed consent to have an audiotape made of their therapies to be used for research. The five patients in this study ranged in age from 29 to 59 years. Although they all suffered from neurotic and personality disorders, they differed substantially in the nature and severity of their presenting complaints and in terms of the psychodynamic issues underlying their problems.

A team of four or five experienced clinicians, who shared the cognitive psychoanalytic theoretical orientation previously described (4), studied the verbatim transcripts of the intake interview and the first 2 therapy hours of each patient. For each, the team developed a written plan formulation by consensus. In addition to a description of the patient's presenting problems and history, each formulation contained 1) the patient's goals for therapy, 2) the inner obstacles (pathogenic beliefs) preventing the attainment of goals, 3) the means by which the patient would work in therapy to disconfirm his or her pathogenic beliefs (tests), and 4) the insights that would be helpful to the patient. The five formulations were quite different, as were their respective plan components.

For each patient, the four plan components were reduced to a series of concise statements. For example, the narrative description of a patient's therapy goals was recast as a list of discrete goals. A similar procedure was used to generate lists of obstructions, tests, and insights for each patient.

Each list was then expanded to include a comparable number of items that the clinical team felt were of

---

Revised version of a paper presented at the annual meeting of the Society for Psychotherapy Research, Sheffield, England, June 18-23, 1983. Received March 30, 1984; revised Sept. 9, 1985, and Jan. 27, 1986; accepted April 30, 1986. From the Department of Psychiatry, Mount Zion Hospital and Medical Center, San Francisco. Address reprint requests to Dr. Rosenberg, Langley Porter Psychiatric Institute, Box 8D, University of California, San Francisco, CA 94143.

Supported in part by NIMH grants MH-34052 and MH-35230, by a grant from the Chapman Research Fund, and by the Research Support Program of Mount Zion Hospital and Medical Center.

Copyright © 1986 American Psychiatric Association.

lesser or no relevance to the patient but that another clinician might include in a formulation. To ensure the plausibility of these items, we developed alternative formulations that would be reasonable from the point of view of our theory and took the items from them in the same manner used for the clinical team's formulation. As an additional safeguard, items that received high-priority ratings for other patients but that, although plausible, were not high priorities for the selected patient were also included as alternatives. This ensured against items being rated on the basis of style or apparent adherence to a certain theoretical perspective rather than on the basis of their relevance to the patient in question. (Detailed lists are available from S.E.R.)

Because of the procedures for developing alternative items, some items appeared on more than one patient's list. Still, it was more the exception than the rule that a given item could be applied to more than one patient and yet appear specific to and reasonable for each patient. Because each formulation—and thus each item derived from the formulation—was patient specific, the vast majority of items were necessarily tailored for and unique to a given patient.

For each of the four plan components, an array of 15–30 items (ranging from highly pertinent to irrelevant) was generated. The clinical team that prepared the plan formulation then rated the degree to which each goal, obstruction, insight, and test was pertinent for that patient. These ratings were made on a 9-point scale ranging from 0 (item not relevant) to 8 (item highly relevant).

For each patient a new team of four judges (the reliability judges) who adhered to our theoretical approach read the same verbatim transcripts that were studied by the clinicians who prepared the plan formulation. Each reliability judge then independently rated (on the 9-point scale previously described) the relevance of each goal, obstruction, insight, and test for the given patient.

*Case example.* Two elderly men sought treatment, each with the complaint of feeling conflict over retirement. Each man's spouse had recently developed a serious illness, and this had led each of the men to consider retiring to be more available to his wife. In many other ways, their histories, complaints, and symptoms were similar. However, the nature of the underlying conflicts of each man was not obvious, and the formulation team extensively deliberated each case. For one of these patients, Mr. A, the formulation team determined that his conflict stemmed from guilt feelings over being more healthy than his wife and over not wanting to retire and devote all of his time to her. The formulation team determined that the second patient, Mr. B, actually wanted to retire but felt undeserving of the freedom such a move would afford him and felt compelled to remain a "worker." The lists of goals developed for these patients shared some items about retirement that were reasonable and potentially relevant for each man. However, the goals relating to this issue that were appropriate for one patient were not necessarily of high priority to the other. Thus, the goal "patient would like to continue working" was rated a 6 (high

TABLE 1. Interjudge Reliabilities for Plan Formulations

Case	Agreement of Reliability Judges		Correlation Between Clinical Team and Reliability Judges (r)
	Average Judge <sup>a</sup>	Pooled Judges <sup>b</sup>	
Patient 1			
Goals	.55	.83	.87
Obstacles	.61	.86	.87
Tests	.14	.39	.70
Insights	.50	.80	.94
Patient 2			
Goals	.88	.97	.86
Obstacles	.52	.81	.73
Tests	.74	.92	.67
Insights	.74	.92	.87
Patient 3			
Goals	.63	.89	.79
Obstacles	.46	.81	.62
Tests	.41	.78	.80
Insights	.80	.95	.97
Patient 4			
Goals	.63	.89	.87
Obstacles	.67	.91	.94
Tests	.56	.87	.91
Insights	.62	.89	.82
Patient 5			
Goals	.77	.93	.88
Obstacles	.73	.92	.90
Tests	.77	.93	.95
Insights	.77	.93	.96

<sup>a</sup>Intraclass correlation for a single, randomly chosen judge.

<sup>b</sup>Intraclass correlation of the pooled judges' ratings (N=4).

priority) for Mr. A but a 2 (low priority) for Mr. B. On the other hand, the obstruction "patient would feel guilty not working" was rated a 7 for Mr. B and a 1 for Mr. A.

## RESULTS

Because a different plan formulation was developed for each patient, there was little overlap of items among patients. As a result, rather than using a Rater by Subject design and reporting an intraclass correlation for the method, we present reliability statistics for each patient.

The interjudge agreement for each of the four plan components (i.e., the agreement among the reliability judges) was assessed by using an intraclass correlation for pooled judges' ratings (5). To assess the degree of overlap between the first (plan formulation) team of clinicians and the second (reliability) team, Pearson correlations between the average team ratings for each item were compared. These data are presented in table 1.

There was a high level of agreement among the reliability judges. They agreed on which goals were priorities for any given patient, which obstacles were particularly operative, which insights were relevant, and which tests were pertinent. There was also good agreement between the two teams of judges. Those goals, obstructions, insights, and tests identified as high priorities for a given patient by the clinical team

were also rated as high priorities by the team of reliability judges.

These results cannot be attributed solely to judges viewing all patients in the same way (i.e., as having essentially the same goals, obstructions, etc.), since certain items that were rated as highly relevant for one patient were rated as irrelevant for another. Rather, our data show that it is possible to establish good interjudge reliabilities for psychodynamic case formulations.

#### COMMENT

Three factors contributed to these successful findings: 1) the raters adhered to the same theory, 2) the narrative formulations were of uniform structure and could be reduced to a series of distinct statements, and 3) clinical judges rated lists of items for pertinence to a particular patient rather than generating their own formulations.

Generating a dynamic formulation from the theoretical vantage point we describe requires clinical experience with the theory and cannot be learned from a training manual. However, although this procedure was developed within our particular conceptual frame-

work, it can be applied by investigators who adhere to different theoretical perspectives.

This procedure does not address the issue of whether clinicians with diverse theoretical viewpoints can generate similar formulations. Rather, it measures the degree to which clinicians who share a theoretical orientation agree on which statements should be included in a formulation for a given patient.

#### REFERENCES

1. Seitz PFD: The consensus problem in psychoanalytic research, in *Methods of Research in Psychotherapy*. Edited by Gottschalk L, Auerbach AH. New York, Appleton-Century-Crofts, 1966
2. Caston J: The reliability of the diagnosis of the patient's unconscious plan, in *The Psychoanalytic Process: Theory, Clinical Observations, and Empirical Research*. Edited by Weiss J, Sampson H, Mount Zion Psychotherapy Research Group. New York, Guilford Press (in press)
3. Levine TJ, Luborsky L: The core conflictual relationship theme: a demonstration of reliable clinical inference by the method of mismatched cases, in *Object and Self: A Developmental Approach*. Edited by Tuttmann S, Kay C, Zimmerman M. New York, International Universities Press, 1981
4. Weiss J, Sampson H, Mount Zion Psychotherapy Research Group: *The Psychoanalytic Process: Theory, Clinical Observations, and Empirical Research*. New York, Guilford Press (in press)
5. Tinsley HE, Weiss DJ: Interrater reliability and agreement of subjective judgments. *J Counseling Psychol* 22:358-376, 1975