

1-1-1978

## The effects of surveillance on clinical ratings.

Suzanne W. Hadley  
*University of Massachusetts Amherst*

Follow this and additional works at: [https://scholarworks.umass.edu/dissertations\\_1](https://scholarworks.umass.edu/dissertations_1)

---

### Recommended Citation

Hadley, Suzanne W., "The effects of surveillance on clinical ratings." (1978). *Doctoral Dissertations 1896 - February 2014*. 1517.  
<https://doi.org/10.7275/aqhv-k478> [https://scholarworks.umass.edu/dissertations\\_1/1517](https://scholarworks.umass.edu/dissertations_1/1517)

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

312066013538166

THE EFFECTS OF SURVEILLANCE ON CLINICAL RATINGS

A Dissertation Presented

By

SUZANNE W. HADLEY

Submitted to the Graduate School of the  
University of Massachusetts in partial fulfillment  
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1978

Psychology

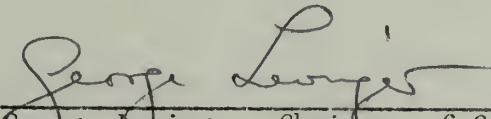
THE EFFECTS OF SURVEILLANCE ON CLINICAL RATINGS

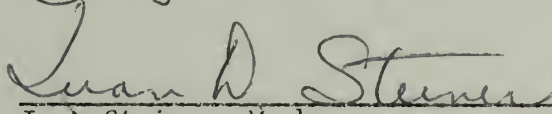
A Dissertation Presented

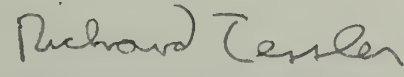
By

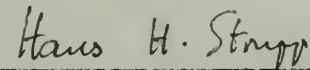
SUZANNE W. HADLEY

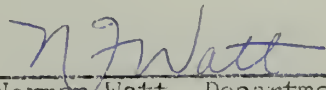
Approved as to style and content by:

  
George Levinger, Chairman of Committee

  
Ivan Steiner, Member

  
Richard Tessler, Member

  
Hans Strupp, Adjunct Member

  
Norman Watt, Department Head  
Department of Psychology

## DEDICATION

This dissertation is dedicated to the memory of Christina.

## ACKNOWLEDGEMENTS

I acknowledge with deep gratitude the guidance and support of all the members of my dissertation committee, and most especially, Dr. George Levinger.

More generally, I acknowledge the special esteem I feel for Dr. Seymour Berger, Dr. George Levinger, Dr. F. J. McGuigan, Dr. Ivan Steiner, and Dr. Hans H. Strupp. Each of these individuals has contributed uniquely to my professional ideals. Their personal warmth, no less than their scholarly excellence, is their distinction.

My thanks to Dr. Robert Vidulich and Robert Kores at Memphis State University and to Dr. Harold Fine and Timothy Pollock at the University of Tennessee for their assistance.

This investigation was carried out while the author was a Research Assistant at Vanderbilt University. Partial support was provided by NIMH Grant #5 R01 MH20369-06, Hans H. Strupp, Principal Investigator.

## ABSTRACT

### The Effects of Surveillance on Clinical Ratings

(February, 1978)

Suzanne W. Hadley, B.S., Middle Tennessee State University

M.A., Hollins College, Ph.D., University of Massachusetts

Directed by: Dr. George Levinger

The data base of psychotherapy research consists in large part of clinical ratings made by observers, often in the actual or implied presence of others. Previous research in social psychology suggests that the presence of others may increase one's evaluation apprehension and thus lead to a tendency to conform with group norms. Social surveillance is frequently operative in clinical rating sessions; however, only scant attention has been given to the possibility that surveillance may affect these ratings.

This experiment varied the conditions under which ratings were made. Ratings in one group were made under conditions of anonymity and minimal surveillance. Ratings in another group were made under implied surveillance by a group of experts in analytic psychotherapy. A third group of subjects made their ratings under conditions of anticipated surveillance by their peers.

It was hypothesized that there would be differences in the ratings made under these three conditions, although the dearth of previous research on surveillance and clinical ratings made it



impossible to specify more precisely the form these differences might take. Evaluation apprehension and conformity--considered possible mediators of the anticipated effects of surveillance on the clinical ratings--were also assessed.

It was found that surveillance by peers led to greater evaluation apprehension and lower rater confidence. On the clinical ratings, high competence raters under conditions of peer surveillance evaluated the therapist less favorably than did high competence raters under conditions of minimal surveillance. These differences were manifested on a number of therapist variables, including personal qualities as well as technical competence.

Differences in the clinical ratings as a function of level of experience, competence, and theoretical orientation of the raters were also found. Conformity in the present investigation was measured by the amount of discrepancy between a subject's own clinical ratings and his estimate of the ratings of experts and peers. Very few group differences in conformity were found.

Implications of the study for previous and future research--both social and clinical--were considered. Particular mention was made of the need for further research into the effects of surveillance on clinical ratings, especially the joint effects of surveillance and certain aspects of a rater's self-concept. The importance of assessing both therapist and patient qualities was also noted.



## TABLE OF CONTENTS

|                         |   |     |
|-------------------------|---|-----|
| DEDICATION              |   | iii |
| ACKNOWLEDGEMENTS        |   | iv  |
| ABSTRACT                |   | v   |
| LIST OF TABLES          |   | ix  |
| LIST OF FIGURES         |   | xi  |
| Chapter                 |   |     |
| I. INTRODUCTION         |   | 1   |
|                         | Overview of Observer Ratings in Clinical Judgments  |     |
|                         | Validity and Reliability of Clinical Ratings        |     |
|                         | Varieties of Influence on Clinical Ratings          |     |
|                         | The Effects of Surveillance on Clinical Ratings     |     |
|                         | A Social Psychological Perspective on               |     |
|                         | Surveillance Effects                                |     |
|                         | The Special Significance of Agreement in            |     |
|                         | Clinical Judgments                                  |     |
|                         | Description and Hypotheses of Present Investigation |     |
| II. PILOT INVESTIGATION |   | 22  |
|                         | Evaluation Apprehension                             |     |
|                         | Clinical Ratings                                    |     |
| III. METHOD             |   | 30  |
|                         | Raters  |     |
|                         | Procedure   |     |
|                         | Assessment Battery                                  |     |
| IV. RESULTS             |   | 36  |
|                         | Overall Strategy and Preliminary Analyses           |     |
|                         | Demographic and Independent Variables               |     |
|                         | Dependent Variables                                 |     |

|   |     |
|---|-----|
| V. DISCUSSION . . . . .                         | 94  |
| Major Results--and "Why?"                       |     |
| Implications for Social Psychology              |     |
| Implications for Clinical Research and Practice |     |
| A Final Caution                                 |     |
| BIBLIOGRAPHY . . . . .                          | 110 |
| APPENDIX A . . . . .                            | 116 |
| APPENDIX B . . . . .                            | 119 |

## LIST OF TABLES

|     |   |    |
|-----|---|----|
| 1.  | Correlation Matrix of Independent Variables . . . . .                             | 37 |
| 2.  | Demographic Data . . . . .  | 41 |
| 3.  | Evaluation Apprehension . . . . .   | 43 |
| 4.  | Self-Perceived Influence of Surveillance . . . . .                                | 44 |
| 5.  | Preference for Anonymity . . . . .  | 45 |
| 6.  | Results of Factor Analysis: Degree of Confidence in<br>Clinical Ratings . . . . . | 46 |
| 7.  | Confidence in Ratings of Prognosis . . . . .                                      | 47 |
| 8.  | Confidence in Ratings of Therapist Activity . . . . .                             | 48 |
| 9.  | Results of Factor Analysis: Clinical Ratings . . . . .                            | 50 |
| 10. | Global Evaluation of Therapist . . . . .  | 52 |
| 11. | General Reaction to Patient . . . . .   | 54 |
| 12. | General Prognosis for Patient . . . . .   | 55 |
| 13. | Patient Anxiety . . . . .   | 56 |
| 14. | Patient Depression . . . . .  | 57 |
| 15. | Patient Adjustment . . . . .  | 58 |
| 16. | Prognosis Without Therapy . . . . .   | 59 |
| 17. | Prognosis With Therapy . . . . .  | 60 |
| 18. | Liking for Patient . . . . .  | 61 |
| 19. | Therapist Activity . . . . .  | 62 |
| 20. | Therapist Supportiveness . . . . .  | 63 |
| 21. | Therapist Competence . . . . .  | 64 |

|     |  |    |
|-----|--|----|
| 22. | Therapist Orientation . . . . .  | 65 |
| 23. | Liking for Therapist . . . . .   | 67 |
| 24. | Resemblance to Therapist . . . . .   | 68 |
| 25. | Determination of Therapy Goals . . . . .                                       | 69 |
| 26. | Self-Other Total Discrepancy Scores . . . . .                                  | 79 |
| 27. | Results of Factor Analysis: Self-Expert Discrepancy .                          | 82 |
| 28. | Results of Factor Analysis: Self-Peer Discrepancy . .                          | 83 |
| 29. | General Reaction to Therapist (Self-Expert Discrepancy)                        | 84 |
| 30. | General Reaction to Patient (Self-Expert Discrepancy)                          | 85 |
| 31. | Therapist Activity (Self-Expert Discrepancy) . . . . .                         | 86 |
| 32. | Patient Pathology (Self-Expert Discrepancy) . . . . .                          | 87 |
| 33. | General Reaction to Therapist (Self-Peer Discrepancy)                          | 88 |
| 34. | General Reaction to Patient (Self-Peer Discrepancy) .                          | 89 |
| 35. | Evaluation of Patient--Therapist Interaction (Self-Peer Discrepancy) . . . . . | 90 |
| 36. | Cell N's for Analysis of Variance Tables . . . . .                             | 91 |

## LIST OF FIGURES

|    |  |    |
|----|--|----|
| 1. | Global Evaluation of Therapist Showing Surveillance--<br>Competence Interaction . . . . .      | 53 |
| 2. | Personal Reaction to Therapist Showing Surveillance--<br>Competence Interaction . . . . .      | 71 |
| 3. | Resemblance to Therapist Showing Surveillance--<br>Competence Interaction . . . . .            | 72 |
| 4. | Ratings of Patient Anxiety Showing Surveillance--<br>Competence Interaction . . . . .          | 73 |
| 5. | Ratings of Prognosis Without Therapy Showing<br>Surveillance--Competence Interaction . . . . . | 74 |

# CHAPTER I

## INTRODUCTION

### Overview of Observer Ratings in Clinical Judgments

Clinical judgments are of considerable importance--both for patients and for therapists and society as well. Judgments of mental illness or health determine who shall--who must--be treated, who shall be confined and who shall go free. Diagnoses of mania, depression, paranoia, or schizophrenia provide the rationale for initiating particular therapeutic interventions which may take several years. Even more serious is the thesis of a number of experts that terms like "mental illness" are in reality little more than societal repression of deviance (Kittrie, 1971; Szasz, 1977).

Furthermore, therapists themselves--their personalities and their techniques--are increasingly the focus of clinical judgments. Clinical training has for years been based on observation and supervision of the trainee by a more experienced therapist who judges the therapeutic competence of the trainee. More recently, based on increased awareness of the necessity for evaluating the therapist's competence, there has begun a movement toward in vivo evaluation of therapists for licensing and certification (Bergin, 1976). These clinical decisions at the



individual level affect public policy and expenditures. Thus, the personal and social implications of clinical judgments are both far-reaching and profound.

Clinical judgments are based primarily on test scores and observer ratings. Human observers, while avowedly subject to more measurement errors than test batteries, are used both to generate and to integrate clinical data on the assumption that observers bring to these tasks unique capabilities not supplied by more "mechanical" methods (Sawyer, 1966). Meehl (1954), for example, suggests that "the brain's 'superiority' shows up heavily at the level of perception itself" (p. 27).

Observer ratings may vary along a number of dimensions. The raters may be altogether extrinsic to the therapy situation or may be intimately involved, as when the patient or therapist makes ratings. (The difference in the demand characteristics and consequently, in the objectivity afforded by these extreme perspectives is obvious.)

Observer ratings have been used both as independent and dependent variables and have included ratings of patient and therapist qualities (Auerbach & Luborsky, 1968); process aspects of psychotherapy (Marsden, 1965; Mintz, Luborsky, & Auerbach; 1971); and psychotherapy outcome (Cristol, 1972). The data generated have represented a variety of theoretical orientations, including behavioral (Lewisohn, Weinstein, & Alper; 1970); Rogerian (Kiesler, 1966); and psychoanalytic (Forer, Farberow, Feifel, Meyer, Sommers, & Tolman; 1961). Similarly, the

ratings have ranged from simple observations of the frequency of behaviors to complex judgments of highly inferential diagnostic and prognostic variables.

### Validity and Reliability of Clinical Ratings

Validity and reliability are fundamental limiting conditions in all measurement, but are of special importance in the assessment of clinical variables. The above-noted distinction among varieties of clinical ratings, ranging from the immediately observable to the highly inferential, has special implications for reliability and validity. Observations of overt behaviors have considerable "face validity" (Hall, 1974) and assessments of validity for such observations may involve nothing more than verifying the accuracy of frequency counts. Assessments of reliability for measures of observable behaviors and the like are, therefore, useful primarily for establishing the stability of these measures across time, stimulus, or rater dimensions.<sup>1</sup>

The role of reliability as an index of stability for immediately observable variables is also applicable for more inferential variables. However, these latter variables are often heavily value-laden (Strupp & Hadley, 1977) and have no immediate referent against which accuracy

---

<sup>1</sup>The terms "reliability" and "agreement" are often used interchangeably in clinical research. However, as several authors (Lawlis & Lu, 1972; Tinsley & Weiss, 1975) have observed, reliability and agreement are not synonymous, but are conceptually and statistically distinct. The present study was concerned with agreement and this term will be used throughout.

can be assessed. Thus, reliability (and inter-rater agreement in particular) fulfills an additional function as a major "limiting factor" in validity (Seitz, 1966; Fleiss, Spitzer, Endicott & Cohen, 1972; Taibelson, Woodruff, Reich, & Wish; 1966).

Although some authors have cautioned that "there is no guarantee that a reliable system is valid" (Spitzer & Fleiss, 1974), it appears that many authors implicitly assume otherwise. Coefficients of inter-rater agreement are frequently presented in the clinical literature with little or no additional information on scale validity given. Indeed, for some of the more highly inferential clinical measures, there may be no external validating criteria available, with the consequence that agreement for these ratings is treated as the functional equivalent of validity. In short, accuracy in clinical ratings is essential. And agreement or consensus is considered a major index of accuracy. Study of the variables which may affect accuracy and agreement is thus of considerable importance.

#### Varieties of Influence on Clinical Ratings

In spite of the manifest importance of observer judgments in clinical practice and research, there has been relatively little investigation of precisely what those judgments comprise. It is only in recent years that systematic study of the clinical observation process itself--of variables other than those supposedly being assessed that may impinge upon and influence clinical ratings--has begun.

Bordin and his associates (1954) were among the first to call attention to general measurement problems in therapy research. They stressed that "the meaningfulness of research in psychotherapeutic interaction depends heavily on the degree to which the measurement issues involved in such research have been dealt with." They concluded that "Many of the results of current research on psychotherapy are open to question because of an apparent failure to recognize this fundamental point" (Bordin, Cutler, Dittmann, Harway, Raush, & Rigler, 1954, p. 82). Bordin et al went on to suggest that among the measurement issues which must be dealt with are the effects of measurement devices, the judges' understanding of their task, the method of presentation of the stimuli, and the training and attitude of the judges.

Zax and Klein (1960) focused specifically on variables which might affect ratings of therapy-induced change. They noted a major limitation to the validity of patient self-report measures which is equally applicable to therapy assessments in general: "It seems likely that the content of such assessments depends greatly upon who asks for it and the circumstances under which it is requested" (p. 444).

Hunt and Jones (1962) reiterated both the necessity and the possibility of studying clinical observations. They argued against the view that clinical judgments, because they are based on human intuition, cannot be scientifically studied. They concluded that, on the contrary, "The clinician of today is on firm ground in viewing

his judgmental processes as a natural behavioral phenomenon open to all the investigative procedures of experimental psychology" (p. 28).

Hunt and Jones suggested that clinical judgments might profitably be viewed as an analogue of psychophysical observations and judgments. Using this approach, they suggested the necessity for studying context effects in the clinical judgment situation, including (1) qualities of the stimulus itself (how much is known about the patient, what kinds of biographical data and/or test scores are available to the observer); (2) the medium by which the stimulus is presented (visual, written transcript, direct observation, etc.); (3) qualities of the rater (professional experience, prior history with the stimulus person); and (4) task characteristics (number and kind of ratings to be made).

Hunt and Jones' suggestions were valuable, but as they themselves acknowledged, were little more than suggestions. Yet as they observed, "ignorance must be pinpointed before it can be remedied" (p. 49). In recent years, an increasing number of studies have been aimed at this "pinpointing of ignorance," which indicates the seriousness with which clinicians currently view context effects. A considerable number of factors have been identified as exercising a significant influence on clinical ratings. The following summary is by no means exhaustive, but is representative of the variety of these factors and their spheres of influence.

Influence of patient variables on clinical ratings. Knowledge of patient variables other than those being assessed is a potent influence



on clinical ratings (Mintz, 1972). Not all clinical ratings are equally vulnerable to this influence, however. There is some indication that ratings which require less observer inference are less subject to bias from this variable than are ratings which require greater inference by the observer (Redfield & Paul, 1976). Further, the potential for bias is greater for ratings of ambiguous behaviors (Johnson & Ryan, 1976) and for ratings which involve more subjective judgments (Shuller & McNamara, 1976).

Knowledge of the severity, consistency, and typicalness of a patient's behavior has been shown to influence clinicians' causal attributions for that behavior (Johnson, Calhoun, & Boardman; 1975). Causal attributions in turn affect observers' ratings of mental illness (and may lead to social rejection of the patient [Calhoun, Selby, & Wroten; 1977]). Strasburger and Jackson (1977) found that different kinds of information regarding personality test responses have different effects on observers' ability to make clinical predictions of other responses made by the target person. These authors observe that "Redundant or irrelevant data tend only to reinforce judges' erroneous inferences" (p. 308).

The setting for an individual's behavior may affect attributions of mental illness (Calhoun et al, 1977); these effects are apparently mediated by the degree of appropriateness of the behavior in the situational context. For example, the mere fact that an individual is observed in a "patient context" leads to more frequent and severe ratings of maladjustment and pathology (Temerlin & Trousdale,



1969; Langer & Abelson, 1974). Diagnostic labels, and importantly, the source of those labels, have been shown repeatedly to affect clinical judgments (Temerlin, 1968; Temerlin & Trousdale, 1969; Lee & Temerlin, 1970). DiNardo (1975) reported that diagnoses made by psychiatrists affect observers' clinical judgments, while diagnoses made by psychologists do not, indicating a professional or status bias.

Finally, a patient's sex or social class (whether obvious or merely inferred) may influence clinical judgments by observers. Sex per se is apparently not as potent a variable as the sex role appropriateness of the behaviors or attitudes observed (Zeldow, 1975; 1976). Lower social class results in judgments of greater pathology and poorer prognoses for a patient (Lee & Temerlin, 1970; DiNardo, 1975).

#### Influence of observer and stimulus variables on clinical ratings.

Although studied less frequently than patient variables, observer variables have also been shown to exert a pervasive influence on clinical judgments. An observer's emotional state may affect both the kind and the intensity of emotion attributed to another (Schiffenbauer, 1974). Zeldow (1975) found that female raters tend to perceive a greater need for therapeutic intervention than do males.

Systematic differences in ratings as a function of observers' theoretical orientation and clinical experience have also been reported. Langer and Abelson (1974) found analytically-oriented clinicians more vulnerable than behavioral therapists to labelling bias in patient evaluations. Kächele, Kühn, Grünzig, and Ohlmeier (1975) found

differences in ratings of process variables of group therapy related to observers' orientation. By contrast, Raskin (1965) found no effects of observer orientation on evaluations of therapist variables.

As to level of experience, Temerlin (1968) reported no differences in susceptibility to labelling bias related to experience. However, Fleiss, Spitzer, and Burdock (1965) found greater expert--observer rating discrepancy for less experienced observers. These authors suggested that observer training may need to be increased to insure obtaining reliable ratings.

Finally, Kiesler (1966) investigated the effects on ratings of a variety of stimulus dimensions, including the medium through which the stimulus is presented to observers. While his results showed no effect on ratings as a function of differences in length of therapy segments rated, he did find that the location in time of the segments does influence ratings.

In sum, the studies described above reveal that a considerable number of factors extraneous to the clinical dimensions being measured may influence clinical ratings. The unqualified acceptance of these ratings as valid and meaningful is thus not justified. Temerlin and Trousdale speak pessimistically of the use of observer ratings in psychodiagnosis (1969, p. 26). More generally, DiNardo has concluded that studies such as those described above "are providing converging lines of evidence that indicate a clinician's assessment of a patient may not always represent the clinical reality presented by the patient" (1975, p. 367).

### The Effects of Surveillance on Clinical Ratings

The variable examined in this study is another kind of contextual variable, namely, the extent to which surveillance by others may influence an observer's clinical ratings. The potential influence of this variable is considerable, for in nearly every situation in which clinical ratings are made, raters know their judgments will be subject to some scrutiny--whether from peers, as when conference ratings are made; from researchers during the process of data analysis; or from administrators who review the findings.

Yet investigations of the impact of surveillance on clinical ratings are rare. Romanczyk, Kent, Diament, and O'Leary (1973) showed that when raters know inter-rater agreement is being assessed, both the agreement and the validity of behavioral observations are affected. Raters' observations showed consistently higher agreement with the observations of an identified assessor than with those of an unidentified assessor. (Subjects were aware of the unique rating scheme used by each assessor.) Agreement with both identified and unidentified assessors under overt assessment conditions was consistently higher than agreement with assessors under covert assessment conditions. Finally, frequency of reported behavior in the covert assessment condition was only 75% of that reported in the overt assessment condition.

Kent, O'Leary, Diament, and Dietz (1974) replicated Romanczyk et al's results in part, showing that when pairs of observers knew agreement within pairs was being assessed, agreement within pairs was

high relative to the agreement within pairs under a surreptitious assessment condition; on the other hand, the agreement between pairs was lower in the overt assessment condition than under surreptitious assessment. The implication was that overt assessment of intra-group agreement increased agreement within the group while decreasing the agreement among the groups. The authors concluded that, "if reliability assessment is to reflect the characteristics of data generated by observers in their routine application of a behavioral code, assessment must be unknown to the observer" (p. 191).

Thus, from this very limited evidence, it appears that in general, different conditions of surveillance of clinical ratings may result in differences both in the level of agreement on clinical ratings and in the content of clinical ratings. The need for further research into the specific nature of these effects is evident.

#### A Social Psychological Perspective on Surveillance Effects

Many of the variables shown to influence clinical ratings are social variables; it is therefore somewhat surprising that there has been so little effort made toward integrating the relevant social and clinical psychological literature. This is particularly true as regards surveillance effects, for two concepts in the social literature, evaluation apprehension and conformity, appear of considerable relevance to the clinical judgment situation.

The concept of evaluation apprehension was introduced by Rosenberg (1965) in the context of a discussion of the phenomenon of dissonance. Rosenberg defined evaluation apprehension as "an active, anxiety-toned concern that he (the subject in a typical psychology experiment) win a positive evaluation from the experimenter, or at least that he provide no grounds for a negative one" (p. 29). Further, it was Rosenberg's view that evaluation apprehension is a potent variable which affects subjects' behavior in many experimental settings, particularly those devoted to the study of dissonance.

Henchy and Glass (1968) invoked evaluation apprehension as an explanation for the phenomenon of social facilitation, in which the "mere presence" of others appears to enhance individual performance at a variety of tasks (Zajonc, 1968). Henchy and Glass suggested that the presence of others is not a social facilitator in and of itself, but rather, facilitates performance only to the extent those others are presumed to be in a position to evaluate the actor's performance.

Similarly, Steiner observed:

Recent evidence suggests that the presence of others is arousing only (or primarily) when the others who are present are believed to be in a position to affect one's outcome. This is the case when others are expected to evaluate one's performance (perhaps adversely), to demonstrate by their own superior proficiency that one's performance is inferior, or to deprive one of desired goals (1972, p. 133).

The link between the social and clinical settings is obvious, for it is precisely such circumstances that characterize many clinical rating sessions. That is, raters often are required to make their



clinical judgments in the presence or under the scrutiny of others. Furthermore, clinical raters are frequently subject to evaluation by those others, or to comparison with them. Thus, evaluation apprehension appears likely to be an operative variable in many clinical rating sessions and, therefore, it seems likely that the social psychological literature on evaluation apprehension might profitably be applied to the clinical judgment setting.

The question then arises as to what effect evaluation apprehension has on clinical ratings. Steiner described one aspect of evaluation apprehension which seems particularly relevant to the clinical setting, namely, its directing or biasing potential:

"evaluation apprehension is not a blind, unprejudiced energizer. It steers behavior as well as propelling it. It motivates people to do those acts which are likely to elicit favorable evaluations" (1972, p. 134). In other words, evaluation apprehension, especially if it is induced by the presence of others, motivates individuals to do or say those things which are likely to gain approval from others.

How does one gain the approval of others, or at least escape a negative evaluation? A number of studies have indicated that in matters of judgment, individuals seeking approval from others tend to conform to the perceived judgments of those others. A considerable literature has developed on the subject of conformity, beginning with the classic research of Allport (1924), Sherif (1947), and Asch (1956). Allport found that when judging odors and weights, individuals in a social setting made more moderate judgments than did individuals



working alone. Allport concluded from these findings that, "To think and to judge with others is to submit oneself unconsciously to their standards" (1924, p. 278).

The moderation of ratings seen in Allport's group-based subjects may have been due to the fact that they had no way of knowing what each other's ratings might be; i.e., they had no information regarding group norms. In such situations, the "safest" strategy (in the sense of avoiding deviancy) would seemingly be to respond with moderate judgments, avoiding the extremes. By contrast, the subjects in the Sherif (1947) and Asch (1956) studies were informed of the judgments of all other group members before they gave their own judgments. These subjects exhibited marked conformity to the judgments of others, often rejecting even the best evidence of their own senses.

Summing up the results of studies of social judgment, Brown (1965) concluded:

There seems to be an almost ineradicable tendency for members of a group to move toward agreement. It occurs when there is no instruction to reach a consensus. It occurs when there is no opportunity to argue. It even occurs, incipiently, when the members do not know one another's opinions but can only guess at them (p. 669).

In sum, taken together, the investigations of evaluation apprehension and conformity suggest that persons in the presence of others who are presumably acting in an evaluative capacity vis-a-vis the subjects will experience evaluation apprehension. Evaluation apprehension, in turn, may motivate the affected individuals to behave so as to gain a favorable evaluation from the others. More

specifically, in matters of judgment, evaluation apprehension may motivate these individuals to conform to the presumed opinions of others.

The Special Significance of Agreement  
in Clinical Judgments

A wide variety of variables related to the tendency for conformity have been studied. Group variables (unanimity, expertise), situational variables (anonymity), and individual variables (sex, self-confidence) have all been shown related in various ways to the tendency to conform (Crutchfield, 1955; Campbell, 1961; Freedman, Carlsmith & Sears, 1970). In general, it would be expected that any variable which increases evaluation apprehension would as a consequence increase the tendency to conform. In this regard, of special relevance for surveillance and clinical judgments are such variables as the expertise of the presumed evaluators, the immediacy of the evaluation, the consequences to oneself or others of the judgments, the degree of anonymity, and raters' experience and competence.

A variable which is perhaps even more important is the function served by uniformity for certain kinds of judgments. Cartwright and Zander note that group pressures for uniformity will be greater as the importance of uniformity per se increases (1968, p. 144). For matters of judgment where there is no direct validating evidence available, conformity or agreement among judgments becomes essential in order to establish social reality (Cartwright & Zander, 1968, p. 142).

"Consensus may be essential because the consensus is the only reality" (Brown, 1965, p. 670). In other words, what is agreed upon is accepted as valid; discrepant opinions are considered to be invalid and there is no external criterion against which "correctness" may otherwise be assessed.

As noted previously (p. 4), agreement is a major index of accuracy for many clinical judgments, particularly those requiring more inference. One rater's assessment of a patient's "ego strength," for example, may show no relation at all to the assessment of another rater; and neither of these assessments may coincide with the results of a pencil and paper test of this variable. Who determines which assessment is "correct?"

Predictions of "improvement" might be seen as having more of an external criterion of correctness; yet even here, "getting better" and "getting worse" are variously defined (Strupp & Hadley, 1977). There is thus no unequivocal standard by which it may be determined that one therapeutic outcome is better than another, that one prediction was superior to another--except insofar as a consensus for what constitutes improvement, etc. may be derived. Agreement in matters of clinical judgment thus appears to be of crucial importance and it follows that pressures for conformity are particularly strong in matters of clinical judgment.

Description and Hypotheses  
of Present Investigation

The research and theories reviewed above suggested that surveillance and presumed evaluation by "significant others" (experts or peers, for example) might lead to the arousal of evaluation apprehension in clinical observers. Further, these studies suggested that if observer subjects experienced sufficient apprehension, and if they knew the probable direction of the judgments with which theirs were to be compared, their judgments might be biased toward those of the significant others (or possibly in the opposite direction, if the significant others constituted a negative reference group). As a consequence, the clinical ratings made by observers operating under different conditions of surveillance might reveal systematic differences in the judgments of those observers (although, given the dearth of previous research on surveillance in clinical settings, it was considered unlikely that the nature of these differences could be specified accurately in advance). It also seemed likely that the rating differences would be increased by any variable which might increase raters' vulnerability to evaluation apprehension, including such variables as immediacy of the evaluation, raters' level of experience, and self-perceived competence.

Based on these general suppositions, the present investigation was designed to assess the effects of variations in surveillance on three dependent variables: (1) Raters' evaluation apprehension;

(2) The content of raters' clinical judgments; and (3) Raters' tendency to conform with others.

The study included three conditions of surveillance designated as follows:

(1) Anonymous Condition--Subjects made their ratings anonymously. They expected their ratings would become part of a large normative data pool and that there would be no scrutiny of any individual's ratings. Thus, little or no surveillance was involved.

(2) Expert Condition--Subjects made their ratings expecting that they would be examined and compared with the ratings made by a group of analytically-oriented experts.

(3) Peer Condition--Subjects made their ratings knowing they would be asked to share and compare their ratings with their peers.

Both Conditions 2 and 3 were designed to approximate surveillance conditions typical of much psychotherapy research.

As applied to the three experimental conditions, the general hypothesis of the study (i.e., surveillance increases evaluation apprehension which in turn leads to group differences in conformity and in the content of clinical ratings) led to the following specific expectations:

HYPOTHESIS 1--Effects of surveillance on evaluation apprehension.

Due to the surveillance of their ratings, raters in the Expert and Peer Conditions would experience more evaluation apprehension than would raters in the Anonymous Condition, as manifested in expressions of greater concern regarding surveillance and lower confidence in ratings for the former two groups.



HYPOTHESIS 2--The effects of surveillance on evaluation apprehension would be greater for certain subgroups of raters within the Expert and Peer Conditions as follows:

HYPOTHESIS 2A--Within the Expert and Peer Conditions, raters with less experience would feel more evaluation apprehension than raters with more experience. No differences as a function of rater experience within the Anonymous Condition were expected.

HYPOTHESIS 2B--Within the Expert and Peer Conditions, the low competence raters would feel more evaluation apprehension than high competence raters. No differences as a function of competence within the Anonymous Condition were expected.

HYPOTHESIS 2C--Because of the alleged surveillance of their ratings by analytic experts, analytically-oriented raters in the Expert Condition would experience more evaluation apprehension than raters of other theoretical orientations within that condition.

HYPOTHESIS 2D--Analytically-oriented raters in the Expert Condition would experience more evaluation apprehension than analytically-oriented raters in either of the other two conditions.

HYPOTHESIS 3--Effects of surveillance on the content of clinical ratings. The clinical judgments of raters in the Expert and Peer Conditions would be significantly different from those of raters in the Anonymous Condition. Furthermore, based on the great number of variables which have been shown to affect clinical judgments (pp. 6-9), it seemed likely that on the clinical ratings there would be a number of interaction effects between surveillance and the other



independent variables--Rater Self-Perceived Competence, Rater Clinical Experience, and Rater Orientation. More precise predictions concerning these group differences were not possible since knowledge of the effects of surveillance is so limited.

Although a complete set of precise predictions was not formulated, results of a pilot investigation (Chapter II) and some earlier investigations suggested that there might be rater orientation main effects and surveillance by orientation interaction effects on the clinical ratings as follows:

HYPOTHESIS 3A--Relative to client-centered raters, analytically-oriented raters, regardless of surveillance, would have a less favorable personal reaction to the patient, would give the patient a less favorable prognosis, and would be more likely to set their own therapy goals for the patient (Sundland & Barker, 1962; Strupp, 1958, 1960).

HYPOTHESIS 3B--Due to the alleged surveillance by analytic experts in the Expert Condition and the presumed desire of analytically-oriented raters to conform to the opinion of analytic experts, the hypothesized differences between analytic and client-centered raters described in HYPOTHESIS 3A would be most marked in the Expert Condition.

HYPOTHESIS 4--Effects of surveillance on conformity. It was also expected that self-expert discrepancy scores would be lower (i.e., conformity would be greater) for analytically-oriented Expert Condition raters than for raters of other orientations in the Expert

Condition and lower than self-expert discrepancies for analytically-oriented raters in the Anonymous or Peer Conditions. More generally, it was expected that on self-other discrepancy scores there would be other interactions of surveillance with the competence, experience, and orientation independent variables and that these discrepancy scores might show some relationship to the raters' own clinical judgments. The form these effects might take was not specified further.

## CHAPTER II

### PILOT INVESTIGATION

A pilot investigation of the general hypotheses outlined above was undertaken with 21 graduate students in clinical psychology at the University of Massachusetts serving as raters. The raters viewed a videotaped segment of one therapy session from the Vanderbilt Psychotherapy Project. They then completed an extensive questionnaire which included items from a variety of sources selected to represent patient, therapist, and therapy variables rated in many clinical settings. The raters were randomly assigned to one of four conditions which varied in the "set" given subjects before they viewed the tape as to the surveillance to which their ratings would be subjected:

(1) Condition 1--Minimal Surveillance--Raters were told that the Vanderbilt Psychotherapy Research Project is in the process of developing a standardized set of clinical tapes and related rating instruments for psychotherapy research and training and their responses, along with many others throughout the country, were being collected as a part of a normative data pool. The emphasis was thus on anonymity and non-surveillance of individual responses.

(2) Condition 2--Unknown Expert Surveillance--Raters were told that as a part of the Vanderbilt Project, their ratings as graduate students were to be compared with those made by a group of unidentified

psychotherapy experts. Emphasis was placed on the fact that there would be a comparison of ratings made by the students as therapists-in-training with the ratings of the experts.

(3) Condition 3--Identified (Analytic) Expert Surveillance--

Raters were told their ratings were to be compared with those made by a group of expert analytic clinicians from the Chicago Psychoanalytic Institute. The attempt was to provide raters with a clearly identifiable reference group whose opinions on theoretical issues are well-known. (Obviously, other reference groups might have been chosen--behavior therapists, client-centered therapists, for example. Further research using such reference groups would be of interest.)

(4) Condition 4--Peer Surveillance--Raters were told that

following their rating task, they would be asked to compare and discuss their ratings with their peers. (Interestingly, this condition was not included in the original experimental design. It was added during the process of data collection when several raters spontaneously observed that immediacy of surveillance might be a much more powerful variable than expertise.)

The data derived from the pilot work were of two forms--the responses to the questionnaire items and, equally important, informal feedback subsequent to debriefing. The informal feedback was noteworthy for the consensus which emerged along these lines:

(1) Agreement with the basic premise that evaluation apprehension may affect clinical ratings; but, (2) Reservations about the potency

of the manipulations in the pilot study for arousing measurable apprehension or measurable bias in clinical ratings.

A number of refinements in the experimental procedures were implemented as a result of the raters' suggestions, the most important of these being the introduction of the peer surveillance condition into the experimental design. Other procedural modifications included in the design of the full-scale research are described below.

The questionnaire data for the pilot raters showed clear inter-group differences, several of them statistically significant and others, although not significant, considered promising for future research with larger N's.

#### Evaluation Apprehension

Of first interest was the degree to which the experimental manipulations had been successful in inducing evaluation apprehension. The only attempt at a formal assessment of evaluation apprehension was the questionnaire item regarding raters' confidence in their ratings. Condition 4 raters, followed by Condition 3 rated themselves as less confident. However, the differences among the conditions were not significant. Most raters described themselves as "somewhat" to "highly" confident. Furthermore, there were no differences in self-confidence as a function of therapy experience, theoretical orientation, or self-evaluations of competence as a therapist.

The verbal report of raters regarding confidence following the rating session paralleled the pattern described above, with raters



in Condition 2 responding in the negative concerning any evaluation apprehension. While the raters in this group agreed that surveillance could affect clinical ratings, they reported that the surveillance aspects of the present study were less than salient to them, largely because the surveillance was removed both in time and distance and the "experts" were unidentified.

By contrast, subjects in Condition 3 and especially those in Condition 4 reported strong awareness of the surveillance and the arousal of feelings of apprehension concerning the evaluation of their ratings. The subjects' comments revealed that in Condition 3 the identifying "analytic" label gave considerable reality to the possibility of surveillance. Subjects in Condition 4 reported they had experienced mild to moderate apprehension over the anticipated sharing of their judgments with their peers. In fact, two of these subjects expressed some hostility over the fact that this had been only an experimental manipulation; however, when invited to share their ratings with their colleagues, they declined.

A number of raters expressed dissatisfaction with the single assessment of confidence, indicating they were far more confident concerning some ratings than others. This was considered a likely cause for the lack of statistically significant inter-group differences. Furthermore, it seemed likely that confidence ratings were at best only an indirect measure of evaluation apprehension. Improved assessment of confidence and evaluation apprehension was implemented in the full-scale study as described below.



### Clinical Ratings

Overall, clinical ratings of the patient were less affected by manipulations in the pilot study than were ratings of the therapist. (This is not particularly surprising since the raters as therapists-in-training might be expected to focus more on the therapist's behavior. Indeed, many raters reported after the session that they had attended primarily, if not exclusively, to the therapist, in part because opportunities for them to observe another therapist at work are rare.)

The greatest number of differences among groups on clinical ratings occurred between Condition 4 and the other three groups. Condition 4 raters rendered the most favorable judgments of the therapy and the therapist. Furthermore, they tended to give among the lowest ratings of the psychological health of the client.

For such a pattern of results to be interpretable under the conformity hypothesis, one must know what these Condition 4 raters perceived their peers' opinions to be. A systematic assessment of perceptions of peers was not a part of the pilot work, but was implemented in the full-scale study. It is a reasonable assumption, however, that graduate students in clinical psychology might have strong feelings concerning the need for and the efficacy of therapy. Thus, in terms of conformity, one would interpret the results for Condition 4 (low ratings of the patient's adjustment; high ratings of the therapist's competence; poor prognosis for the patient without

therapy) as a response to the assumed consensus among clinical students that therapy is a valid and efficacious technique for the amelioration of genuine psychological difficulties.

In contrast to Condition 4, Condition 2 raters had an unfavorable reaction to the therapist and his techniques. It appeared possible that the likelihood of surveillance of their ratings by experts induced in raters a more strict, indeed a negative judgmental perspective. Following this line of reasoning, a conformity interpretation of the results would suggest that raters implicitly adopted a strict evaluative standard which they assumed would approximate that of the "experts," possibly in order to demonstrate that nothing was being "put over" on them--that they were able to detect the technical errors in the therapy observed as well as the experts. Such an interpretation seemed less than convincing, however, in light of the failure of the experimental manipulations to induce any obvious evaluation apprehension. Thus, this condition was deleted from the full-scale study.

Raters in Condition 3 (analytic expert) described the therapist as more supportive, more empathic, more encouraging of change, and more helpful to the patient in recognizing his feelings than did raters in Condition 2. The reason for this difference may lie in part with the particular therapist viewed. This therapist was a psychoanalytic psychiatrist, who while he did not use any markedly analytic techniques in the segment viewed, maintained a relatively low-key, nondirective stance. Such a therapeutic approach would

likely be rated as passive by non-analysts, but as optimally active by analysts. If this interpretation of the results is valid, then activity level may be one of the primary distinguishing characteristics of the analytic surveillance condition.

Although Strupp (1958) found that analytically-oriented therapists are more prone to elucidate therapeutic goals for their patients than are client-centered therapists, the item relating to therapy goals in the pilot study failed to distinguish Condition 3 raters from the other groups. Thus, this item was reworded in the full-scale study.

Condition 3 raters did not differ from Condition 2 raters in their overall ratings of the therapist's adequacy nor in the extent to which they would approximate his style. Thus, these raters seemed to distinguish between evaluating the therapist's activity and expressing any tendency or desire to emulate him. By contrast, subjects in Condition 4, who also described the therapist higher on support, exploration, and the like rated the therapist as more adequate and reported they would conduct the interview in a manner more similar to the therapist viewed than did raters in Condition 2.

There were no differences in inter-rater agreement among the four conditions. This was the only attempt at a measure of conformity in the pilot study. Upon further reflection, it seemed that this was an inadequate measure of the kind of conformity hypothesized to

have occurred, since only in Condition 4 were fellow-raters the agents of surveillance. More direct and expanded measures of conformity were developed for the full-scale investigation, as described below.

Statistical tests across conditions on raters grouped by variables such as experience, self-ratings of competence, and theoretical orientation showed no significant differences in clinical judgments, although the breakdown by orientation showed some differences approaching significance. These variables were hypothesized to interact with the experimental conditions; however, the N in the pilot study was too small to permit any assessment of interactions.

## CHAPTER III

### METHOD

#### Raters

Seventy-five graduate students in clinical psychology and related fields were recruited from three Tennessee schools.<sup>2</sup> Potential raters were initially contacted by letter and then by phone to schedule appointments. The minimal criterion for raters was a working knowledge of clinical assessment and therapy techniques. Raters were paid \$5.00 each for the time required, usually less than one hour.

Twenty-five raters were assigned to each of the three treatment groups. Raters from each of the three schools were distributed proportionally among the three groups. The mean number of raters per session was 3.26.

#### Procedure

Raters in each of the three conditions--Anonymous, Expert, and Peer--were given identical instructions regarding the overall

---

<sup>2</sup>Approval was given to the project by the Department Head or Director of Clinical Training of the schools. An advanced graduate student served as a liaison at two of the distant sites. Each of these students was paid for the time and effort required.



purpose of the rating session as follows: The Vanderbilt Psychotherapy Project, an ongoing, long-term study of psychotherapy outcome and process was briefly described, including the fact that clinical ratings constitute a large and significant part of the data base. Raters were reminded of the desideratum that these ratings, insofar as possible, be free of bias, including bias which might result from familiarity with the therapists or patients. Thus, the purpose of obtaining ratings at various locations, including the raters' own schools, was to obtain ratings unbiased by this factor.

At this point, different instructions related to the experimental manipulations were given to each of the three conditions. Raters in the Anonymous Condition were told that the ratings were being obtained in order to assess an "average" clinician's responses and, therefore, that their ratings, along with those of many others, would be fed into a normative data pool.

Raters in the Expert Condition were told that their ratings were being obtained specifically for comparison with the ratings made by a group of expert analytic therapists, including the Principal Investigator of the Vanderbilt Project (identified by name), an acknowledged authority in analytic psychotherapy whose name and reputation are well-known to clinical graduate students.

Raters in the Peer Condition were told that immediately after making their ratings, they would be asked to share and compare ratings with their peers--that this procedure had been found to generate additional useful information.

Finally, raters in all three conditions were given identical instructions regarding the details of the procedure. They were told they would see a 20-minute videotaped segment of an early therapy session and the patient was briefly described. Raters were instructed to attend in general to the patient, the therapist, and the therapy process. They were urged to take notes of anything within these three domains which they felt was noteworthy and were told they would be given rating sheets following the tape. Finally, the Expert and Peer Condition raters were reminded of the comparisons to be made of their ratings with the group of analytic experts and with their peers, respectively.

At the conclusion of the videotaped therapy segment, raters were asked to complete a rating sheet which included assessments of thirteen clinical variables and the raters' degree of confidence in each (see Appendix A). The composition of the questionnaire is described more fully below.

Raters in the Anonymous Condition were told specifically not to provide their name or any other identifying information on the clinical rating sheet. By contrast, as a part of the experimental manipulations, raters in the Expert and Peer Conditions were required to provide their name, school, and year at the top of the form.

After the raters had provided their own assessments, the forms were collected and estimates of clinical ratings made by others were assessed following a procedure similar to that outlined in Levinger and Schneider (1969). Raters were first asked to complete an identical

set of clinical ratings according to their best estimate, based on their own knowledge of analytic psychotherapy, of the ratings most likely made by a group of expert analytically-oriented therapists. (Expert Condition raters were asked to complete the questionnaire as they believed the analytically-oriented Vanderbilt group had done.) These forms were collected.

Next, all raters were asked to respond to the clinical rating sheet with their best estimate of the ratings made by most of their peers. These sheets were collected. Finally, all raters completed a brief personal data sheet (Appendix B). This form included several demographic variables and three items designed to measure evaluation apprehension. As with the clinical rating sheets, Expert and Peer Condition raters were required to identify their personal data sheets by name. When all the ratings had been collected, raters were paid for their time. The Vanderbilt Psychotherapy Project and the surveillance study in particular were then described in detail.

#### Assessment Battery

The clinical assessment questionnaire consisted of ratings of thirteen variables. Patient variables were selected as representative of those frequently assessed in psychotherapy research. Ratings of several therapist variables, an important, but hitherto neglected area of methodological inquiry, were also obtained. The thirteen variables rated and the scale for each were as follows:

1. Patient Anxiety  
1 = Virtually none  
7 = A great deal
2. Patient Depression  
1 = Not at all depressed  
7 = Severely depressed
3. Patient Overall Adjustment  
1 = Very poor  
7 = Excellent
4. Prognosis Without Therapy  
1 = Got much worse  
7 = Got much better
5. Prognosis With Therapy  
1 = Got much worse  
7 = Got much better
6. Liking for Patient  
1 = Strongly negative  
7 = Strongly positive
7. Therapist Activity  
1 = Extremely passive  
7 = Extremely active
8. Therapist Supportiveness  
1 = Non-supportive  
7 = Very supportive
9. Therapist Competence  
1 = Incompetent  
7 = Highly competent
10. Therapist Orientation  
\_\_\_\_ Behavioral  
\_\_\_\_ Psychoanalytic  
\_\_\_\_ Client-centered  
\_\_\_\_ Eclectic
11. Liking for Therapist  
1 = Strongly negative  
7 = Strongly positive
12. Resemblance to Therapist  
1 = Markedly dissimilar  
7 = Very similar
13. Determination of Therapy Goals  
1 = Leave goals to patient's discretion  
7 = Determine goals based on my perceptions and evaluation of patient

Confidence ratings on a 7-point Likert scale were made after all the clinical variables had been rated. No confidence assessments were made for the four clinical variables designed to assess raters' personal values or preferences (liking for the patient, liking for the therapist, similarity to the therapist, and method of determining therapy goals).

The personal data sheet included demographic variables such as age and sex and other variables expected to interact with the

surveillance manipulations (years in graduate school, experience as a therapist, theoretical orientation, and self-perceptions of competence as a therapist<sup>3</sup>).

The last three items on the personal data sheet were designed to assess various aspects of evaluation apprehension. These items included a rating of degree of apprehension over the possibility of surveillance and queries on perceived influence of surveillance on ratings and (for Expert and Peer Condition raters only) preference for anonymity in making the ratings.

---

<sup>3</sup>In order to avoid confusion between them, the two "orientation" variables (one an independent variable pertaining to the rater and one a dependent variable assessment of the rater's perception of the therapist observed) will be referred to henceforth as "Rater Orientation" and "Therapist Orientation" respectively. The same procedure will be followed for "Rater Self-Perceived Competence" and "Therapist Competence."



## CHAPTER IV

### RESULTS

#### Overall Strategy and Preliminary Analyses

Independent variables. A correlation matrix of the ordinal independent variables was constructed. Inspection of the intercorrelations among these variables revealed that for all raters (and within each of the individual treatment groups as well [Table 1]), there was a significant positive correlation between number of years in graduate school and amount of clinical experience. Since level of therapist experience was considered most relevant to the study, this variable was retained and years in school was deleted from subsequent analyses. Thus, in addition to the primary independent variable--surveillance--there were two ordinal independent variables (Rater Self-Perceived Competence and Rater Clinical Experience) and one nominal independent variable (Rater Theoretical Orientation) which were analyzed for effects on the dependent variables.

Dependent variables. There were four different sets of dependent variables:

- (1) Measures of evaluation apprehension (3 items);
- (2) Measures of confidence in the clinical ratings (9 items);
- (3) Clinical ratings (13 items); and

TABLE 1  
CORRELATION MATRIX OF INDEPENDENT VARIABLES

|                             | Therapist Experience | Competence         |
|-----------------------------|----------------------|--------------------|
| <u>Years in School</u>      |                      |                    |
| All raters (70 d.f.)        | .63 <sup>a</sup>     | .43 <sup>a</sup>   |
| (Anonymous) (23 d.f.)       | (.66) <sup>a</sup>   | (.41) <sup>b</sup> |
| (Expert) (22 d.f.)          | (.55) <sup>a</sup>   | (.34)              |
| (Peer) (22 d.f.)            | (.62) <sup>a</sup>   | (.50) <sup>b</sup> |
| <u>Therapist Experience</u> |                      |                    |
| All raters (70 d.f.)        | . . .                | .50 <sup>b</sup>   |
| (Anonymous) (23 d.f.)       | . . .                | (.45) <sup>b</sup> |
| (Expert) (22 d.f.)          | . . .                | (.48) <sup>b</sup> |
| (Peer) (22 d.f.)            | . . .                | (.51) <sup>b</sup> |

<sup>a</sup> $p < .01$

<sup>b</sup> $p < .05$

(4) Self-other discrepancy scores (an index of conformity; 13 each self-expert and self-peer discrepancy scores).

The first three sets of scores (evaluation apprehension, confidence scores, and the clinical ratings) were measured directly on the questionnaire. Self-other discrepancy scores were calculated as follows: For each rater, thirteen Expert Discrepancy (ED) and thirteen Peer Discrepancy (PD) scores were computed by subtracting the rater's

estimates of experts' and peers' clinical ratings, respectively, from the rater's own clinical ratings. For example:

$$\frac{\text{ED Score for Patient Anxiety}}{\text{Patient Anxiety}} = \text{Rater's Judgment of Anxiety} - \text{Rater's Estimate of Experts' Judgment of Anxiety}$$

$$\frac{\text{PD Score for Patient Anxiety}}{\text{Patient Anxiety}} = \text{Rater's Judgment of Anxiety} - \text{Rater's Estimate of Peers' Judgment of Anxiety}$$

For reasons to be described below, the confidence ratings, clinical ratings, and discrepancy scores were factor analyzed and factor scores for each rater were calculated based on a weighted sum of the z-scores of all items with an item--factor correlation of .50 or above. The item composition of the varimax rotation of the confidence, clinical rating, and discrepancy factors appears in the discussion of results for each of these measures.

The primary statistical analyses consisted of two-way univariate analyses of variance for surveillance by Self-Perceived Competence, by Rater Experience, and by Rater Orientation,<sup>4</sup> with the factor scores as the dependent variables. The separate two-way data analyses, while limited to examination of first-order interactions, were used because of insufficient cell numbers for three or four-way ANOVA's.

---

<sup>4</sup>Because of the unexpectedly empty analytic orientation cell in the Anonymous Condition, two way ANOVA's for a complete factorial could not be carried out for surveillance by orientation. As an approximation to the complete factorial, two separate sets of analyses for surveillance by orientation were carried out--one for three conditions by three orientations, deleting all analytically-oriented raters and the other for two conditions by four orientations, deleting all Anonymous Condition raters. The results were generally identical.

Multivariate analysis of variance was considered, but rejected as inappropriate because of the unique nature of the major dependent variables in the study--the clinical ratings. MANOVA and similar analyses are appropriately applied to several dependent variables when there is good reason to believe those measures may tap a meaningful larger unit (Bock, 1975). This was not the case for the clinical ratings, each of which was selected and measured for its own unique clinical importance. Thus, taken as a whole, the clinical ratings did not constitute a meaningful unit.

Nonetheless, some grouping of the numerous individual scores was necessary, both as a practical matter and to reduce the likelihood of significant results occurring by chance. As noted by Kaplan and Litrownik (1977), factor analysis is analogous to multiple discriminant analysis--except that factor analysis focuses on individual variance and discriminant analysis on group variance. These theoretical considerations, and a number of practical considerations which made factor analysis the most feasible procedure, led to the factor analysis--analysis of variance strategy.

Pair-wise post-hoc comparisons using Duncan's New Multiple Range Test (1955; 1957) and Kramer's adaptation (1956) for unequal N's (a more conservative test than Duncan's 1957 unequal N test) were carried out for variables on which the ANOVA indicated there was a significant overall difference.

### Demographic and Independent Variables

Means (or frequency distributions, as appropriate) for the three surveillance groups on the demographic and independent variables appear in Table 2. There were no significant differences among the groups except with respect to Rater Clinical Experience, where the Peer Condition raters ( $\bar{M} = 2.4$ ) were found to be more experienced than Expert Condition raters ( $\bar{M} = 1.5$ ).

The rating of experience with analytic therapy showed little inter-rater discriminative power and was not used in further analyses. Both the Rater Clinical Experience and the Rater Self-Perceived Competence variables were collapsed into three categories (Low, Medium and High) as follows:

|   |        |   |               |                          |
|---|--------|---|---------------|--------------------------|
| <u>Rater Clinical Experience:</u>       | Low    | ≤ | 50 hours      |                          |
|   | Medium | = | 51--200 hours |                          |
|   | High   | ≥ | 201 hours     |                          |
| <u>Rater Self-Perceived Competence:</u> | Low    | = | 1--3          |                          |
|   | Medium | = | 4             | (1 = Not very competent) |
|   | High   | = | 5--7          | (7 = Fully competent)    |

Rater Orientation was analyzed by the four categories from which raters could select: behavioral, analytic, client-centered, and eclectic.

### Dependent Variables

Evaluation apprehension. Hypothesis 1 predicted that raters in the Expert and Peer Conditions would experience more evaluation apprehension than would raters in the Anonymous Condition. Further,



TABLE 2  
DEMOGRAPHIC DATA

| Variable  | Surveillance Condition       |                           |                         |
|---|------------------------------|---------------------------|-------------------------|
|   | Anonymous<br>( <u>N</u> =25) | Expert<br>( <u>N</u> =25) | Peer<br>( <u>N</u> =25) |
| <u>Sex</u>  |                              |                           |                         |
| Female  | 9                            | 13                        | 9                       |
| Male  | 16                           | 12                        | 16                      |
| <u>Age</u> ( <u>M</u> )                                   |                              |                           |                         |
|   | 27                           | 25                        | 26                      |
| <u>Years in Graduate School</u> ( <u>M</u> )              |                              |                           |                         |
| (Range: 0.5--5.0)   | 2.0                          | 1.6                       | 2.2                     |
| <u>Experience as a Therapist</u> ( <u>M</u> )             |                              |                           |                         |
| (Range: 1 = Under 50 hours<br>6 = Over 1,500 hours)       | 2.1                          | 1.5                       | 2.4 <sup>a</sup>        |
| <u>Experience with Analytic Therapy</u> ( <u>M</u> )      |                              |                           |                         |
| (Range: 1 = None<br>4 = Great Deal)                       | 2.3                          | 2.0                       | 2.1                     |
| <u>Orientation</u>  |                              |                           |                         |
| Behavioral  | 3                            | 5                         | 3                       |
| Analytic  | 0                            | 3                         | 3                       |
| Client-Centered   | 3                            | 7                         | 3                       |
| Eclectic  | 19                           | 10                        | 16                      |
| <u>Self-rated Competence as a Therapist</u> ( <u>M</u> )  |                              |                           |                         |
| (Range: 1 = Not Very<br>Competent<br>7 = Fully Competent) | 3.7                          | 3.6                       | 4.3                     |

<sup>a</sup>Peer--Expert Difference Significant ( $p < .05$ )

Hypothesis 2 specified that certain sub-groups of raters within the Expert and Peer Conditions would experience more evaluation apprehension than others. There were four measures of evaluation apprehension:

- (1) Degree of apprehension experienced (one item);
- (2) Perceived influence of surveillance (one item);
- (3) Preference for anonymity in making ratings (one item);
- (4) Degree of confidence in clinical ratings (two factor scores).

Means and frequencies for the first three measures appear in Tables 3--5. (Cell N's for Table 3 and for all other ANOVA tables appear in Table 36.) Peer Condition raters ( $\bar{M} = 3.03$ ) were more apprehensive than Anonymous Condition raters ( $\bar{M} = 1.94$ ). Evaluation apprehension for Expert Condition raters fell mid-way between the two extremes and was not significantly different from either. There were no other significant group differences in apprehension nor were there any group differences on perceived influence of surveillance or preference for anonymity (Tables 4 and 5). The one exception to this was that raters at the medium level of competence more frequently expressed a preference for anonymity.

Confidence in clinical ratings. Factor analysis of the degree of confidence scores revealed two significant factors, labelled "Confidence in Ratings of Prognosis" and "Confidence in Ratings of Therapist Activity." The items which make up each factor appear in

TABLE 3  
EVALUATION APPREHENSION<sup>a</sup>

| Surveillance <sup>b</sup> |        |      | Competence  |        |      |           |
|---------------------------|--------|------|-------------|--------|------|-----------|
| Anonymous                 | Expert | Peer | Low         | Medium | High |           |
| 1.94                      | 2.70   | 3.03 | 2.57        | 2.74   | 2.36 |           |
| Experience                |        |      | Orientation |        |      |           |
| Low                       | Medium | High |             | Beh.   | Anl. | Cli. Ecl. |
| 2.62                      | 2.41   | 2.32 | (3 x 3)     | 1.93   | .... | 2.81 2.46 |
|                           |        |      | (2 x 4)     | 2.07   | 3.17 | 3.38 2.61 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 2.45      | 1.88 | 1.50 | 2.09   | 3.14 | 2.86 | 3.17 | 3.20 | 2.71 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 2.11      | 2.11 | 1.86 | 2.75   | 2.50 | 2.00 | 3.00 | 2.62 | 3.10 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 1.67      | .... | 1.67 | 2.16 | 1.80   | 2.67 | 3.43 | 2.40 | 2.33 | 3.67 | 3.33 | 2.81 |

<sup>a</sup>Means not underscored by same line significantly differ  $p < .01$ .

<sup>b</sup> $F = 3.28$ ;  $p = .043$ ; 2/66 d.f.

TABLE 4  
SELF-PERCEIVED INFLUENCE OF SURVEILLANCE

| Ratings Influenced? | Surveillance Condition    |          |                 |          |       |
|---------------------|---------------------------|----------|-----------------|----------|-------|
|                     | Anonymous                 | Expert   | Peer            | Total    |       |
| Yes                 | 3                         | 2        | 5               | 10       |       |
| No                  | 22                        | 23       | 20              | 65       |       |
| Total               | 25                        | 25       | 25              | 75       |       |
|                     | Self-Perceived Competence |          |                 |          |       |
|                     | Low                       | Medium   | High            | Total    |       |
| Yes                 | 4                         | 3        | 3               | 10       |       |
| No                  | 23                        | 17       | 25              | 65       |       |
| Total               | 27                        | 20       | 28              | 75       |       |
|                     | Rater Experience          |          |                 |          |       |
|                     | Low                       | Medium   | High            | Total    |       |
| Yes                 | 5                         | 2        | 3               | 10       |       |
| No                  | 27                        | 21       | 17              | 65       |       |
| Total               | 32                        | 23       | 20              | 75       |       |
|                     | Rater Orientation         |          |                 |          |       |
|                     | Behavioral                | Analytic | Client-Centered | Eclectic | Total |
| Yes                 | 0                         | 0        | 5               | 5        | 10    |
| No                  | 11                        | 6        | 8               | 40       | 65    |
| Total               | 11                        | 6        | 13              | 45       | 75    |

TABLE 5  
PREFERENCE FOR ANONYMITY

| Prefer Anonymity?                      | Surveillance Condition |          |                 |          |       |
|--|------------------------|----------|-----------------|----------|-------|
|  | Anonymous              | Expert   | Peer            | Total    |       |
| Yes                                    | ...                    | 4        | 1               | 5        |       |
| No                                     | ...                    | 5        | 5               | 10       |       |
| Don't Care                             | ...                    | 16       | 19              | 35       |       |
| Total                                  | ...                    | 25       | 25              | 50       |       |
| Self-Perceived Competence <sup>a</sup> |                        |          |                 |          |       |
|  | Low                    | Medium   | High            | Total    |       |
| Yes                                    | 0                      | 4        | 1               | 5        |       |
| No                                     | 5                      | 2        | 3               | 10       |       |
| Don't Care                             | 16                     | 6        | 13              | 35       |       |
| Total                                  | 21                     | 12       | 17              | 50       |       |
| Rater Experience                       |                        |          |                 |          |       |
|  | Low                    | Medium   | High            | Total    |       |
| Yes                                    | 2                      | 2        | 1               | 5        |       |
| No                                     | 5                      | 0        | 5               | 10       |       |
| Don't Care                             | 16                     | 12       | 7               | 35       |       |
| Total                                  | 23                     | 14       | 13              | 50       |       |
| Rater Orientation                      |                        |          |                 |          |       |
|  | Behavioral             | Analytic | Client-Centered | Eclectic | Total |
| Yes                                    | 0                      | 1        | 1               | 3        | 5     |
| No                                     | 2                      | 1        | 1               | 6        | 10    |
| Don't Care                             | 6                      | 4        | 8               | 17       | 35    |
| Total                                  | 8                      | 6        | 10              | 26       | 50    |

<sup>a</sup> $\chi^2$  Square = 10.0896;  $p < .05$ .



Table 6. (It should be remembered that these items were measures of the degree of confidence in the respective clinical judgments and not the clinical judgments themselves.)

TABLE 6  
RESULTS OF FACTOR ANALYSIS:  
DEGREE OF CONFIDENCE IN CLINICAL RATINGS

| Factor Label                                  | Confidence in Ratings of: | Weight | % Explained Variance | % Total Variance |
|---|---------------------------|--------|----------------------|------------------|
| "Confidence In Ratings Of Prognosis"          | Patient Depression        | .503   | 47                   | 25               |
|   | Patient Adjustment        | .553   |                      |                  |
|   | Prognosis Without Therapy | .791   |                      |                  |
|   | Prognosis With Therapy    | .730   |                      |                  |
|   | Therapist Competence      | .525   |                      |                  |
| "Confidence In Ratings Of Therapist Activity" | Patient Anxiety           | .588   | 32                   | 25               |
|   | Therapist Activity        | .888   |                      |                  |
|   | Therapist Supportiveness  | .830   |                      |                  |
|   |                           |        | 79% <sup>a</sup>     | 50%              |

<sup>a</sup>Second-order factors accounted for remainder.

Group means for the confidence factors appear in Tables 7 and 8. Peer Condition raters ( $\bar{M} = -1.13$ ) were significantly less confident regarding Ratings of Prognosis than were raters in either the Expert ( $\bar{M} = 0.79$ ) or Anonymous ( $\bar{M} = 0.43$ ) Conditions and the latter two conditions were not significantly different from each other. There

TABLE 7  
CONFIDENCE IN RATINGS OF PROGNOSIS<sup>a</sup>

| Surveillance <sup>b</sup> |           |       | Competence    |        |       |       |
|---------------------------|-----------|-------|---------------|--------|-------|-------|
| Expert                    | Anonymous | Peer  | Low           | Medium | High  |       |
| 0.79                      | 0.43      | -1.13 | -0.44         | -0.29  | 0.83  |       |
| Experience                |           |       | Orientation   |        |       |       |
| Low                       | Medium    | High  | Beh.          | Anl.   | Cli.  | Ecl.  |
| -0.20                     | 0         | 0.56  | (3 x 3) -0.49 | ....   | 0.33  | 0     |
|                           |           |       | (2 x 4) 0     | 0.12   | -0.30 | -0.16 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer  |       |       |
|-----------|------|------|--------|------|------|-------|-------|-------|
| Low       | Med. | High | Low    | Med. | High | Low   | Med.  | High  |
| 0         | 0.38 | 0.99 | 0      | 0.52 | 1.94 | -1.17 | -1.78 | -0.45 |

Surveillance--Experience

| Anonymous |       |      | Expert |      |      | Peer  |       |       |
|-----------|-------|------|--------|------|------|-------|-------|-------|
| Low       | Med.  | High | Low    | Med. | High | Low   | Med.  | High  |
| 0.83      | -0.27 | 0.45 | 0.14   | 1.19 | 1.69 | -1.57 | -0.84 | -0.45 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer  |       |       |       |
|-----------|------|------|------|--------|------|------|------|-------|-------|-------|-------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh.  | Anl.  | Cli.  | Ecl.  |
| -1.58     | .... | 1.59 | 0.43 | 0.30   | 2.04 | 0.27 | 0.53 | -0.20 | -1.80 | -0.87 | -0.85 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 5.74$ ;  $p = .005$ ; 2/65 d.f.

TABLE 8  
CONFIDENCE IN RATINGS OF THERAPIST ACTIVITY<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |             |
|--------------|--------|------|-------------|--------|------|-------------|
| Anonymous    | Expert | Peer | Low         | Medium | High |             |
| 0            | 0.25   | 0    | -0.58       | 0.12   | 0.72 |             |
| Experience   |        |      | Orientation |        |      |             |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. Ecl.   |
| -0.38        | 0.13   | 0.69 | (3 x 3)     | 0.35   | .... | 0 -0.25     |
|              |        |      | (2 x 4)     | 0.29   | 0.87 | -0.30 -0.18 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |       |      |
|-----------|------|------|--------|------|------|------|-------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med.  | High |
| -1.20     | 0.51 | 0.84 | -0.77  | 0.49 | 1.04 | 0.23 | -0.64 | 0.27 |

Surveillance--Experience

| Anonymous |      |       | Expert |      |      | Peer  |       |      |
|-----------|------|-------|--------|------|------|-------|-------|------|
| Low       | Med. | High  | Low    | Med. | High | Low   | Med.  | High |
| -0.31     | -.17 | -0.40 | -0.64  | 1.09 | 1.44 | -0.19 | -0.88 | 1.03 |

Surveillance--Orientation

| Anonymous |      |      |       | Expert |      |      |       | Peer |      |       |      |
|-----------|------|------|-------|--------|------|------|-------|------|------|-------|------|
| Beh.      | Anl. | Cli. | Ecl.  | Beh.   | Anl. | Cli. | Ecl.  | Beh. | Anl. | Cli.  | Ecl. |
| 0.41      | .... | 0.68 | -0.40 | 0.22   | 0.95 | 0.14 | -0.41 | 0.36 | 0.79 | -0.75 | 0    |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

were no other main effects nor were there any interaction effects for Confidence in Ratings of Prognosis. There were no group differences on Confidence in Ratings of Therapist Activity.

Based on the variables described above--one a direct query concerning evaluation apprehension and the other a cluster of confidence ratings which accounted for nearly 50 percent of the explained variance--it is concluded that anticipated peer surveillance is related to greater apprehension in raters and less confidence in the clinical judgments made. Hypothesis 1 is thus verified for Peer Condition raters, but not for raters in the Expert Condition. Hypothesis 2 was not verified.

The unique contribution of surveillance to these dependent variables is underscored by the fact that there were no interactions of surveillance with the other independent variables on evaluation apprehension or confidence. The low confidence scores for Peer Condition raters are all the more striking when viewed in light of the high level of clinical experience for this group (Table 2).

Clinical ratings. In Hypothesis 3, it was predicted that as a result of the surveillance manipulations, there would be differences in the clinical judgments made by raters in each of the three conditions. Further, it was suggested that there would be a number of interactions between surveillance and each of the other independent variables (Rater Self-Perceived Competence, Level of Clinical Experience, and Theoretical Orientation). In order to obtain the most precise and complete information regarding group differences in the clinical

ratings, analyses were carried out on the factor scores and on the scores for the individual items which made up each factor. The item composition of the clinical rating factors appears in Table 9. These factors were labeled "Global Evaluation of the Therapist," "General Reaction to the Patient," and "General Prognosis for the Patient."

TABLE 9  
RESULTS OF FACTOR ANALYSIS:  
CLINICAL RATINGS

| Factor Label                           | Items                        | Weight | % Explained<br>Variance | % Total<br>Variance |
|--|------------------------------|--------|-------------------------|---------------------|
| "Global<br>Evaluation<br>Of Therapist" | Prognosis With<br>Therapy    | .761   | 47                      | 29                  |
|  | Therapist Activity           | .501   |                         |                     |
|  | Therapist<br>Supportiveness  | .692   |                         |                     |
|  | Therapist<br>Competence      | .834   |                         |                     |
|  | Liking For Therapist         | .832   |                         |                     |
|  | Resemblance To<br>Therapist  | .675   |                         |                     |
| "General<br>Reaction To<br>Patient"    | Patient Depression           | -.646  | 21                      | 13                  |
|  | Patient Adjustment           | .682   |                         |                     |
|  | Liking For Patient           | .642   |                         |                     |
| "General<br>Prognosis<br>For Patient"  | Patient Anxiety              | .779   | 17                      | 10                  |
|  | Prognosis Without<br>Therapy | -.590  |                         |                     |
|  |                              |        | 85%                     | 52%                 |



Group means for the three clinical rating factors and for the thirteen individual clinical ratings appear in Tables 10--25. Although there were no main effects for surveillance, there were a number of surveillance interaction effects with the other three independent variables as well as a number of main effects for the latter variables. Hypothesis 3 is thus verified with respect to interaction effects, but is not verified with respect to main effects for surveillance. The significant group differences on the clinical ratings are described in the following pages.

Interactions of surveillance and rater competence. The surveillance by competence interaction produced by far the greatest number of group differences on clinical ratings, with the major interaction effect found on Global Evaluation of the Therapist (Table 10). This factor, which accounted for nearly 50 percent of the explained variance in clinical ratings, included six of the thirteen individual clinical ratings. (Higher scores indicate a more favorable evaluation of the therapist.)

As seen in Figure 1, the experimental treatments produced significant differences in evaluations of the therapist for high competence raters. Peer Condition high competence raters had a significantly less favorable attitude toward the therapist ( $\bar{M} = -1.54$ ) than did Anonymous Condition high competence raters ( $\bar{M} = 3.29$ ). Expert Condition high competence raters were not significantly different from either extreme.

TABLE 10  
GLOBAL EVALUATION OF THERAPIST<sup>a</sup>

| Surveillance |        |      | Competence    |        |       |       |
|--------------|--------|------|---------------|--------|-------|-------|
| Anonymous    | Expert | Peer | Low           | Medium | High  |       |
| 0.50         | 0.60   | 0    | -0.22         | 0.44   | 0.88  |       |
| Experience   |        |      | Orientation   |        |       |       |
| Low          | Medium | High | Beh.          | Anl.   | Cli.  | Ecl.  |
| -0.37        | 0.42   | 0.22 | (3 x 3) -0.86 | ....   | -0.64 | 0     |
|              |        |      | (2 x 4) -0.17 | 1.06   | 0.81  | -0.43 |

Surveillance--Competence<sup>b</sup>

| Peer/<br>High | Anon/<br>Med. | Expt/<br>Low | Anon/<br>Low | Peer/<br>Low | Expt/<br>High | Peer/<br>Med. | Expt/<br>Med. | Anon/<br>High |
|---------------|---------------|--------------|--------------|--------------|---------------|---------------|---------------|---------------|
| -1.54         | -1.34         | -0.84        | -0.43        | 0.61         | 0.89          | 0.90          | 1.76          | 3.29          |

Surveillance--Experience

| Anonymous |       |      | Expert |      |       | Peer  |       |       |
|-----------|-------|------|--------|------|-------|-------|-------|-------|
| Low       | Med.  | High | Low    | Med. | High  | Low   | Med.  | High  |
| -0.88     | -0.11 | 1.89 | 0      | 1.79 | -0.36 | -0.21 | -0.41 | -0.88 |

Surveillance--Orientation

| Anonymous |      |       |      | Expert |      |      |      | Peer  |       |      |       |
|-----------|------|-------|------|--------|------|------|------|-------|-------|------|-------|
| Beh.      | Anl. | Cli.  | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh.  | Anl.  | Cli. | Ecl.  |
| -2.25     | .... | -3.53 | 1.14 | 0      | 2.77 | 0    | 0.11 | -0.30 | -0.65 | 1.62 | -0.97 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 3.517$ ;  $p = .012$ ; 4/66 d.f.

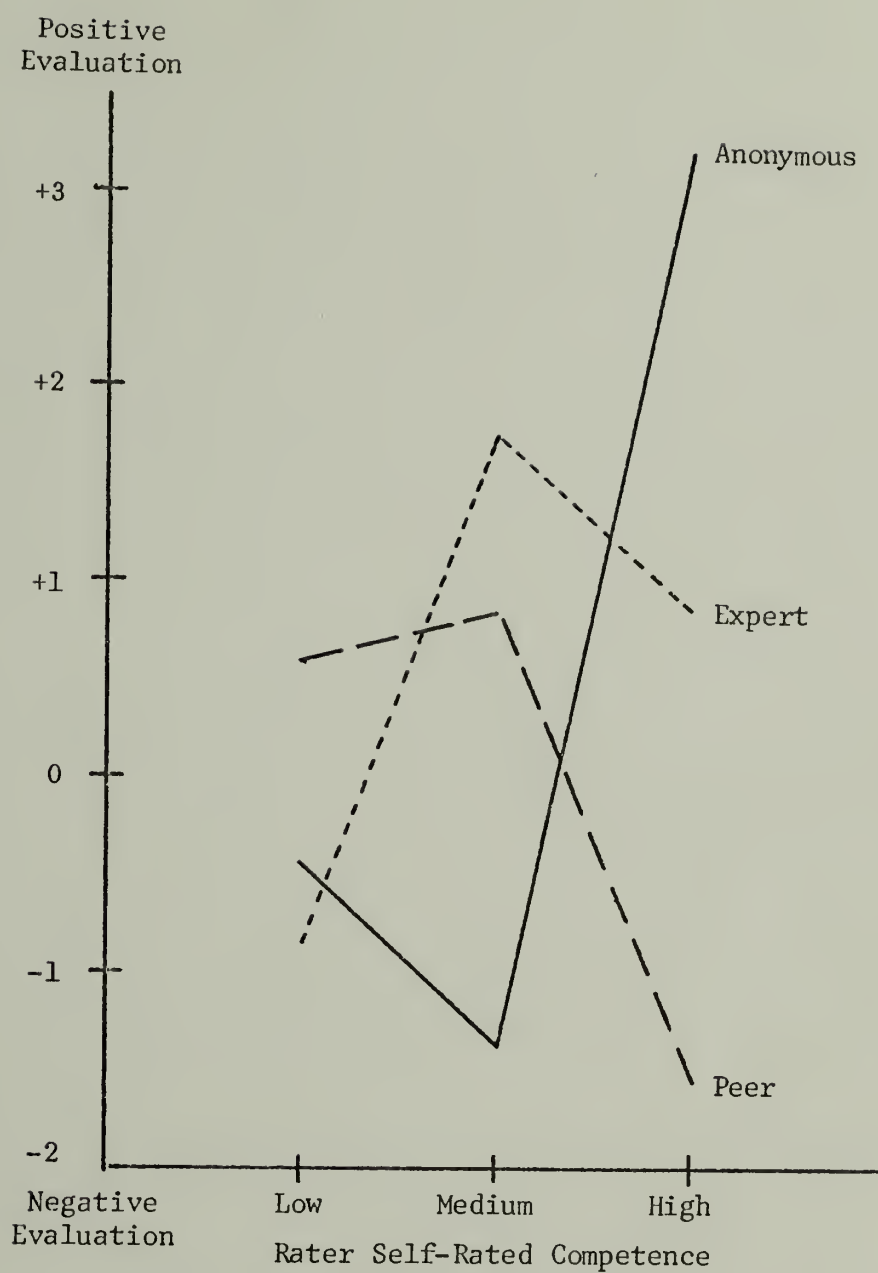


Figure 1. Global evaluation of therapist showing surveillance--competence interaction.

TABLE 11  
GENERAL REACTION TO PATIENT<sup>a</sup>

| Surveillance |        |       | Competence <sup>b</sup> |       |      |      |      |
|--------------|--------|-------|-------------------------|-------|------|------|------|
| Anonymous    | Expert | Peer  | Medium                  | Low   |      | High |      |
| 0            | 0.29   | -0.12 | -0.37                   | -0.25 |      | 0.73 |      |
| Experience   |        |       | Orientation             |       |      |      |      |
| Low          | Medium | High  |                         | Beh.  | Anl. | Ecl. | Cli. |
| -0.24        | 0      | 0.55  | (3 x 3)                 | 0.23  | .... | 0    | 0    |
|              |        |       | (2 x 4)                 | 0.46  | 0    | 0.17 | 0.39 |

Surveillance--Competence

| Anonymous |       |      | Expert |      |      | Peer |       |      |
|-----------|-------|------|--------|------|------|------|-------|------|
| Low       | Med.  | High | Low    | Med. | High | Low  | Med.  | High |
| -0.19     | -0.82 | 0.81 | -0.55  | 0.24 | 1.18 | 0    | -0.54 | 0.20 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| -0.98     | 0    | 0.71 | 0      | 0    | 0.87 | 0.19 | 0.25 | 0    |

Surveillance--Orientation

| Anonymous |      |       |       | Expert |      |      |      | Peer  |       |      |      |
|-----------|------|-------|-------|--------|------|------|------|-------|-------|------|------|
| Beh.      | Anl. | Cli.  | Ecl.  | Beh.   | Anl. | Cli. | Ecl. | Beh.  | Anl.  | Cli. | Ecl. |
| 0.23      | .... | -0.79 | -0.11 | -0.65  | 0.69 | 0.33 | 0.27 | -0.27 | -0.58 | 0.45 | 0    |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 4.764$ ;  $p = .012$ ; 2/66 d.f.

TABLE 12  
GENERAL PROGNOSIS FOR PATIENT<sup>a</sup>

| Surveillance            |        |       | Competence  |      |        |      |       |  |
|-------------------------|--------|-------|-------------|------|--------|------|-------|--|
| Anonymous               | Expert | Peer  | Low         |      | Medium |      | High  |  |
| -0.26                   | 0      | 0.12  | 0.24        |      | 0      |      | -0.42 |  |
| Experience <sup>b</sup> |        |       | Orientation |      |        |      |       |  |
| Low                     | Medium | High  | Beh.        |      | Anl.   | Cli. | Ecl.  |  |
| 0.43                    | -0.22  | -0.65 | (3 x 3)     | 0.24 | ....   | 0.22 | -0.25 |  |
|                         |        |       | (2 x 4)     | 0.58 | 0.31   | 0.33 | -0.28 |  |

| Surveillance--Competence |      |       |        |       |       |      |       |      |
|--------------------------|------|-------|--------|-------|-------|------|-------|------|
| Anonymous                |      |       | Expert |       |       | Peer |       |      |
| Low                      | Med. | High  | Low    | Med.  | High  | Low  | Med.  | High |
| -0.24                    | 0.52 | -1.04 | 0.58   | -0.34 | -0.32 | 0.39 | -0.14 | 0    |

| Surveillance--Experience |       |       |        |       |       |      |      |       |
|--------------------------|-------|-------|--------|-------|-------|------|------|-------|
| Anonymous                |       |       | Expert |       |       | Peer |      |       |
| Low                      | Med.  | High  | Low    | Med.  | High  | Low  | Med. | High  |
| 0.20                     | -0.50 | -0.30 | 0.47   | -0.22 | -1.47 | 0.62 | 0    | -0.18 |

| Surveillance--Orientation |      |      |       |        |      |      |       |      |      |      |      |
|---------------------------|------|------|-------|--------|------|------|-------|------|------|------|------|
| Anonymous                 |      |      |       | Expert |      |      |       | Peer |      |      |      |
| Beh.                      | Anl. | Cli. | Ecl.  | Beh.   | Anl. | Cli. | Ecl.  | Beh. | Anl. | Cli. | Ecl. |
| -0.44                     | .... | 0    | -0.18 | 0.50   | 0.43 | 0.50 | -0.55 | 0.66 | 0.19 | 0.16 | 0    |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 6.137$ ;  $p = .004$ ; 2/66 d.f.



TABLE 13  
PATIENT ANXIETY<sup>a</sup>

| Surveillance            |        |      | Competence  |        |      |      |      |
|-------------------------|--------|------|-------------|--------|------|------|------|
| Anonymous               | Expert | Peer | Low         | Medium |      | High |      |
| 4.17                    | 4.28   | 4.71 | 4.69        | 4.36   |      | 4.11 |      |
| Experience <sup>b</sup> |        |      | Orientation |        |      |      |      |
| Low                     | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 5.02                    | 4.31   | 3.74 | (3 x 3)     | 4.47   | .... | 4.75 | 4.32 |
|                         |        |      | (2 x 4)     | 4.53   | 4.83 | 4.95 | 4.38 |

Surveillance--Competence<sup>c</sup>

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.18      | 5.00 | 3.33 | 4.54   | 4.28 | 4.00 | 5.33 | 3.80 | 5.00 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.78      | 4.11 | 3.71 | 4.58   | 4.33 | 3.00 | 5.71 | 4.50 | 4.50 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 4.33      | .... | 4.33 | 4.21 | 4.40   | 4.67 | 4.57 | 4.00 | 4.67 | 5.00 | 5.33 | 4.75 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 2.865$ ;  $p = .029$ ; 4/66 d.f.    <sup>c</sup> $F = 5.594$ ;  $p = .006$ ; 2/66 d.f.

TABLE 14  
PATIENT DEPRESSION<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium |      | High |      |
| 3.88         | 4.00   | 4.00 | 4.14        | 3.88   |      | 3.85 |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 4.14         | 3.94   | 3.71 | (3 x 3)     | 3.87   | .... | 4.40 | 3.89 |
|              |        |      | (2 x 4)     | 4.30   | 4.00 | 3.93 | 3.94 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.82      | 4.00 | 3.83 | 4.27   | 3.86 | 3.86 | 4.33 | 3.80 | 3.86 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.22      | 3.78 | 3.57 | 4.08   | 4.17 | 3.67 | 4.14 | 3.88 | 3.90 |

Surveillance--Orientation<sup>b</sup>

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Beh.      | Cli. | Ecl. | Beh.   | Cli. | Ecl. | Beh. | Cli. | Ecl. |
| 3.00      | 5.33 | 3.88 | 4.60   | 3.86 | 4.00 | 4.00 | 4.00 | 3.79 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 2.914$ ;  $p = .028$ ; 4/60 d.f.

TABLE 15  
PATIENT ADJUSTMENT<sup>a</sup>

| Surveillance |        |      | Competence           |        |      |      |      |
|--------------|--------|------|----------------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low                  | Medium | High |      |      |
| 3.65         | 3.84   | 3.51 | 3.58                 | 3.44   | 3.98 |      |      |
| Experience   |        |      | Orientation          |        |      |      |      |
| Low          | Medium | High |                      | Beh.   | Anl. | Cli. | Ecl. |
| 3.50         | 3.51   | 4.02 | (3 x 3) <sup>b</sup> | 3.24   | .... | 4.19 | 3.65 |
|              |        |      | (2 x 4)              | 3.20   | 3.67 | 4.12 | 4.72 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.54      | 3.25 | 4.17 | 3.36   | 3.86 | 4.28 | 3.83 | 3.20 | 3.50 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.11      | 3.33 | 4.57 | 3.69   | 3.83 | 4.00 | 3.71 | 3.38 | 3.50 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 3.33      | .... | 4.33 | 3.53 | 4.50   | 4.00 | 3.57 | 4.00 | 3.00 | 3.33 | 4.67 | 3.44 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 4.423$ ;  $p = .016$ ; 2/60 d.f.

TABLE 16  
PROGNOSIS WITHOUT THERAPY<sup>a</sup>

| Surveillance |        |      | Competence   |        |      |      |
|--------------|--------|------|--------------|--------|------|------|
| Anonymous    | Expert | Peer | Low          | Medium | High |      |
| 3.91         | 3.67   | 3.85 | 3.64         | 3.68   | 4.11 |      |
| Experience   |        |      | Orientation  |        |      |      |
| Low          | Medium | High | Beh.         | Anl.   | Cli. | Ecl. |
| 3.66         | 3.99   | 4.11 | (3 x 3) 3.66 | ....   | 3.71 | 4.03 |
|              |        |      | (2 x 4) 3.00 | 3.67   | 3.74 | 4.13 |

| Surveillance--Competence <sup>b</sup> |               |               |               |              |              |               |               |               |
|---------------------------------------|---------------|---------------|---------------|--------------|--------------|---------------|---------------|---------------|
| Expt/<br>Low                          | Peer/<br>Med. | Anon/<br>Med. | Expt/<br>High | Anon/<br>Low | Peer/<br>Low | Expt/<br>Med. | Peer/<br>High | Anon/<br>High |
| 3.00                                  | 3.40          | 3.50          | 3.86          | 3.91         | 4.00         | 4.14          | 4.14          | 4.33          |

| Surveillance--Experience <sup>c</sup> |      |      |        |      |      |      |      |      |
|---------------------------------------|------|------|--------|------|------|------|------|------|
| Anonymous                             |      |      | Expert |      |      | Peer |      |      |
| Low                                   | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.78                                  | 4.22 | 3.57 | 3.19   | 4.00 | 4.67 | 4.00 | 3.75 | 4.10 |

| Surveillance--Orientation <sup>d</sup> |      |      |        |      |      |      |      |      |
|--|------|------|--------|------|------|------|------|------|
| Anonymous                              |      |      | Expert |      |      | Peer |      |      |
| Beh.                                   | Cli. | Ecl. | Beh.   | Cli. | Ecl. | Beh. | Cli. | Ecl. |
| 4.33                                   | 3.67 | 3.84 | 3.00   | 3.14 | 4.20 | 3.00 | 4.33 | 4.06 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 2.634$ ;  $p = .041$ ; 4/66 d.f.    <sup>c</sup> $F = 2.590$ ;  $p = .044$ ; 4/66 d.f.

<sup>d</sup> $F = 2.547$ ;  $p = .048$ ; 4/60 d.f.

TABLE 17  
PROGNOSIS WITH THERAPY<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium | High |      |      |
| 5.07         | 5.37   | 5.24 | 4.96        | 5.29   | 5.41 |      |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 5.16         | 5.08   | 5.38 | (3 x 3)     | 5.09   | .... | 4.84 | 5.25 |
|              |        |      | (2 x 4)     | 5.30   | 5.33 | 5.10 | 5.29 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.91      | 4.62 | 5.67 | 4.82   | 5.86 | 5.43 | 5.17 | 5.40 | 5.14 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 5.00      | 4.78 | 5.28 | 5.19   | 5.33 | 5.67 | 5.28 | 5.12 | 5.20 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 4.67      | .... | 4.33 | 5.16 | 5.60   | 5.33 | 4.86 | 5.40 | 5.00 | 5.33 | 5.33 | 5.19 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .



TABLE 18  
 LIKING FOR PATIENT<sup>a</sup>

| Surveillance |        |      | Competence <sup>b</sup> |        |      |      |      |
|--------------|--------|------|-------------------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low                     | Medium | High |      |      |
| 4.32         | 4.85   | 4.56 | 4.37                    | 4.06   | 5.29 |      |      |
| Experience   |        |      | Orientation             |        |      |      |      |
| Low          | Medium | High |                         | Beh.   | Anl. | Cli. | Ecl. |
| 4.51         | 4.55   | 4.76 | (3 x 3)                 | 4.58   | .... | 4.27 | 4.60 |
|              |        |      | (2 x 4)                 | 4.70   | 4.67 | 4.57 | 4.79 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.18      | 3.62 | 5.17 | 4.27   | 4.57 | 5.71 | 4.67 | 4.00 | 5.00 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.78      | 4.78 | 4.14 | 4.75   | 4.50 | 5.33 | 5.00 | 4.38 | 4.80 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 4.33      | .... | 3.67 | 4.32 | 4.40   | 5.00 | 5.14 | 4.60 | 5.00 | 4.33 | 4.00 | 4.88 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 7.692$ ;  $p = .001$ ; 2/66 d.f.

TABLE 19  
THERAPIST ACTIVITY<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium |      | High |      |
| 4.37         | 4.51   | 4.42 | 4.46        | 4.09   |      | 4.75 |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 4.08         | 4.46   | 4.58 | (3 x 3)     | 3.87   | .... | 4.03 | 4.34 |
|              |        |      | (2 x 4)     | 3.80   | 5.33 | 4.55 | 4.22 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.45      | 3.50 | 5.17 | 4.09   | 4.57 | 4.86 | 4.83 | 4.20 | 4.21 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.00      | 4.22 | 4.86 | 4.25   | 4.17 | 4.00 | 4.00 | 4.00 | 4.90 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 4.00      | .... | 3.00 | 4.58 | 4.60   | 5.67 | 4.43 | 4.00 | 3.00 | 5.00 | 4.67 | 4.44 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

TABLE 20  
THERAPIST SUPPORTIVENESS<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer |             |        |      |      |      |
|              |        |      | Low         | Medium |      | High |      |
| 3.97         | 3.80   | 3.43 | 3.52        | 3.81   |      | 3.86 |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 3.46         | 3.87   | 3.58 | (3 x 3)     | 3.09   | .... | 3.67 | 3.57 |
|              |        |      | (2 x 4)     | 3.30   | 4.33 | 4.00 | 3.28 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.36      | 3.88 | 4.67 | 3.54   | 4.14 | 3.71 | 3.67 | 3.40 | 3.21 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.33      | 4.11 | 4.14 | 3.62   | 4.50 | 3.00 | 3.43 | 3.00 | 3.60 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 2.67      | .... | 3.00 | 4.16 | 3.60   | 5.00 | 4.00 | 3.30 | 3.00 | 3.67 | 4.00 | 3.25 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

TABLE 21  
THERAPIST COMPETENCE<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium |      | High |      |
| 4.98         | 5.08   | 4.89 | 4.79        | 5.02   |      | 5.13 |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 4.71         | 5.04   | 4.90 | (3 x 3)     | 4.73   | .... | 4.78 | 4.85 |
|              |        |      | (2 x 4)     | 5.27   | 5.00 | 5.33 | 4.65 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.73      | 4.38 | 5.83 | 4.82   | 5.28 | 5.14 | 4.83 | 5.40 | 4.43 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.56      | 4.78 | 5.43 | 5.00   | 5.33 | 4.67 | 4.57 | 5.00 | 4.60 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 3.67      | .... | 3.67 | 5.26 | 5.20   | 5.67 | 5.00 | 4.80 | 5.33 | 4.33 | 5.67 | 4.50 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

TABLE 22  
ORIENTATION OF THERAPIST OBSERVED

| Therapist's Orientation | Surveillance Condition    |        |      |       |
|-------------------------|---------------------------|--------|------|-------|
|                         | Anonymous                 | Expert | Peer | Total |
| Behavioral              | 0                         | 0      | 0    | 0     |
| Analytic                | 8                         | 9      | 6    | 23    |
| Cli.-Ctrd.              | 5                         | 8      | 12   | 25    |
| Eclectic                | 12                        | 8      | 7    | 27    |
| Total                   | 25                        | 25     | 25   | 75    |
|                         | Self-Perceived Competence |        |      |       |
|                         | Low                       | Medium | High | Total |
| Behavioral              | 0                         | 0      | 0    | 0     |
| Analytic                | 8                         | 7      | 10   | 25    |
| Cli.-Ctrd.              | 7                         | 9      | 10   | 26    |
| Eclectic                | 12                        | 4      | 8    | 24    |
| Total                   | 27                        | 20     | 28   | 75    |
|                         | Rater Experience          |        |      |       |
|                         | Low                       | Medium | High | Total |
| Behavioral              | 0                         | 0      | 0    | 0     |
| Analytic                | 10                        | 7      | 6    | 23    |
| Cli.-Ctrd.              | 14                        | 9      | 6    | 29    |
| Eclectic                | 8                         | 7      | 8    | 23    |
| Total                   | 32                        | 23     | 20   | 75    |



TABLE 22--Continued

| Therapist's<br>Orientation | Rater Orientation |          |                     |          |       |
|----------------------------|-------------------|----------|---------------------|----------|-------|
|                            | Behavioral        | Analytic | Client-<br>Centered | Eclectic | Total |
| Behavioral                 | 0                 | 0        | 0                   | 0        | 0     |
| Analytic                   | 3                 | 3        | 5                   | 13       | 24    |
| Cli.-Ctrd.                 | 5                 | 2        | 6                   | 13       | 26    |
| Eclectic                   | 3                 | 1        | 2                   | 19       | 25    |
| Total                      | 11                | 6        | 13                  | 45       | 75    |

TABLE 23  
 LIKING FOR THERAPIST<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium |      | High |      |
| 3.67         | 3.61   | 3.48 | 3.40        | 3.76   |      | 3.63 |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 3.29         | 3.78   | 3.36 | (3 x 3)     | 3.33   | .... | 2.97 | 3.56 |
|              |        |      | (2 x 4)     | 3.67   | 3.50 | 3.28 | 3.39 |

Surveillance--Competence<sup>b</sup>

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.27      | 3.12 | 4.67 | 3.27   | 4.14 | 3.43 | 3.67 | 4.00 | 2.78 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.00      | 3.56 | 4.28 | 3.44   | 4.17 | 3.00 | 3.43 | 3.62 | 2.80 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 2.67      | .... | 2.33 | 3.89 | 3.00   | 4.33 | 3.57 | 3.60 | 4.33 | 2.67 | 3.00 | 3.19 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 2.820$ ;  $p = .031$ ; 4/66 d.f.

TABLE 24  
RESEMBLANCE TO THERAPIST<sup>a</sup>

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer |             |        |      |      |      |
| 4.28         | 3.92   | 3.73 | Low         | Medium | High |      |      |
|              |        |      | 3.89        | 3.95   | 4.09 |      |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             |        |      |      |      |
| 3.88         | 3.85   | 3.82 |             | Beh.   | Anl. | Cli. | Ecl. |
|              |        |      | (3 x 3)     | 3.38   | .... | 4.08 | 3.81 |
|              |        |      | (2 x 4)     | 3.07   | 4.00 | 4.78 | 3.51 |

Surveillance---Competence<sup>b</sup>

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 4.00      | 3.50 | 5.33 | 3.18   | 4.14 | 4.43 | 4.50 | 4.20 | 2.50 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.89      | 3.89 | 4.86 | 3.62   | 4.17 | 4.00 | 4.14 | 3.50 | 2.60 |

Surveillance--Orientation<sup>c</sup>

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 4.00      | .... | 2.67 | 4.42 | 2.80   | 4.67 | 3.57 | 4.20 | 3.33 | 3.33 | 6.00 | 2.81 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 4.254$ ;  $p = .004$ ; 4/66 d.f. <sup>c</sup> $F = 4.116$ ;  $p = .005$ ; 4/60 d.f.

TABLE 25  
DETERMINATION OF THERAPY GOALS<sup>a</sup>

| Surveillance |        |      | Competence |        |      |
|--------------|--------|------|------------|--------|------|
| Anonymous    | Expert | Peer | Low        | Medium | High |
| 3.89         | 3.79   | 4.13 | 3.85       | 4.37   | 3.59 |

| Experience |        |      | Orientation <sup>b</sup> |      |      |      |
|------------|--------|------|--------------------------|------|------|------|
| Low        | Medium | High | Beh.                     | Anl. | Cli. | Ecl. |
| 3.71       | 4.13   | 3.95 | (3 x 3) 4.02             | .... | 3.16 | 3.89 |
|            |        |      | (2 x 4) 4.20             | 5.00 | 2.74 | 3.91 |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.45      | 4.38 | 3.83 | 4.09   | 4.14 | 3.14 | 4.00 | 4.60 | 3.78 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 3.67      | 3.78 | 4.14 | 3.75   | 4.00 | 4.00 | 3.71 | 4.62 | 3.70 |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 3.67      | .... | 4.00 | 3.84 | 4.40   | 5.00 | 3.14 | 3.70 | 4.00 | 5.00 | 2.33 | 4.12 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 3.795$ ;  $p = .017$ ;  $3/42$  d.f.

An identical interaction pattern was found on two individual clinical ratings, liking for the therapist and degree of resemblance to the therapist, both of which were weighted in the Global Evaluation of the Therapist. The overall ANOVA for the interaction was significant on the liking score, but there were no significant pair-wise differences (Table 23; Figure 2). For resemblance, however, Peer Condition high competence raters reported significantly less resemblance than did Anonymous high competence raters ( $\underline{M}$ 's = 2.50 and 5.33, respectively) (Table 24; Figure 3).

The surveillance--competence interaction ANOVA was also significant for ratings of patient anxiety (Table 13) and prognosis without therapy (Table 16). For anxiety, no specific pair-wise differences were found; however, the pattern of effects resembled that seen for Global Evaluation of the Therapist, i.e., surveillance manipulations produced effects at high levels of competence, with Peer Condition high competence raters ( $\underline{M}$  = 5.00) evaluating the patient highest on anxiety and Anonymous Condition high competence raters ( $\underline{M}$  = 3.33) rating the patient as least anxious (Figure 4).

For prognosis without therapy, the surveillance--competence interaction was more complex (Figure 5). As with anxiety, the most favorable evaluation of the patient (i.e., best prognosis without therapy) was given by high competence raters under Anonymous conditions. ( $\underline{M}$  = 4.33). The least favorable prognosis was given by Expert Condition low competence raters ( $\underline{M}$  = 3.00).



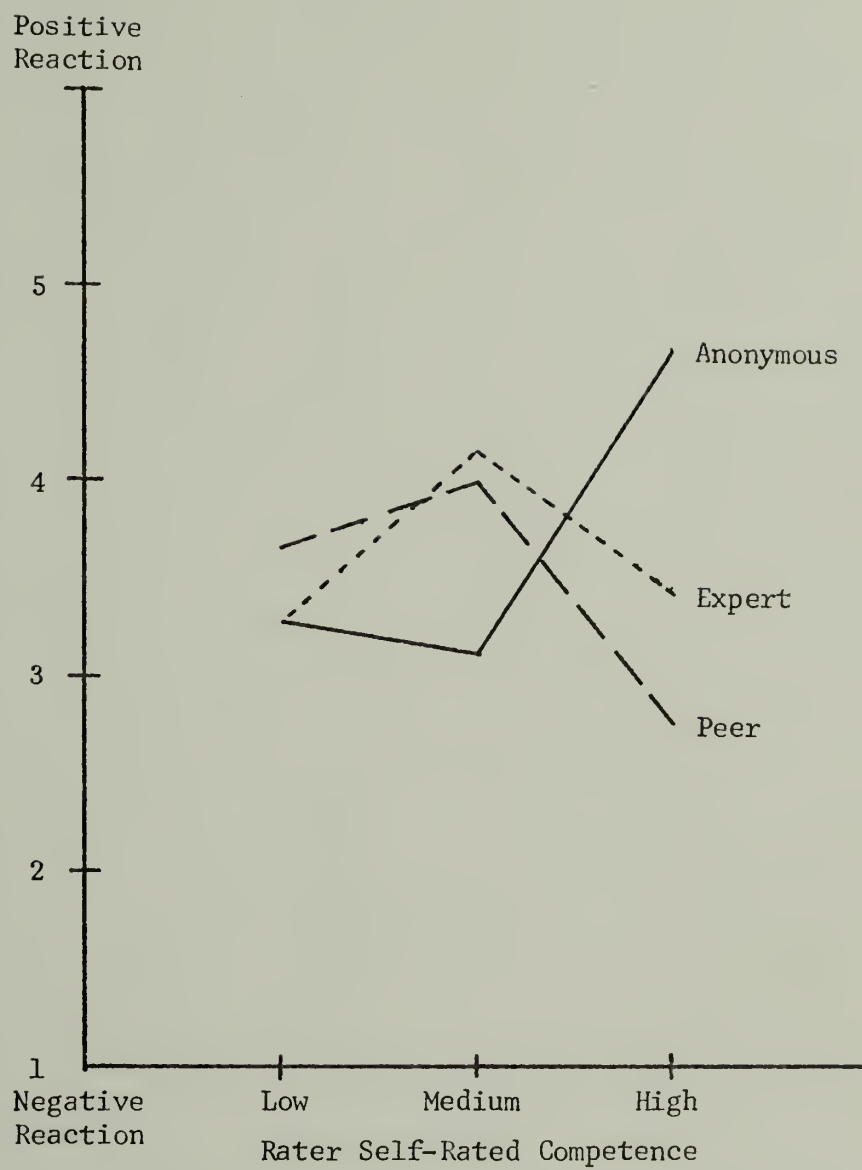


Figure 2. Personal reaction to therapist showing surveillance--competence interaction.

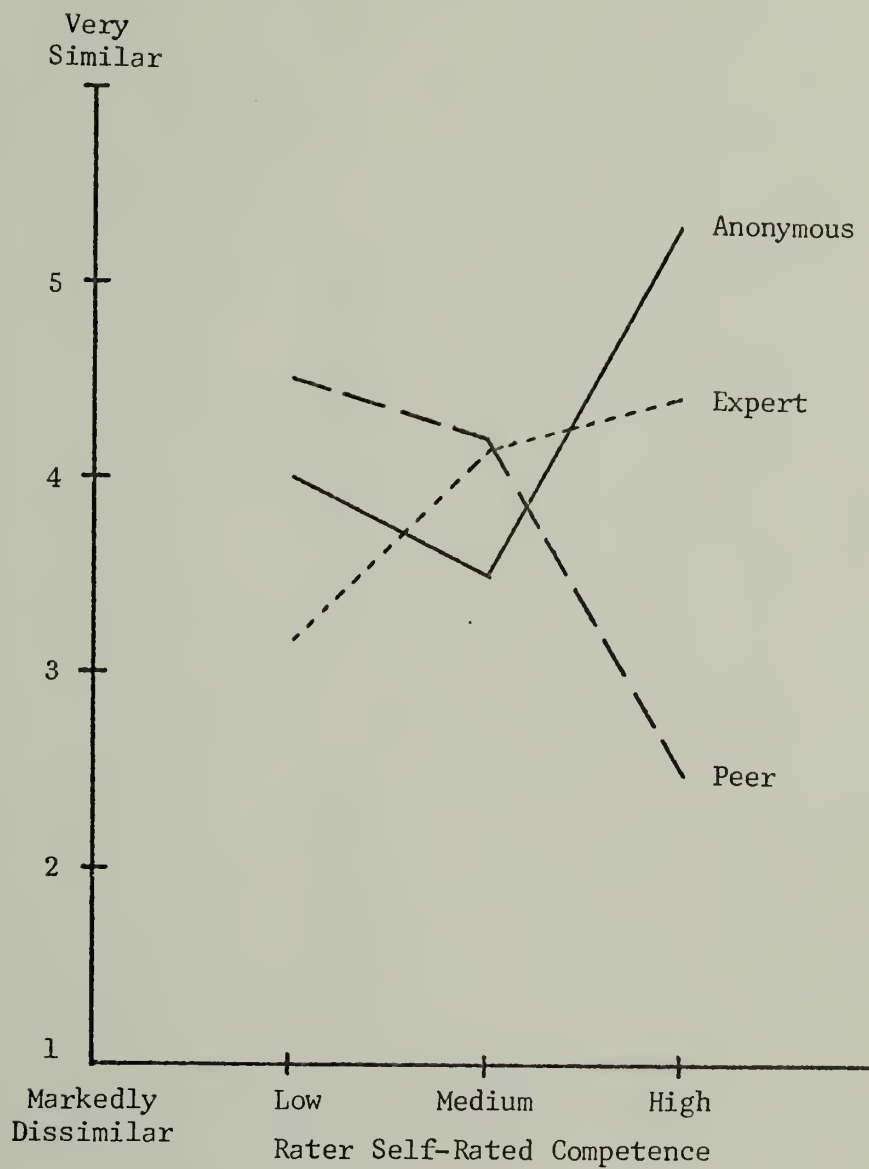


Figure 3. Resemblance to therapist showing surveillance--competence interaction.

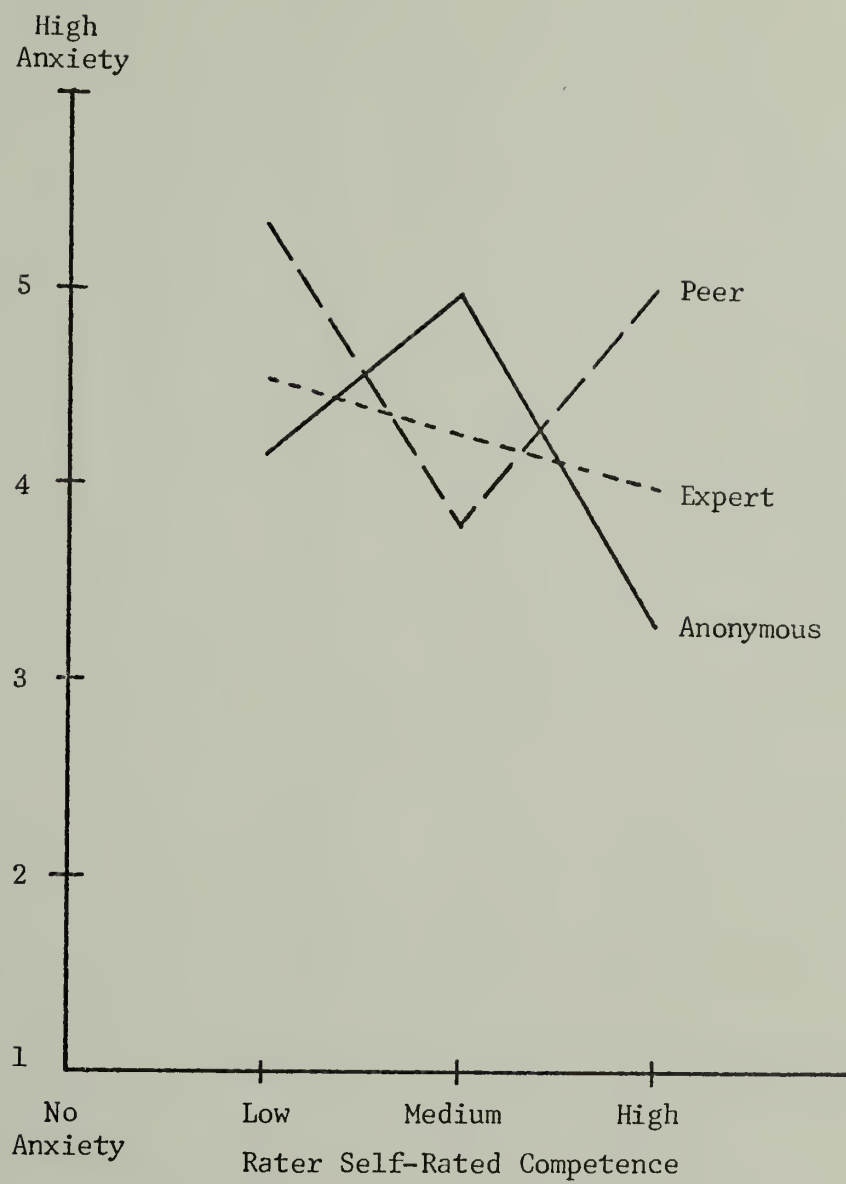


Figure 4. Ratings of patient anxiety showing surveillance--competence interaction.

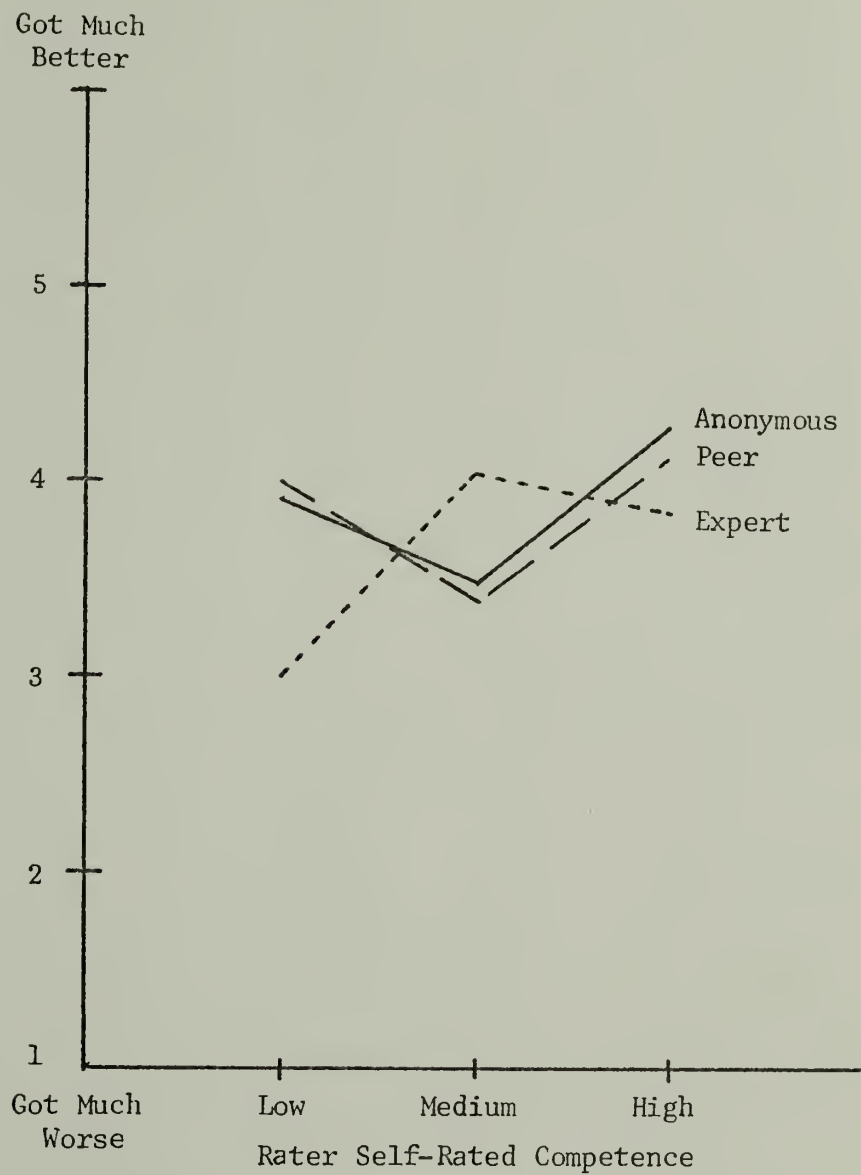


Figure 5. Ratings of prognosis without therapy showing surveillance--competence interaction.

In summary, the major finding concerning the impact of surveillance and competence on clinical ratings is as follows: Under conditions of minimal surveillance, high competence raters give favorable evaluations of both the patient and therapist, but especially the therapist. Specifically, relative to all other surveillance--competence combinations, high competence raters under minimal surveillance describe the patient as less anxious and as having a better prognosis without therapy; they evaluate the therapist and his techniques more favorably on a variety of dimensions. Under peer surveillance conditions, however, the judgments of high competence raters--particularly, judgments of the therapist--show a marked reversal, with peer surveillance high competence raters, relative to all other groups, expressing the most negative evaluation of the therapist and his techniques on a number of variables. It appears that raters who describe themselves as highly competent are uniquely susceptible to surveillance of their clinical ratings--and that the major impact of the surveillance on clinical judgments is manifested on evaluations of the therapist.

Interaction of surveillance and rater experience. Additional insight into the main effect of experience on General Prognosis for the Patient described below is provided by the significant interaction of surveillance and experience on prognosis without therapy (Table 16). No significant pair-wise comparisons were found; however, the most extreme ratings on this variable occurred within the Expert Condition,



with highly experienced raters giving the most favorable prognosis without therapy and low experienced raters giving the least favorable prognosis.

Interaction of surveillance and rater orientation. The surveillance by orientation interaction predicted in Hypothesis 3B specified that because of the analytic expert surveillance in the Expert Condition, the main effects for orientation predicted in Hypothesis 3A would be most pronounced within the Expert Condition. The analysis of variance indicated a significant surveillance--orientation interaction on ratings of patient depression, prognosis without therapy, and resemblance to the therapist. Although several of the pair-wise comparisons were significant at the .05 level, none reached significance at the .01 level. Furthermore, none of these differences indicated that analytically and client-centered raters in the Expert Condition were markedly extreme in their ratings. Therefore, Hypothesis 3B with respect to the clinical ratings of analytic and client-centered raters was not verified.

Main effects of rater competence. A significant main effect for competence was found on ratings of General Reaction to the Patient. High competence raters ( $M = 0.73$ ) had a more favorable general reaction to the patient than raters in either of the other two competence groups (medium =  $-0.37$ ; low =  $-0.25$ ). The latter two groups did not differ from each other. A main effect for competence was also found on the single item, liking for the patient (Table 23). High competence

ratars ( $\underline{M} = 5.29$ ) liked the patient significantly more than did medium ( $\underline{M} = 4.06$ ) or low ( $\underline{M} = 4.37$ ) competence ratars and the latter two groups did not differ from each other.

Main effects of rater experience. A significant main effect for level of experience was found on General Prognosis for the Patient, which subsumed high patient anxiety and poor prognosis without therapy (Table 12). The most negative assessments of the patient were given by the least experienced ratars ( $\underline{M} = 0.43$ ); the most positive assessments were given by the most experienced ratars ( $\underline{M} = -0.65$ ). The same pattern was found on the patient anxiety rating (Table 13), with low experience ratars ( $\underline{M} = 5.02$ ) rating the patient more anxious than did high experience ratars ( $\underline{M} = 3.74$ ).

Main effects of rater orientation. In Hypothesis 3A, it was predicted that relative to client-centered ratars, analytically-oriented ratars, regardless of the surveillance conditions, would have a less favorable personal reaction to the patient, would give a less favorable prognosis for the patient, and would report themselves as more likely to set their own therapy goals for the patient. No orientation main effects were found on the two patient ratings; however, with respect to therapy goals, analytically-oriented ratars described themselves as more likely to set their own goals ( $\underline{M} = 5.00$ ) and client-centered ratars were most likely to leave the determination of goals up to the patient ( $\underline{M} = 2.74$ ). Hypothesis 3A was thus verified with respect to ratars' method of setting therapy goals.

In summary, relative to the other observer variables measured in this study, rater orientation appears to have the least effect on clinical ratings, with no main effects on clinical rating factor scores and no interactions with the experimental treatments. The one main effect which was found, while of clinical interest, appears to be as much a measure of an enduring observer technical preference as it is a judgment of a given therapeutic interaction and thus, is less than central to the concerns of the present study.

Self-other discrepancies. Beyond the very general prediction of Hypothesis 4, no specific hypotheses regarding conformity in clinical ratings were formulated for this study. However, the literature on evaluation apprehension reviewed earlier indicated that evaluation apprehension, if induced by surveillance, might affect raters' clinical judgments by motivating the raters to conform with the judgments they considered the agents of surveillance would make. In the present study, the agents of surveillance were a group of analytically-oriented psychotherapy experts for the Expert Condition and the raters' peers for the Peer Condition. Conformity was defined as a reduction in discrepancy between a rater's own clinical judgments and the clinical judgments the rater estimated most analytic experts or most of his peers would make. Thus, as described previously, sets of Self-Expert Discrepancy (ED) and Self-Peer Discrepancy (PD) scores were calculated for each rater by subtracting his estimates

of the thirteen ratings made by analytic experts and the thirteen ratings made by peers on each of the clinical variables from his own ratings of those variables.

Raters' general tendency for conformity. As an index of the general tendency for conformity without regard to the kind of clinical variable being assessed, a Total Ed score and a Total PD score were calculated for each rater by summing his ED scores and his PD scores over all thirteen clinical variables. The mean Total ED score was compared with the mean Total PD score within each of the three experimental conditions (Table 26). For all three groups, mean PD scores were significantly lower than mean ED scores, indicating there was less self-peer discrepancy than self-expert discrepancy, regardless of surveillance.

TABLE 26  
SELF-OTHER TOTAL DISCREPANCY SCORES

| Score                                     | Surveillance Condition |                     |                   |
|---|------------------------|---------------------|-------------------|
|   | Anonymous <sup>a</sup> | Expert <sup>b</sup> | Peer <sup>c</sup> |
| Self-Expert Discrepancy (ED) <sup>d</sup> | 17.68                  | 17.52               | 14.48             |
| Self-Peer Discrepancy (PD)                | 10.12                  | 10.68               | 10.80             |

<sup>a</sup>Anonymous ED - PD  $\underline{t} = 4.21$ ;  $p < .005$ ; 24 d.f.

<sup>b</sup>Expert ED - PD  $\underline{t} = 4.07$ ;  $p < .005$ ; 24 d.f.

<sup>c</sup>Peer ED - PD  $\underline{t} = 2.25$ ;  $p < .01$ ; 24 d.f.

<sup>d</sup>Anonymous ED - Peer ED  $\underline{t} = 1.73$ ;  $p < .05$ ; 48 d.f.

Mean PD scores for the three experimental conditions were virtually identical ( $\bar{M}$ 's = 11.12, 10.68, 10.80, respectively); however, there was a significant difference among the experimental groups on ED Total scores. The ED Total score for the Peer Condition was significantly lower ( $\bar{M}$  = 14.48) than for the Anonymous Condition ( $\bar{M}$  = 17.68). Comparison of the mean ED scores for the Peer and Expert Conditions (Expert Condition  $\bar{M}$  = 17.52) barely missed reaching significance at the .05 level.

These results were puzzling and ran directly counter to a simple conformity interpretation. In the first place, despite the fact that the Peer Condition was distinguished by immediate peer surveillance, and despite the fact that Peer Condition raters described themselves as more apprehensive regarding surveillance than did raters in either of the other experimental conditions, there were no differences among the experimental conditions on Self-Peer Discrepancy scores. On the other hand, there was a significant difference among the experimental conditions on the Self-Expert Total Discrepancy score. However, the raters manifesting the least amount of discrepancy with experts were not the Expert Condition raters, as a conformity interpretation would require, but raters in the Peer Condition. Thus, it appears that insofar as a general tendency for conformity is concerned, there was no evidence in this study that clinical raters under different conditions of surveillance tend to conform to their respective agents of surveillance. Hypothesis 4 was, therefore, not confirmed.



Discrepancy factor scores. As described above, the analysis of self-other Total Discrepancy scores gave no indication that surveillance affects general conformity on clinical ratings. However, it seemed possible that there might be some effects of surveillance on certain sub-sets of clinical ratings. To examine this possibility, factor analyses of the ED and PD scores were carried out and group scores for the resulting factors were compared. Item composition of the ED and PD factors (with the direction of the self-other discrepancy taken into account) appears in Tables 27 and 28. The composition of the two sets of factors is remarkably similar, especially the composition of the first two factors. (It should be remembered that these factors comprise discrepancy scores for the clinical ratings listed, and not the clinical ratings themselves.)

Group means for the ED and PD factor scores appear in Tables 29 through 35. The results of the statistical analyses of these scores were essentially negative as regards any relationship of surveillance (or any of the other independent variables) to conformity. The analyses of variance indicated several significant main effects and interaction effects. However, no significant pairwise differences were found, except on the self-peer discrepancy factor labelled "Evaluation of the Patient--Therapist Interaction." Analytic and eclectic raters gave a generally more favorable evaluation of the interaction than they estimated their peers were likely to give, while behavioral and client-centered raters gave a less favorable evaluation than they expected their peers to give.

TABLE 27  
RESULTS OF FACTOR ANALYSIS:  
SELF-EXPERT DISCREPANCY

| Factor Label                          | Discrepancy on<br>Ratings of: | Weight | % Explained<br>Variance | % Total<br>Variance |
|---------------------------------------|-------------------------------|--------|-------------------------|---------------------|
| "General<br>Reaction To<br>Therapist" | Prognosis With<br>Therapy     | .825   | 12                      | 23                  |
|                                       | Therapist<br>Competence       | .827   |                         |                     |
|                                       | Liking For Therapist          | .857   |                         |                     |
|                                       | Resemblance To<br>Therapist   | .861   |                         |                     |
| "General<br>Reaction To<br>Patient"   | Patient Adjustment            | .797   | 11                      | 13                  |
|                                       | Prognosis Without<br>Therapy  | .697   |                         |                     |
|                                       | Liking For Patient            | .688   |                         |                     |
| "Therapist<br>Activity"               | Therapist Activity            | .795   | 9                       | 16                  |
|                                       | Therapist<br>Supportiveness   | .809   |                         |                     |
|                                       | Therapy Goals                 | -.506  |                         |                     |
| "Patient<br>Pathology"                | Patient Anxiety               | .786   | 7                       | 9                   |
|                                       | Patient Depression            | .762   |                         |                     |
|                                       |                               |        | 39% <sup>a</sup>        | 61%                 |

<sup>a</sup>Second order factors accounted for remainder.

TABLE 28  
RESULTS OF FACTOR ANALYSIS:  
SELF-PEER DISCREPANCY

| Factor Label                          | Discrepancy on<br>Ratings of: | Weight | % Explained<br>Variance | % Total<br>Variance |
|---------------------------------------|-------------------------------|--------|-------------------------|---------------------|
| "General<br>Reaction To<br>Therapist" | Prognosis With<br>Therapy     | .681   | 47                      | 24                  |
|                                       | Therapist Activity            | .578   |                         |                     |
|                                       | Therapist<br>Competence       | .810   |                         |                     |
|                                       | Liking For Therapist          | .611   |                         |                     |
|                                       | Resemblance To<br>Therapist   | .639   |                         |                     |
| "General<br>Reaction To<br>Patient"   | Patient Anxiety               | -.530  | 33                      | 17                  |
|                                       | Patient Adjustment            | .659   |                         |                     |
|                                       | Prognosis Without<br>Therapy  | .705   |                         |                     |
|                                       | Liking For Patient            | .671   |                         |                     |
| "Evaluation<br>Of<br>Interaction"     | Patient Depression            | .552   | 20                      | 10                  |
|                                       | Therapist<br>Supportiveness   | .702   |                         |                     |
|                                       | Liking For Therapist          | .561   |                         |                     |
|                                       | Therapy Goals                 | .719   |                         |                     |
|                                       |                               |        | 100%                    | 51%                 |

In short, the investigation into the possibility of surveillance condition differences in conformity for specific kinds of clinical judgments produced no evidence for these differences. It appears that the negative conclusion drawn above with respect to

TABLE 29  
GENERAL REACTION TO THERAPIST<sup>a</sup>  
(SELF-EXPERT DISCREPANCY)

| Surveillance |        |      | Competence  |        |      |       |      |
|--------------|--------|------|-------------|--------|------|-------|------|
| Anonymous    | Expert | Peer | Low         | Medium | High |       |      |
| 0.80         | -0.70  | 0.84 | -0.59       | 0.51   | 1.03 |       |      |
| Experience   |        |      | Orientation |        |      |       |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli.  | Ecl. |
| 0            | 0.55   | 0    | (3 x 3)     | 0.19   | .... | -1.74 | 0.42 |
|              |        |      | (2 x 4)     | 0.44   | 0.41 | -0.87 | 0    |

Surveillance--Competence<sup>b</sup>

| Anonymous |       |      | Expert |       |       | Peer |      |       |
|-----------|-------|------|--------|-------|-------|------|------|-------|
| Low       | Med.  | High | Low    | Med.  | High  | Low  | Med. | High  |
| -0.69     | -0.31 | 3.40 | -1.77  | -0.23 | -0.10 | 0.68 | 2.07 | -0.21 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |       | Peer |      |       |
|-----------|------|------|--------|------|-------|------|------|-------|
| Low       | Med. | High | Low    | Med. | High  | Low  | Med. | High  |
| -0.29     | 0.18 | 1.62 | -1.29  | 0    | -0.43 | 1.66 | 1.46 | -1.18 |

Surveillance--Orientation

| Anonymous |      |       |      | Expert |      |       |      | Peer |      |      |      |
|-----------|------|-------|------|--------|------|-------|------|------|------|------|------|
| Beh.      | Anl. | Cli.  | Ecl. | Beh.   | Anl. | Cli.  | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| -0.32     | .... | -3.49 | 1.14 | -1.59  | 0.51 | -2.20 | 0    | 2.47 | 0.30 | 0.46 | 0.12 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 2.691$ ;  $p = .038$ ; 4/66 d.f.

TABLE 30

GENERAL REACTION TO PATIENT<sup>a</sup>  
(SELF-EXPERT DISCREPANCY)

| Surveillance            |        |      | Competence <sup>b</sup> |        |       |      |      |
|-------------------------|--------|------|-------------------------|--------|-------|------|------|
| Anonymous               | Expert | Peer | Low                     | Medium | High  |      |      |
| 0                       | -0.15  | 0    | 0                       | -0.57  | 0.66  |      |      |
| Experience <sup>c</sup> |        |      | Orientation             |        |       |      |      |
| Low                     | Medium | High |                         | Beh.   | Anl.  | Cli. | Ecl. |
| -0.59                   | 0.58   | 0.31 | (3 x 3)                 | 0.14   | ....  | 0.30 | 0.12 |
|                         |        |      | (2 x 4)                 | 0      | -1.26 | 0.20 | 0.26 |

## Surveillance--Competence

| Anonymous |       |      | Expert |       |      | Peer |       |      |
|-----------|-------|------|--------|-------|------|------|-------|------|
| Low       | Med.  | High | Low    | Med.  | High | Low  | Med.  | High |
| 0         | -0.89 | 1.08 | -0.28  | -0.68 | 0.53 | 0    | -0.14 | 0.37 |

## Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer  |      |      |
|-----------|------|------|--------|------|------|-------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low   | Med. | High |
| -1.13     | 1.00 | 0.20 | -0.48  | 0.26 | 0.62 | -0.16 | 0.48 | 0.23 |

## Surveillance--Orientation

| Anonymous |      |      |       | Expert |       |       |      | Peer  |       |      |      |
|-----------|------|------|-------|--------|-------|-------|------|-------|-------|------|------|
| Beh.      | Anl. | Cli. | Ecl.  | Beh.   | Anl.  | Cli.  | Ecl. | Beh.  | Anl.  | Cli. | Ecl. |
| 0.56      | .... | 0.52 | -0.16 | 0      | -1.57 | -0.22 | 0.16 | -0.24 | -0.95 | 0.62 | 0.36 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 3.324$ ;  $p = .041$ ; 2/66 d.f. <sup>c</sup> $F = 3.130$ ;  $p = .049$ ; 2/66 d.f.



TABLE 31  
THERAPIST ACTIVITY<sup>a</sup>  
(SELF-EXPERT DISCREPANCY)

| Surveillance |        |      | Competence  |        |      |       |       |  |
|--------------|--------|------|-------------|--------|------|-------|-------|--|
| Anonymous    | Expert | Peer | Low         | Medium | High |       |       |  |
| 0.26         | -0.50  | 0    | -0.22       | -0.20  | 0.11 |       |       |  |
| Experience   |        |      | Orientation |        |      |       |       |  |
| Low          | Medium | High |             | Beh.   | Anl. | Cli.  | Ecl.  |  |
| -0.29        | 0      | 0    | (3 x 3)     | 0      | .... | -0.34 | 0     |  |
|              |        |      | (2 x 4)     | 0      | 0    | 0     | -0.35 |  |

Surveillance--Competence

| Anonymous |      |      | Expert |       |       | Peer  |       |      |
|-----------|------|------|--------|-------|-------|-------|-------|------|
| Low       | Med. | High | Low    | Med.  | High  | Low   | Med.  | High |
| -0.34     | 0.20 | 0.24 | -0.54  | -0.20 | -0.77 | -0.48 | -0.61 | 0.85 |

Surveillance--Experience

| Anonymous |      |      | Expert |       |       | Peer  |       |      |
|-----------|------|------|--------|-------|-------|-------|-------|------|
| Low       | Med. | High | Low    | Med.  | High  | Low   | Med.  | High |
| -0.17     | 0.46 | 0.60 | 0.40   | -0.25 | -1.64 | -0.29 | -0.33 | 1.06 |

Surveillance--Orientation

| Anonymous |      |       |      | Expert |      |       |       | Peer |      |      |      |
|-----------|------|-------|------|--------|------|-------|-------|------|------|------|------|
| Beh.      | Anl. | Cli.  | Ecl. | Beh.   | Anl. | Cli.  | Ecl.  | Beh. | Anl. | Cli. | Ecl. |
| 0.15      | .... | -1.18 | 0.52 | -0.67  | 0    | -0.12 | -0.87 | 0.72 | 0.13 | 0.27 | 0.16 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

TABLE 32  
 PATIENT PATHOLOGY<sup>a</sup>  
 (SELF-EXPERT DISCREPANCY)

| Surveillance <sup>b</sup> |        |      | Competence  |        |      |      |      |
|---------------------------|--------|------|-------------|--------|------|------|------|
| Anonymous                 | Expert | Peer | Low         | Medium |      | High |      |
| -0.31                     | -0.37  | 0.55 | 0           | 0      |      | 0    |      |
| Experience                |        |      | Orientation |        |      |      |      |
| Low                       | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 0.21                      | -0.11  | 0    | (3 x 3)     | 0.28   | .... | 0.19 | 0    |
|                           |        |      | (2 x 4)     | 0.18   | 0.29 | 0.13 | 0.11 |

Surveillance--Competence

| Anonymous |      |       | Expert |       |       | Peer |      |      |
|-----------|------|-------|--------|-------|-------|------|------|------|
| Low       | Med. | High  | Low    | Med.  | High  | Low  | Med. | High |
| -0.40     | 0    | -0.47 | -0.28  | -0.20 | -0.63 | 0.74 | 0    | 0.85 |

Surveillance--Experience

| Anonymous |       |       | Expert |      |       | Peer |      |      |
|-----------|-------|-------|--------|------|-------|------|------|------|
| Low       | Med.  | High  | Low    | Med. | High  | Low  | Med. | High |
| 0         | -0.41 | -0.62 | -0.46  | 0    | -0.41 | 1.04 | 0.15 | 0.81 |

Surveillance--Orientation

| Anonymous |      |      |       | Expert |       |       |       | Peer |      |      |      |
|-----------|------|------|-------|--------|-------|-------|-------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl.  | Beh.   | Anl.  | Cli.  | Ecl.  | Beh. | Anl. | Cli. | Ecl. |
| -1.22     | .... | 0.31 | -0.26 | -0.11  | -0.13 | -0.49 | -0.45 | 0.48 | 0.72 | 0.75 | 0.68 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 3.78$ ;  $p = 0.27$ ; 2/66 d.f.

TABLE 33  
GENERAL REACTION TO THERAPIST<sup>a</sup>  
(SELF-PEER DISCREPANCY)

| Surveillance |        |       | Competence    |        |       |      |
|--------------|--------|-------|---------------|--------|-------|------|
| Anonymous    | Expert | Peer  | Low           | Medium | High  |      |
| 0.36         | 0.34   | -0.23 | -0.24         | 0.59   | -0.35 |      |
| Experience   |        |       | Orientation   |        |       |      |
| Low          | Medium | High  | Beh.          | Anl.   | Cli.  | Ecl. |
| -0.42        | 0.67   | 0.26  | (3 x 3) -0.84 | ....   | -0.45 | 0.45 |
|              |        |       | (2 x 4) -1.24 | -0.35  | 0.18  | 0.22 |

Surveillance--Competence<sup>b</sup>

| Anonymous |       |      | Expert |      |       | Peer  |      |       |
|-----------|-------|------|--------|------|-------|-------|------|-------|
| Low       | Med.  | High | Low    | Med. | High  | Low   | Med. | High  |
| 1.56      | -1.11 | 0.64 | -0.38  | 1.56 | -0.17 | -0.47 | 1.33 | -1.53 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer  |       |       |
|-----------|------|------|--------|------|------|-------|-------|-------|
| Low       | Med. | High | Low    | Med. | High | Low   | Med.  | High  |
| 0         | 0.31 | 1.37 | -0.45  | 1.90 | 0.45 | -0.78 | -0.21 | -1.05 |

Surveillance--Orientation

| Anonymous |      |       |      | Expert |       |      |      | Peer  |       |      |       |
|-----------|------|-------|------|--------|-------|------|------|-------|-------|------|-------|
| Beh.      | Anl. | Cli.  | Ecl. | Beh.   | Anl.  | Cli. | Ecl. | Beh.  | Anl.  | Cli. | Ecl.  |
| 0         | .... | -1.69 | 0.91 | -1.66  | -0.27 | 0.20 | 1.33 | -0.82 | -0.42 | 0.15 | -0.90 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 3.383$ ;  $p = .014$ ; 4/66 d.f.

TABLE 34  
GENERAL REACTION TO PATIENT<sup>a</sup>  
(SELF-PEER DISCREPANCY)

| Surveillance |        |      | Competence  |        |       |      |      |
|--------------|--------|------|-------------|--------|-------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium | High  |      |      |
| -0.16        | 0.32   | 0    | 0           | -0.51  | 0.66  |      |      |
| Experience   |        |      | Orientation |        |       |      |      |
| Low          | Medium | High |             | Beh.   | Anl.  | Cli. | Ecl. |
| -0.46        | 0      | 0.78 | (3 x 3)     | 0      | ....  | 0.25 | 0.25 |
|              |        |      | (2 x 4)     | 0      | -1.35 | 0.19 | 0.59 |

Surveillance--Competence

| Anonymous |       |      | Expert |       |      | Peer |       |      |
|-----------|-------|------|--------|-------|------|------|-------|------|
| Low       | Med.  | High | Low    | Med.  | High | Low  | Med.  | High |
| -0.44     | -0.98 | 0.94 | 0      | -0.15 | 1.13 | 0.60 | -0.40 | 0    |

Surveillance--Experience

| Anonymous |      |      | Expert |       |      | Peer  |      |      |
|-----------|------|------|--------|-------|------|-------|------|------|
| Low       | Med. | High | Low    | Med.  | High | Low   | Med. | High |
| -1.19     | 0    | 0.44 | 0.14   | -0.20 | 1.91 | -0.32 | 0.32 | 0    |

Surveillance--Orientation

| Anonymous |      |      |       | Expert |       |      |      | Peer  |       |      |      |
|-----------|------|------|-------|--------|-------|------|------|-------|-------|------|------|
| Beh.      | Anl. | Cli. | Ecl.  | Beh.   | Anl.  | Cli. | Ecl. | Beh.  | Anl.  | Cli. | Ecl. |
| 0         | .... | 0.37 | -0.42 | 0.82   | -2.02 | 0    | 0.94 | -0.94 | -0.69 | 0.47 | 0.24 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

TABLE 35  
EVALUATION OF PATIENT--THERAPIST INTERACTION<sup>a</sup>  
(SELF-PEER DISCREPANCY)

| Surveillance |        |       | Competence                 |        |       |      |
|--------------|--------|-------|----------------------------|--------|-------|------|
| Anonymous    | Expert | Peer  | Low                        | Medium | High  |      |
| 0            | 0.39   | -0.23 | 0.16                       | 0.51   | -0.53 |      |
| Experience   |        |       | Orientation                |        |       |      |
| Low          | Medium | High  | Cli.                       | Beh.   | Ecl.  | Anl. |
| -0.30        | 0.39   | 0.13  | (3 x 3) <sup>b</sup> -1.16 | -0.96  | 0.40  | .... |
|              |        |       | (2 x 4) <sup>c</sup> -1.24 | -1.36  | 0.47  | 1.42 |

Surveillance--Competence

| Anonymous |      |       | Expert |      |       | Peer |      |       |
|-----------|------|-------|--------|------|-------|------|------|-------|
| Low       | Med. | High  | Low    | Med. | High  | Low  | Med. | High  |
| 0.36      | 0.16 | -0.61 | 0      | 1.42 | -0.26 | 0    | 0    | -0.73 |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer  |       |      |
|-----------|------|------|--------|------|------|-------|-------|------|
| Low       | Med. | High | Low    | Med. | High | Low   | Med.  | High |
| -0.18     | 0.32 | 0    | -0.15  | 1.62 | 0.32 | -0.58 | -0.77 | 0    |

Surveillance--Orientation

| Anonymous |      |       |      | Expert |      |       |      | Peer  |      |       |       |
|-----------|------|-------|------|--------|------|-------|------|-------|------|-------|-------|
| Beh.      | Anl. | Cli.  | Ecl. | Beh.   | Anl. | Cli.  | Ecl. | Beh.  | Anl. | Cli.  | Ecl.  |
| -0.16     | .... | -0.98 | 0.26 | -0.97  | 1.54 | -0.30 | 1.07 | -1.74 | 1.31 | -2.19 | -0.13 |

<sup>a</sup>Means not underscored by same line differ significantly  $p < .01$ .

<sup>b</sup> $F = 4.365$ ;  $p = .017$ ; 2/60 d.f. <sup>c</sup> $F = 5.040$ ;  $p = .005$ ; 3/42 d.f.

TABLE 36  
CELL N'S FOR ANALYSIS OF VARIANCE TABLES

| Surveillance |        |      | Competence  |        |      |      |      |
|--------------|--------|------|-------------|--------|------|------|------|
| Anonymous    | Expert | Peer | Low         | Medium | High |      |      |
| 25           | 25     | 25   | 27          | 20     | 28   |      |      |
| Experience   |        |      | Orientation |        |      |      |      |
| Low          | Medium | High |             | Beh.   | Anl. | Cli. | Ecl. |
| 32           | 23     | 20   | (3 x 3)     | 11     | 0    | 13   | 45   |
|              |        |      | (2 x 4)     | 8      | 6    | 10   | 26   |

Surveillance--Competence

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 11        | 8    | 6    | 11     | 7    | 7    | 6    | 5    | 14   |

Surveillance--Experience

| Anonymous |      |      | Expert |      |      | Peer |      |      |
|-----------|------|------|--------|------|------|------|------|------|
| Low       | Med. | High | Low    | Med. | High | Low  | Med. | High |
| 9         | 9    | 7    | 16     | 6    | 3    | 7    | 8    | 10   |

Surveillance--Orientation

| Anonymous |      |      |      | Expert |      |      |      | Peer |      |      |      |
|-----------|------|------|------|--------|------|------|------|------|------|------|------|
| Beh.      | Anl. | Cli. | Ecl. | Beh.   | Anl. | Cli. | Ecl. | Beh. | Anl. | Cli. | Ecl. |
| 3         | 0    | 3    | 19   | 5      | 3    | 7    | 10   | 3    | 3    | 3    | 16   |



effects of surveillance on Total Discrepancy scores is equally valid for the more specific factor scores. Thus, it appears that despite the apparent relevance of the social and clinical literature, an unqualified extrapolation of social psychological principles of conformity to the clinical judgment situation is not justified. Reasons for the lack of evidence of conformity in this study are examined in the discussion.

Summary. To recapitulate, the major findings of this study with regard to the experimental hypotheses were as follows:

(1) Hypothesis 1 specified that raters under expert or peer surveillance would experience more evaluation apprehension concerning their clinical ratings than would raters under conditions of minimal surveillance. Hypothesis 1 was verified with respect to conditions of peer surveillance, but not with respect to expert surveillance.

(2) Hypothesis 2 specified that with respect to evaluation apprehension, there would be interaction effects between surveillance and the other independent variables--Rater Self-Perceived Competence, Rater Clinical Experience, and Rater Orientation. Hypothesis 2 was not confirmed.

(3) Hypothesis 3 specified that there would be a number of group differences due to surveillance main effects and interaction effects on the clinical ratings. Hypothesis 3 was confirmed with respect to a number of group comparisons.

(4) Hypotheses 3A and 3B specified that there would be group differences due to Rater Orientation main effects and a surveillance--orientation interaction effect on certain clinical ratings. The hypothesis of orientation main effects was verified for some clinical ratings, but not for others. The hypothesis with regard to the surveillance--orientation interaction was not verified.

(5) Hypothesis 4 predicted a surveillance--orientation interaction in conformity. Hypothesis 4 was not confirmed, nor were other major differences in conformity found.

## CHAPTER V

### DISCUSSION

It appears no overstatement to conclude that this investigation is as notable for what it did not find as for what it did find. The implications of several of the results--the positive and particularly, some of the negative--are noteworthy. They point up the need for additional investigation of a number of points. Accordingly, this discussion will focus on (1) an exploration of possible causes for the results--both positive and negative--and (2) an exploration of implications and significance of the results for the social psychology literature on evaluation apprehension and conformity and the implications for psychotherapy research.

#### Major Results--and "Why?"

Overall, there were group differences found on evaluation apprehension and on the clinical ratings. However, a straightforward relationship between these two effects was not evident. Specifically, peer surveillance resulted in greater evaluation apprehension and lower rater confidence. However, peer surveillance effects on the clinical ratings were evident only for high competence raters. Further, high competence raters under peer surveillance did not describe themselves as more apprehensive than did other raters. There

was little indication of any effects on raters' tendency for conformity. With this overview in mind, possible reasons for the effects on evaluation apprehension and on the clinical ratings will be examined in greater detail.

Evaluation apprehension. As in the pilot investigation, informal comments of the raters verified that surveillance of clinical judgments may create or exacerbate evaluation apprehension. One rater in the Expert Condition, for example, evidenced considerable hesitancy to identify his questionnaire by name, asking twice if it was necessary that he do so.

Evaluation apprehension was greater in the Peer surveillance condition than for either of the other two conditions, confirming results of the pilot investigation. Apprehension for the Expert Condition raters was higher than for raters in the Anonymous Condition; however, the difference was not significant. Proximity and immediacy of surveillance, rather than the expertise of the agents of surveillance, thus appear to be the crucial factors in the arousal of evaluation apprehension.

Nonetheless, expertise should not be dismissed as having no effects. It seems quite likely that immediate expert surveillance would arouse more evaluation apprehension than immediate peer surveillance. Informal comments by the subjects bore out this supposition. One rater in the Anonymous Condition, where no mention of surveillance had been made, spontaneously asked in a half-joking manner if the Director of Clinical Training at that university would see the rating

sheets. At another university, raters in the Expert surveillance condition mentioned they would have experienced a great deal more apprehension had one particular faculty member at that university been given access to their rating sheets.

The fact that surveillance may contribute to evaluation apprehension and lower confidence in clinical raters is an interesting phenomenon in its own right, particularly from the social psychology perspective. However, considering that there were no corresponding main effects of surveillance on clinical ratings, the usefulness of the concept of evaluation apprehension as a mediating or explanatory variable for the observed effects of surveillance on clinical ratings appears questionable. This issue will be dealt with further below.

Surveillance--competence interaction on clinical ratings. Notwithstanding the apparent lack of a three-way interrelationship among surveillance, evaluation apprehension, and clinical ratings, the fact remains that surveillance did affect clinical ratings, although the effects of surveillance were manifested only in interaction with one of the other three independent variables, most notably, Rater Self-Perceived Competence.

Parenthetically, it should be noted that in contrast to surveillance, each of the other three independent variables--Rater Competence, Experience, and Orientation--was found to have significant main effects on the clinical ratings. There is thus some indication that when each independent variable is considered alone, surveillance



may not have as great an impact on clinical ratings as any of the other independent variables.

The major result of this study was the finding of a fairly extensive interaction effect involving surveillance and Rater Self-Perceived Competence. As described earlier, the effects of this interaction were manifested primarily on ratings of the therapist rather than the patient (which, incidentally, underscores the importance of assessing therapist variables in methodological research). The general form of the surveillance--competence interaction effect was that high competence raters making clinical judgments under conditions of anonymity and minimal surveillance gave a strongly favorable evaluation of the therapist; however, under peer surveillance conditions, evaluations of the therapist made by high competence raters were strongly negative. The expert surveillance manipulation did not produce significant effects on high competence raters, nor were low or medium competence raters differentially affected by varying conditions of surveillance.

Why should evaluations of a therapist made by high competence raters be uniquely susceptible to peer surveillance? Common sense and the social psychological literature on conformity would more likely have predicted that low competence raters, because of their predicted higher levels of evaluation apprehension (Campbell, 1961), would be most vulnerable. The reason for the seemingly counter-intuitive interaction effect on the clinical ratings was not to be found in evaluation apprehension or conformity, since peer surveillance



high competence raters did not report more apprehension nor did they conform more to their estimates of their peers' clinical ratings than did other groups of raters.

A possible explanation for the results is found in the fact that competence, as assessed in this investigation, was self-perceived competence; i.e., it was a highly subjective personal judgment. The subjective nature of the competence measure was underscored by the fact that competence was only moderately related ( $r = .50$ ) to the measure of Rater Clinical Experience--seemingly a more objective index of competence. (In this regard, it is useful to recall that the label of "High Competence" was applied to those who scored themselves 5-plus on the "Present Competence as a Therapist" scale where 1 = "Not Very Competent" and 7 = "Fully Competent.")

When these raters made clinical judgments of the videotaped therapist, knowing their judgments would be scrutinized by their peers, they evaluated that therapist negatively, while raters who described themselves as less competent did not respond differentially to peer surveillance. Given that competence was a subjective judgment, it then begins to make some sense that for persons who describe themselves as highly competent, clinical skill is an important part of their self-identity. Such raters would thus be more affected by scrutiny and consequently, might respond with more critical judgments of the therapist (criticism being a valued skill in much graduate training). By contrast, less competent raters might be more likely to

have a "who cares?" unthreatened attitude, seeing the rating task as reflecting little on their sense of worth.

If this line of reasoning is valid, it seems that evaluation apprehension must be reintroduced into the picture--that high competence peer surveillance raters must have been more apprehensive--and that apprehension mediated the group differences in the clinical ratings. As noted previously, the results of this study indicate this was not the case. However, remembering once again the subjective nature of the competence assessment and the subjective nature of the evaluation apprehension assessments, perhaps it is not surprising that the high competence raters, because of their self-concept of high competence, would not explicitly describe themselves as "more apprehensive" or "less confident" regarding their ratings.

The explanation thus boils down to two major alternatives--(1) that high competence Peer Condition raters really did experience more evaluation apprehension than other raters but the measures in this study were not adequate to tap it; (2) that evaluation apprehension is simply not an accurate or complete explanation of why these raters responded as they did on the clinical ratings. Logic seems to favor the former possibility, but in the absence of a more objective measure of evaluation apprehension, the issue cannot be resolved. Considering the possible implications of the surveillance--competence interaction for clinical practice and research, additional investigation of the phenomenon, including further study of the role of evaluation apprehension, seems indicated.

One additional point which should be made regarding the surveillance--competence effect has to do with the precise nature of that effect. There is no objective criterion by which it may be determined if the negative evaluation of the therapist by high competence peer surveillance raters represents a "bias" in their clinical ratings away from some objective reality or is in fact a reflection of the greater care with which their judgments were made. Resolution of these kinds of issues must await generally accepted objective criteria and measures of clinical skill. In the meantime, results like those described above, in which evaluations of a therapist's skill and personality are found to be influenced by observer variables, bear further study (see the discussion of clinical implications below).

### Implications for Social Psychology

General implications for social psychology. A distinguishing feature of this investigation was the attempt at an integration of what seemed to be related areas of social and clinical psychology. Thus, in part, the results of the study have some bearing on the general utility of this approach.

Integrating the social and clinical perspectives was found to be advantageous in several respects. For clinical psychology, the social literature provided the basis for hypothesizing and demonstrating the effects on clinical ratings of a new variable--surveillance. For social psychology, the study provided an opportunity to test new hypotheses and to verify previous findings regarding

evaluation apprehension and conformity. Furthermore, the fact that the study involved the extrapolation of social psychological principles to a new arena--clinical judgments--provided an opportunity to test the limits of applicability of some social psychological data.

Implications with regard to group differences in conformity. There were two major implications of the results of this study for social psychology--one bearing on group differences in conformity--the other bearing on the relationship of conformity to evaluation apprehension. With regard to the first issue, the results of the study showed no evidence of group differences in conformity, whether as a function of surveillance or any of the other independent variables. The lack of group differences was surprising, particularly in view of the "social reality" theory described earlier (p. 15), which seemed to indicate (1) that conformity would be a likely occurrence in clinical judgments and (2) that surveillance would increase the likelihood that conformity would occur.

There are several possible explanations for the negative results. One possibility is that for one of a number of reasons conformity is simply not likely to occur in most clinical judgment settings. For example, the inherent importance of clinical judgments--the implications of these judgments, particularly for patients--may militate against manifestations of conformity such as those observed in less critical matters (judgments of odors or length of lines). Further, it may be that the theory of social reality is not applicable to clinical judgments because discrepancy and divergence are tolerated

more than might be expected. Or, it could be that conformity as manifested in much social psychology research may not occur in the clinical context because the procedures of the social and clinical judgment situations are inherently quite different. (It might be possible to structure a clinical judgment situation so as to demonstrate that conformity may occur [i.e., raters could be required to announce their ratings publicly, etc.] However, such artificial situations would bear little resemblance to the typical clinical setting and demonstrations of conformity under these circumstances would have little meaning for clinical research.)

Another possible reason for the negative results with respect to conformity is that certain unique factors in the present study precluded the manifestation or the measurement of conformity. While there is no direct evidence that this was the case, the possibility must at least be considered. There is at least one aspect of the study that might support such an interpretation. The amount of self-expert discrepancy was significantly greater than the amount of self-peer discrepancy for all raters, regardless of surveillance. This may indicate that at least as far as graduate student raters are concerned, anyone described as an "expert" constitutes a more negative referent than one's graduate student peers. Thus, it is possible that a "ceiling effect" may have operated with respect to self-expert discrepancy scores, obscuring actual group differences in conformity due to surveillance. (However, in research relevant to



this issue, Gomes-Schwartz [1977] found no evidence of differences in conformity as a function of the professional status of the reference rater.)

Another possibility is that the procedures themselves, or the method of calculating conformity, contributed to the negative results. Thus, the fact that raters were required to give both their own judgments and their estimates of the judgments of others (and the fact that these two scores made up the discrepancy/conformity score) may have increased raters' awareness of the self-other comparison and thereby decreased the likelihood of conformity for any experimental group.

One other factor which could have contributed to the lack of group differences in conformity should be noted. This factor, which would apply to the general clinical context and to this particular study as well, related to the possibility that conformity in clinical ratings may occur, but as a function of several variables, none of which was isolated in this study. This possibility is suggested by the fact that the results of this study with respect to conformity were negative even for the variable which seemed most likely to produce main effects--Rater Orientation. Further, the possibility that a number of influential variables could affect conformity in clinical judgments is given added credence when it is recalled that making clinical ratings is no trivial task for therapists and therapists-in-training, but is more likely to be regarded as a serious undertaking which may reflect on one's professional ability.



Testing the accuracy of these suppositions would require complex and large designs since the potentially influential variables would have to be identified and adequately controlled; and research designs would have to provide for sufficient numbers of raters to test the relevant interactions. Unless such procedures are implemented, however, it seems unlikely that reliable group differences in conformity in clinical judgments will be found.

Implications with regard to the apprehension--conformity link. The fact that the experimental treatment groups were significantly different on measures of evaluation apprehension but did not differ in conformity (with one exception, the Self-Expert Total Discrepancy scores) indicates that the link between evaluation apprehension and conformity is not as strong or straightforward as expected.<sup>5</sup>

A re-examination of the social psychological literature indicated that at least thus far, the link between evaluation apprehension and conformity is largely an untested assumption. Thus, the likelihood that there is a clear-cut causal relationship

---

<sup>5</sup>A tangential but relevant point noted earlier (p. 99) relates to the measures of evaluation apprehension used in this study. That is, there is always the possibility that the kind of assessment of evaluation apprehension is related to the likelihood of finding relationships between this variable and others. As regards self-described evaluation apprehension, there is first the matter of awareness and then the matter of willingness to acknowledge evaluation apprehension. Thus, not only for this study, but for others as well, the possibility that self-described evaluation apprehension does not assess the same variable as a more objective measure might must always be considered.

between evaluation apprehension and conformity may have been stated more strongly in the introduction to this study than the empirical data support. It seems intuitively quite reasonable that evaluation apprehension may lead to conformity, but the actual strength of association between the two has yet to be empirically demonstrated. Further research on this point is indicated.

Indeed, for researchers interested in conformity a number of issues for further study have been identified. As far as clinical research is concerned, there was no evidence of a relationship between conformity and clinical ratings, which seems to indicate that the concept of conformity is not useful in explaining variables which may affect clinical ratings.

#### Implications for Clinical Research and Practice

Surveillance--competence interaction. The surveillance--competence interaction effect on clinical ratings which was described in detail earlier has at least three major implications for the clinical field which should be highlighted. Several of the important points have been noted previously. They are summarized below.

Surveillance and clinical ratings. The demonstrated impact of surveillance, a manifestly pervasive variable in many clinical judgment settings, indicates that greater attention should be given to context effects in general and to the effects of surveillance in particular. The need for further research into parameters of the influence of surveillance is indicated, e.g., the kinds of raters

who are most vulnerable to surveillance; the intervening variables (evaluation apprehension, etc.) involved; and the nature of the effect itself (is it "bias" or increased accuracy?).

Self-perceived competence and clinical ratings. Like surveillance, raters' competence has only infrequently been studied in clinical research. This study indicated there are a number of effects of competence on clinical ratings. Main effects of competence on clinical ratings are discussed further below.

The impact of competence--an observer variable--reaffirms the need for further investigation of observer variables in general. More specifically, the unique nature of the measure of competence in this study--the fact that it was a subjective measure which was found to exert an influence on a variety of clinical ratings--indicates that the rater's self-concept is a potentially important influence on clinical ratings which deserves closer scrutiny.

Assessment of therapist variables. The substantial interaction effect on ratings of the therapist in this study suggests strongly that whenever possible, clinical methodological investigations should include assessments of experimental effects on ratings of the therapist as well as the patient.

This suggestion takes on additional significance in light of the recent increased recognition of the importance of the therapist's personality and technique in effecting therapeutic change. Issues relating to the possibility of negative effects from psychotherapy are crucial in this regard (Strupp, Hadley, & Gomes-Schwartz, 1977).

More generally, remuneration for psychotherapy by insurance companies and other third-parties has heightened awareness of the need for improved evaluations of psychotherapy process and outcome (Hadley & Strupp, 1977). The assessment of therapist variables is an integral part of such evaluations. Licensing boards, professional associations, and commercial testing interests are developing measures of therapist competence, most of which rely heavily on ratings by observers. Study of variables which influence these ratings is clearly of considerable importance.

Main effects of competence, experience, and orientation. The main effects observed in this study both expand and confirm the results of previous research. Main effects of Self-Perceived Competence were found on ratings of the patient. Evaluations of the patient made by high competence raters were more favorable than evaluations by less competent raters. These patient variable main effects, together with the therapist variable surveillance--competence interaction effects observed, speak strongly of the desirability of further study of competence.

Main effects for rater experience also occurred on patient ratings, with more experienced raters giving the patient a more favorable general prognosis than did less experienced raters. The presence of main effects for experience runs counter to some earlier work (Temerlin, 1968) which showed little or no evidence of effects. The presence of experience main effects in the present study may be due to the fact that the measure of experience was directly related

to the amount of clinical work (i.e., number of hours experience as a therapist) rather than being a more gross estimate such as years in practice. Another possibility is that the relative level of experience is particularly important for raters at generally lower levels such as the graduate student raters in this study.

Rater Theoretical Orientation was found to be the least influential variable in this study, which confirms earlier findings (Raskin, 1965). The strongest effect observed was on the preferred method of determining therapy goals, which confirms Strupp's (1960) and Sundland and Barker's (1962) data.

Of general clinical interest. There were a number of tangential but intriguing aspects of the clinical ratings which should be noted. The first of these relates to the clustering of patient and therapist ratings which was observed in the factor analysis. Table 9 shows that liking for the therapist as a person clustered with a variety of ratings of his technical skill. For ratings of the patient, liking clustered with absence of depression and overall adjustment, but not with anxiety or general prognosis. Prognosis with therapy clustered with the therapist and not with other patient ratings.

There were at least two other intriguing aspects of the ratings. First, ratings of liking for the patient were generally higher than liking for the therapist (Tables 18 and 23). Also, not surprisingly, raters generally described the patient's prognosis with therapy as more favorable than his prognosis without therapy (Tables 16 and 17).



### A Final Caution

The customary caution on generalizing results is particularly important for this study. Specific note of this fact was made at several points throughout. In general, it should be recalled that only one patient and one therapist were rated. Some of the results such as the differences in liking or prognosis described above may well be largely a function of unique qualities in these individuals. More importantly, it seems likely that some aspects of the results could be unique to graduate student raters. The impact of self-perceived competence on clinical ratings and the negative evaluations of the therapist which resulted from peer surveillance are two such results. Further research will of course be necessary to establish the limits of applicability of the data. In the meantime, the results should be generalized with care.



## BIBLIOGRAPHY

- Allport, F. H. Social psychology. Boston: Houghton, 1924.
- Asch, S. E. Studies of independence and conformity: A minority of one against a unanimous majority. Psychological Monographs, 1956, 70 (9, Whole No. 416).
- Auerbach, A. H., & Luborsky, L. Accuracy of judgments of psychotherapy and the nature of the "good hour." In Schlein, J. M., Hunt, H. F., Matarazzo, J. D., & Savage, C. (Eds.), Research in psychotherapy, Vol. 3, Washington, D. C.: American Psychological Association, 1968.
- Bergin, A. E. Identifying the pseudoshrink. In H. H. Strupp (Chair), Psychonoxious therapy: Clinical, theoretical, and research perspectives. Symposium presented at the meeting of the American Psychological Association, Washington, D. C., September, 1976.
- Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.
- Bordin, E. S. Research strategies in psychotherapy. New York: John Wiley & Sons, 1974.
- Bordin, E. S., Cutler, R. L., Dittman, A. T., Harway, W. I., Raush, H., & Rigler, D. Measurement problems in process research on psychotherapy. Journal of Consulting Psychology, 1954, 18, 79-82.
- Brown, R. Social psychology. New York: Free Press, 1965.
- Calhoun, L. G., Selby, J. W., & Wroten, J. D. Situational constraint and type of causal explanation: The effects on perceived "mental illness" and social rejection. Journal of Research in Personality, 1977, 11, 95-100.
- Campbell, D. T. Conformity in psychology's theories of acquired behavioral dispositions. In Berg, I. A., & Bass, B. M. (Eds.), Conformity and deviation. New York: Harper & Brothers, 1961.
- Cartwright, D., & Zander, A. Pressures to uniformity in groups. In Cartwright, D., & Zander, A. (Eds.), Group dynamics. New York: Harper & Row, Publishers, 1968.
- Cristol, A. H. Studies of outcome of psychotherapy. Comprehensive Psychiatry, 1972, 13, 189-200.

- Crutchfield, R. S. Conformity and character. American Psychologist, 1955, 10, 191-198.
- DiNardo, P. A. Social class and diagnostic suggestion as variables in clinical judgment. Journal of Consulting and Clinical Psychology, 1975, 43, 363-368.
- Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42.
- Duncan, D. B. Multiple range tests for correlated and heteroscedastic means. Biometrics, 1957, 13, 164-176.
- Fleiss, J. L., Spitzer, R. L., & Burdock, E. I. Estimating accuracy of judgment using recorded interviews. Archives of General Psychiatry, 1965, 12, 562-567.
- Fleiss, J. L., Spitzer, R. L., Endicott, J., & Cohen, J. Quantification of agreement in multiple psychiatric diagnosis. Archives of General Psychiatry, 1972, 26, 168-171.
- Forer, B. R., Farberow, N. L., Feifel, H., Meyer, M. M., Sommers, V. S., & Tolman, R. S. Clinical perception of the therapeutic transaction. Journal of Consulting Psychology, 1961, 25, 93-101.
- Freedman, J. L., Carlsmith, J. M., & Sears, D. O. Social psychology. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.
- Gomes-Schwartz, B. Development of an instrument to assess predictors of negative change in psychotherapy. Unpublished manuscript, Vanderbilt University, 1977.
- Hadley, S. W., & Strupp, H. H. The evaluation of psychotherapy: Naiveté or necessity? Professional Psychology, 1977, in press.
- Hall, J. N. Inter-rater reliability of ward rating scales. British Journal of Psychiatry, 1974, 125, 248-255.
- Helzer, J. E., Robins, L. N., Taibelson, M., Woodruff, R. A., Reich, T., & Wish, E. D. Reliability of psychiatric diagnosis: A methodological review. Archives of General Psychiatry, 1977, 34, 129-133.
- Henchy, T., & Glass, D. C. Evaluation apprehension and the social facilitation of dominant and subordinate responses. Journal of Personality and Social Psychology, 1968, 10, 446-454.
- Hunt, W. A., & Jones, N. F. The experimental investigation of clinical judgment. In Bachrach, A. J. (Ed.), Experimental foundations of clinical psychology. New York: Basic Books, Inc., 1962.

- Johnson, R. E., Calhoun, L. G., & Boardman, W. K. The effect of severity, consistency, and typicalness information on clinicians' causal attributions. Journal of Clinical Psychology, 1975, 31, 600-604.
- Johnson, R. W., & Ryan, B. J. Observer recorder error as affected by different tasks and different expectancy inducements. Journal of Research in Personality, 1976, 10, 201-214.
- Kächele, H., Kühn, H., Grünzig, H., & Ohlmeier, D. Problems of clinical ratings of psychoanalytic group processes. Group Therapy and Group Dynamics, 1975, 9, 285-306.
- Kaplan, R. M., & Litrownik, A. J. Some statistical methods for the assessment of multiple outcome criteria in behavioral research. Behavior Therapy, 1977, 8, 383-392.
- Kent, R. N., O'Leary, K. D., Diament, C., & Dietz, A. Expectation biases in observational evaluation of therapeutic change. Journal of Consulting and Clinical Psychology, 1974, 42, 774-780.
- Kiesler, D. J. Basic methodologic issues implicit in psychotherapy process research. American Journal of Psychotherapy, 1966, 20, 135-155.
- Kittrie, N. N. The right to be different: Deviance and enforced therapy. Baltimore, Maryland: Johns Hopkins University Press, 1971.
- Kramer, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. Biometrics, 1956, 12, 307-310.
- Langer, E. J., & Abelson, R. P. A patient by any other name . . . : Clinical group difference in labeling bias. Journal of Consulting and Clinical Psychology, 1974, 42, 4-9.
- Lawlis, G. F., & Lu, E. Judgment of counseling process: Reliability, agreement, and error. Psychological Bulletin, 1972, 78, 17-20.
- Lee, S. D., & Temerlin, M. K. Social class, diagnosis, and prognosis for psychotherapy. Psychotherapy: Theory, Research, and Practice, 1970, 7, 181-185.
- Levinger, G., & Schneider, D. J. Test of the "risk is a value" hypothesis. Journal of Personality and Social Psychology, 1969, 11, 165-169.
- Lewisohn, P. M., Weinstein, M. S., & Alper, T. A behavioral approach to the group treatment of depressed persons: A methodological contribution. Journal of Clinical Psychology, 1970, 26, 525-532.

- Marsden, G. Content-analysis studies of therapeutic interviews: 1954 to 1964. Psychological Bulletin, 1965, 63, 298-321.
- Meehl, P. E. Clinical vs. statistical prediction. Minneapolis: Colwell Press, 1954.
- Mintz, J. What is "success" in psychotherapy? Journal of Abnormal Psychology, 1972, 80, 11-19.
- Mintz, J., Luborsky, L., & Auerbach, A. H. Dimensions of psychotherapy: A factor-analytic study of ratings of psychotherapy sessions. Journal of Consulting and Clinical Psychology, 1971, 36, 106-120.
- Raskin, N. J. The psychotherapy research project of the American Academy of Psychotherapists. APA Convention Proceedings, 1965, 253-254.
- Redfield, J., & Paul, G. L. Bias in behavioral observation as a function of observer familiarity with subjects and typicality of behavior. Journal of Consulting and Clinical Psychology, 1976, 44, 156.
- Romanczyk, R. G., Kent, R. N., Diament, C., & O'Leary, K. D. Measuring the reliability of observational data: A reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.
- Rosenberg, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1965, 1, 28-42.
- Sawyer, J. Measurement and prediction, clinical and statistical. Psychological Bulletin, 1966, 66, 178-200.
- Schiffenbauer, A. Effect of observer's emotional state on judgments of the emotional state of others. Journal of Personality and Social Psychology, 1974, 30, 31-35.
- Schwartz, C. C., Myers, J. K., & Astrachan, B. M. The outcome study in psychiatric evaluation research. Archives of General Psychiatry, 1973, 29, 98-102.
- Seitz, P. F. D. The consensus problem in psychoanalytic research. In Gottschalk, L., & Auerbach, A. (Eds.), Methods of research in psychotherapy. New York: Appleton-Century-Crofts, 1966.
- Sherif, M. Group influences upon the formation of norms and attitudes. In Newcomb, T. M., & Hartley, E. L. (Eds.), Readings in social psychology. New York: Holt, 1947.



- Shuller, D. Y., & McNamara, J. R. Expectancy factors in behavioral observation. Behavior Therapy, 1976, 7, 519-527.
- Spitzer, R. L., & Fleiss, J. L. A reanalysis of the reliability of psychiatric diagnosis. British Journal of Psychiatry, 1974, 125, 341-347.
- Steiner, I. D. Group process and productivity. New York: Academic Press, 1972.
- Strasburger, E. L., & Jackson, D. N. Improving accuracy in a clinical judgmental task. Journal of Consulting and Clinical Psychology, 1977, 45, 303-309.
- Strupp, H. H. The performance of psychoanalytic and client-centered therapists in an initial interview. Journal of Consulting Psychology, 1958, 22, 265-274.
- Strupp, H. H. Psychotherapists in action. New York: Grune & Stratton, 1960.
- Strupp, H. H., & Hadley, S. W. A tripartite model of mental health and therapeutic outcomes. American Psychologist, 1977, 32, 187-196.
- Strupp, H. H., Hadley, S. W., & Gomes-Schwartz, B. Psychotherapy for better or worse: An analysis of the problem of negative effects. New York: Jason Aronson, Inc., 1977.
- Sundland, D. M., & Barker, E. N. The orientations of psychotherapists. Journal of Consulting Psychology, 1962, 26, 201-212.
- Szasz, T. S. Psychiatric slavery. New York: The Free Press, 1977.
- Temerlin, M. K. Suggestion effects in psychiatric diagnosis. Journal of Nervous and Mental Disease, 1968, 147, 349-353.
- Temerlin, M. K., & Trousdale, W. W. The social psychology of clinical diagnosis. Psychotherapy: Theory, Research, and Practice, 1969, 6, 24-29.
- Tinsley, H. E. A., & Weiss, D. J. Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology, 1975, 22, 358-376.
- Zajonc, R. B. Social facilitation. Science, 1965, 149, 269-274.
- Zax, M., & Klein, A. Measurement of personality and behavior changes following psychotherapy. Psychological Bulletin, 1960, 57, 435-448.

Zeldow, P. B. Clinical judgment: A search for sex differences. Psychological Reports, 1975, 37, 1135-1142.

Zeldow, P. B. Effects of nonpathological sex role stereotypes on student evaluations of psychiatric patients. Journal of Consulting and Clinical Psychology, 1976, 44, 304.



## APPENDIX A

### Vanderbilt Psychotherapy Project Clinical Observations Questionnaire

Please circle the number which best represents your opinion. (Indicate half steps by ".5" if you wish.) Answer each question in sequence. Ignore the column at the right-hand side for the time being.

1. How much anxiety is this client experiencing?

1 2 3 4 5 6 7  
Virtually A great  
none deal

1 2 3 4 5 6 7

2. How depressed is he?

1 2 3 4 5 6 7  
Not at all Severely  
depressed depressed

1 2 3 4 5 6 7

3. How would you describe the client's overall adjustment?

1 2 3 4 5 6 7  
Very Excellent  
poor

1 2 3 4 5 6 7

4. If the client discontinued therapy immediately following the interview you saw, what do you believe occurred in terms of his overall adjustment over the next year?

1 2 3 4 5 6 7  
Got much Got much  
worse better

1 2 3 4 5 6 7

5. If the client continued with 25 hours of therapy with the therapist you observed, what do you believe occurred in the client's overall adjustment over the next year?

1 2 3 4 5 6 7  
Got much Got much  
worse better

1 2 3 4 5 6 7

6. How would you describe your general personal reaction to the client?

1 2 3 4 5 6 7  
Strongly Strongly  
negative positive

7. How active was the therapist in this session?

1 2 3 4 5 6 7  
Extremely Extremely  
passive active

1 2 3 4 5 6 7

8. How supportive was the therapist?

1 2 3 4 5 6 7  
Non- Very  
supportive supportive

1 2 3 4 5 6 7

9. How would you evaluate the therapist's overall competence?

1 2 3 4 5 6 7  
Incompetent Highly  
competent

1 2 3 4 5 6 7

10. What do you believe this therapist's predominant theoretical orientation to be? Please check one answer.

☐ Behavioral  
☐ Psychoanalytic  
☐ Client-centered  
☐ Eclectic

1 2 3 4 5 6 7

11. How would you describe your general personal reaction to the therapist?

1 2 3 4 5 6 7  
Strongly Strongly  
negative positive

12. If you were working with the same client, how much would your therapeutic technique resemble that of the therapist you observed?

|            |   |   |         |   |   |   |
|------------|---|---|---------|---|---|---|
| 1          | 2 | 3 | 4       | 5 | 6 | 7 |
| Markedly   |   |   | Very    |   |   |   |
| dissimilar |   |   | similar |   |   |   |

13. If you were the therapist, how would you determine the goals of therapy?

|             |   |   |                   |   |   |   |
|-------------|---|---|-------------------|---|---|---|
| 1           | 2 | 3 | 4                 | 5 | 6 | 7 |
| Leave goals |   |   | Determine goals   |   |   |   |
| to client's |   |   | based on my       |   |   |   |
| discretion  |   |   | perceptions       |   |   |   |
|             |   |   | and evaluation of |   |   |   |
|             |   |   | client            |   |   |   |

Now that you have rated all the items above, please go back and for each question where a scale to the right appears, rate your degree of confidence in your original response. Use the following scale definitions:

1 = Not at all confident

7 = Very confident

Please do not change any of your original responses.

APPENDIX B

Personal Data Questionnaire

1. Age \_\_\_\_\_ Sex \_\_\_\_\_
2. Number years (months) completed in graduate school \_\_\_\_\_
3. Approximate number hours experience as a therapist \_\_\_\_\_
4. How much experience (training, personal therapy) have you had with psychoanalytic principles and techniques?  

☐ Virtually none  
☐ A little  
☐ Moderate amount  
☐ Great deal
5. Which best describes your predominant theoretical clinical orientation? Please check one only.  

☐ Behavioral  
☐ Analytic  
☐ Client-centered  
☐ Eclectic
6. How would you rate your present competence as a therapist?  

|           |   |   |   |           |   |   |
|-----------|---|---|---|-----------|---|---|
| 1         | 2 | 3 | 4 | 5         | 6 | 7 |
| Not very  |   |   |   | Fully     |   |   |
| competent |   |   |   | competent |   |   |
7. How much apprehension did you experience over the possibility of your ratings being seen and evaluated by others?  

|      |   |   |   |   |         |   |
|------|---|---|---|---|---------|---|
| 1    | 2 | 3 | 4 | 5 | 6       | 7 |
| None |   |   |   |   | A great |   |
|      |   |   |   |   | deal    |   |

8. Do you believe your ratings were influenced to any degree by the possibility of their being seen by others?

\_\_\_\_\_ Yes                  \_\_\_\_\_ No

If "yes," how?

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

9. Would you have preferred that your ratings be anonymous?

\_\_\_\_\_ Yes                  \_\_\_\_\_ No                  \_\_\_\_\_ Don't Care





