12-1986

Experiencing Scale Discrimination between More and Less Productive Psychotherapy Sessions

Drevis L. Hager
Western Michigan University

Follow this and additional works at: http://scholarworks.wmich.edu/dissertations

Part of the Educational Psychology Commons

Recommended Citation
Dissertations, 2292.
http://scholarworks.wmich.edu/dissertations/2292

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
EXPERIENCING SCALE DISCRIMINATION BETWEEN MORE AND LESS PRODUCTIVE PSYCHOTHERAPY SESSIONS

by

Drevis L. Hager

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Counselor Education
and Counseling Psychology

Western Michigan University
Kalamazoo, Michigan
December 1986

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
EXPERIENCING SCALE DISCRIMINATION BETWEEN MORE AND LESS PRODUCTIVE PSYCHOTHERAPY SESSIONS

Drevis L. Hager, Ed.D.
Western Michigan University, 1986

Previous studies have failed to detect differences in experiencing (EXP) between more and less productive psychotherapy sessions. This study explored the possible relationship between EXP and session productivity and examined the efficacies of the conventional versus an alternative method of applying the EXP Scale.

Six client-therapist pairs audio recorded 52 psychotherapy hours, and after each independently completed a session productivity questionnaire. More and less productive sessions were selected based on the clients', therapists', and the clients' and therapists' combined responses. EXP ratings were made of 8-minute audio recorded segments, first using conventional mode and peak scores, and then using alternative frequency and duration scores for each scale level occurring in each segment.

There were no significant differences in EXP between more and less productive sessions when all of the EXP variables were simultaneously tested using Hotelling's $T^2$. A step-wise discriminant analysis detected differences between sessions judged by both clients and therapists as more and less productive sessions when the frequencies of EXP Levels 2, 1, and 3, plus the mode formed the discriminant function ($p = .003$). There were significant relationships between
the mode scores of segments and the levels with the greatest frequencies ($r = .56, p < .001$) and cumulative durations ($r = .32, p < .05$). There was no significant difference between the conventional and alternative methods in ability to discriminate between more and less productive sessions, although the alternative variables contributed more weight to the significant discriminant function.

It was concluded that the alternative method provides valuable information, and further study is necessary before conclusions can be made regarding EXP and session productivity.
INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.
- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.
- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17"x 23" black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6"x 9" black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
Hager, Drevis Lee

EXPERIENCING SCALE DISCRIMINATION BETWEEN MORE AND LESS PRODUCTIVE PSYCHOTHERAPY SESSIONS

Western Michigan University

University Microfilms International

300 N. Zeeb Road, Ann Arbor, MI 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received
16. Other

University Microfilms International

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the clients and therapists who participated in this study. Thanks are also due to Anna Marie McClintock, Francie Dalton, and Barb Knott for their transcribing of the tape recorded sessions; James Guinee and Susan Janke for the many hours spent making Experiencing Scale ratings; Lee Pakko for her typing expertise and helpful formatting suggestions; and Dr. Ralph Trimble for making available the resources of the University of Illinois Counseling Center.

I also give thanks to Mary Zirpoli, Larry Beer, and Mike Baker, who have helped me to make it through this ordeal by simply being who they are; Dr. Ed Trembley for stimulating my interest in the helpful ingredients of psychotherapy; and Dr. Jim Bosco for continually asking that I make some sense. Special thanks go to Dr. Bill Carlson for his superior guidance during the past 5 years. As a teacher and supervisor he has taught me to be a therapist, as a mentor he has taught me the meaning of professionalism, and as a friend he has taught me much about life.

Finally, I give my deep appreciation to my dearest friend, Linda Michalik Hager, whose boundless love has brought an enduring contentment to my life, even when having to work on "the dissertation."

Drevis L. Hager
## TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................. ii  
LIST OF TABLES ....................................................... vi  
LIST OF FIGURES ...................................................... viii  

### CHAPTER

I. INTRODUCTION .................................................. 1  
   Definition of Key Terms ........................................... 5  

II. REVIEW OF THE LITERATURE ................................. 7  
   Origins of the Experiencing Scale ............................... 7  
   Process and Experiencing Scale Validation Studies .......... 12  
   Process Scale Studies ............................................. 12  
   Early Experiencing Scale Studies .............................. 19  
   Experiencing and Session Productivity .................... 33  
   The Problem and the Research Focus ......................... 46  
   Session Productivity ............................................. 46  
   Data Form ........................................................ 48  
   The Present Research Focus ................................... 55  
   Research Questions .............................................. 58  
   Experiencing and Session Productivity .................... 59  
   Conventional and Alternative Methods ....................... 60  

III. METHOD .......................................................... 63  
   Subjects .......................................................... 63  
   Therapists ......................................................... 63  

iii  

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table of Contents—Continued

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection of Clients</td>
<td>65</td>
</tr>
<tr>
<td>Clients</td>
<td>65</td>
</tr>
<tr>
<td>Procedure</td>
<td>67</td>
</tr>
<tr>
<td>Selection and Preparation of Segments</td>
<td>69</td>
</tr>
<tr>
<td>Raters and Rating</td>
<td>72</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>79</td>
</tr>
<tr>
<td>Experiencing and Session Productivity</td>
<td>79</td>
</tr>
<tr>
<td>Conventional and Alternative Methods</td>
<td>80</td>
</tr>
<tr>
<td>Multivariate Mathematical Assumptions</td>
<td>82</td>
</tr>
<tr>
<td><strong>IV. RESULTS</strong></td>
<td>85</td>
</tr>
<tr>
<td>Question 1</td>
<td>85</td>
</tr>
<tr>
<td>Null Hypothesis 1A</td>
<td>85</td>
</tr>
<tr>
<td>Null Hypothesis 1B</td>
<td>87</td>
</tr>
<tr>
<td>Null Hypothesis 1C</td>
<td>91</td>
</tr>
<tr>
<td>Question 2</td>
<td>98</td>
</tr>
<tr>
<td>Null Hypothesis 2A</td>
<td>98</td>
</tr>
<tr>
<td>Null Hypothesis 2B</td>
<td>100</td>
</tr>
<tr>
<td>Null Hypothesis 2C</td>
<td>100</td>
</tr>
<tr>
<td>Question 3</td>
<td>101</td>
</tr>
<tr>
<td>Question 4</td>
<td>102</td>
</tr>
<tr>
<td>Null Hypothesis 4A</td>
<td>113</td>
</tr>
<tr>
<td>Null Hypothesis 4B</td>
<td>113</td>
</tr>
<tr>
<td>Null Hypothesis 4C</td>
<td>114</td>
</tr>
</tbody>
</table>
# Table of Contents—Continued

## CHAPTER

V. DISCUSSION .................................................. 115

- Discussion of the Results .................................. 115
- Experiencing and Session Productivity ................. 115
- Conventional and Alternative Methods ................. 117
- Issues and Limitations .................................... 121
- Recommendations for Future Research ................ 126
  
  
  Experiencing and Session Productivity ................... 126
  
  
  Alternative Method of Summarizing
  Experiencing ................................................... 128

Summary ......................................................... 129

## APPENDICES .................................................. 133

- A. Client Initial Symptom/Complaint Self-Report
  Ratings and Posttherapy Change Scores ................. 134
- B. Consent Form for Research Participants .............. 138
- C. Pre- and Postcounseling Symptom List ................ 140
- D. Postsession Questionnaire ............................. 142
- E. Therapist Data Sheet .................................... 144
- F. Code of Ethics .......................................... 146
- G. Directions for Running Ratings ...................... 148

## BIBLIOGRAPHY ............................................... 150
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Therapists' Characteristics</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>Client Characteristics</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>Temporal Location of Taped Sessions</td>
<td>68</td>
</tr>
<tr>
<td>4</td>
<td>Pearson $r$ Correlations Between Raters' and Standard Ratings of Training Segments</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>Pearson $r$ Correlations for Raters A and B During Training and Research Segments</td>
<td>75</td>
</tr>
<tr>
<td>6</td>
<td>Reliabilities for Alternative Ratings</td>
<td>77</td>
</tr>
<tr>
<td>7</td>
<td>Means, Standard Deviations, and $t$ Values for Client-Determined More (MP) and Less Productive (LP) Sessions</td>
<td>86</td>
</tr>
<tr>
<td>8</td>
<td>Means, Standard Deviations, and $t$ Values for Therapist-Determined More (MP) and Less Productive (LP) Sessions</td>
<td>90</td>
</tr>
<tr>
<td>9</td>
<td>Means, Standard Deviations, and $t$ Values for Combined-Determined More (MP) and Less Productive (LP) Sessions</td>
<td>94</td>
</tr>
<tr>
<td>10</td>
<td>Standardized Canonical Discriminant Function Coefficients for Client-, Therapist-, and Combined-Determined Productivity With Accompanying Probability Values for the Discriminant Functions</td>
<td>99</td>
</tr>
<tr>
<td>11</td>
<td>Pearson $r$ Correlation coefficients for Levels With the Highest Frequency ($f$), Total Duration ($td$), and Average Duration ($ad$) With the Mode</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>Variables Selected With Step-Wise Discriminant Analysis Procedure With Accompanying Standardized Canonical Discriminant Function Coefficients and Discriminant Function Probability Values</td>
<td>102</td>
</tr>
<tr>
<td>13</td>
<td>Chi-Squared Values and Probability Values Associated With Discriminant Analyses of More and Less Productive Sessions Based on Conventional and Alternative Variables</td>
<td>103</td>
</tr>
</tbody>
</table>
List of Tables--Continued

14. Percentage of More Productive, Less Productive, and Total Sessions Correctly Classified by Functions Based on Conventional Versus Alternative Variables .................................................. 111

15. Frequencies (and Percentages) of Cases Classified Correctly and Incorrectly for Discriminant Functions Based on Conventional and Alternative Variables ..................................... 112

16. Chi-Squared Values and Probability Statements for McNemar's Test of Differences Between Correctly and Incorrectly Classified Sessions .............................. 113
## LIST OF FIGURES

1. Mean Mode and Peak Ratings for Client-Determined More (MP) and Less Productive (LP) Sessions .......... 87

2. Mean Frequencies of EXP States for Client-Determined More (MP) and Less Productive (LP) Sessions .......... 88

3. Mean Durations of EXP States for Client-Determined More (MP) and Less Productive (LP) Sessions .......... 89

4. Mean Mode and Peak Ratings for Therapist-Determined More (MP) and Less Productive (LP) Sessions .......... 91

5. Mean Frequencies of EXP States for Therapist-Determined More (MP) and Less Productive (LP) Sessions .......... 92

6. Mean Durations of EXP States for Therapist-Determined More (MP) and Less Productive (LP) Sessions .......... 93

7. Mean Mode and Peak Ratings for Combined-Determined More (MP) and Less Productive (LP) Sessions .......... 95

8. Mean Frequencies of EXP States for Combined-Determined More (MP) and Less Productive (LP) Sessions .......... 96

9. Mean Durations of EXP States for Combined-Determined More (MP) and Less Productive (LP) Sessions .......... 97

10. Frequency Distributions of Client-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Conventional Variables ............... 105

11. Frequency Distributions of Client-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Alternative Variables ............... 106

12. Frequency Distributions of Therapist-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Conventional Variables ............... 107
List of Figures--Continued

13. Frequency Distributions of Therapist-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Alternative Variables ............... 108

14. Frequency Distributions of Combined-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Conventional Variables ............... 109

15. Frequency Distributions of Combined-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Alternative Variables ............... 110
CHAPTER I

INTRODUCTION

The search for factors associated with client improvements in psychotherapy has inspired psychotherapy researchers to investigate an array of variables related to treatment conditions, therapists, clients, and client-therapist combinations. One of the more promising client process variables, experiencing (EXP), has developed from the works of Rogers (1958, 1959a, 1961a, 1961b; Walker, Rablen, & Rogers, 1960) and Gendlin (1958, 1961, 1962, 1964, 1969; Gendlin & Zimring, 1955). According to Luborsky and Spence (1978), "more work has been devoted to the concept of experiencing than any other process variable" (p. 339).

Experiencing refers to the manner which an individual attends to the continuous flow of sensory data known as feeling. Experiencing is conceptualized as a continuum, at one end the individual is unaware of his or her feelings, and at the other extreme the individual's attention is focused directly on his or her feelings, and uses this awareness to arrive at resolutions and new understandings. It was proposed by both Rogers (1958) and Gendlin (1964) that psychotherapeutic improvement is evidenced by progression along this continuum toward greater utilization of one's feelings as a referent for meaning and action.
Experiencing is measured with the Experiencing Scale (Klein, Mathieu, Gendlin, & Kiesler, 1969), a 7-point rating scale used to infer the client's in-session depth of self-exploration and involvement with his or her feelings. Judges are trained to apply the scale to the client verbalizations evidenced on audio tape segments and/or transcripts of segments sampled from therapy interviews. A number of studies indicate that EXP is positively associated with case outcome; clients who improve the most have been found to exhibit higher levels of EXP at varying points in treatment, including during the initial session (Kiesler, 1971; Ryan, 1966; Tomlinson, 1967; Tomlinson & Hart, 1962; van der Veen, 1967; Walker et al., 1960). Although developed from within client-centered theory as a tool to investigate individual psychotherapy, the scale has also been used to investigate psychoanalytic psychotherapy (Auerbach & Luborsky, 1968; Cartwright, 1966; Elliott, Cline, & Shulman, 1982), Gestalt therapy (Greenberg, 1980, 1983; Greenberg & Rice, 1981; Greenberg & Webster, 1982), group therapy (Lewis & Beck, 1983), and the reporting of dreams (Hendricks & Cartwright, 1978).

A number of authorities have indicated that one of the more promising avenues for psychotherapy research involves the analysis of productive sessions and significant intrasession events (Auerbach & Luborsky, 1968; Elliott, 1983a, 1984; Gendlin, 1986; Mahrer & Nadler, 1986; Marmor, Wilner, & Horowitz, 1984; Orlinsky & Howard, 1967, 1975; Rice & Greenberg, 1984; Stiles, 1980). This represents a movement from past studies primarily emphasizing various correlates of case outcome, to a current emphasis on the productive microscopic
processes (microprocesses) that occur within and between sessions. The EXP Scale has been used to investigate within- and between-session processes in a number of studies (Auerbach & Luborsky, 1968; Elliott, 1983b; Elliott et al., 1982; Greenberg, 1980, 1983; Greenberg & Rice, 1981; Pollak, 1973). Some of these have demonstrated EXP trends associated with significant events within single sessions (Elliott, 1983b; Greenberg, 1980, 1983; Greenberg & Rice, 1981), while three studies have tested for differences between more and less productive sessions (Auerbach & Luborsky, 1968; Elliott et al., 1982; Pollak, 1973). Studies of more and less productive sessions, however, have failed to demonstrate significant differences in EXP. Perhaps one reason for the lack of significant differences in two of these studies (Auerbach & Luborsky, 1968; Pollak, 1973) is that the judgments of session productivity were not made by the actual therapy participants. The Elliott et al. (1982) study did utilize the therapist and client as judges of session productivity, and the differences in EXP between more and less productive sessions approached significance (p < .10). However, the reason for the lack of significant differences may have been the use of very brief (30 second) audio tape segments as units of analysis (the typical segment is 4 to 8 minutes long).

To date there has been no attempt to test for differences in EXP between more and less productive sessions with the actual therapy participants as judges of session productivity and with the segment length being within the standard 4 to 8 minutes. The approach used in the present study is to utilize both clients and therapists as

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
judges of session productivity and to use sufficiently long (8-minute) tape segments in order to address the question: Are there differences in EXP between more and less productive sessions?

This study also has a second focus on issues associated with the conventional method of summarizing data yielded with the EXP Scale. In brief, these issues include various interpretation problems which arise from a generally narrow band of results, questions about what dimensions the scale actually measures, and some inherent difficulties in using the scale at the microprocess level. An alternative method of summarizing EXP data is introduced as a possible antidote to the limitations inherent in the conventional method. Three research questions are then asked which address the relationship between these two methods: (a) Is there a relationship between the data produced by the conventional and alternative methods? (b) Which combination of conventional and alternative EXP variables optimally discriminates between more and less productive sessions? (c) Is there a difference between the conventional and alternative methods in ability to discriminate between more and less productive sessions?

Concordant with the dual focus of this study, the purposes of this study are twofold: first, to determine if the quality of client EXP differs depending on whether a session is regarded as productive or not and, second, to determine if either the conventional or alternative method of summarizing EXP generates more useful data. The value of this study is that it will contribute to the body of existing knowledge on session productivity and experiencing, and this in turn may support the theoretical stance that higher levels of EXP
represent more productive therapy behaviors. This study will also contribute to the existing knowledge of the EXP Scale's qualities, and may provide a more useful method of summarizing EXP.

Definition of Key Terms

The key terms used repeatedly in this study are defined here. 

**Experiencing (EXP)** is "the extent to which the ongoing, bodily, felt flow of experiencing is the basic datum of an individual's awareness and communications about the self and the extent to which this inner datum is integral to action and thought" (Mathieu-Coughlin & Klein, 1984, p. 213). Experiencing is conceptualized as a continuum, at one end the individual is unaware of his or her feelings, and at the other extreme the individual's attention is focused directly on his or her feelings, and uses this awareness to arrive at resolutions and new understandings.

The **Experiencing Scale** is a rating instrument used to infer a client's level of EXP. The scale has seven operational anchors; descriptions of each anchor are given in Chapter II.

The **conventional method** is the usual method of applying the EXP Scale as prescribed by the authors of the scale (Klein et al., 1969). Raters are trained to assign two scores to each segment of audio tape (or transcript): a **mode** score and a **peak** score. The mode is the most pervasive scale level evidenced by a client, and the peak is the highest level attained by the client.

The **alternative method** is the application of the EXP Scale as introduced in the present study. Raters are trained to assign an EXP
Scale level to each client statement within a segment. The scores consist of the frequency of occurrence and the duration in seconds of each scale level evidenced in the segment.
CHAPTER II

REVIEW OF THE LITERATURE

Origins of the Experiencing Scale

Hart (1970) has described the history of client-centered therapy as characterized by three eras: nondirective psychotherapy (1940-1950), reflective psychotherapy (1950-1957), and experiential psychotherapy (1957 to date). It is from this third era that the experiencing construct emerged and became central to client-centered theory. The works of Rogers (1958, 1959a, 1961a, 1961b; Walker et al., 1960) and Gendlin (1958, 1961, 1962, 1964, 1969; Gendlin & Zimring, 1955) have woven together to create the EXP construct.

Gendlin and Zimring (1955) were the first client-centered therapists to speculate on the vicissitudes of the client's internal experience which constitute the course of psychotherapeutic change. Rogers (1958, 1961a, 1961b) then described his process notion of psychotherapy: a conception which also focused on the phenomenology of the changing client.

Rogers (1958) envisioned the improving client as moving through a sequence of up to seven stages, with each stage marked by changes on a number of factors or strands. He described a total of seven strands which cut across the seven stages of change. These strands included the individual's manner of relating, communication of self, degree of incongruence, relationship to problems, construing of
experience, relationship to feelings, and manner of experiencing. Rogers proposed that as one progresses upward on the continuum of change stages, the qualities inherent in the above strands move from a general state of fixity, stasis, or rigidity, to increasing degrees of flow or changingness. The person becomes increasingly aware of the myriad processes occurring within, and these awarenesses become the basis for the creation of new meanings. Rogers also regarded this general process as constituting improvement in psychotherapy.

At the lower end of this continuum of stages, "the individual is largely unaware of his feeling life . . . . Feelings may at times be exhibited in ways which seem quite obvious to the observer, but they are unrecognized as such by the individual" (Rogers, 1959a, p. 99). Toward the middle of the continuum,

We find feelings and personal meanings described as present objects, owned by the self . . . . There is often a dim recognition that feelings previously denied to awareness may break through and be experienced in the present, but this seems to be a frightening possibility. (p. 99)

At the upper end, "new feelings are experienced with richness and immediacy, and this experiencing is used as a clear and definite referent from which further meanings may be drawn" (p. 99).

The Process Scale (Rogers, 1959a; Walker et al., 1960) was an initial attempt to measure these process movements by applying a procedure to rate interview transcripts and audio recordings. The scale was organized around Rogers's seven stages of change and required raters to arrive at judgments of where the client's verbalizations fell on this continuum of change stages.
At the same time that Rogers was developing his process conception, Gendlin (1958, 1961, 1962, 1964) concentrated his efforts on more fully describing the experiencing dimension of this process and its quality. Gendlin's experiencing and Rogers's process conceptions were parallel and nearly identical. Both Gendlin and Rogers emphasized that effective therapy was accompanied by progressive qualitative shifts in the client's experience of his or her feeling life. The primary difference between the two is that Rogers more strongly emphasized how the person's feelings, constructs, and various awarenesses all entered into his or her notion of self, whereas Gendlin more strongly attended to the shifts in actual body sensations which constitute the person's feelings, along with the meanings which emerged from these shifts. Gendlin focused on the more narrowly circumscribed manner of experiencing factor, while Rogers's approach emphasized seven interrelated factors in change. Gendlin has also developed experiential psychotherapy (1961, 1964, 1979) as a general approach and experiential focusing (1969, 1981) as a central technique in this approach.

Gendlin and Tomlinson (1962) revised the Process Scale, thus creating the Experiencing Scale. This new scale was based on a single factor which largely reflected a blend of Rogers's manner of experiencing and relationship to feelings strands, and thus simplified the rating procedure by reducing the number of rating dimensions from seven to one. The current version of the EXP Scale is the result of two additional revisions (Klein et al., 1969, Mathieu & Klein, 1963). Mathieu-Coughlin and Klein (1984) now define EXP as
"the extent to which the ongoing, bodily, felt flow of experiencing is the basic datum of an individual's awareness and communications about the self and the extent to which this inner datum is integral to action and thought" (p. 213). The scale incorporates seven anchors quite similar to Rogers's initial conception.

At the lower level on the continuum of experiencing, discourse is markedly impersonal or superficial. Moving up the scale, there is a progression from simple, limited, or externalized self references to inwardly elaborated descriptions of feelings. At higher levels, feelings are explored. Then, new aspects of experiencing emerge from what is directly sensed but at first unclear. (Mathieu-Coughlin & Klein, 1984, p. 213)

[At Stage 1] the chief characteristic... is that the content or manner of expression is impersonal. In some cases the content is intrinsically impersonal, being a very abstract, general, superficial, or journalistic account of events or ideas with no personal referent established. In other cases, despite the personal nature of the content, the speaker's involvement with the content is impersonal, so that he reveals nothing important about himself and his remarks could as well be about a stranger or an object. (Klein et al., 1969, p. 56)

[At Stage 2] the association between the speaker and the content is explicit. Either the speaker is the central character in the narrative or his interest is clear. The speakers' involvement, however, does not go beyond the specific situation or content. All comments, associations, reactions, and remarks serve to get the story or idea across but do not refer to or define the speaker's feelings. (Klein et al., 1969, pp. 56-57)

[At Stage 3] the content is a narrative or a description of the speaker in external or behavioral terms with added comments on his feelings or private experiences. These remarks are limited to the events or situation described, giving the narrative a personal touch without describing the speaker more generally. Self-descriptions restricted to a specific situation or role are also at stage three. (Klein et al., 1969, p. 58)

[At Stage 4] the content is a clear presentation of the speaker's feelings, giving his personal, internal perspective or feelings about himself. Feelings or the
experience of events, rather than the events themselves, are the subject of the discourse. By attending to and presenting his experiencing, the speaker communicates what it is like to be him. These interior views are presented, listed, or described, but are not interrelated or used as the basis for systematic self-examination or formulation. (Klein et al., 1969, p. 59)

[At Stage 5] the content is a purposeful exploration of the speaker's feelings and experiencing. There are two necessary components. First, the speaker must pose or define a problem or proposition about himself explicitly in terms or feelings. The problem or proposition may involve the origin, sequence, or implications of feelings or relate feelings to other private processes. Second, he must explore to work with the problem in a personal way. The exploration or elaboration must be clearly related to the initial proposition and must contain inner references so that it functions to expand the speaker's awareness of his experiencing. Both components, the problem and the elaboration, must be present. (Klein et al., 1969, p. 60)

[At Stage 6] the content is a synthesis of readily accessible, newly recognized, or more fully realized feelings and experiences to produce personally meaningful structures or to resolve issues. . . . He communicates a new or enriched self-experiencing and the experiential impact of the changes in his attitudes or feelings about himself. . . . Apart from the specific content, the speaker conveys a sense of active, immediate involvement in an experientially anchored issue with evidence of its resolution or acceptance. (Klein et al., 1969, p. 61)

[At Stage 7] the content reveals the speaker's expanding awareness of his immediately present feelings and internal processes. He demonstrates clearly that he can move from one inner reference to another, altering and modifying his conceptions of himself, his feelings, his private reactions to his thoughts or actions in terms of their immediately felt nuances as they occur in the present experiential moment, so that each new level of self-awareness functions as a springboard for further exploration. (Klein et al., 1969, p. 62)

An elaborate standardized procedure has been developed to train raters to make EXP Scale ratings (Klein et al., 1969). Raters are taught to assign two scores to a segment: a mode score and a peak score. The mode represents the rater's judgment of the client's

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
overall level of EXP within the segment, while the peak is the highest level attained by the client, regardless of how long the client remained at that level.

Interrater reliabilities for the scale have ranged widely but have generally been acceptable. In seven studies reported by Klein et al. (1969), Ebel interclass reliabilities (Ebel, 1951; Guilford, 1954) ranged from .76 to .91 for mode ratings, and from .75 to .92 for peak ratings. Other reported interrater reliabilities are Pearson product-moment correlation coefficients of .53 to .67 (Tomlinson, 1962); .72 to .81 (Greenberg, 1980; Greenberg & Rice, 1981); .89, .87, and .75 (Greenberg, 1983); and .86 (Greenberg & Webster, 1982). Kiesler (1971) reported Ebel interclass reliabilities of .76 and .79, while others have reported reliabilities (type not specified) of .48, .22, and .42 (Auerbach & Luborsky, 1968); .58 (van der Veen, 1967b); .91 to .94 (Kiesler, 1970); .83 (Lansford & Bordin, 1983); and .79 and .87 (Pollak, 1973).

Process and Experiencing Scale Validation Studies

Process Scale Studies

A number of validation studies were conducted during the late 1950s and early 1960s which investigated the relationships between Rogers's Process Scale and case outcome. Tomlinson (1959) applied an early draft of the Process Scale (Rogers & Rablen, 1958) to tape segments taken from an earlier therapy study at the University of Chicago Counseling Center (Halkides, 1958). Nine segments were
randomly drawn from each of 40 sessions: two sessions for each of 20 clients. The length of the segments was not temporally defined, but instead consisted of a client statement of at least two syllables plus the following therapist response. The sessions sampled were one "early" and one "late" therapy interview for each client. Tomlinson did not specify exactly which interviews were selected or how they were determined. The clients were classified as either more successful \((n = 10)\) or less successful \((n = 10)\) on the basis of therapist ratings of client improvement and pre-post differences on a battery of psychological tests. The interrater reliabilities for the various pair combinations of three judges were at best modest and ranged from .47 to .63 (Pearson \(r\)).

Significant differences in process were found between the more successful and less successful cases, with the more successful cases receiving higher process ratings at both early and late interviews. There were no differences in process between early and late interviews for either more or less successful cases. These findings were concordant with Tomlinson's (1959) prediction that higher levels of process would be associated with better outcome, but did not support Rogers's (1958) hypothesis that successful therapy is evidenced by a client's progression to higher process stages.

Although statistically significant, the differences in mean process scores between the more and less successful cases were relatively small: 3.21 versus 2.89 for the early interviews and 3.26 versus 2.90 for the later interviews. This result poses an interpretation problem, for although in theory the process concept is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
continuous, at the operational level the Process Scale has discrete verbal anchors. Thus it is difficult to arrive at firm conceptual conclusions based on differences of less than one full scale stage.

Walker et al. (1960) applied the Process Scale to transcripts of segments extracted from six therapy cases at the University of Chicago Counseling Center. The six cases were selected from two earlier studies and were chosen because they represented three marked progress and three minimal progress cases. Four pages of transcript were selected for rating from each case: two of the pages taken from two early sessions (usually the second and third sessions) and the other two pages from two late sessions (usually the two sessions prior to the last session). Thus a total of 24 transcript pages were rated by two raters who achieved an interrater reliability of $r = .83$. The results indicated that the marked progress cases had made significantly higher gains in process level from earlier sessions to later sessions than did the minimal progress cases. The gain in process level was determined by the difference between the mean process level of the two early sessions and the mean process level of the two later sessions. The mean gain for the marked progress group was 1.93 levels, while the mean gain for the minimal progress group was 0.30 levels. The conclusion drawn was that those who evidence more improvement also make greater gains in terms of process levels.

Some shortcomings of this study which might limit confidence in the above conclusion include an absence of stated criteria for the selection of transcript portions from each session, a lack of explicitly defined a priori criteria for the selection of cases, and no
reporting of the mean process levels evidenced by the groups (only
the early-late difference scores were reported). Also, one of the
minimal progress cases was seen only for three sessions; and conse­
quently, the two early sessions consisted of Sessions 1 and 2, and
the later sessions were Sessions 2 and 3. This arrangement clearly
confounded the early-late difference score for this case and also the
minimal progress group due to Session 2 being treated as both an
early and a late session.

Tomlinson and Hart (1962) made Process Scale ratings of 10
therapy cases at the University of Chicago Counseling Center. For
each case 9 two-minute segments were extracted from tape recordings
of the second session and the second to last session. A total of 180
two-minute segments were randomly assigned for rating by two raters
who reached an interjudge reliability of .65 (Pearson r). A signifi­
cant difference in level of process ratings was found between the
five cases judged to be more successful and the five cases judged to
be less successful, with the more successful clients receiving higher
ratings. Therapy success was determined on the basis of the thera­
pists' ratings of outcome, the clients' ratings of outcome, and the
clients' self-concept Q-sorts. A significant difference was also
found between process scores of earlier versus later sessions, with
the later sessions receiving higher process ratings. Although this
difference was statistically significant, the early-late difference
scores were relatively small (an average of 0.7 levels for one rater
and 0.3 for the other), thus making this result difficult to inter­
pret conceptually. It was hypothesized that more successful cases
would evidence greater gains in process scores between early and later sessions than would less successful cases; however, the data did not support this hypothesis. There was also no statistical support for the hypothesis that the later half of a session would be higher in process than the first half of a session.

Cartwright (1966) compared data from two psychoanalytic therapy cases and two client-centered therapy cases that were closely matched on a number of client and therapist variables. Transcripts were made of the first, last, and every fifth session for each of the four cases. Using this method, a total of 18 psychoanalytic sessions and 17 client-centered sessions were transcribed. The transcripts were divided into units three pages long, and each unit was rated by one of two raters. The scores given to the units which comprised a session were then averaged together to produce a score for that session.

The results indicated that the range of process levels was narrower for the two psychoanalytic cases (2.7 to 3.4 for one case, 3.0 to 3.8 for the other) than for the client-centered cases (2.7 to 4.2; 3.6 to 5.5). Also, the amount of change in process level from the first session to the highest session was greater for the client-centered cases (1.5 levels for both cases) than for the two psychoanalytic cases (0.2 levels for one case and 0.8 levels for the other). No comparison was made of the mean process levels between the psychoanalytic and client-centered cases. Cartwright (1966) concluded that changes on the Process Scale "seem to show that it reflects a process peculiar to the client-centered technique"
Some factors which might limit confidence in Cartwright's (1966) conclusion include having an $N$ of two per group, no report of statistical tests of significance, and one of the raters (Cartwright) rating two of the cases and another rater rating the other two. No mention is made of which cases each rater rated and thus it may be that the differences found between the psychoanalytic and client-centered cases actually reflect differences between raters.

Van der Veen (1967a) tested Rogers's (1957, 1959a, 1959b) hypothesis that therapy patients will encounter improvement and predictable changes in process level if therapists communicated to their patients personal genuineness and congruence, unconditional positive regard, and empathic understanding. As a measure of patient process level, van der Veen did not use the full Process Scale, but instead used three scales derived from three of the seven strands of Rogers's (1958) model of therapeutic change: a problem expression scale, an intrapersonal exploration scale, and a manner of relating scale; each having seven anchors which reflected Rogers's stages of change. Outcome was measured by a combination of clinical judgment and changes on the Minnesota Multiphasic Personality Inventory (MMPI), Rorschach, an anxiety scale, Thematic Apperception Test (TAT), self-concept Q-sort, and Wechsler Adult Intelligence Scale (WAIS). Congruence, empathy, and positive regard were inferred by raters and also by the patients' and therapists' responses to the Therapist Relationship Inventory (Barrett-Lennard, 1962).
The subjects were 15 hospitalized schizophrenic patients and their therapists. Five sessions were selected from each case: one session from the beginning, one session from each quarter of the case length (the 25% point, the 50% point, and the 75% point), and the last session. Two 4-minute segments were extracted from each session, with one segment randomly selected from the first third of the session, and the other segment randomly selected from the last third of the session. Thus a total of 150 segments were extracted and rated on patient level of problem expression, intrapersonal exploration, and manner of relating, and also on therapist congruence, empathy, and positive regard. The Therapist Relationship Inventory was administered at the third month of therapy. At the end of therapy each patient was assigned an outcome score.

The results indicated that changes in process level (problem expression, intrapersonal exploration, and manner of relating) over the course of therapy were not related to therapist conditions nor to outcome. Instead, it was found that outcome ratings were significantly correlated with the patients' overall mean levels of problem expression and intrapersonal exploration, with higher levels of these behaviors associated with greater outcome. Outcome ratings were also significantly correlated with therapists' levels of empathy, with higher levels of empathy being associated with greater outcome.

Van der Veen (1967a) gave some possible explanations for the absence of significant correlations between process change and therapist conditions and outcome. The interrater reliabilities (type not specified) were relatively low (.44 for problem expression, .42 for
manner of relating, and .58 for intrapersonal exploration), thus indicating inconsistencies in the use of the scales by the raters. He also suggested that "it is possible that the theory of linear patient movement oversimplifies the change process" and that "the lack of clear-cut results in this area suggests a need for caution in the conceptualization of the change process in the patient" (p. 298).

**Early Experiencing Scale Studies**

The initial EXP Scale validation studies began in the 1960s, at the same time that the Process Scale studies were occurring. A number of investigators focused on relationships between EXP Scale scores and a variety of diagnostic and outcome variables. These initial validation studies are reviewed below.

Tomlinson (1962) selected four of the seven strands of Rogers's (1959b) Process Scale, with the aim of operationalizing each into a separate rating scale. It was hoped that by treating each strand separately the raters would encounter a less complex task, would achieve greater interrater agreement, and the scales would become more sensitive to process differences. One of the scales was the EXP Scale as developed by Gendlin and Tomlinson (1962). The other three were the Personal Constructs, Problem Expression, and Manner of Relating scales. Tomlinson tested for differences on these scales between early and late sessions, more and less successful outcome, and neurotic and schizophrenic patients. The neurotic group consisted of 12 clients from the University of Chicago Counseling Center and two clients from the Stanford University Counseling Center; the
schizophrenic group consisted of 14 patients from the Mendota State Hospital in Wisconsin.

Two audio recordings of therapy sessions were collected from each of the neurotic and schizophrenic cases, with the recordings for the neurotic group selected from the second and second-from-last sessions, and those for the schizophrenic group from the fifth and fifth-from-last sessions. A 4-minute segment was randomly extracted from the first and last third of each selected session, and these segments were rated on all four scales by each of four raters. Pearson r interrater correlations for the four scales ranged from .44 to .74, with EXP correlations being .53, .58, .61, .61, .63, and .67 for all rater pair combinations. By using data from a large number of psychological tests, the cases were classified as having a more successful or less successful outcome.

Significant differences were found on the Problem Expression and Personal Constructs scales between the neurotic and schizophrenic cases, with the neurotics having higher scores. There were significant differences on the Personal Constructs and Manner of Relating scales for the therapy success X early-late interview interaction. The more successful cases began therapy at lower process levels than the less successful cases, but at the end of therapy the more successful cases had higher levels than the less successful cases.

There was an additional significant difference on the Personal Constructs scale for the therapy success X neurotic-schizophrenic interaction, with the less successful neurotic and schizophrenic cases having equivalent mean scores, but with the more successful neurotic
cases scoring higher than the less successful neurotics, and the more successful schizophrenics scoring lower than their less successful counterparts. No significant main or interaction effects were found on the EXP Scale, although the EXP patterns tended to parallel the above findings.

Kiesler, Mathieu, and Klein (1964) examined the effects of different audio tape segment lengths on rater reliabilities and EXP scores. The data for their study came from two tape recorded therapy sessions, one randomly selected from the first five sessions and the other from the last five, for each of 21 cases: 7 hospitalized schizophrenics, 7 neurotics from the University of Chicago Counseling Center, and 7 "normal subjects who met with a therapist for quasi-therapy sessions" (p. 351). A 2-minute segment was selected from each tape, then a 4-minute segment was built for each tape by adding 1 minute onto both sides of the 2-minute segment. This segment expanding technique was repeated twice more so that four segments were constructed for each session, with the segment lengths being 2, 4, 8, and 16 minutes, and with each segment forming the nucleus of the next larger segment. Four groups of raters (with four raters per group) then made EXP Scale ratings of the segments, with each rating group rating all the segments of a given segment length.

There were no significant differences in interrater reliabilities as a function of segment length; the range of reliability coefficients (Ebel interclass $r_{kk}$) was narrow and the reliabilities were relatively high (from .85 to .91). The rate-rerate reliabilities ranged from .81 to .93, with a median of .80. No significant
differences in rate-rerate reliabilities were found between segment length groups. There were no significant differences in the range or standard deviations of EXP ratings between segment length groups. There were also no significant differences in mode or peak means between the three diagnostic groups.

When the data from all three diagnostic groups were analyzed as a whole, a significant difference in mode EXP scores was found between early and late sessions, with the mean EXP level being lower for the later sessions. This finding is in the opposite direction of other studies finding differences between early and late sessions (Tomlinson & Hart, 1962; Walker et al., 1960). There was no difference found for EXP peaks between early and late sessions. Kiesler et al. (1964) concluded: "Apparently, modal rather than peak ratings provide the more sensitive index of process changes" (p. 354).

Perhaps the most important findings of this study were differences in both mode and peak EXP means between the different segment length groups. The general trend was for the longer segments to receive higher EXP ratings, although this trend was not perfectly linear due to the 4-minute group receiving the lowest EXP scores. Each of the segment length groups was significantly different from the others except for the difference in mode scores between the 8-minute and 16-minute groups.

Kiesler et al. (1964) suggested two possible reasons for the relationship between segment length and absolute EXP level:

One possibility is that longer segments provide a greater opportunity for the development and resolution of feeling themes necessary for higher stage ratings. An alternative
possibility is that the higher stages of the scale require more complicated decisions on the part of the rater which necessitate a longer exposure to the material. (p. 355)

It may also be that these results reflect confoundment due to utilizing a separate group of raters for each segment length, thus the differences between segment length groups may have been influenced by differences between raters. Yet another possibility is that differing segment lengths influence the degree of rater subjectivity involved in the rating task. For instance, longer segment lengths might allow raters to become more emotionally involved with the material of the segment and thus lead raters to project this emotional involvement and assign higher EXP ratings.

Schoeninger (1965) recruited 32 volunteers from an undergraduate psychology class to take part in an analogue study of "short-term therapy." The therapists were graduate students in clinical psychology. Schoeninger set out to test for the effects of client pretherapy experiencing training and therapist self-disclosure on client experiencing.

One half of the "clients" were given a pretherapy orientation which included instructions on how to focus on emotions (experiencing training) and the other one half received a pretherapy orientation without the experiencing training. Each client then participated in three 60-minute therapy sessions. With one half of the clients, the therapists were instructed to assume a "self-disclosing" style in which (in addition to reflection of feelings) they were to reveal information about their own personalities, hypotheses they had about the client, and their own feelings about the client-therapist.
relationship. For the other one half of the clients, the therapists assumed a "nondisclosing" role which primarily involved reflection of feelings, asking questions, and mild interpretations. After all three sessions were finished, each client completed a modified version of Barrett-Lennard's (1962) Relationship Inventory, which measured the client's perceptions of the therapist's qualities. One hundred and ninety-two 6-minute tape segments were generated by sampling the third and eighth 6-minute interval from each session. Four raters rated these segments and achieved Ebel $r_{kk}$ interrater reliabilities of .76 (modes) and .78 (peaks).

There were no significant main differences in EXP between the self-disclosing and non-self-disclosing conditions nor between the EXP training versus no EXP training conditions. There was, however, a significant interaction effect in which the clients who received the EXP training had higher EXP peaks, but only for the segments extracted from the early portion of the session. There was a significant difference in EXP between the early and late portions of the interview, with the earlier portions receiving higher levels of mode and peak EXP. There were also significant correlations between both mode and peak EXP during the second of the three sessions and the clients' ratings of the therapists' levels of comfort and empathy, and a significant correlation between mode EXP in the second session and the client's ratings of the therapists' genuineness.

Kiesler, Klein, and Mathieu (1965) investigated the influence of the location of the tape segment within the therapy session on EXP ratings. One session from each of 24 therapy cases (8 schizophrenic,
8 neurotic, and 8 normal) was selected for analysis. One half of the sessions from each diagnostic group were randomly drawn from the first five sessions, and the rest from the last five sessions. Each session was divided into five 8-minute segments and each of these 120 segments was then rated independently by four raters who achieved interrater reliabilities (Ebel $r_{kk}$) of .85 (modes) and .87 (peaks).

No significant main effects were found in regard to the five segment location groups; however, there was a significant interaction effect between segment location and diagnostic group. The neurotic cases began the interview and maintained significantly higher EXP levels than schizophrenics and normals, and also gained higher levels in each consecutive segment of the session. The schizophrenics and normals began the interview at approximately the same EXP level, but the normals then gained in EXP during the second 8-minute segment and then declined in EXP through the last three segments, while the schizophrenics dropped in EXP during the second segment but then increased in EXP during the last three segments. No differences were found between the 12 cases contributing to the early session group and the 12 which comprised the late session group.

Kiesler, Mathieu, and Klein (1967a) reanalyzed the data taken from the Kiesler et al. (1965) study with the aim of detecting any effects that the clients' or therapists' verbalization rates and speech patterns might have on EXP ratings. The segments were rerated using Saslow, Matarazzo, and Guze's (1955) interaction-chronograph rating method which assesses the proportions of client-therapist speech and silence during therapy interviews. The EXP ratings were
then compared to the interaction-chronograph ratings. Kiesler et al. (1967a) reported that:

(a) Pearson correlations between EXP and the IC variables revealed little evidence that EXP ratings are systematically influenced or biased by patient or therapist formal speech patterns; (b) when the covariance of each IC variable was partialled [sic] out, the EXP differences obtained for the original unadjusted EXP scores were not significantly altered. (p. 224)

These results indicate that EXP Scale scores are relatively independent of verbalization rates and reflect the quality, and not the quantity, of the client's verbalizations.

Ryan (1966) explored client characteristics and outcome as related to EXP, and also tested for differences in EXP at three points in the overall therapy process, rather than simply early-late differences. Ratings were made on 8-minute segments taken from the second, middle, and next to last interviews of each of 32 clients seen at the University of Illinois Counseling Center. The four raters achieved interrater reliabilities (Ebel $r_{kk}$) of .76 (modes) and .77 (peaks). Prior to the beginning of therapy and at the end of therapy each client was administered the 16 Personality Factors (16 PF), Gilbert Self-Interview, and the TAT as measures of the client's personality. Following the third session, and again at the close of counseling, each counselor completed a Personality Description Inventory which is a form used to rate the client on a number of dimensions such as motivation for therapy and severity of maladjustment. The Hunt-Kogan Movement Scale (Hunt & Kogan, 1950), a measure of client improvement, was completed by each counselor at the close of therapy.
The results indicated a number of significant correlations in regard to EXP and client characteristics. High EXP ratings at the middle point of therapy were generally indicative of clients who were initially less anxious (16 PF), remained in therapy longer, had fairly good self-concepts (Personality Description Inventory), and showed greater initial faith in psychotherapy (Personality Description Inventory). Higher EXP scores during the late session were indicative of heightened anxiety at the end of therapy (16PF), an unfavorable self-concept (Personality Description Inventory), and an orientation to fantasy (Personality Description Inventory). Somewhat contrary to these findings, high EXP during the late sessions was also related to initial low internal stress and to decreases in internal stress and conflict as measured by TAT ratings.

In regard to EXP and the therapy process, the EXP scores for the middle sessions were significantly and positively correlated with outcome as defined by scores on the Hunt-Kogan Movement Scale; however, there were no significant correlations between outcome and early and late EXP scores. A significant interaction effect was found when the sample was divided into more successful (n = 18) and less successful (n = 14) groups on the basis of Hunt-Kogan scores, with the more successful clients beginning with lower EXP scores in the early interview, then rising at the middle interview point, then dropping at the late interview. The pattern for the less successful clients was the reverse, with relatively high EXP scores in the early interview, low at the middle interview, and up again at the late interview. Although this interaction was statistically significant,
all of the means of the more productive and less productive groups at all three interview points were within one full EXP stage, thus indicating small conceptual differences between means and posing difficulties in interpreting the results.

Van der Veen (1967b) investigated the relationship between individual patients and therapists in regard to therapists' empathy and congruence and patients' levels of EXP and degree of problem expression. Through a unique arrangement at a psychiatric hospital, schizophrenic patients were given the opportunity to initiate therapy sessions with any of eight therapists whenever they felt the need. This allowed many patients to see many therapists concurrently. From the 25 patients seen using this procedure, the audio taped sessions of three patients were selected for analysis. Each of the three had seen the same five of the eight therapists on at least two occasions. Three 4-minute segments were extracted from each of two interviews for each patient (one segment randomly selected from each third of the interview), with the targeted sessions being "as close as possible to the first and fourth meeting for each pair" (p. 354). Two raters rated the therapists' levels of congruence and empathy, while two others rated the patients' degree of problem expression and EXP. A total of 90 segments were rated with correlations between the rater pairs being .46 for congruence, .55 for empathy, .46 for problem expression, and .58 for EXP. These relatively modest reliabilities indicate considerable variation in each rater's use of the scales.

Significant differences in client level of problem expression and EXP were found between patients and between therapists. There
was a significant interaction effect on EXP as a function of particular patient and therapist pairs. Also significant were differences in therapist congruence and empathy between therapists and between clients. Interpreted, this means that individual patients display typical process levels which are different from one another, and which are raised or lowered as a function of interacting with certain therapists, with some therapists consistently eliciting higher process levels than others. Similarly, individual therapists display typical congruence and empathy levels which are different from other therapists, and which are modified as a function of interacting with particular patients, with some patients consistently eliciting higher levels of empathy and congruence. An additional finding was that the amount of change in patient process levels from the first sampled session to the next was significantly and positively correlated with the therapist's mean level of congruence and empathy.

Tomlinson (1967) examined the relationship between outcome and process scores for 12 schizophrenic patients: six selected as clear examples of more successful cases, and another six as less successful. The outcome groups were determined by a combination of clinical judgments and an extensive battery of tests. Two 4-minute segments were systematically extracted from each of three interviews for each patient: the fifth, fifth from the end, and an intermediate interview. Four raters then rated each segment using the EXP Scale and three other process scales derived from Rogers's formulation: a problem expression scale, a personal constructs scale, and a relationship scale. Interrater reliabilities (Ebel $r_{kk}$) were .88 for
problem expression, .88 for personal constructs, .88 for relationship, and .89 for EXP.

The data were analyzed in regard to each scale separately and also with all four scales averaged as a composite process score. No main effects were found due to patient success or to interview location; however, the interaction of success and location was significant for the EXP, personal constructs, and relationship scales, and also with the composite process score. The patterns were parallel for all the measures, with the more successful patients beginning therapy at lower process levels than their less successful counterparts, but then gaining at the middle interview, and again at the late interview. The less successful patients had the reverse pattern: Although beginning at higher levels than the more successful patients, they declined to levels below the more successful group at the intermediate interview, and then declined further at the late interview. These results were different from Ryan's (1966) study of more and less successful neurotic clients at three interview points, perhaps indicating differences in process patterns for different diagnostic groups.

Kiesler (1971) investigated the relationship between outcome and EXP scores across the first 30 sessions for 26 neurotic clients and 12 hospitalized schizophrenics. The basic data consisted of previously recorded therapy sessions from the University of Illinois Counseling Center (initially used in a study by Hunt, Ewing, Laforge, & Gilbert, 1959) and the Mendota State Hospital in Wisconsin (initially used by Rogers, Gendlin, Kiesler, & Truax, 1967). Four-minute
segments were randomly extracted from the later one half of each of the 1,140 sessions sampled, and EXP ratings were made by two sets of raters; one set rated the segments of sessions taken from the Illinois study (neurotics), while the other set rated the segments of the sessions taken from the Wisconsin study (schizophrenics). These later ratings were made as part of the Rogers et al. (1967) study prior to Kiesler's (1971) investigation. Ebel \textit{k_k} interrater reliability for EXP mode ratings were .76 for the Wisconsin raters and .79 for the Illinois raters. The two groups of patients were divided into more successful and less successful outcome groups on the basis of scores obtained in the original Rogers et al. (1967) and Hunt et al. (1959) studies, with six schizophrenics and 13 neurotics in each of the more and less successful groups. EXP ratings were then compared for the schizophrenic versus neurotic groups, and the more successful versus less successful groups.

A significant difference was found between neurotic and schizophrenic patients, with the neurotics consistently receiving higher EXP ratings. These results should be interpreted, however, in light of the rater arrangements for these two groups. The differences may have been influenced by the fact that the ratings were made by two different groups of raters at two different points in time. The difference in EXP between the more and less successful groups was also significant, with the more successful patients (regardless of diagnosis) receiving higher ratings. There were no significant differences for the interaction of diagnosis and success. Contrary to hypothesized, there was no significant difference found in the amount
of EXP change across sessions between the more and less successful patients, nor between more successful neurotics and more successful schizophrenics. As with other studies using the EXP and Process scales (Schoeninger, 1965; Tomlinson, 1959, 1962; Tomlinson & Hart, 1962; Ryan, 1966), although the differences between groups were statistically significant, the absolute values of differences in EXP means between these groups were small (0.26 levels between more and less successful patients; 0.69 levels between neurotics and schizophrenics), thus making conceptual interpretations of these results difficult.

The above validation studies have established the EXP Scale (and Process Scale) as an instrument that can be reliably applied to interview materials, and that appears to measure an important dimension in psychotherapy. The findings of Kiesler (1971), Ryan (1966), Tomlinson (1967), Tomlinson and Hart (1962), van der Veen (1967a), and Walker et al. (1960) indicate that the dimension measured by the EXP Scale is closely associated with case outcome. Tomlinson (1962), however, failed to find significant differences between more and less successful outcome groups.

The relationship between EXP and early/late points in therapy is less clear. Tomlinson and Hart (1962) found that clients scored higher in process at later points in therapy than at earlier points; however, Kiesler et al. (1965) and Tomlinson (1962) failed to find such a relationship, and Kiesler et al. (1964) found the reverse to be true: clients were found to have decreased in EXP at later points in therapy. Some investigators have found distinct early-late or
early-middle-late patterns of EXP for clients judged to have more versus less successful outcomes (Ryan, 1966; Tomlinson, 1967; Walker et al., 1960); however, a greater number have failed to demonstrate this relationship (Kiesler, 1971; Tomlinson, 1962; Tomlinson & Hart, 1962; van der Veen, 1967a). The above suggests that the dimension measured by the EXP Scale tends not to change appreciably over time and that it is closely linked to case outcome. This has led some authorities (Gendlin, Beebe, Cassens, Klein, & Oberlander, 1968; Kiesler, 1971) to speculate that EXP might be a relatively enduring personality trait and/or that changes in EXP over time may not naturally accompany successful psychotherapy. If this is so then Rogers's (1958) process theory of therapy is in question. Another possibility, however, is that the EXP Scale tends to capture the trait dimensions of a client's experiencing, but is unable to adequately detect whatever "state" fluctuations might be associated with client change.

Experiencing and Session Productivity

A number of authorities have indicated that one of the more promising avenues for psychotherapy research involves the analysis of significant sessions and/or specific intrasession events (Auerbach & Luborsky, 1968; Elliott, 1983a, 1984; Gendlin, 1986; Mahrer & Nadler, 1986; Marmar et al., 1984; Orlinsky & Howard, 1967, 1975; Rice & Greenberg, 1984; Stiles, 1980). This trend is, in part, in response to Kiesler's (1966) delineation of the "uniformity myths" that exist in psychotherapy research: the erroneous beliefs about psychotherapy
processes as being largely homogeneous across clients, therapists, sessions, and intrasession interactions. The current movement from global process-outcome studies to microprocess studies also reflects attempts to generate research results which more meaningfully translate into psychotherapeutic practice (Elliott, 1983a). Whereas the earlier process-outcome EXP studies were aimed at establishing the EXP Scale as a valid instrument, studies which aim at identifying the EXP characteristics of productive sessions and/or significant within-session events aid in bringing research results into closer proximity of the practicing psychotherapist's moment-to-moment perspective. The EXP Scale has been used to examine this within- and between-session productivity in a number of studies (Auerbach & Luborsky, 1968; Elliott, 1983b; Elliott et al., 1982; Greenberg, 1980, 1983; Greenberg & Rice, 1981; Pollak, 1973). These studies are reviewed below.

Auerbach and Luborsky (1968) investigated differences on a variety of patient and therapist variables between sessions rated as better and poorer. Each of 15 psychoanalytic-oriented psychotherapists tape recorded sessions with two patients. The recordings of two successive sessions from the early phase of each therapy (usually Sessions 3 and 4) were presented to three raters who rated the entire session using 24 rating scales, 12 which addressed the therapist's behavior and another 12 which focused on the patient, one of which was the EXP Scale. The interrater reliabilities ranged widely from a low of -.13 to a high of .76, with reliabilities for EXP being .48, .22, and .42 for all three rater pair combinations. After the
sessions were rated, 10 sessions were selected as examples of better
hours and 11 sessions were classified as poorer on the basis of
ratings made on one of the therapist variables: "Therapist Responds
Effectively to Patient's Main Communication." These two groups were
then used to test for differences on the remaining 23 patient and
therapist variables.

A number of therapist variables were found to significantly
differ between the better and poorer hours; they were: (a) therapist
skill, (b) therapist empathy, (c) therapist unconditional positive
regard, (d) therapist maturity and security, (e) therapist warmth,
(f) therapist creativity, and (g) therapist emphasis on the un-
conscious. No patient variables were found to differentiate between
the better and poorer hours.

There are a number of explanations for the wide ranging and
generally modest interrater reliabilities in this study. The ratings
were made on entire sessions, thus forcing each judge to summarize an
abundance of information with only a single score for each scale.
There were no transcripts for approximately one half of the sessions;
in these cases each rater made notes as the tape recording played,
and these notes formed the basis of the various ratings, thus intro-
ducing heightened opportunities for each rater's subjectivity to
interfere with the rating task. Also, it appears that the EXP Scale
was not applied in its usual form: Auerbach and Luborsky (1968)
reported that each variable was rated on a 5-point scale, but the EXP
Scale has 7 points.
It is curious that no patient variables were found to differentiate between better and poorer sessions, yet seven of the therapist variables did. The most plausible explanation for this finding is that these seven variables are highly correlated with the criterion variable (therapist responds effectively to patient's main communication), and may actually be components of a single underlying construct. Thus it may be that the seven therapist variables (plus the criterion variable) are not so much descriptive of better therapy hours as they are indicative of generally agreed-upon characteristics of good therapist behavior.

Pollak (1973) investigated the relationships between session productivity and client experiencing, perceptual concreteness, and amount of affect evidenced in the session. Ten "working" (productive), 10 "resistant," and 5 "middle range" sessions were selected for evaluation from a pool of 363 sessions taken from a psychoanalytic case study reported by Dahl (1972). The distinction between working, resistant, and middle range sessions was initially made by Dahl on the basis of a factor analysis of the content of each session. Four factors contributed to the analytic work score: anxiety, and talk about family, sex, and dreams. Two other factors were combined as a measure of resistance.

Two 8-minute audio tape segments were extracted from each of the 25 sessions and were rated for EXP, perceptual concreteness, and affect expression. Equal numbers of segments were extracted from the second, third, fourth, and fifth 8-minute blocks. The three EXP raters attained interrater reliabilities of .79 (modes) and .87
(peaks). The EXP ratings of the two segments from each session were averaged, thus producing one mode and one peak score for each session. A productivity score was calculated for each session by subtracting the session's resistance score from the session's analytic work score (already provided by Dahl, 1972). Correlation coefficients were then calculated between the productivity scores and the EXP, affect, and perceptual concreteness scores.

There were no significant correlations between productivity and EXP or affect expression. The Pearson correlation coefficients for productivity and EXP modes and peaks were .18 and .17, respectively. There was, however, a significant correlation (r = .83) between perceptual concreteness and productivity.

In discussing possible reasons for the failure of EXP to significantly correlate with productivity, Pollak (1973) pointed out that the EXP Scale has emerged from within client-centered theory, while the productivity ratings and the therapy itself were psychoanalytically based. Thus what is considered to be productive for client-centered therapy may not be productive in psychoanalytic therapy. Pollak also pointed out that the EXP scores were all within a limited range, regardless of whether the session was a working or resistant one. This suggested that perhaps the EXP Scale was measuring the patient's habitual or trait EXP, and not whatever microscopic deviations might be associated with the patient's working or resistance. Pollak suggested that "because the EXP Scale was not constructed to detect movement over small intervals, the Scale cannot discriminate exceedingly small but significant experiential change
over sessions" (p. 98), and that the failure to find a relationship between EXP and productivity may have been due "to the inability of the Scale to detect small experiential change" (p. 98).

Greenberg (1980, 1983) and Greenberg and Rice (1981) used the EXP Scale to describe significant events in Gestalt therapy. In examining the resolution of intrapsychic conflicts or splits, Greenberg (1980) had raters rate segments extracted from nine sessions (three sessions from each of three clients) where a resolution of a split had allegedly occurred. The Gestalt two-chair method was used in each of these instances, which requires the client to converse from both sides of the split by imagining the antagonist in one chair, conversing with it, and then changing to the opposite chair and repeating the process from the opposing perspective.

In addition to EXP, ratings were made of the client's voice quality (Rice & Wagstaff, 1967), which is classified as externalized, limited, focused, or emotional. Focused and emotional voice quality has been associated with positive outcome (Rice & Wagstaff, 1967). The tape recordings were divided into 2-minute segments and were then rated, with ratings made of the client's statements while in the "experiencing" chair and also while in the "other" chair, thus producing two sets of scores per segment. The EXP scores consisted of the peaks, not the modes, for each segment. The interrater reliabilities (Pearson r) for EXP ranged from .72 to .81.

The data were examined with the intent of identifying any trends in EXP and voice quality in the segments prior to, during, and after the alleged split resolution. In regard to EXP peaks, the
experiencing chair ratings tended to be at Level 4 or above, while the other chair ratings were initially lower but later increased to levels equivalent to the experiencing chair. The client voice qualities while in the other chair also tended to shift from externalized during the segments leading up to the split resolution, to focused and emotional at the "merging" point (where EXP in the other chair raised to the level of the experiencing chair). Greenberg (1980) suggested that two phases were evident: "a pre-resolution phase, prior to the increase in the other chair, and a resolution phase in which both chairs tend to increase in depth of experiencing" (p. 146). This pattern, however, was less evident for one of the three clients.

A limitation of this study stems from the sole use of resolution events as data sources. Having not included a comparison group (for example, a group of nonresolution events), there is no way of knowing if the EXP and voice quality patterns are truly unique to instances of split resolution. Another limitation of this study is that no mention was made of the criteria used to select specific sessions for analysis.

Apparently using the same three subjects as the above 1980 study, Greenberg and Rice (1981) examined the effects of Gestalt two-chair interventions versus empathic interventions on EXP and voice quality. Each client was seen for 12 sessions. For each client, in three of the sessions (randomly predetermined) the therapist responded to client "splits" by introducing the Gestalt two-chair technique, and in another three sessions (randomly predetermined) the therapist "actively empathized with the presented split" (p. 33). An
example of a client split would be the statement "I don't want to do this but I feel I have to" (p. 32). The Gestalt events and empathy events were divided into several 4-minute segments and these segments were rated on peak EXP and voice quality. Greenberg and Rice were unclear if an "event" meant an entire session or a portion of a session, if all the segments which comprised a session were rated or if ratings were made only on specific segments, and if the number of segments rated for each type of event varied or were held constant. The interrater reliabilities (Pearson r) for EXP peaks ranged from .72 to .81. The dependent variable was the number of segments which evidenced peak EXP levels of five or above.

Significant differences were found between the Gestalt events and the empathy events in the number of segments with peak EXP levels of five or above, with the Gestalt events having the greater frequencies. No significant differences were found in client voice quality. These results seem to indicate that two-chair interventions are effective in increasing client EXP peaks at points where a split emerges; however, this conclusion should be accepted cautiously in light of the above mentioned lack of clarity in method.

In a third study of intrapsychic conflict resolution using the Gestalt two-chair method, Greenberg (1983) compared 14 instances of conflict resolution with 14 nonresolution events. Six therapists trained in Gestalt therapy contributed the tape samples for analysis. An instance was considered a resolution event if both client and therapist rated the session a 5 on a 7-point conflict resolution scale, if the client reported a reduction of at least 5 points on the
Target Complaints Discomfort Box Scale (Battle, Imber, Hoehn-Saric, Stone, Nash, & Frank, 1966), and if the client had reached a peak of 6 on the EXP Scale. The instances which did not meet the above criteria were classified as nonresolution events. Each event was segmented into many two-page transcript units, and each unit was given two EXP ratings; one modal rating for each "chair" in the two-chair dialogue. The two EXP raters achieved an interrater reliability (Pearson r) of .87. The clients' two-chair dialogues were also rated on the client Voice Quality system and on the Structural Analysis of Social Behavior (SASB) system (Benjamin, 1974), which is a classification system for measuring dialogue quality.

Each resolution event was divided into three phases: an opposition phase, a merging phase, and an integration phase. The opposition phase began at the point where the two-chair dialogue began and continued until "affiliation" was expressed by the two chairs as defined by scores on the SASB. The merging phase began at this point and continued until the client reached an EXP level of 6, which then served as a marker for the beginning of the integration phase. The integration phase lasted until the two-chair exercise was completed. The scores on each of the transcript units which comprised each phase were then averaged together to produce overall scores for each phase. The nonresolution events, by definition, did not contain the SASB and EXP scores necessary for clear division into phases. Instead these events were sectioned into phases by superimposing the proportions of each phase typical of the resolution events over the course of each nonresolution event. Thus, the first 60% of each nonresolution
transcript was labeled as the opposing phase, the next 30% as the merging phase, and the final 10% as the integration phase. These "phases" of the resolution and nonresolution groups were compared for EXP, voice quality, and scores on the SASB.

When the two groups were examined for the opposition phase, no differences were found for EXP, voice quality, or scores on the SASB. During the merging phase, significant differences were found between the resolution and nonresolution events on EXP, voice quality, and "affiliation" on the SASB, with higher levels of these variables found in the resolution events. Within the resolution group, there were significant differences in EXP and focused-emotional voice between the opposing and merging phases, with the merging phase receiving higher ratings. Additionally, there were significant differences in EXP between the two chairs during the opposing phase, but not during the merging phase. This suggests that the process of intrapsychic conflict resolution, when using the two chair method, involves a process of emotional deepening, with one side of the split beginning at a more shallow emotional level, but then deepening to a level equivalent to that of the other side of the split.

Greenberg (1983) did not report any statistical tests between the integration phase and other phases, or between the two groups during the integration phase. However, examination of the table of means and standard deviations indicates that differences existed that may have been significant. Perhaps one reason why these comparisons were not reported is because EXP scores were used as a criterion for determining resolution events and the integration phase, and thus
testing for differences on EXP would have been unnecessary.

In a comprehensive analysis of 10 sessions from a single case, Elliott et. al., (1982) applied 41 process rating scales to video tape recordings of therapy events judged by the client and therapist as helpful or nonhelpful. The therapist in this case was classified as generally psychodynamic in orientation. Immediately following each session the client and therapist independently completed a Therapy Session Report (Orlinsky & Howard, 1975), which is a measure of each participant's impressions of the session. The client and therapist also reviewed the video tape of the session and rated each therapist response on a 9-point helpfulness rating scale. The therapist and client each independently selected what they considered to be the two most helpful therapist responses. All therapist responses which scored in the "neutral" or "hinder"ing" end of the scale were selected as examples of nonhelpful events. Thirty seconds of "presegment" and "postsegment" client talk were then edited onto the segment of therapist talk, thus producing the following sequence for each taped event: a 30-second presegment of client talk, followed by the therapist intervention, followed by a 30-second postsegment of client talk. These samples were then rated by 13 raters on a battery of process rating scales which included 25 therapist scales and 16 client scales, one of which was the EXP Scale. Interrater reliabilities ranged from .47 to .92. No reliabilities were reported for the EXP Scale, but because the ratings were made by two of the scale's authors (Klein and Mathieu-Coughlin), high if not perfect agreement could be assumed. When all 10 sessions had been completed, the
therapist and client reexamined the video tapes and retrospectively rated the therapist's responses.

The results were numerous and complex. Generally, the therapist qualities or behaviors most associated with the therapist response helpfulness ratings were nonverbal expressiveness, empathy, "positivity," and a "helpful experiencing" factor (direct reference to the client's experiencing process and modeling of the experiencing process). The behaviors associated with low helpfulness ratings were the therapist's closed questions and disagreement with the client, and the client's requests for help and disagreement with the therapist. Peak EXP of the presegment portion was the only client variable which correlated with the client's postsession evaluations of the therapist's effectiveness. Presegment EXP (peak) and client "agreement" were the only client variables significantly related to the therapist's postsession ratings of therapist helpfulness. These correlations were not apparent for the client's and therapist's retrospective (after the 10 sessions) evaluations. The client's EXP peaks during the postsegments were significantly correlated with the therapist's levels of helpful experiencing, depth, and empathy during the period preceding the postsegment.

In regard to the ratings of each session's effectiveness (from the Therapy Session Report), the helpful experiencing (of the therapist) factor was significantly correlated with the client's ratings of session effectiveness. The correlations between the client's ratings of session effectiveness and client EXP and agreement approached significance (p < .10). The therapist's use of closed
questions was negatively correlated with the therapist's ratings of session effectiveness. No relationship between EXP and the therapist's ratings of session effectiveness was found.

In a microscopic process analysis of a "significant event" taken from a single session, Elliott (1983b) applied 12 therapist and 5 client process measures (one of which was the EXP Scale) to each client and therapist response contained on a transcript and tape segment of the event. The event was selected from a population of 200 other recorded therapy events because it contained a therapist response which was rated highest on helpfulness by both client and therapist.

Elliott's (1983b) analysis of the event was qualitative, using the quantitative ratings as anchors for the ongoing description of the event. The client's EXP levels began relatively high at Levels 4 (mode) and 5 (peak), then dropped and maintained primarily at Level 3 through the majority of the event, but then climbed to Level 6 at the point of the client's last statement. Other process measures (e.g., voice quality) also tended to indicate client-therapist interactions which "deepened" as the interaction progressed. Due to the lack of a comparison event, it is difficult to arrive at any firm conclusions based on this study regarding the role of EXP during significant therapy events. Nevertheless, this study was an important one in regard to the use of the EXP Scale, for Elliott had applied the scale to each of the client's statements and thus demonstrated new avenues of application.
The Problem and the Research Focus

Session Productivity

The most consistent differences in EXP have been found in macroscopic outcome studies of more and less improved outcome groups (Kiesler, 1971; Ryan, 1966; Tomlinson, 1959, 1967; Tomlinson & Hart, 1962; van der Veen, 1967a; Walker et al., 1960). These studies have revealed that more successful clients evidence higher EXP levels than less successful clients. Recent studies have tended to move from these macroscopic outcome studies to detailed investigations of microprocesses which occur within single sessions. Some of these recent studies have demonstrated microprocess trends in EXP which correspond with alleged significant change events (Elliott, 1983b, Greenberg, 1980, 1983; Greenberg & Rice, 1981).

What has been called the "productive session" or "good hour" can be conceptualized as an intermediate point between significant intrasession events and positive outcome. It might be a reasonable prediction, then, that differences in EXP would also exist between sessions judged to be more and less productive. Despite indications that EXP is related to client productivity at the micro and macro ends of this process continuum, there has yet to emerge any evidence which establishes a link between EXP and session productivity. Those who have compared more and less productive sessions have failed to find significant differences in EXP (Auerbach & Luborsky, 1968; Elliott et al., 1982; Pollak, 1973). Clinical consensus suggests, however, that real differences in productivity exist between
sessions, and that clients and therapists can recognize when a session has seemed productive or not. Empirical support for this notion can be found in Stiles's (1980; Stiles & Snow, 1984) work, who has demonstrated that both clients' and therapists' judgments of session impact do tend to vary from one session to the next.

Perhaps one reason that Auerbach and Luborsky (1968) and Pollak (1973) did not detect differences in EXP associated with session productivity was because the ratings of productivity were made by people other than the actual therapy participants. It may be that clients and therapists are in the unique position of more accurately judging the impact and value of a given session, or that their judgments of session productivity are more closely associated with the dimension(s) measured by the EXP Scale. The results of the Elliott et al. (1982) study hint that this may be the case: Although not statistically significant, the correlation between EXP and the client's ratings of session effectiveness did approach significance ($p < .10$). It may be, then, that a relationship does exist between EXP and the therapy participants' judgments of session productivity. Perhaps the most obvious methodological limitation of the Elliott et al. study which may have limited the strength of the above correlation was the use of 30-second segments of client speech as scoring units. This is an unusually short amount of time; most EXP studies use segments between 4 and 8 minutes long. The use of such a brief unit may have limited the range of EXP measured. The question could legitimately be asked: Might differences in EXP be detected between more and less productive sessions if productivity was assessed by the
actual therapy participants and if the scoring units were a standard length? The present study was designed, in part, to investigate this possibility by comparing more and less productive sessions as assessed by both clients and therapists, and by utilizing 8-minute segments as the scoring units. The following research question is asked in the present study in regard to EXP and session productivity: Are there any differences in EXP between more productive and less productive sessions?

Data Form

In addition to investigating possible EXP differences between more and less productive sessions, a second focus in the present study addresses some issues which arise from the conventional method of summarizing EXP data: (a) confused interpretations which arise from a generally narrow band of data, (b) the possibility that the mode score may be measuring a trait or personality dimension and not a state of process dimension, (c) the lack of a clear operational definition of the mode score, (d) difficulty in testing the assumption that EXP and therapeutic productivity are parallel continua, and (g) new directions in the use of running ratings and some advantages and limitations which are associated with these ratings. Below is a delineation of each of these issues, followed by a proposed remedy for these problems which takes the form of an alternative method of summarizing EXP data, which is then followed by three additional research questions regarding these conventional and alternative methods of summarizing EXP.
Narrow Band of Results

In many of the EXP and Process Scale studies the differences between group means have been relatively small even though the differences were statistically significant. A number of investigators have reported significant differences between means of from one to two scale levels (Kiesler et al., 1964, 1965; Tomlinson, 1967; van der Veen, 1967b; Walker et al., 1960), and more have found significant differences between means of less than one full scale level (Kiesler, 1971; Ryan, 1966; Schoeninger, 1965; Tomlinson, 1959, 1962; Tomlinson & Hart, 1962; Walker et al., 1960). This narrow band of results often poses interpretation problems: Because the scale utilizes discrete operational anchors for determining values, and because these values are then grouped and averaged, the final group means are typically in the form of fractions of a level. These minute differences are difficult to conceptually interpret, especially when the differences are less than one full scale level. For example, Kiesler (1971) reported that the more successful clients attained significantly higher EXP scores ($\bar{x} = 2.37$) than less successful clients ($\bar{x} = 2.09$). If "2" on the EXP Scale means that the content of the client's communications consists of "external events" and "behavioral and intellectual self-description" (Klein et al., 1969, p. 64), and "3" indicates "personal reactions to external events; limited self-descriptions; behavioral descriptions of feelings" (p. 64), then it is extremely difficult to arrive at conclusions on what "2.09" or "2.37" mean, and what the conceptual
difference between these two points is. In reference to this problem of less than one level differences, Gendlin et al. (1968) have acknowledged that "even when statistically significant, the increase was so small as to be psychologically insignificant" (p. 225).

Gendlin et al. (1968) have suggested that one way to deal with this problem is to further refine the scale so that detailed judgments about fractions of a level can be made. Although this idea seems to have intuitive merit, there has yet to be developed such a system, perhaps due to the complexities that fraction-of-a-level judgments would require. A simpler way of remedying this problem would be to maintain the scale as it is, but then to not average together the EXP levels of a number of segments. This way each level remains discrete and fractions of a level are not created. Each scale level would be treated as a category, and the frequency of occurrence and duration of each scale level evidenced within the segment would be the quantified dependent variables.

**Trait Versus State**

Rogers's (1958) original theoretical formulation regarding the process of psychotherapy predicted that successful clients would tend to exhibit progressive changes along the process continuum as time goes on. Research with both the Process and EXP Scales, however, has tended to disconfirm this notion. Although some investigators have reported changes in process and EXP over time (Ryan, 1966; Tomlinson, 1967; Walker et al., 1960), more have failed to demonstrate this phenomenon (Kiesler, 1971; Tomlinson, 1962; Tomlinson & Hart, 1962;
van der Veen, 1967a). Kiesler et al. (1964) have even found overall higher levels of EXP at earlier points in therapy than at later points. This lack of firm evidence, combined with some fairly consistent results indicating that more and less successful outcome groups can be differentiated with the scale (Kiesler, 1971; Ryan, 1966; Tomlinson, 1967; Tomlinson & Hart, 1962; van der Veen, 1967a; Walker et al., 1960), has led some authorities to speculate that EXP may be a relatively enduring personality trait and that changes in EXP over time may not necessarily accompany successful therapy (Gendlin et al., 1968; Kiesler, 1971). If this is so, this might also explain why differences in EXP have not been found between more and less productive sessions. As Pollak (1973) stated:

The absence of statistical support in this research for a relationship between experiencing and productivity may have been due to the likelihood that level of experiencing is a deeply engrained trait that does not change as a function of therapy, or to the inability of the Scale to detect small experiential change. (p. 98)

It may be that both of Pollak's explanations are true: EXP may have both trait and state dimensions, and each person may have a typical manner of experiencing from which he or she deviates periodically. If so, then the problem may lie in the EXP mode score and its inability to detect these deviations. In effect, the mode may be primarily measuring the trait dimension of EXP, while at the same time missing the state dimensions. The peak score does appear to measure a degree of deviation from this alleged trait experiencing, yet it measures only one deviate: the highest level attained in the segment.
If it is true that the mode is neglecting the state dimensions of EXP, then this may also explain the lack of consistent findings regarding changes in EXP over time and also between more and less productive sessions. In order for this problem to be more fully investigated, an alternative method of summarizing EXP is needed that would detect minute changes in EXP from one segment to the next. The approach alluded to previously would be such a method. By summarizing EXP in terms of the frequency and duration of each EXP level occurring within the segment, a profile of the client's experiencing would emerge. This profile would provide information not only about the most pervasive level and the highest level attained, but also about all other levels evidenced (and not evidenced) within the segment.

The Mode Score

The mode is defined by Klein et al. (1969) as "the rating that characterizes the overall, general or average scale level of the segment or unit" (p. 65). Although this definition may make immediate intuitive sense, after some consideration it becomes increasingly unclear. Klein et al. (1969) do not specify in the Training Manual or elsewhere exactly what is meant by the "overall, general, or average" level. It may mean the level at which the client spends the most time in the segment, or it may mean the level which occurs with the greatest frequency, or it may mean both or neither of these. Whereas the Training Manual contains elaborate descriptions and definitions of each EXP level, a clear operational definition of the mode
is lacking. One is therefore left to subjectively arrive at conclusions as to the "overall, general or average" level. It would be reasonable to assume that raters translate the phrase "overall, general or average" however it makes sense to them, which may tend to vary from one rater to the next.

**Experiencing/Productivity Assumption**

Because higher levels of EXP have been associated with better therapy outcome, there exists the assumption that higher EXP levels are basically more therapeutic or productive than lower levels. Some recent studies such as Greenberg's (1983) study of intrapsychic conflict resolution or Elliott's (1983) analysis of a significant therapy event utilize shifts on the scale as definition points for these alleged resolutions or significant events. By doing this, they are demonstrating an acceptance of the notion that certain levels of EXP are indeed indicative of productive therapy behavior, and perhaps also that the EXP continuum parallels a hypothetical productivity continuum. So far, however, there is no evidence that might indicate if each successive level is indicative of increasingly productive therapy behavior, nor if specific levels tend to be more often associated with productive therapy behaviors.

Perhaps the most apparent reason for this lack of evidence is the inherent difficulty in arriving at it via the conventional method of summarizing EXP data. With the conventional method there is no way to quantify each level which occurred (or did not occur) within the segment, and thus comparisons of each level between productive
and nonproductive events cannot be made. It may be that higher (or lower) durations or frequencies of occurrence of certain levels may accompany more or less productive sessions, yet the conventional method may tend to obscure these relationships because the mode score is a global estimate which is then typically averaged into a group of other mode scores. Consider the following hypothetical example: If more productive sessions tended to have higher frequencies of Level 4 statements, and Level 4 also tended to be accompanied by higher frequencies of Level 2 statements, then a certain number of more productive segments may carry modes of 4, and others Level 2, which when averaged together might then produce a group mean of 3. If less productive sessions then tend to have modes of 3, then this relationship between EXP and session productivity would remain obscured by virtue of the data form alone. A method that would highlight the relative occurrence of each level in segments judged to be more and less productive might then shed light on possible relationships between each EXP level and productive in-session behaviors.

**Running Ratings**

As mentioned earlier, a current direction in psychotherapy research is to more closely examine various in-session microprocesses. This approach has been taken with the EXP Scale (Elliott, 1983b; Elliott et al., 1982; Greenberg, 1980, 1983; Greenberg & Rice, 1981; Pollak, 1973), and this has led some to use statement-by-statement ratings ("running ratings") as a method of describing microscopic changes in EXP (Elliott, 1983b; Klein, Mathieu-Coughlin, & Kiesler,
in press; Mathieu-Coughlin & Klein, 1984). This new direction may be the beginning of an answer to Pollak's (1973) criticism that "the Experiencing Scale cannot discriminate small experiential changes and needs further refinement" (p. 110).

Whereas the use of running ratings has the advantage of capturing the fluctuations inherent in a given segment, there is currently no method available that would allow for grouping of running ratings from a number of segments, which would then allow for comparisons with other groupings. The advantage of using mode and peak scores is that they can be grouped and compared to other groups. The disadvantage of the mode is that, as explained earlier, it tends to overlook the fluctuating nature of EXP. What is needed is a method of summarizing EXP that would capitalize on the advantages and minimize the disadvantages of both running ratings and mode-peak ratings. Such a method would capture the within segment variability while at the same time allowing for the grouping and comparing of data. Klein et al. (in press), in reference to running ratings as a method of providing detailed EXP profiles, have stated: "We think this level of detail is possible, but will require more study and development of the training, scoring, and data analysis procedures" (p. 20).

The Present Research Focus

In the present study an alternative method of summarizing data is proposed that allows for grouping of running rating scores and thus comparisons between groups. This alternative method also captures whatever EXP variability exists in a given segment and thus
addresses whatever state dimensions of EXP might exist. The method involves two steps: First, raters assign EXP scores to each client statement; second, raters then measure the length of time (in seconds) elapsed during each client statement. The EXP levels themselves retain their status of having numbers attached to them, but the numbers are used for identification purposes only (not for quantification). Mathematical averaging is not attempted with the level labels, thus eliminating the problem of confused interpretation of results. Instead, the quantified dependent variables are the duration and frequency of occurrence of each EXP level evidenced within the segment. The segment thus receives one frequency score and one duration score for each of seven levels. Both frequency and duration are ratio scale measurements, have equal intervals between points, and duration has the advantage of being a continuous data form (data in fractions of a second are both possible and meaningful). Most statistical procedures may be legitimately performed with these data forms, and results can be more meaningfully interpreted. With this method it would also be possible to examine whatever patterns and relationships exist between each EXP level and more and less productive sessions.

This alternate method requires a slight conceptual modification. An assumption behind the conventional method is that the mode rating is representative of the client's overall experiencing within the segment. Put differently, the entire segment is assumed to be primarily of a certain level. The alternative method requires that each segment be regarded as containing a potential array of EXP levels.
The client is seen as continually moving in and out of different EXP levels. This is similar to Horowitz's (1979; Horowitz, Marmar, & Wilner, 1979) concept of states: "a recurrent pattern of experience and of behavior that is both verbal and nonverbal" (Horowitz et al., 1979, p. 92). The use of total seconds duration as a dependent variable is also similar to one of the methods used by Horowitz (1979; Horowitz et al., 1979; Marmar et al., 1984) to analyze a patient's states and state transitions in psychotherapy. Although this method was developed from within psychoanalytic theory to describe states different from EXP, the method itself offers a number of advantages. With Horowitz's method each of a number of client states within the interview segment are identified by raters and subsequently measured for seconds duration. The total seconds for each state is then used to describe the nature of the segment. Comparisons between different session segments and their respective states may then be made. This approach of summarizing the quality of a segment in terms of seconds duration has also been successfully utilized by Matarazzo, Wiens, Matarazzo, and Saslow (1968) as a means to analyze patient silence/speech rates within and between sessions.

Henceforth the term "states" will be used in place of "level" whenever EXP is discussed in the context of this alternative method. Conceptualizing EXP levels in terms of states also brings the data derived from the EXP Scale into closer proximity of the theoretical spirit of experiencing. As Klein et al. (1969) explained:

Experiencing is a process, not a fixed trait. Its measurement deals with the quality of an individual's communicated experience of himself in a given aspect of his life at a
certain moment; that is, how immediate, personal, rich, changing, inclusive or expansive it seems. Although some individuals may show typical experiencing in certain settings or in discussing certain subjects, theoretically there is ample room for change and variation. (pp. 6-7)

As it is, the conventional method yields a description of the client's experiencing within a given segment as predominantly static and of one type (mode rating), with attention paid to only one deviation from this point (peak rating). A profile of the frequency and duration of each EXP level evidenced within the segment would more closely resemble the theoretical notion that experiencing is an ongoing and changing process; that a person's experiencing within any given segment might well be composed of a number of EXP states, with some states occurring more frequently and/or persisting longer than others.

In this study a comparison is made between the conventional and the alternative methods of summarizing EXP data. The data yielded by the two methods are compared in terms of their similarities and/or differences, and also in terms of their respective abilities to discriminate between more productive and less productive sessions.

Research Questions

Four general research questions are asked in the present study; these questions are listed below. Each general question is followed by a number of subordinate questions and null hypotheses. The questions and hypotheses are divided into two groups: those that address relationships between EXP and session productivity, and those that
address differences between the conventional and alternative methods of summarizing EXP data.

**Experiencing and Session Productivity**

**Question 1**

Are there differences in EXP between more and less productive sessions?

Since this investigator is particularly interested in differences between more and less productive sessions as judged by the therapy participants, the following subordinate questions apply:

Question (Q) 1A: Are there differences in EXP between sessions judged by clients as more and less productive?

Null Hypothesis (H₀) 1A: There are no differences in EXP between sessions judged by clients to be more and less productive.

Q 1B: Are there differences in EXP between sessions judged by therapists as more and less productive?

H₀ 1B: There are no differences in EXP between sessions judged by therapists as more and less productive.

Q 1C: Are there differences in EXP between sessions judged by both clients and therapists as more and less productive?

H₀ 1C: There are no differences in EXP between sessions judged by both clients and therapists as more and less productive.
Conventional and Alternative Methods

Question 2

Question 2 emerges directly from the previously mentioned criticism that the conventional mode score lacks a clear operational definition. The following is asked in order to gain a better idea of how Klein's et al. (1969, p. 65) definition of the mode as "the overall, general or average" level translates into the language of frequency and duration: Is there a relationship between the conventional mode rating and the alternative frequency and duration ratings?

Q 2A: Is there a relationship between the mode rating of a segment and the EXP state with the highest frequency of occurrence within the segment?

H₀ 2A: There is no relationship between the mode rating of a segment and the EXP state with the highest frequency of occurrence within the segment.

Q 2B: Is there a relationship between the mode rating of a segment and the EXP state with the greatest total seconds duration within the segment?

H₀ 2B: There is no relationship between the mode rating of a segment and the EXP state with the greatest total seconds duration within the segment.

In addition to the frequency and total duration of each EXP state, it may be that the mode is related to the EXP state with the greatest average duration (total seconds duration/frequency), hence

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the following question:

Q 2C: Is there a relationship between the mode rating of a segment and the EXP state with the greatest average seconds duration within the segment?

H_0 2C: There is no relationship between the mode rating of a segment and the EXP state with the greatest average seconds duration within the segment.

**Question 3**

Which combination of conventional and alternative variables best discriminates between more and less productive sessions?

Q 3A: Which combination of conventional and alternative variables best discriminates between sessions judged by clients as more and less productive?

Q 3B: Which combination of conventional and alternative variables best discriminates between sessions judged by therapists as more and less productive?

Q 3C: Which combination of conventional and alternative variables best discriminates between sessions judged by both clients and therapists as more and less productive?

Significance tests are not used to answer Questions 3A, 3B, and 3C, thus null hypotheses are not formulated here.

**Question 4**

Is there a difference between the conventional and alternative methods in ability to discriminate between more and less productive
sessions?

Q 4A: Is there a difference between the conventional and alternative methods in ability to discriminate between sessions judged by clients as more and less productive?

$H_0$ 4A: There is no difference between the conventional and alternative methods in ability to discriminate between sessions judged by clients as more and less productive.

Q 4B: Is there a difference between the conventional and alternative methods in ability to discriminate between sessions judged by therapists as more and less productive?

$H_0$ 4B: There is no difference between the conventional and alternative methods in ability to discriminate between sessions judged by therapists as more and less productive.

Q 4C: Is there a difference between the conventional and alternative methods in ability to discriminate between sessions judged by both clients and therapists as more and less productive?

$H_0$ 4C: There is no difference between conventional and alternative methods in ability to discriminate between sessions judged by both clients and therapists as more and less productive.
CHAPTER III

METHOD

Subjects

The subjects consisted of six client-therapist pairs from a counseling center at a large Midwest university. The six pairs consisted of six clients who met with a total of four therapists, with two of the therapists meeting with one client each, and the other two meeting with two clients each. Initially, five therapists and seven clients were identified as participants; however, shortly after the beginning of the project one of the client-therapist pairs terminated therapy and this pair was therefore eliminated from the pool.

Therapists

Of the participating four therapists, two were male, two were female; one was a predoctoral intern and the others were doctoral level psychologists with 1, 2, and 21 years of postdegree experience (see Table 1). Their theoretical orientations were varied, but all tended to report client-centered and/or psychodynamic influences. Three of the therapists received their doctoral training in counseling psychology, and the fourth in clinical psychology (Table 1).

All four therapists received a brief orientation to the general purpose of the study. The therapists were aware that the
<table>
<thead>
<tr>
<th>Therapist</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of experience</td>
<td>1</td>
<td>21 intern</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Degree program</td>
<td>Counseling Psychology</td>
<td>Clinical Psychology</td>
<td>Counseling Psychology</td>
<td>Counseling Psychology</td>
</tr>
<tr>
<td>Rank-order of typical experiential focus</td>
<td>1. Feelings</td>
<td>1. Feelings</td>
<td>1. Feelings</td>
<td>1. Feelings</td>
</tr>
<tr>
<td></td>
<td>3. Thoughts</td>
<td>3. Thoughts</td>
<td>3. Thoughts</td>
<td>3. Thoughts</td>
</tr>
<tr>
<td>Rank-order of typical temporal focus</td>
<td>1. Past</td>
<td>1. Present</td>
<td>1. Present</td>
<td>1. Present</td>
</tr>
<tr>
<td>Rank-order of typical content focus</td>
<td>1. Relationships with others</td>
<td>1. Experience of self others</td>
<td>1. Relationships with others</td>
<td>1. Experience of self others</td>
</tr>
<tr>
<td></td>
<td>2. Relationship with therapist</td>
<td>2. Relationships with others</td>
<td>2. Experience of self others</td>
<td>2. Relationships with others</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Experience of situations</td>
<td>3. Relationship with therapist</td>
<td>3. Experience of situations</td>
</tr>
</tbody>
</table>
investigator would be studying differences between more and less productive sessions, between two methods of summarizing data derived from sessions, and that the EXP Scale was to be used as the primary dependent variable. None of the therapists were familiar with the EXP Scale.

Selection of Clients

Each therapist was asked to participate with at least one of their clients. The therapists were given the option of selecting the clients they wished to participate with, as long as the clients met the following criteria: (a) the client was working on personal issues (not primarily vocational or educational), and (b) there were no plans to terminate therapy within 8 weeks of the beginning of the project. It was assumed that this relatively wide latitude in client selection would enable each therapist to select appropriate clients based on various therapy-related considerations (e.g., each client's trust level, security in the therapeutic relationship, tolerance for intrusion, appreciation of clinical research, etc.). The primary limitation of this method was that it was not random and it may have skewed the character of the subject pool.

Clients

An outline of the basic client characteristics is found in Table 2. They consisted of one male and five female university students, three undergraduate and three graduate students ranging in age from 20 to 36 years old. The initial presenting complaints were wide
Table 2

Client Characteristics

<table>
<thead>
<tr>
<th>Client</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Age</td>
<td>23</td>
<td>31</td>
<td>20</td>
<td>27</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td>Year in school</td>
<td>senior</td>
<td>Grad: M.A.</td>
<td>Junior</td>
<td>Grad: Ph.D.</td>
<td>Senior</td>
<td>Grad: M.A.</td>
</tr>
</tbody>
</table>
ranging but most tended to report depression or anxiety, and/or deficits in self-esteem and interpersonal relationships. At the close of treatment all reported reductions in a number of symptoms/complaints; however, four of the six also reported near equal numbers of increases in other symptoms/complaints. The remaining two reported clear reductions in symptom/complaint severity with minimal or no increases in additional symptoms/complaints (Appendix A).

Procedure

Approval of the research project was attained from the Human Subjects Committees at Western Michigan University and at the site of data collection. The clients who were identified as potential participants were then briefed on the general nature of the study and informed consent was attained (Appendix B). At the beginning of therapy each client completed a pretherapy symptom measure which is routinely used at the counseling center (Appendix C). This same measure was administered again at the close of treatment.

Each therapist then made audio tape recordings of every session that occurred with that particular client(s) within a 14-week block. The clients were each at different points in therapy at the time that the study was conducted; the sessions for Clients A and B came from a middle portion of therapy, and the sessions for Clients C, D, E, and F came from the end portions (Table 3). The number of sessions recorded by each client-therapist pair ranged from 5 to 12, with the mean being 8.67 sessions (Table 3).
Table 3
Temporal Location of Taped Sessions

<table>
<thead>
<tr>
<th>Client</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sessions with therapist</td>
<td>28</td>
<td>24</td>
<td>13</td>
<td>30</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>Number of sessions taped</td>
<td>11</td>
<td>12</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Began taping at session number</td>
<td>13</td>
<td>4</td>
<td>7</td>
<td>18</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Ended taping at session number</td>
<td>23</td>
<td>15</td>
<td>13</td>
<td>27</td>
<td>22</td>
<td>24</td>
</tr>
</tbody>
</table>

At the close of each session, each therapist and client independently completed a brief questionnaire designed to identify their perceptions of the session's productivity. The questionnaire was adapted from Stiles's (1980) Session Evaluation Questionnaire (SEQ). A number of postsession questionnaires are currently available (Greenberg & Rice, 1981; Orlinsky & Howard, 1975; Rand, 1979; Stiles, 1980). Among them, Orlinsky and Howard's is the most comprehensive, but is also extremely long and time consuming to complete. Because of the time demands placed on the therapists at a busy university counseling center, a postsession questionnaire was needed that could tap the participants' impressions of session productivity, and do so in a brief amount of time. Stiles has devised a relatively short postsession questionnaire which reliably measures the participants'
immediate impressions of the session's impact. The SEQ consists of 22 bipolar adjective pairs in a semantic differential format. By factor analyzing a large pool of scores, Stiles (1980) identified three factors that are measured by the SEQ: depth/value, smoothness/ease, and postsession positive feelings. The postsession questionnaire used in this study consisted of the five adjective pairs which comprise Stiles's depth/value factor (Appendix D). This depth/value factor appeared to this investigator to characterize a popular notion of session productivity. This modified version of the SEQ also had the advantage of being quickly completed by both therapists and clients. The scores derived from this questionnaire were then used to identify the more and less productive sessions. At the end of the course of therapy each therapist completed a therapist data sheet (Appendix E).

**Selection and Preparation of Segments**

At the end of the course of each treatment (or the 14-week time block, whichever came first), a number of session recordings were selected for segment extraction and rating, determined as follows:

1. The most productive and least productive sessions as determined by the clients' ratings on the postsession questionnaire. This was done by selecting the upper 25% and the lower 25% of the sessions for each client. The decision to use 25% as the cutoff was arbitrary. Using this approach each client typically contributed an equivalent number of more and less productive sessions, and the absolute number that each client contributed was proportional to the
number of sessions recorded. An exception to this was whenever two sessions had received tied session productivity scores, and both sessions fell on the border of the upper or lower 25%. In this case both sessions were included in the sample, even if it created a disproportionate amount of more or less productive sessions for that particular client. As examples of how this procedure worked, if a given client had rated a total of 12 sessions, then the top 3 and the bottom 3 rated sessions were selected. If the 25% point fell between two whole numbers (11 sessions multiplied by .25 produces 2.75 sessions), then the figure was rounded to the nearest whole number (3 more productive and 3 less productive sessions).

2. The most productive and least productive sessions as determined by the therapists' ratings on the postsession questionnaire. The upper and lower 25% of the sessions were selected using the above mentioned procedure as it applied to each therapist's ratings.

3. The most productive and least productive sessions as determined by both the clients' and therapists' ratings. The goal was to select those sessions which both participants considered to be more or less productive. To do this, a session was selected if it fell into an arbitrary upper 33% range or lower 33% range of both the client's and therapist's productivity ratings. This way only those sessions which received shared endorsements as more and less productive sessions were selected. This approach was preferred over summing the clients' and therapists' ratings and then selecting the upper and lower 25%, which may have produced artificial estimates of productivity.
Using the procedures outlined above, a total of 40 sessions were selected from a population of 52 sessions, with 14 more productive and 14 less productive sessions as judged by clients, 13 more productive and 13 less productive as determined by the therapists, and 9 more productive and 6 less productive sessions as determined by the clients' and therapists' combined judgments. (The discrepancy between the number of client-determined and therapist-determined sessions was due to instances of tied scores on the postsession questionnaire.)

An 8-minute segment was extracted from each of the selected sessions. Because EXP has been shown to vary in predictable trends within a given session (Kiesler et al., 1965), random sampling of interview segment location was not done. Instead, the segments were extracted from the beginning of the 30th minute until the end of the 37th minute. This 8-minute portion of the interview was selected because it typically yields a greater range of EXP levels than earlier points in the interview (Kiesler et al., 1965), and more often than not is typified by the client's "working" than by session "warm-up" or planning of future sessions. The segments were then given code numbers and were edited onto a master tape in random order. Transcripts of each segment were made and were proofread and corrected three times so that the transcripts were verbatim including assorted paralanguage (um, m-hm, etc.). All identifying material such as names, work sites, and course titles were deleted from the tapes and transcripts. Schoeninger, Klein, and Mathieu (1967) have found correlations of .82 and .89 (mode and peak ratings,
respectively) when correlating ratings of segments with and without therapists' verbalizations edited out. They concluded that it is unnecessary to edit out the therapist's responses; therefore, they were kept as part of the transcripts and tapes.

Raters and Rating

Kiesler (1970) has documented near equivalent accuracy and reliabilities for EXP ratings made by both clinically naive (undergraduate) and sophisticated (psychologist) raters. Based on this finding a decision was made to employ university students as raters. Four undergraduates, one male and three females, were selected to be trained in the use of the EXP Scale. Three of the trainees were psychology majors, one was a health education major, and all were also "paraprofessionals" at the counseling center. The procedure for trainee selection consisted primarily of an interview during which each individual was also asked to listen to benign tape segments of the clients' voices and indicate if he or she had recognized any of the voices. None of the interviewees recognized any of the voices. Upon selection each trainee was instructed on the importance of maintaining confidentiality and was required to sign a confidentiality oath (Appendix F). The trainees were kept completely blind to the nature of the study and to any information regarding EXP Scale correlates from previous studies.
Conventional Rating

The four individuals were taught to make the conventional mode and peak EXP ratings by participating in the training procedure prescribed by Klein et al. (1969) which consists of eight 2-1/2 hour sessions of structured training in which the trainees practice making ratings on sample tapes compiled by Klein et al. (1969). After four of the sessions, one of the raters discontinued her participation in the study. The remaining three individuals completed the structured training. At the close of the training phase, the two individuals who had evidenced the greatest proficiency levels were selected for further participation in the study. The two raters were selected on the basis of correlations between their ratings of the last 30 of 90 training segments and the Training Manual's (Klein et al., 1969) ratings of the same segments. As Table 4 illustrates, Raters A and B evidenced the highest average correlations with the Manual's ratings; thus they were selected as the raters for this study.

Table 4
Pearson r Correlations Between Raters' and Standard Ratings of Training Segments

<table>
<thead>
<tr>
<th>Rater</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard mode</td>
<td>.81</td>
<td>.85</td>
<td>.75</td>
</tr>
<tr>
<td>Standard peak</td>
<td>.82</td>
<td>.75</td>
<td>.77</td>
</tr>
</tbody>
</table>
The two raters were required to independently rate the 40 randomly presented tape segments and accompanying transcripts. The raters were together in the same room as the tape recording of each segment played, and they followed along on their respective transcripts. Barriers were constructed so that the raters were unable to see each other's transcripts or ratings. The raters were also not allowed to communicate with one another during this rating task.

The interrater reliability coefficients for this first rating phase were low to moderate (Pearson $r = .29$ for modes and $.50$ for peaks). These coefficients indicated a lack of consistent agreement between the raters. The investigator met with the raters with the goal of arriving at conclusions about why such a large discrepancy existed between the two raters, and between the interrater reliabilities for the training segments and the actual research segments. It was concluded that the raters had become overconfident and had decreased their reliance on the Manual as the authoritative reference in instances of confusion. Based on this conclusion, the investigator decided that the raters should rerate the segments. Explicit instructions were given to the raters to utilize the Manual as a reference whenever there was the slightest doubt as to the rating of a particular segment. The second rating trial produced considerably improved interrater reliabilities (Pearson $r = .83$ for modes and $.80$ for peaks). As would be expected, the majority of the rate-rerate reliabilities were low to moderate, thus indicating some overlap in ratings between the first and second trials, yet also indicating considerable changes in many of the ratings (Table 5). Rater B's
rate-rerate reliabilities were lower than rater A's, thus indicating a greater number of changed ratings for Rater B (Table 5). Because the second rating trial produced the better interrater reliabilities, the scores from this trial were used and the scores from the first trial were ignored. According to convention, the two raters' ratings were averaged in instances of interrater discrepancy, and this became the score for that particular segment.

Table 5
Pearson r Correlations for Raters A and B During Training and Research Segments

<table>
<thead>
<tr>
<th></th>
<th>Mode</th>
<th>Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training segments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rater A with standard</td>
<td>.81</td>
<td>.82</td>
</tr>
<tr>
<td>Rater B with standard</td>
<td>.85</td>
<td>.75</td>
</tr>
<tr>
<td>Interrater</td>
<td>.80</td>
<td>.74</td>
</tr>
<tr>
<td><strong>Research segments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1 interrater</td>
<td>.29</td>
<td>.50</td>
</tr>
<tr>
<td>Trial 2 interrater</td>
<td>.83</td>
<td>.80</td>
</tr>
<tr>
<td>Rater A rate-rerate</td>
<td>.45</td>
<td>.77</td>
</tr>
<tr>
<td>Rater B rate-rerate</td>
<td>.39</td>
<td>.60</td>
</tr>
</tbody>
</table>

**Alternative Rating**

The second phase of rating began by retraining the raters to make running ratings. Because the raters were already familiar with
the EXP Scale levels, this training procedure was considerably shorter than the first phase of training. The raters were trained to rate each client verbalization which evidenced a distinct EXP level (see Appendix G for instructions to raters). Because there are currently no standardized materials for training raters to make running ratings, a series of interview transcripts were taken from three sources in which running ratings accompanied the transcripts (Klein et al., 1969; Klein et al., in press; Mathieu-Coughlin & Klein, 1984). The authors' ratings were covered and the raters rated each group of transcripts. The coverings were removed after the raters had rated a group of transcripts, their answers were compared to the authors', and discrepancies between the raters' and authors' ratings were then discussed. Using this method a total of 140 client statements or "states" were rated in which raters A and B achieved 69% and 68% perfect agreement, respectively, with the authors' ratings. Percentage perfect agreement was used as the primary measure of reliability for this phase rather than correlation coefficients because of the alternative rating assumption that each level is a nominal category. Pearson r_s between the raters' and authors' ratings are presented in Table 6, however, for comparison purposes.

Based on these figures, it was decided that adequate interrater and rater-standard reliability had been established. The tape segments were rerandomized onto new master tapes and were then given to the raters, who then rated every ratable statement that the client made within each segment. A statement was marked "NR" (not ratable) if the statement was too brief to be clearly at one level or another
Table 6
Reliabilities for Alternative Ratings

<table>
<thead>
<tr>
<th>Rater</th>
<th>%</th>
<th>( r )</th>
<th>%</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>69%</td>
<td>.80</td>
<td>68%</td>
<td>.76</td>
</tr>
<tr>
<td>A</td>
<td>59%</td>
<td>.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(see Appendix F). Initially, each rater rated the transcript independent of the other. Because the numbers attached to the EXP levels were being treated as nominal categories with this alternative method, the averaging of EXP level ratings in cases of interrater discrepancy was not possible. Instead, after independent ratings were made, the raters were instructed to confer with each other in instances of interrater discrepancy with the mandate of reaching agreement on each rating. This procedure was followed until all transcripts were rated. The two raters achieved 79% perfect agreement on 570 client statements rated. This figure includes all "NR" ratings on not-ratable statements. For comparison purposes, the Pearson \( r \) coefficient was .68 for 537 client statements (NRs not counted).

After the raters had assigned an EXP rating to each client statement, an assistant timed each statement with a stopwatch and recorded the duration scores on each transcript. Each ratable client
verbalization thus acquired both an EXP label and a duration. Because client statements are often mixed with silences, and because each client statement had a duration assigned to it, the following set of rules were generated to determine when silences were to be included as part of an EXP state, and hence duration rating:

1. When the client made two adjacent statements, interrupted by silence, and the two statements had equivalent EXP ratings. In this case both statements plus the silence constituted a single EXP state.

2. When the client made a string of statements, interrupted by silences, and all statements had equivalent EXP ratings. In this case the string of statements and silences constituted a single EXP state.

3. When the client's statement was followed by silence, a therapist response, and then an additional client statement, and the two client statements had equivalent ratings. In this case the first client statement and the following silence (up to the point of the therapist's response) constituted a single EXP state.

4. When the client's statement was followed by a therapist response, and then by a silence, followed by another client statement, and the two client statements had equivalent EXP ratings. In this case the silence preceding the second client response and the second response itself constituted a single EXP state.

Any silences that did not conform to the above rules were not included as part of EXP states. The time that the therapists spent talking was also not included as part of the duration measurements.
Data Analysis

The above procedures yielded 16 scores per segment: a mode rating, a peak rating, total seconds duration for each of 7 EXP levels, and the frequency of occurrence for each of 7 EXP levels. The frequencies and durations for Levels 6 and 7 were zero for all 40 segments; consequently, the frequency and duration data for these levels were excluded from the analysis. This reduced the total number of dependent variables in this study from 16 to 12, and for the alternative method from 14 to 10.

Experiencing and Session Productivity

The primary statistical test used to answer Question 1 (Are there differences in EXP between more and less productive sessions?) consisted of Hotelling's $T^2$ statistic (Harris, 1975; Morrison, 1976; Winer, 1962). Hotelling's $T^2$ allows for a test for the omnibus null hypothesis that there are no differences between more and less productive sessions by simultaneously testing for differences on all dependent variables, and thereby also controlling for the increased possibility of Type I errors (false rejection of $H_0$) due to the use of multiple dependent variables. Univariate $t$ tests were then used to test for differences on each variable. In order to control for the effects of multiple tests on Type I error rate, Bonferroni critical values (Huitema, 1980) were used with the univariate $t$ tests. Standardized discriminant function coefficients, which are derived from a discriminant analysis procedure (Klecka, 1975, 1980; Tatsuoka,
1970, 1971) were also used to elaborate on the characteristics of the two groups.

The above statistical procedure was performed a total of three times, once for the more and less productive sessions as defined by the clients \( (N = 28) \), once for those defined by the therapists \( (N = 26) \), and again for the combined judgments of session productivity \( (N = 15) \).

**Conventional and Alternative Methods**

The statistical procedure used to answer Question 2 (Is there a relationship between the conventional mode rating and the alternative frequency and duration ratings?) consisted of a Pearson product-moment correlation coefficient between the mode of each segment and the EXP level with the highest frequency and/or duration ratings for each segment. A total of three Pearson \( r_s \) were computed: (a) the mode of the segment by the level with the greatest frequency of occurrence within the segment, (b) the mode of the segment by the level with the greatest total duration in the segment, and (c) the mode by the segment with the level with the greatest average duration within the segment. Data from all 40 segments were entered into each correlation.

Question 3 (Which combination of conventional and alternative variables best discriminates between more and less productive sessions?) was answered with a discriminant analysis procedure (Klecka, 1975, 1980; Tatsuoka, 1970, 1971) which constructs a linear combination of weighted variables such that the linear combination optimally
discriminates between the two groups. By utilizing a "stepwise" procedure for the selection of variables for inclusion into the linear combination (or "discriminant function"), the variable which contributes the most discriminative power is selected. Next, the second-best contributor (when combined with the first best) is identified and entered into the equation, followed by the third best contributor (when combined with the first two), and so forth until a variable is found not to appreciably contribute to the discriminant function; the variables which follow are then discarded as non-discriminators. Thus a weighted combination of variables is generated which together optimally discriminate between the two groups. This procedure was conducted three times, once for each of the client, therapist, and combined definitions of session productivity.

A series of three statistical operations were utilized to answer Question 4 (Is there a difference between the conventional and alternative methods in ability to discriminate between more and less productive sessions?). The first operation involved generating a discriminant function for conventional data (mode and peak as variables), and an additional discriminant function for alternative data (frequencies and durations as variables). One method of assessing the utility and strength of the discriminant functions is to statistically reclassify the original cases (segments) into the original groups according to how well each case "resembles" the discriminant function's mathematical definition of each group (Klecka, 1975). The second step, therefore, was to reclassify each segment according to the functions generated for the conventional and alternative methods.
If one discriminant function or the other did a better job of correctly reclassifying the segments into their original more or less productive groups, then that particular discriminant function could be said to be the better discriminator. The third step utilized McNemar's test (Bishop, Fienberg, & Holland, 1975) to compare the frequencies of segments that were correctly and incorrectly classified by the discriminant functions based on the conventional versus alternative variables.

**Multivariate Mathematical Assumptions**

Discriminant analysis and Hotelling's $T^2$ share similar mathematical assumptions because they are based on the same linear model. Klecka (1980) stated that one of the requirements for using discriminant analysis is that the number of discriminating variables must be at least 2 less than the total number of observations. Tatsuoka (1970), however, has stated that the total $N$ should be at least 2 times the number of discriminating variables. All of the analyses conducted in the present study satisfy Klecka's definition, and two of three of the analyses satisfy Tatsuoka's. The "combined" ratings analysis ($N = 15$) did not satisfy Tatsuoka's stated requirements. In regard to Hotelling's $T^2$ there is no clear consensus on the minimal number of cases needed, but as Hotelling's $T^2$ is quite similar mathematically to discriminant analysis, it might be assumed that similar rules-of-thumb would apply (Richardson, 1986; Wasserman, 1986).

Another requirement of discriminant analysis is that "no discriminating variable may be a linear combination of other
discriminating variables" (Klecka, 1980, p. 11). It is because of this requirement that average duration was not used as a discriminating variable, as it is a linear combination of total duration and frequency.

Harris (1975) and Klecka (1980) stated that the covariance matrices for the two groups should be approximately equivalent, hence Box's $M$ (Klecka, 1975) was used to test for equivalency of the covariance matrices. The matrices were found not to differ when they were based on the conventional variables. Box's $M$ could not be computed on the group covariance matrices based on the alternative variables, due to one or both of the groups having a "singular" status, presumably due to a number of zero scores on one or more alternative variables. Basically, what this means is that the variances for these variables were zero, and computing Box's $M$ with such an arrangement is the matrix algebra equivalent of dividing by zero (Tatsuoka, 1971). Therefore, a test of equivalency of the covariance matrices based on the alternative variables was not done. It is assumed, then, that the group covariance matrices are approximately equal, but this assumption is tentative in regard to the alternative variables.

Other assumptions regarding these multivariate statistics are that the discriminating variables are at least interval level measurements (Klecka, 1980), that no observation belongs to more than one group (Tatsuoka, 1970), and that the groups were drawn from populations with multivariate normal distributions (Harris, 1975; Klecka, 1980). Of these last three assumptions, the former two are satisfied in the present study, and the latter is assumed to be true.
The assumption that the populations have multivariate normal distributions is made possible by invoking the multivariate central limit theorem which suggests that a normal distribution is approached as the sample size becomes sufficiently large (Morrison, 1976; Thompson, 1984). Currently, however, there is no accepted rule for determining when a sample is "sufficiently large" enough to invoke the multivariate central limit theorem (Harris, 1975; Thompson, 1984).
CHAPTER IV

RESULTS

The results are presented in this chapter and are organized according to the research questions and null hypotheses addressed.

Question 1

The following results address Question 1: Are there differences in EXP between more and less productive sessions? Each null hypothesis is listed and the results which address it follow.

Null Hypothesis 1A

Null Hypothesis 1A states that there are no differences in EXP between sessions judged by clients as more and less productive. Table 7 contains the means, standard deviations, and \( t \) values for each univariate \( t \) test and for Hotelling's \( T^2 \) test of differences on all variables. Neither Hotelling's \( T^2 \) nor the univariate tests reveal statistically significant differences, thus the null hypothesis fails to be rejected. Figure 1 illustrates the mean mode and peak scores, and Figures 2 and 3 present the mean frequencies and durations, respectively, for the client-determined more and less productive sessions. As can be seen, the means on all dimensions parallel each other very closely; indicating no appreciable differences.
Table 7
Means, Standard Deviations, and t Values for Client-Determined More (MP) and Less Productive (LP) Sessions

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\bar{X}_{LP}$</th>
<th>SD$_{LP}$</th>
<th>$\bar{X}_{MP}$</th>
<th>SD$_{MP}$</th>
<th>t value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>2.50</td>
<td>0.71</td>
<td>2.39</td>
<td>0.71</td>
<td>0.40</td>
</tr>
<tr>
<td>Peak</td>
<td>3.46</td>
<td>0.95</td>
<td>3.29</td>
<td>0.64</td>
<td>0.58</td>
</tr>
<tr>
<td>F1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.14</td>
<td>0.36</td>
<td>1.47</td>
</tr>
<tr>
<td>F2</td>
<td>7.86</td>
<td>4.26</td>
<td>8.64</td>
<td>5.67</td>
<td>0.41</td>
</tr>
<tr>
<td>F3</td>
<td>3.43</td>
<td>2.10</td>
<td>3.64</td>
<td>2.76</td>
<td>0.22</td>
</tr>
<tr>
<td>F4</td>
<td>0.71</td>
<td>0.91</td>
<td>0.71</td>
<td>0.73</td>
<td>0.00</td>
</tr>
<tr>
<td>F5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>0.26</td>
<td>1.00</td>
</tr>
<tr>
<td>D1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.48</td>
<td>1.26</td>
<td>1.42</td>
</tr>
<tr>
<td>D2</td>
<td>194.23</td>
<td>117.09</td>
<td>191.40</td>
<td>149.76</td>
<td>0.00</td>
</tr>
<tr>
<td>D3</td>
<td>130.06</td>
<td>97.77</td>
<td>106.39</td>
<td>106.40</td>
<td>0.62</td>
</tr>
<tr>
<td>D4</td>
<td>35.00</td>
<td>46.75</td>
<td>45.27</td>
<td>59.78</td>
<td>0.51</td>
</tr>
<tr>
<td>D5</td>
<td>0.00</td>
<td>0.00</td>
<td>3.03</td>
<td>11.33</td>
<td>1.00</td>
</tr>
</tbody>
</table>

All
(Hotelling's $I^2$) $F = .0025$

Note. F = frequency; D = duration.

$^a$The Bonferroni critical $t$ value for the univariate tests = 3.14 (26 degrees of freedom). The critical $F$ value for Hotelling's $I^2 = 2.60$ (12 and 16 degrees of freedom). None of the comparisons is statistically significant.
Figure 1. Mean Mode and Peak Ratings for Client-Determined More (MP) and Less Productive (LP) Sessions.

Null Hypothesis 1B

Null Hypothesis 1B states that there are no differences in EXP between sessions judged by therapists as more and less productive. Table 8 contains the means, standard deviations, and $t$ values for each univariate $t$ test and for Hotelling's $T^2$ test of differences on all variables. Neither Hotelling's $T^2$ nor the univariate tests reveal statistically significant differences, thus the null hypothesis fails to be rejected. Figure 4 illustrates the mean mode and peak scores, and Figures 5 and 6 present the mean frequencies and durations, respectively, for the therapist-determined more and less
Figure 2. Mean Frequencies of EXP States for Client-Determined More (MP) and Less Productive (LP) Sessions.
Figure 3. Mean Durations of EXP States for Client-Determined More (MP) and Less Productive (LP) Sessions.
Table 8
Means, Standard Deviations, and \( t \) Values for Therapist-Determined More (MP) and Less Productive (LP) Sessions

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \bar{X}_{LP} )</th>
<th>SD_{LP}</th>
<th>( \bar{X}_{MP} )</th>
<th>SD_{MP}</th>
<th>( t ) value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>2.46</td>
<td>0.85</td>
<td>2.26</td>
<td>0.70</td>
<td>0.63</td>
</tr>
<tr>
<td>Peak</td>
<td>3.42</td>
<td>0.81</td>
<td>3.46</td>
<td>0.92</td>
<td>0.10</td>
</tr>
<tr>
<td>F1</td>
<td>0.08</td>
<td>0.28</td>
<td>0.15</td>
<td>0.38</td>
<td>0.59</td>
</tr>
<tr>
<td>F2</td>
<td>7.08</td>
<td>5.68</td>
<td>10.31</td>
<td>6.20</td>
<td>1.39</td>
</tr>
<tr>
<td>F3</td>
<td>3.38</td>
<td>3.04</td>
<td>3.62</td>
<td>2.57</td>
<td>0.20</td>
</tr>
<tr>
<td>F4</td>
<td>0.62</td>
<td>0.96</td>
<td>0.46</td>
<td>0.66</td>
<td>0.48</td>
</tr>
<tr>
<td>F5</td>
<td>0.08</td>
<td>0.28</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>D1</td>
<td>0.77</td>
<td>2.77</td>
<td>0.52</td>
<td>1.31</td>
<td>0.30</td>
</tr>
<tr>
<td>D2</td>
<td>191.80</td>
<td>125.25</td>
<td>187.89</td>
<td>128.11</td>
<td>0.10</td>
</tr>
<tr>
<td>D3</td>
<td>156.84</td>
<td>122.79</td>
<td>104.08</td>
<td>107.92</td>
<td>1.16</td>
</tr>
<tr>
<td>D4</td>
<td>25.43</td>
<td>40.21</td>
<td>29.34</td>
<td>49.91</td>
<td>0.22</td>
</tr>
<tr>
<td>D5</td>
<td>3.26</td>
<td>11.76</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

All (Hotelling's \( T^2 \)) \( F = .0279 \)

Note. F = frequency; D = duration.

<sup>a</sup>The Bonferroni critical \( t \) value for the univariate tests = 3.17 (24 degrees of freedom). The critical \( F \) value for Hotelling's \( T^2 \) = 2.64 (12 and 14 degrees of freedom). None of the comparisons is statistically significant.
productive sessions. At first glance it may appear that significant differences might exist for the frequency of Level 2 and the duration of Level 3; however, the variability on these dimensions is large enough so that the differences are not statistically significant.

Null Hypothesis 1C

Null Hypothesis 1C states that there are no differences in EXP between sessions judged by both clients and therapists as more and less productive. Table 9 contains the means, standard deviations, and t values for each univariate t test and for Hotelling's $T^2$ test of differences on all variables. Neither Hotelling's $T^2$ nor the

Figure 4. Mean Mode and Peak Ratings for Therapist-Determined More (MP) and Less Productive (LP) Sessions.
Figure 5. Mean Frequencies of EXP States for Therapist-Determined More (MP) and Less Productive (LP) Sessions.
Figure 6. Mean Durations of EXP States for Therapist-Determined More (MP) and Less Productive (LP) Sessions.
Table 9

Means, Standard Deviations, and $t$ Values for Combined-Determined More (MP) and Less Productive (LP) Sessions

<table>
<thead>
<tr>
<th>Variable</th>
<th>$X_{LP}$</th>
<th>SD$_{LP}$</th>
<th>$X_{MP}$</th>
<th>SD$_{MP}$</th>
<th>$t$ value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>2.33</td>
<td>0.41</td>
<td>2.11</td>
<td>0.22</td>
<td>1.37</td>
</tr>
<tr>
<td>Peak</td>
<td>3.50</td>
<td>0.63</td>
<td>3.22</td>
<td>0.44</td>
<td>1.01</td>
</tr>
<tr>
<td>F1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
<td>0.44</td>
<td>1.22</td>
</tr>
<tr>
<td>F2</td>
<td>6.33</td>
<td>1.21</td>
<td>10.33</td>
<td>4.97</td>
<td>1.91</td>
</tr>
<tr>
<td>F3</td>
<td>3.50</td>
<td>2.35</td>
<td>3.78</td>
<td>2.95</td>
<td>0.20</td>
</tr>
<tr>
<td>F4</td>
<td>0.67</td>
<td>1.21</td>
<td>0.44</td>
<td>0.53</td>
<td>0.49</td>
</tr>
<tr>
<td>F5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.74</td>
<td>1.54</td>
<td>1.17</td>
</tr>
<tr>
<td>D2</td>
<td>192.88</td>
<td>86.96</td>
<td>182.92</td>
<td>110.25</td>
<td>0.17</td>
</tr>
<tr>
<td>D3</td>
<td>134.05</td>
<td>82.91</td>
<td>81.96</td>
<td>92.41</td>
<td>1.11</td>
</tr>
<tr>
<td>D4</td>
<td>28.33</td>
<td>49.51</td>
<td>29.80</td>
<td>47.10</td>
<td>0.00</td>
</tr>
<tr>
<td>D5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

All (Hotelling's $T^2$) $F = .0025$

Note. $F =$ frequency; $D =$ duration.

$^a$The Bonferroni critical $t$ value for the univariate tests = 3.47 (13 degrees of freedom). The critical $F$ value for Hotelling's $T^2 = 3.49$ (12 and 3 degrees of freedom). None of the comparisons is statistically significant.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
univariate tests reveal statistically significant differences; thus the null hypothesis fails to be rejected. Figure 7 illustrates the mean mode and peak scores, and Figures 8 and 9 present the mean frequencies and durations, respectively, for the combined-determined more and less productive sessions. As is the case with the therapist-determined sessions, there initially appears to be appreciable differences for the frequency of Level 2 and the duration of Level 3. Similarly, however, the variability on these dimensions is large enough so that the differences are not statistically significant.

![Figure 7. Mean Mode and Peak Ratings for Combined-Determined More (MP) and Less Productive (LP) Sessions.](image)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 8. Mean Frequencies of EXP States for Combined-Determined More (MP) and Less Productive (LP) Sessions.
Figure 9. Mean Durations of EXP States for Combined-Determined More (MP) and Less Productive (LP) Sessions.
Table 10 contains the standardized canonical discriminant function coefficients attached to each variable for the client-, therapist-, and combined-determined discriminant functions. Also shown in Table 10 are probability values indicating the power of each discriminant function in detecting differences between groups. None of these discriminant functions significantly discriminated between the two groups. The absolute values of the coefficients can be interpreted as the relative amount that each variable contributes to the discriminant function; the sign indicates the direction of the contribution (+ = more productive, - = less productive).

Question 2

The following results address Question 2: Is there a relationship between the conventional mode rating and the alternative frequency and duration ratings? Each null hypothesis is listed and the results which address it follow.

Null Hypothesis 2A

Null Hypothesis 2A states that there is no relationship between the mode rating of a segment and the EXP state with the highest frequency of occurrence within the segment. A Pearson $r$ of .56 exists between the mode and the level with the highest frequency (Table 11). This correlation is statistically significant ($p < .001$); thus the null hypothesis is rejected.
Table 10

Standardized Canonical Discriminant Function Coefficients for Client-, Therapist-, and Combined-Determined Productivity With Accompanying Probability Values for the Discriminant Functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>Discriminant function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Client-determined&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mode</td>
<td>-0.026</td>
</tr>
<tr>
<td>Peak</td>
<td>-0.326</td>
</tr>
<tr>
<td>F1</td>
<td>0.592</td>
</tr>
<tr>
<td>F2</td>
<td>0.502</td>
</tr>
<tr>
<td>F3</td>
<td>2.009</td>
</tr>
<tr>
<td>F4</td>
<td>-1.002</td>
</tr>
<tr>
<td>F5</td>
<td>1.196</td>
</tr>
<tr>
<td>D1</td>
<td>0.607</td>
</tr>
<tr>
<td>D2</td>
<td>1.739</td>
</tr>
<tr>
<td>D3</td>
<td>-0.272</td>
</tr>
<tr>
<td>D4</td>
<td>2.070</td>
</tr>
<tr>
<td>D5</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<sup>a</sup><sub>P</sub> = .26.  <sup>b</sup><sub>P</sub> = .58.  <sup>c</sup><sub>P</sub> = .11.
**Table 11**

*Pearson r Correlation Coefficients for Levels With the Highest Frequency (f), Total Duration (td), and Average Duration (ad) With the Mode*

<table>
<thead>
<tr>
<th></th>
<th>r with mode</th>
<th>% shared variance (r²)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level with highest f</td>
<td>.56</td>
<td>31.4%</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Level with highest td</td>
<td>.32</td>
<td>10.2%</td>
<td>&lt;.050</td>
</tr>
<tr>
<td>Level with highest ad</td>
<td>.20</td>
<td>4.0%</td>
<td>&gt;.100</td>
</tr>
</tbody>
</table>

**Null Hypothesis 2B**

Null Hypothesis 2B states that there is no relationship between the mode rating of a segment and the EXP state with the greatest total duration within the segment. A Pearson r of .32 exists between the mode and the level with the greatest total duration (Table 11). This correlation is statistically significant (p < .05); thus the null hypothesis is rejected.

**Null Hypothesis 2C**

Null Hypothesis 2C states that there is no relationship between the mode rating of a segment and the EXP state with the greatest average duration within the segment. A Pearson r of .20 exists between the mode and the level with the greatest total duration (Table 11). This correlation is not statistically significant (p > .10); thus the null hypothesis fails to be rejected. Table 11 also contains the percentage of variance shared by the mode ratings and
each of the other ratings. This percentage is an estimate of the degree of overlap between the two variables being correlated (Hopkins & Glass, 1978).

**Question 3**

The following results address Question 3: Which combination of conventional and alternative variables best discriminates between more and less productive sessions? The results which address Questions 3A, 3B, and 3C are all contained in Table 12, which lists the variables selected by the discriminant analysis step-wise procedure as the combinations of variables which produced the optimum discrimination between the groups. Table 12 also contains the standardized canonical discriminant function coefficients attached to each of the variables and a probability value for each of the three step-wise produced discriminant functions. The reader will note that the probability values for these step-wise discriminant functions are lower than those reported in Table 10; the reason being that the step-wise procedure excludes variables if they hinder or do not add to the power of the discriminant function.

Two variables are included in the discriminant function for the client-defined sessions; they are frequency for Level 1 and duration for Level 5. This discriminant function does not discriminate significantly between groups ($x^2 = 3.20; p = .20$). Only one variable is included in the discriminant function for the therapist-determined sessions, frequency for Level 2. This discriminant function is also nonsignificant ($x^2 = 1.81; p = .18$). Four variables contribute to
Table 12

Variables Selected With Step-Wise Discriminant Analysis Procedure With Accompanying Standardized Canonical Discriminant Function Coefficients and Discriminant Function Probability Values

<table>
<thead>
<tr>
<th>Source</th>
<th>1st var. selected</th>
<th>2nd var. selected</th>
<th>3rd var. selected</th>
<th>4th var. selected</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>F1</td>
<td>D5</td>
<td></td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>0.8528</td>
<td>0.6277</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Therapist</td>
<td>F2</td>
<td></td>
<td></td>
<td></td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>F2</td>
<td>F1</td>
<td>F3</td>
<td>Mode</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td>1.4569</td>
<td>1.5762</td>
<td>1.6943</td>
<td>-1.6943</td>
<td></td>
</tr>
</tbody>
</table>

Note. F = frequency; D = duration.

The discriminant function for the combined-defined sessions; they are frequency of Levels 2, 1, and 3 and the mode. This particular combination of variables produces a discriminant function which significantly discriminates between the more and less productive sessions as defined by the clients' and therapists' combined ratings of session productivity ($\chi^2 = 15.95; p = .003$). Although not formally utilized in the analysis of Question 1, this significant discriminant function reveals differences between more and less productive sessions.

Question 4

The following results address Question 4: Is there a difference between the conventional and alternative methods in ability to
discriminate between more and less productive sessions? Two discriminant functions were constructed for each of three session types (client-, therapist-, and combined-determined productivity), with one of the discriminant functions composed of the conventional variables, and the other of alternative variables. Table 13 gives the chi-squared values and the probability values of each of the discriminant functions. None of these functions is significant.

Table 13
Chi-Squared Values and Probability Values Associated With Discriminant Analyses of More and Less Productive Sessions Based on Conventional and Alternative Variables

<table>
<thead>
<tr>
<th>Variable type</th>
<th>Chi squared</th>
<th>df(^a)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client-determined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td>0.325</td>
<td>2</td>
<td>.85</td>
</tr>
<tr>
<td>Alternative</td>
<td>13.502</td>
<td>9</td>
<td>.14</td>
</tr>
<tr>
<td>Therapist-determined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td>1.501</td>
<td>2</td>
<td>.47</td>
</tr>
<tr>
<td>Alternative</td>
<td>7.281</td>
<td>9</td>
<td>.61</td>
</tr>
<tr>
<td>Combined-determined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td>1.857</td>
<td>2</td>
<td>.40</td>
</tr>
<tr>
<td>Alternative</td>
<td>11.167</td>
<td>8</td>
<td>.19</td>
</tr>
</tbody>
</table>

\(^a\)The chi-squared critical values at the .05 level = 5.99 for 2 degrees of freedom, 16.92 for 9 degrees of freedom, and 15.51 for 8 degrees of freedom.
Figure 10 illustrates the distribution of client-determined sessions along the discriminant function based on the conventional variables. The less productive sessions are represented on the upper distribution, and the more productive on the lower distribution. Each group's centroid (central value of the discriminant function) is indicated by the number (a "1" or a "2") at the bottom of each graph.

Figure 11 illustrates the distribution of client-determined sessions based on the alternative variables. Comparison of Figures 10 and 11 reveals greater distance between more and less productive groups on the discriminant function for the alternative variables. Figures 12 and 13 illustrate similar distributions for the therapist-determined sessions, and Figures 14 and 15 for the combined-determined sessions. As is the case with the client-determined sessions, examination of Figures 12-15 reveals greater distances between the more and less productive groups on the alternative-based discriminant functions.

Each session was reclassified according to the mathematical definition of each group based on the discriminant functions. Table 14 lists the percentage of correct classifications for each discriminant function, broken down into more productive, less productive, and total correctly classified. Table 15 displays the frequencies of correct and incorrect classifications and also illustrates the groups compared with McNemar's test. McNemar's test was used to test for differences between the frequency of sessions correctly classified by the alternative variables but incorrectly classified by the conventional versus the frequency of sessions correctly classified by the
Figure 10. Frequency Distributions of Client-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Conventional Variables.  
Note. Each group's centroid is indicated by the placement of the solitary "1" or "2" at the bottom of each histogram.
Figure 11. Frequency Distributions of Client-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Alternative Variables.

Note. Each group's centroid is indicated by the placement of the solitary "1" or "2" at the bottom of each histogram.
Figure 12. Frequency Distributions of Therapist-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Conventional Variables. 

Note. Each group's centroid is indicated by the placement of the solitary "1" or "2" at the bottom of each histogram.
Figure 13. Frequency Distributions of Therapist-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Alternative Variables.

Note. Each group's centroid is indicated by the placement of the solitary "1" or "2" at the bottom of each histogram.
Figure 14. Frequency Distributions of Combined-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Conventional Variables.

Note. Each group's centroid is indicated by the placement of the solitary "1" or "2" at the bottom of each histogram.
Figure 15. Frequency Distributions of Combined-Determined Less Productive (Upper Histogram) and More Productive (Lower Histogram) Sessions Along Discriminant Function Based on Alternative Variables.

Note. Each group's centroid is indicated by the placement of the solitary "1" or "2" at the bottom of each histogram.
Table 14

Percentage of More Productive, Less Productive, and Total Sessions Correctly Classified by Functions Based on Conventional Versus Alternative Variables

<table>
<thead>
<tr>
<th>Discriminant function</th>
<th>Conventional %</th>
<th>Alternative %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Client-determined</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More productive</td>
<td>57.1</td>
<td>78.6</td>
</tr>
<tr>
<td>Less Productive</td>
<td>50.0</td>
<td>78.6</td>
</tr>
<tr>
<td>Total</td>
<td>53.6</td>
<td>78.6</td>
</tr>
<tr>
<td><strong>Therapist-determined</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More productive</td>
<td>69.2</td>
<td>84.6</td>
</tr>
<tr>
<td>Less productive</td>
<td>53.8</td>
<td>84.6</td>
</tr>
<tr>
<td>Total</td>
<td>61.5</td>
<td>84.6</td>
</tr>
<tr>
<td><strong>Combined-determined</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More productive</td>
<td>77.8</td>
<td>88.9</td>
</tr>
<tr>
<td>Less productive</td>
<td>50.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>66.7</td>
<td>93.3</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 15

Frequencies (and Percentages) of Cases Classified Correctly and Incorrectly for Discriminant Functions Based on Conventional and Alternative Variables

<table>
<thead>
<tr>
<th></th>
<th>Conventional</th>
<th>Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correctly Classified</td>
<td>Incorrectly Classified</td>
</tr>
<tr>
<td>Correctly Classified</td>
<td>11 (39.3%)</td>
<td>4 (14.3%)</td>
</tr>
<tr>
<td>Incorrectly Classified</td>
<td>11 (39.3%)</td>
<td>2 (7.1%)</td>
</tr>
<tr>
<td></td>
<td>22 (78.6%)</td>
<td>6 (21.4%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Therapist-determined:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correctly Classified</td>
</tr>
<tr>
<td>Correctly Classified</td>
<td>13 (50.0%)</td>
</tr>
<tr>
<td>Incorrectly Classified</td>
<td>9 (34.6%)</td>
</tr>
<tr>
<td></td>
<td>22 (84.6%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Combined-determined:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correctly Classified</td>
</tr>
<tr>
<td>Correctly Classified</td>
<td>9 (60.0%)</td>
</tr>
<tr>
<td>Incorrectly Classified</td>
<td>5 (33.3%)</td>
</tr>
<tr>
<td></td>
<td>14 (93.3%)</td>
</tr>
</tbody>
</table>

Note. Arrows indicate frequencies used to test for differences with McNemar's test.
conventional variables but incorrectly classified by the alternative variables (see Table 15).

Null Hypothesis 4A

Null Hypothesis 4A states that there is no difference between the conventional and alternative methods in ability to discriminate between sessions judged by clients as more and less productive. The chi-squared value for McNemar's test is not significant (see Table 16); thus the null hypothesis fails to be rejected.

Table 16
Chi-Squared Values and Probability Statements for McNemar's Test of Differences Between Correctly and Incorrectly Classified Sessions

<table>
<thead>
<tr>
<th>Session type</th>
<th>Chi-squared value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client-determined</td>
<td>2.40</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>Therapist-determined</td>
<td>2.08</td>
<td>&gt;.10</td>
</tr>
<tr>
<td>Combined-determined</td>
<td>1.50</td>
<td>&gt;.10</td>
</tr>
</tbody>
</table>

Critical value = 3.84, determined by 1 degree of freedom.

Null Hypothesis 4B

Null Hypothesis 4B states that there is no difference between the conventional and alternative methods in ability to discriminate between sessions judged by therapists as more and less productive.
The chi-squared value for McNemar's test is not significant (see Table 16); thus the null hypothesis fails to be rejected.

**Null Hypothesis 4C**

Null Hypothesis 4C states that there is no difference between the conventional and alternative methods in ability to discriminate between sessions judged by both clients and therapists as more and less productive. The chi-squared value for McNemar's test is not significant (see Table 16); thus the null hypothesis fails to be rejected.
A discussion of the present study is presented in this chapter. A review and discussion of the results is contained in the first section, issues and limitations associated with the study are addressed in the second section, possible directions for future research are offered in the third section, and a summary of the study concludes the chapter.

Discussion of the Results

Experiencing and Session Productivity

As indicated in the previous chapter, there were no significant differences revealed for any of the tests used to answer Question 1: Are there differences in EXP between more and less productive sessions? These results are concordant with those of Auerbach and Luborsky (1968), Elliott et al. (1982), and Pollak (1973), who also failed to detect significant differences in EXP between more and less productive sessions. In Chapter II it was proposed that the above studies may have failed to detect significant differences due to (a) the use of other than the actual therapy participants as judges of productivity and/or (b) the use of less than standard length interview segments. The present study utilized the therapy participants as judges of productivity and also 8-minute segments. This
arrangement also failed to detect significant differences when testing for differences between more and less productive sessions; therefore, it might be concluded that there are no differences in EXP between more and less productive sessions, even when standard-length segments are used and the therapy participants determine which sessions were more and less productive. One might be fairly confident in this conclusion if not for the results of the step-wise discriminant analysis used to answer Question 3C, which indicated significant differences in EXP between more and less productive sessions when both client and therapist ratings were used to define productivity and when F2, F1, F3, and the mode were combined to form the discriminant function. Given the discrepant data, it may be premature to formulate conclusions regarding the relationship, or lack thereof, between EXP and session productivity.

Although not formally intended to answer Question 1, examination of the step-wise discriminant analysis used to answer Question 3 provides an understanding of the differences between the two groups: The more productive sessions tended to have greater frequencies of Levels 2, 1, and 3, while the less productive sessions tended to receive higher mode ratings (see Tables 9 and 12). None of the durations were found to substantially contribute to this discriminant function. Higher frequency of occurrence also indicates a greater number of shifts from one state to another. Therefore, it may be that the clients more readily moved from one state to another in the more productive sessions, and/or that the clients and therapists evidenced higher rates of turn-taking.
The results of this discriminant analysis also suggest that the EXP continuum of levels may not parallel an hypothetical productivity continuum. The less productive sessions evidenced higher average modes; this is actually the reverse of what would be expected given the theoretical notion that higher levels of EXP are indicative of more productive therapy behavior. The more productive sessions evidenced greater frequencies of Levels 1, 2, and 3; which are on the lower level of the EXP continuum. Theoretically, it would be expected that if more productive sessions were to evidence greater frequencies of any levels, they would be toward the upper end of the continuum, and if greater frequencies of lower EXP levels were found for a particular group, they would be expected to belong to the less productive sessions. The reverse was actually true in this case. Thus, these results contradict what would theoretically be expected, and this casts doubt on the idea that EXP and productivity are parallel continua.

Conventional and Alternative Methods

Question 2 asked Is there a relationship between the conventional mode rating and the alternative frequency and duration ratings? This question emerged from the criticism that the conventional mode score lacks a clear operational definition. By examining the strength of relationship between the mode and the levels with the highest frequency, cumulative duration, and average duration, some conclusions can be reached regarding how these particular raters may have interpreted the definition of the mode as the "overall, general
or average scale level of the segment or unit" (Klein et al., 1969, p. 65). As mentioned in the previous chapter, the correlation between the mode and the level with the highest frequency of occurrence was statistically significant ($r = .56, p < .001$), as was also the correlation between the mode and the level with the highest cumulative duration ($r = .32, p < .05$). This indicates that there were relationships between the levels with the highest frequencies and cumulative durations and the modes that the raters assigned to the segments. Ideally, however, one would want the strength of relationship to be stronger than these coefficients indicate. Consideration of the percentage of shared variance ($r^2$) for each of these two comparisons indicates that roughly 31.4% of the segments had modes similar to the level with the highest frequency, and 10.2% had modes similar to the level with the highest cumulative duration. Thus, although these coefficients were statistically significant, there existed considerable variability indicating inconsistencies in the raters' interpretations of the definition of the mode as the "overall, general, or average scale level of the segment or unit" (Klein et al., 1969, p. 65).

Question 3 asked Which combination of conventional and alternative variables best discriminates between more and less productive sessions? The discriminant analysis step-wise procedures used to answer this question revealed that the combination of the frequency of Level 1 and the duration of Level 5 best discriminated between the client-defined more and less productive sessions ($p = .20$), the frequency of Level 2 alone best discriminated between the therapist-
determined sessions ($p = .18$), and the combination of the frequencies of Levels 2, 1, and 3 plus the mode best discriminated between the combined-determined sessions ($p = .003$). For the former two discriminant functions, however, the differences between groups were not significant. Inclusion of the alternative variables enhanced the EXP Scales' ability to detect differences between the groups. This was especially so for the combined-determined discriminant function, it having reached significance even prior to the step-wise inclusion of the mode ($p = .02$ prior to inclusion of the mode).

Question 4 asked Is there a difference between the conventional and alternative methods in ability to discriminate between more and less productive sessions? Each null hypothesis failed to be rejected due to McNemar's test not revealing significant differences between the frequency of sessions correctly classified by the alternative method but incorrectly classified by the conventional method versus the frequency of sessions correctly classified by the conventional method but incorrectly classified by the alternative method. The nonsignificance of the McNemar's tests is perplexing given the percentages and frequencies of sessions correctly and incorrectly classified as illustrated in Tables 14 and 15. One explanation for this is that the chi-squared distribution (utilized by McNemar's test) is such that it is more apt to obscure differences when smaller sample sizes are used (Richardson, 1986). Although the differences were nonsignificant, the discriminant functions based on the alternative variables consistently reclassified the cases into their original groups at higher rates than the conventional-based discriminant
functions. The percentages correctly classified for the client-, therapist-, and combined-determined sessions were 78.6%, 84.6%, and 93.3%, respectively for the discriminant functions based on the alternative variables, as compared to 53.6%, 61.5%, and 66.7% for the discriminant functions based on the conventional variables. This indicates that the alternative-based discriminant functions more adequately defined each group.

Although the results of the tests used to answer Question 3 were not formally utilized to answer Question 4, they do indicate that the alternative variables provided more valuable information than the conventional variables. None of the conventional variables were selected for inclusion into the step-wise discriminant functions for the client- and therapist-determined sessions. The mode was the final variable included with three of the alternative variables as part of the step-wise discriminant function for the combined-determined sessions; however, the inclusion of the mode was not necessary in order to reach statistical significance, as the function reached a probability of .02 prior to the mode's inclusion. Thus the frequencies of Levels 2, 1, and 3 combined to form a discriminant function which significantly discriminated between the two groups. The mode was able to contribute to the discrimination between the groups only when combined with these three alternative variables.

The alternative method of summarizing EXP also appears to have an advantage over the conventional method in regard to sensitivity to state fluctuations. Examination of Figures 1 through 9 reveals typical EXP patterns for this particular group of clients regardless
of whether the sessions are classified as more or less productive. In each figure Levels 2 and 3 are clearly predominant, and this is also reflected in the means of the mode scores, which are all between 2.11 and 2.5. The predominance of Levels 2 and 3 in the present study is consistent with many other studies which have reported mean modes which organize around Levels 2 and 3 (Kiesler, 1971, Kiesler et al. 1964, 1965, 1967b; Pollak, 1973; Richert, 1976; Ryan, 1966; Schoeninger, 1965; Tomlinson, 1959, 1967; Tomlinson & Hart, 1962; van der Veen, 1967b), which suggests that Levels 2 and 3 are typical EXP levels for a number of people. Despite the fact that more and less productive sessions evidenced similar patterns in regard to the modes, peaks, frequencies, and durations of EXP, in one instance the alternative variables alone were able to detect differences associated with session productivity. This suggests that the alternative variables are more sensitive to minute state fluctuations.

**Issues and Limitations**

Perhaps the most apparent limitation of the present study is the relatively few numbers of subjects used. This influenced the data particularly in regard to the number of more or less productive sessions generated. Had a larger subject pool been utilized then the number of observations would have increased also. The most common issue associated with the use of smaller N designs is the increased possibility of Type II errors (false nonrejection of the null hypothesis). This is especially so with the use of multiple dependent variables and, as mentioned earlier in this chapter, was
probably the case in the use of McNemar's test of differences between correctly and incorrectly classified sessions based on the conventional and alternative discriminant functions. In order to illustrate this point, the original data set of frequencies of sessions classified correctly and incorrectly was artificially inflated by multiplying each frequency by 2, and McNemar's test was then rerun on this artificial data set. The chi-squared values for all three analyses (client-, therapist-, and combined-determined sessions) increased to the point of statistical significance ($\chi^2 = 4.80, 5.04,$ and $4.08$ for client-, therapist-, and combined-determined sessions, respectively; $\chi^2 = 3.84$ needed for .05 probability level). Thus, had the number of observations been twice as large, and had the proportions of correctly and incorrectly classified sessions remained constant, then the frequencies would have been significantly different. A second issue associated with the relatively few number of subjects used is the decreased ability to generalize results.

A limitation related to that mentioned above is the non-randomized selection of subjects. The subject selection procedure used produced one male and five female subjects, with one half of them graduate students and the other one half undergraduates. These proportions are not representative of the proportions of students seen at the university counseling center from which this sample was drawn and are probably not representative of the students seen at other university counseling centers. Because the therapists were allowed relatively wide latitude in subject selection, they may have selected subjects of a particular sort (e.g., clients whom they felt
particularly comfortable with), which may have influenced the nature of the research data. Reconsideration of the data contained in Appendix A suggests that four of the six clients did not substantially improve. Thus it could be argued that the sample was largely representative of less successful cases, who have been shown in previous studies to demonstrate lower levels of EXP (Kiesler, 1971; Ryan, 1966; Tomlinson, 1967; Tomlinson & Hart, 1962; van der veen, 1967a; Walker et al., 1962). This is supported by the observation that no client achieved frequency and duration scores for Levels 6 and 7.

Other issues arise from the selection of session segments. Gendlin (1986) has proposed that the traditional technique of labeling a set proportion of a sample as "better" and an equal proportion as "worse" may not accurately reflect the true distribution of better or worse cases or sessions. It may be that the sessions selected as the upper and lower 25% of each case may not accurately describe the nature of all the sessions contained in each group. For instance, it may have been that only one or two sessions within the group of 14 client-determined more productive sessions were worthy of being labeled "more productive."

Elliott (1983a) has pointed out that the one shortcoming of using postsession questionnaires to identify significant sessions is that such questionnaires do not help to pinpoint the exact events which contribute to sessions' significance. Thus it is not known whether the high and low ranked sessions received such ratings due to the overall nature of the sessions or to one or more specific events
within the sessions. If the latter, then there is no way of knowing if the significant intrasession events were captured by the 8-minute segments. This might have diluted the EXP ratings of the segments. It might also have been that the postsession questionnaire measured only the clients' and therapists' immediate impressions of the session; their impressions might have been different had they been asked to give retrospective impressions one week later.

Another limitation related to segment selection arises from the sampling of taped sessions from differing phases of therapy for each client. As mentioned in Chapter III, the sessions sampled for two of the clients came from the middle portion of therapy, while the sessions for the other four came from the end portions of therapy. The entire course of therapy was not sampled for any of the six clients. This may have restricted the range of session productivity sampled, as the most or least productive sessions could have occurred outside of the sampling window. The introduction of the taping and postsession questionnaire at the middle of treatment could have also disturbed or distorted subsequent client-therapist interactions, which in turn could have effected the data. Whereas no clients or therapists reported any deleterious effects from these procedures during the debriefing interviews, the possibility exists that the introduction of taping and postsession questionnaires altered the client-therapist relationships.

As described in Chapter III, the raters attained less than adequate interrater reliabilities for the first trial of conventional ratings. Because the raters were then instructed to rerate the
segments, the question is raised if the initial exposure to the segments may have influenced the raters' second trial of ratings, specifically in terms of the first trial having a practice effect on the second trial. Whereas the dual exposure to the segments may have influenced the raters in other ways, it does not appear that the raters were influenced in the direction of their original ratings, as evidenced by relatively low to modest rate-rerate reliabilities and substantially improved interrater reliabilities. Just as the raters' dual exposure during the conventional ratings may have confounded the second trial of ratings, exposure to the same tapes for both conventional and alternative ratings may have influenced the alternative ratings.

Three issues emerge from the design and data analysis of the present study. First, because the independent variable (session productivity) was not experimentally manipulated, the data produced were essentially correlational and consequently inferences regarding causality may not be made. Second, because of the "singular" nature of a number of the covariance matrices for the alternative variables, Box's $M$ could not be computed to test for equality of the groups' covariance matrices; thus equivalency was assumed without the benefit of statistical support. Third, the practice of averaging together of scores extracted from the sessions of different clients may have diluted or obscured other relationships between EXP and session productivity. Kiesler (1966) has presented a number of "uniformity myths" that exist in psychotherapy research, of which the "patient uniformity assumption" states that patients "are more alike than they
are different" (p. 110). The present study has been conducted with the assumption that EXP levels within a particular session type (more or less productive) would be more similar than different from one client to the next. However, it could be that each individual evidences unique EXP patterns for more and less productive sessions, and this individuality might also be related to case outcome. If this were so, then the averaging together of EXP scores of a group of individuals would have obscured these individual patterns.

Recommendations for Future Research

The results of the step-wise discriminant analysis between the combined-determined more and less productive sessions suggest differences in EXP and also between the conventional and alternative methods of summarizing EXP. Therefore, it would be a mistake to abandon further exploration of either the relationship between EXP and session productivity or the alternative method of summarizing EXP. This section contains an outline of recommendations for future research, particularly in regard to EXP and session productivity and the alternative method.

Experiencing and Session Productivity

The results of the step-wise discriminant analysis between the combined-determined more and less productive sessions hint at an inverse relationship between EXP and productivity, with more productive sessions having higher frequencies of lower level EXP states and less productive sessions exhibiting higher modes. Further study on
the relationship between session productivity and EXP would help to either confirm or disconfirm the current belief that higher levels of EXP are indicative of productive therapy behavior.

In regard to the defining of session productivity, the use of the actual therapy participants may have merit when considering the above mentioned significant result, and also considering the lack of significant differences in the studies which did not utilize the therapy participants as judges of productivity (Auerbach & Luborsky, 1968; Pollak, 1973). This is especially true when the combination of client and therapist ratings of productivity are used to form the more and less productive groups. In the present study significant differences were detected between the combined-determined more and less productive segments despite that the total number of observations was small relative to the client-determined and therapist-determined segments ($N = 15$ versus $N = 28$ and $26$, respectively). Generally, the power of a given statistical test to detect differences between groups increases as $N$ increases; thus if all three methods of defining productivity did so equally, then one would expect to find differences primarily for the comparison with the greater number of observations. That the results did not conform to this expectation is perhaps testimony to the efficacy of the combined-determined method of separating more and less productive groups. Future investigators of session productivity will want to take care to include a combined-determined definition of session productivity as part of the research design.
As mentioned earlier in this chapter, one of the limitations of the present study was an inability to rest assured that whatever had contributed to a given session's productivity status was actually captured within the 8-minute segment. One remedy would be to evaluate the entire session. Another possibility would be to systematically select multiple segments from each session, thereby increasing the possibility that any significant intrasession events would be contained within the sample. A more elegant procedure would be to use Elliott's (1983a, 1983b, 1984) method of utilizing the client and therapist as reporters. With this method the client and therapist review a tape of the session and indicate to the researcher the exact locations of significant intrasession events. With this method the researcher could be confident that the events studied are actually the phenomena of interest.

Alternative Method of Summarizing Experiencing

It is recommended here that the alternative method of summarizing EXP be utilized in future research, as it has certain advantages over the conventional method such as having clearer operational definition and avoiding results in terms of fractions of a level; and it also provides information beyond that produced by the conventional method. Use of the alternative method may also help to clarify the relationship between productive therapy behavior and each EXP level, and might also represent a method of quantifying running ratings.

Because of the multivariate character of the alternative method, future research should include a greater number of observations than
utilized in the present study in order to decrease the possibility of Type II errors. It is advantageous to utilize a step-wise discriminant analysis procedure when testing for differences between groups so that those variables which would otherwise hinder the discrimination power of the discriminant function would be excluded from the function.

There was considerable within groups variability on a number of the alternative variables, perhaps indicating differences between subjects in terms of habitual experiencing, and also in terms of individual differences for more and less productive sessions. Future research might utilize the alternative method to compare single cases, as it may be that differential patterns of frequencies and durations of EXP levels exist, and that such patterns would also be tied to case outcome. These hypothesized differential patterns might also extend across the entire course of therapy. If so, utilizing the alternative method in sequential analyses of more and less successful cases would provide new information on the vicissitudes of EXP levels as therapy progresses. This information would, in turn, support or refute Rogers's (1958) process theory of therapy.

Summary

Previous studies have failed to detect differences in EXP between more and less productive psychotherapy sessions (Auerbach & Luborsky, 1968; Elliott et al., 1982; Pollak, 1983). It was proposed in the present study that perhaps these investigators failed to detect differences due to either (a) the use of other than the
therapy participants as judges of session productivity or (b) the use of less than standard lengths of taped segments. The present study utilized the actual therapy participants as the judges of productivity and also used standard length 8-minute tape segments as data sources in order to address the question: Are there differences in EXP between more and less productive sessions?

This study also had a second focus on issues associated with the conventional method of summarizing data yielded with the EXP Scale. These issues include various interpretation problems which arise from a generally narrow band of results, questions about what dimensions the scale actually measures, and some inherent difficulties in using the scale at the microprocess level. An alternative method of summarizing EXP data was introduced in the present study as a possible antidote to the limitations inherent in the conventional method, and three research questions were then asked which addressed the relationship between these two methods: (a) Is there a relationship between the conventional mode rating and the alternative frequency and duration ratings? (b) Which combination of conventional and alternative variables best discriminates between more and less productive sessions? (c) Is there a difference between the conventional and alternative methods in ability to discriminate between more and less productive sessions?

In order to answer the above questions, six client-therapist pairs audio recorded every session that occurred within a 14-week block. At the close of each session the clients and therapists independently completed a postsession questionnaire which was used to
determine session productivity. Fourteen more productive and 14 less productive sessions were selected to be rated with the EXP Scale based on the clients' ratings on the postsession questionnaire. Thirteen more productive and 13 less productive sessions were selected on the basis of the therapists' ratings, and 9 more productive and 6 less productive sessions were selected on the basis of the clients' and therapists' combined ratings. An 8-minute segment was systematically selected from each of the above sessions, and two raters rated the segments with the EXP Scale, first using the conventional method and again using the alternative method. The conventional data consisted of mode and a peak scores, and the alternative data consisted of the frequencies and durations of each scale level within the segment.

The data were tested for differences between more and less productive sessions by utilizing Hotelling's $T^2$ to simultaneously test for differences on all variables, and also univariate $t$ tests on each dependent variable. There were no significant differences detected.

Significant correlations were found between the mode and the levels with the highest frequency of occurrence ($r = .56, p < .001$) and cumulative seconds duration ($r = .32, p < .05$). There was no significant relationship between the mode and the level with the highest average duration (cumulative duration/frequency). This indicated a degree of similarity between the mode and the levels with the highest frequency and cumulative duration; however, the strength of relationship was weak enough to indicate inconsistencies in the raters' interpretations of the operational definition of the mode as
the "overall, general, or average scale level of the segment or unit" (Klein et al., 1969, p. 65).

A step-wise discriminant analysis procedure was used to detect the particular combination of variables which best discriminated between the more and less productive sessions. The variables which best discriminated between the client-defined sessions were the frequency of Level 1 and the duration for Level 5. One variable alone best discriminated between the therapist-defined sessions, frequency for Level 2. These two discriminant functions were nonsignificant ($p = .20$ and $.18$, respectively). Four variables contributed to the discriminant function for the combined-defined sessions, they were the frequencies of Levels 2, 1, and 3, plus the mode. This last discriminant function significantly discriminated between the combined-determined more and less productive sessions ($p = .003$). Although not formally utilized to test for differences in EXP between more and less productive sessions, this significant discriminant function revealed differences between more and less productive sessions, and these differences were contrary to what would theoretically be expected.

There were no significant differences between the conventional and alternative methods in ability to discriminate between the more and less productive sessions, although the alternative-based discriminant functions consistently classified the sessions into their original groups at higher rates than the conventional-based discriminant functions. The alternative variables also contributed more weight to the significant step-wise discriminant function.
Appendix A

Client Initial Symptom/Complaint Self-Report Ratings and Posttherapy Change Scores
<table>
<thead>
<tr>
<th>Problem area</th>
<th>Client A</th>
<th>Client B</th>
<th>Client C</th>
<th>Client D</th>
<th>Client E</th>
<th>Client F</th>
<th>Client G</th>
<th>Client H</th>
<th>Client I</th>
<th>Client J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
</tr>
<tr>
<td>1. Schoolwork and grades</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>2. Procrastination, getting motivated, managing time</td>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>+1</td>
</tr>
<tr>
<td>3. Test anxiety</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>4. Stage fright, speaking anxiety</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5. Decision about major/career</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>+1</td>
</tr>
<tr>
<td>6. Relationship with friends</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>0</td>
<td>+2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7. Relationship with roommate</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8. Relationship with romantic partner</td>
<td>0</td>
<td>+2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>-2</td>
<td>1</td>
<td>-1</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>9. Relationship with family and parents</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>+2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>+1</td>
</tr>
<tr>
<td>10. Loss/death of significant person</td>
<td>1</td>
<td>+1</td>
<td>2</td>
<td>+1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Problem area</td>
<td>Client A</td>
<td>Client B</td>
<td>Client C</td>
<td>Client D</td>
<td>Client E</td>
<td>Client F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Sexual matters</td>
<td>0</td>
<td>0</td>
<td>0 +1</td>
<td>0</td>
<td>1 -1</td>
<td>2 -2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Gay/lesbian issues</td>
<td>0</td>
<td>0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Shyness, being assertive</td>
<td>0</td>
<td>0</td>
<td>2 0</td>
<td>0 0</td>
<td>0 +2</td>
<td>2 -2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Self-esteem, self-confidence</td>
<td>1 -1</td>
<td>2 0</td>
<td>3 -3</td>
<td>1 +2</td>
<td>2 0</td>
<td>3 -2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Loneliness, homesickness</td>
<td>1 +1</td>
<td>2 -2</td>
<td>1 0</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Depression</td>
<td>2 0</td>
<td>1 +1</td>
<td>2 -2</td>
<td>0 0</td>
<td>1 0</td>
<td>2 -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Anxiety, fears, worries</td>
<td>2 -1</td>
<td>2 0</td>
<td>3 -3</td>
<td>2 -1</td>
<td>3 -1</td>
<td>2 -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Irritable, angry, hostile feelings</td>
<td>0 0</td>
<td>1 0</td>
<td>1 -1</td>
<td>0 +1</td>
<td>0 +1</td>
<td>1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Physical stress (headaches, stomach</td>
<td>2 -2</td>
<td>2 -1</td>
<td>2 -1</td>
<td>0 0</td>
<td>2 -1</td>
<td>2 -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pains, muscle tension, etc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Sleep problems</td>
<td>0 +1</td>
<td>1 0</td>
<td>2 -2</td>
<td>2 -1</td>
<td>2 -2</td>
<td>3 -3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Eating problems</td>
<td>0 +1</td>
<td>0 0</td>
<td>2 -2</td>
<td>0 0</td>
<td>2 -2</td>
<td>1 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem area</td>
<td>Client A</td>
<td>Client B</td>
<td>Client C</td>
<td>Client D</td>
<td>Client E</td>
<td>Client F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td>Initial</td>
<td>Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Alcohol and/or drugs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. Suicidal feelings/behavior</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Adjustment to the university</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of problem areas evidencing decreased severity</td>
<td>6</td>
<td>4</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of problem areas evidencing increased severity</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of problem areas evidencing no change in severity</td>
<td>13</td>
<td>15</td>
<td>12</td>
<td>14</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. 0 = not a problem, 1 = mild, 2 = moderate, and 3 = severe. The sign in front of each change score indicates the direction of change (+ = increased severity, - = decreased severity).
Appendix B

Consent Form for Research Participants
Consent Form for Research Participants

A research project is being undertaken at the University of Illinois Counseling Center by Drevis L. Hager of the Western Michigan University Department of Counselor Education and Counseling Psychology. The aim of this project is to better understand the processes involved in counseling and what occurs in certain kinds of sessions. Another aim of the project is to improve a method of measuring what happens in counseling. To do this, counselors and clients who agree to participate will fill out a short questionnaire at the end of each session; this will take approximately 30 seconds. Also, each counseling session will be recorded on audiotape and a number of segments from these tape recordings will be statistically analyzed.

If you should agree to participate, all information on the tapes will be kept confidential and your name will not be revealed to anyone other than the above named researcher. The audio tapes will be erased after the data are extracted.

I, (your name) ___________, understand the nature of the research to be undertaken and I agree that information obtained during the course of my counseling may be used for research purposes. This permission is given with the understanding that participation is voluntary, that all information collected is confidential, that this information will be treated in a professional manner, and that safeguards will be taken to ensure my anonymity. I also understand that my counselor will not have access to any information collected about me in this research project, that participation or nonparticipation in this project will not affect the quality or quantity of services provided to me, and that I have the right to discontinue participation in the project without penalty.

Signed _____________________________________________

Address ____________________________________________________________________________

Telephone # __________________________________________________________________________

Date ____________________________________________________________________________

If at any time you have any further questions about this project and your participation, you may contact:

Drevis Hager
University of Illinois Counseling Center
212 F.H.T. Student Services Building
610 E. John St.
Champaign, IL 61820
(217) 333-3704
Appendix C

Pre- and Postcounseling Symptom List
**PLEASE RATE ANY OF THE PROBLEM AREAS THAT ARE OF CONCERN TO YOU BY CIRCLING THE APPROPRIATE NUMBER (0, 1, 2 or 3).**

<table>
<thead>
<tr>
<th></th>
<th>Not a Problem</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
<th>Rank the three most important problems (see instructions below)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Schoolwork and grades</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>Procrastination, getting motivated, managing time</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3.</td>
<td>Test anxiety</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4.</td>
<td>Stage fright, speaking anxiety</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5.</td>
<td>Decision about major/career</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6.</td>
<td>Relationship with friends</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7.</td>
<td>Relationship with roommate</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8.</td>
<td>Relationship with romantic partner</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9.</td>
<td>Relationship with family and parents</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10.</td>
<td>Loss/death of significant person</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11.</td>
<td>Sexual matters</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12.</td>
<td>Gay/lesbian issues</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>13.</td>
<td>Shyness, being assertive</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>14.</td>
<td>Self-esteem, self-confidence</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>15.</td>
<td>Loneliness, homesickness</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>16.</td>
<td>Depression</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>17.</td>
<td>Anxiety, fears, worries</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>18.</td>
<td>Irritable, angry, hostile feelings</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>19.</td>
<td>Physical stress (headaches, stomach pains, muscle tension, etc.)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>20.</td>
<td>Sleep problems</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>21.</td>
<td>Eating problems</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>22.</td>
<td>Alcohol and/or drugs</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>23.</td>
<td>Suicidal feelings/behavior</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>24.</td>
<td>Adjustment to the university</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>25.</td>
<td>Other (please specify)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Now using the boxes to the right, please put a "1" in the box next to the most important issue or problem area. Then put a "2" in the box next to the second most important area and a "3" next to the third most important. PLEASE RANK NO MORE THAN THREE PROBLEM AREAS. (If there is only one issue, please stop at "1").
Appendix D

Postsession Questionnaire
Please rate this session by placing an "X" at the appropriate point on each line.

This session was:

<table>
<thead>
<tr>
<th>X X X X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>Good</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X X X X X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valuable</td>
<td>Worthless</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X X X X X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shallow</td>
<td>Deep</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X X X X X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>Empty</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X X X X X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Special</td>
<td>Ordinary</td>
</tr>
</tbody>
</table>

Today's date ____________________________

Your code ____________________________

Appendix E

Therapist Data Sheet
Your code: ____________ Date: ____________

Years of post-doctoral experience doing therapy (if intern put "0"): ______

Name of degree program (Counseling Psy., Clinical Psych., etc):

____________________________________________________________________

Please characterize your typical way of working with a client by rank-ordering the top three dimensions in each grouping.

I try to focus mostly on the client's:

_____ thoughts
_____ feelings
_____ perceptions
_____ sensations
_____ behaviors

In regard to:

_____ past events
_____ present events
_____ future events

The content of the sessions is usually about:

_____ the person's experience of him/herself
_____ the person's experience of environmental situations
_____ the person's relationships with others
_____ the person's relationship with me
_____ other

Please rank-order the top three theoretical positions which have most influenced the way you work.

_____ Freudian
_____ Adlerian
_____ Jungian
_____ Sullivanian
_____ Gestalt
_____ R.E.T.

_____ Behavioral
_____ Cognitive/Behavioral
_____ Cognitive
_____ Client-Centered
_____ Existential
_____ Experiential
_____ Other

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Appendix F

Code of Ethics
Code of Ethics

Anyone who listens to personal interviews, particularly therapy interviews, is the recipient of professional confidences and is expected to maintain professional conduct and treat all data as strictly confidential. It is essential that you consider all tape-recorded or transcribed interview material as private and personal communication from a vulnerable individual who would not willingly share his confidences with unfriendly strangers. Those who were interviewed understood that interview materials would be used only for serious scientific purposes, that those using the materials would maintain professional conduct and ethics, treating their confidences with respect.

This means that you have two professional responsibilities. You are to maintain at all times a serious, respectful attitude toward the confidences that you are receiving and toward the individuals who have been willing to share their private selves for scientific purposes. Most important, you are not to discuss or refer to the content of the materials with anyone who is not directly connected with your rating task. The reason for this rule is important. Any violation not only will be unethical and abuse professional agreements, but will also place the psychotherapeutic professions and therapy research generally in an unfavorable light. Any indiscretion on your part might deter people from seeking counseling or therapeutic help or from serving in research projects because they understandably fear that their privacy might be infringed.

I understand the importance of maintaining confidentiality and I agree to not talk about any of the contents of any of the tapes that I will hear with anyone not directly involved in this research project. I understand that if I should want to discuss the contents of any of the tapes, that I should consult with Drevis Hager, and no one else. I also understand that I will be held fully accountable for maintaining confidence.

Signature ____________________________

Print your name ____________________________

Date ____________________________

Adapted from The Experiencing Scale: A Research and Training Manual (p. 55) by M. Klein, P. Mathieu, E. Gendlin, & D. Kiesler, 1969, Madison: University of Wisconsin, Bureau of Audio-Visual Instruction.
Appendix G

Directions for Running Ratings
Directions for Running Ratings

1. Assign an EXP score to each client statement block. A statement block is any statement that stands alone (is not contained as part of the therapist's response).

2. As always, consider the statement in context. (If the client makes a Level 1 statement, but it is used to elaborate on a statement at Stage 6, then the entire statement is at Level 6.)

3. A client statement block can have more than one level attached to it provided that there is a clear shift from one level to another, and if the shift is not simply an elaboration of a particular level.

4. Very short statements (such as "yah," "a-ha," "that's it," etc.) can be rated provided that it is reasonably clear that the client is experiencing at a particular level. Clues to the client's possible experiencing can be found by examining the preceding client and therapist statements. Do not assign a score to a very short statement if you cannot be reasonably sure about the level. If the statement is not ratable, mark it "NR."

5. Make your ratings independent of each other. Use a pencil. Clearly indicate where you think a shift in levels occurs by drawing a line so that it can be clearly seen where one level stops and another begins.

6. Compile a summary sheet for that segment.

7. After you have rated the entire segment and all of your ratings are final, then confer with each other about discrepancies and jointly create a master rating that you both can agree on. This master rating must include all levels accompanied by lines which clearly indicate where one level stops and another begins. Make this final rating with a colored pen.
BIBLIOGRAPHY


Richardson, B. (1986, July 30). Personal communication. University of Illinois, Department of Mathematics & Statistics, Urbana, IL.


psychotherapy with schizophrenics. Madison: University of Wisconsin Press.


Wasserman, S. (1986, July 30). Personal communication. University of Illinois, Department of Psychology, Urbana, IL.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.