MEASURING TREATMENT ADHERENCE OF EXPERIENTIAL SYSTEMIC THERAPY AND SUPPORTED FEEDBACK THERAPY IN THE ALCOHOL RECOVERY PROJECT.

by

JAMES EDWARD THOMPSON

B.A., Simon Fraser University, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Counselling Psychology

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1995

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Counselling Psychology

The University of British Columbia
Vancouver, Canada

Date April 27/95

ABSTRACT

The purpose of this study was to answer three research questions concerning the measurement of treatment adherence of Experiential Systemic Therapy (ExST) and Supported Feedback Therapy (SFT) in the Alcohol Recovery Project. First, is the Adherence Rating Scale Reliable and Valid? Second, to what extent do the therapists adhere to and emulate the respective treatments described in the ExST and SFT manuals? Third, are individual and couple's ExST treatment, as performed, indistinguishable from one another and different than SFT treatment?

Observer ratings of 120 video taped sessions were used to establish reliability and validity of tating scale, and to measure treatment adherence in terms of extensiveness, emulation and distinctiveness. A cluster analysis of the session profiles was conducted, establishing clear support for adherence and emulation. The clusters differentiated between ExST and SFT treatment modalities for 97% of the sessions, and did not differentiate between individual and couple's ExST. Treatment modality was also shown to be highly related to cluster indicating high levels of adherence.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGMENTS

I would like to take this opportunity to acknowledge the professional and personal support I have received while working on this thesis.

Professional support has included recommendations, editing information, and lots of guidance from my thesis committee, John D. Friesen, Robert F. Conry and Beth Haverkamp. Special thanks must go to Dr. Conry for his help in establishing the study's methodology and in the analysis of the data. The development of the rating scale required the involvement of John D. Friesen, Robert F. Conry, Darryl Grigg, Warren Weir, Cheryl Bates, Dan Mitchell and Jennifer Newman, for which I am grateful. I also want to thank Donna Leigh Brown and Lorna McDougall for their participation in the rating of tapes. Of course, many thanks go to the therapists and staff of TARP for their work in taping, sorting and storing the sessions.

Personal support has also been vital in my getting through this study. I'd like to thank the men in my men's group for their patience and encouragement. My son Jesse has also been supportive and accomodating through all the time and energy that this work has taken, and has helped me remember that life is for living, not just for measuring. Special love and thanks go to Donna Leigh Brown for her willingness to listen, her confidence in my ability to persevere and her warm heart. Finally I'd like to say thanks and farewell to my mom who died in the middle of this project, but who has provided a lifetime of support and helped me learn to have faith in myself.

Measuring Treatment Adherence of Experiential Systemic Therapy and Supported Feedback Therapy in The Alcohol Recovery Project.

Chapter 1: Introduction

In comparative outcome research, the outcomes of two or more distinct therapies are compared with one another to determine the relative effectiveness of treatment (Elkin et al., 1989). Obviously, it is crucial to specify what procedures are planned to be used in each therapeutic modality to produce a particular outcome (Kazdin, 1987; Luborsky & DeRubeis, 1984; Shapiro & Shapiro, 1987). However, in order to measure the effectiveness of a treatment, one concern that has recently begun to be seriously studied is whether the therapies were actually delivered as described (Hill, O'Grady, & Elkin, 1992; Quay, 1979). As Peterson, Homer and Wonderlich (1982) put it, "inadequate assessment of the independent variable may thus render conclusions about the functional relationship between the dependent target behavior and the independent treatment variable suspect" (p. 478). Beutler (1993) commented that there is often "a discrepancy between what therapists say they do and what observations indicate they actually do" (p.409). Moncher and Prinz (1991) recently reviewed 359 treatment outcome studies and found that the majority (55%) essentially ignored the issue of *treatment fidelity* (or *treatment adherence*), although they commented that this is rapidly becoming unacceptable.

According to the literature, there are three important questions to answer concerning treatment delivery that are critical to support comparative outcome research. First, how much of what intervention is being done by a therapist in each session? Second, how closely does the therapist approach the ideal of the

particular intended treatment? And third, what elements that are unique or common to each treatment format as practiced are a part of the treatment protocol? The combined answers to these three questions provide the most complete measure of the adherence of the treatment as delivered to the treatment protocols. Only rarely, if ever, are all three of these issues dealt with in outcome research.

The Alcohol Recovery Project is a major study comparing two treatment modalities, Experiential Systemic Therapy (ExST) (Friesen, Grigg, Peel, & Newman, 1989) and Supported Feedback Therapy (SFT) (Grigg, Friesen, Weir, & Bate, 1989). The key interventions of ExST were developed to be used in both individual and couples therapy, so both approaches were studied. Therefore there are three experimental treatment modalities explored in this study: (a) individual ExST; (b) couple's ExST; and (c) SFT.

The purpose of this investigation is to measure the level of treatment adherence of the three modalities in the project, in terms of *extensiveness* (how much of what), *emulation* (how close to the therapeutic model), and *discrimination* (unique and common elements). A thorough measure of treatment adherence is essential to establishing support for any findings indicating outcome differences.

This area of research has generated a plethora of terminology and occasionally contrasting definitions for specific terms. A review of the related literature is essential to help define the many overlapping terms used. Also, this review is important in order to clarify the selection of critical concepts involved in formulating the specific research questions and hypotheses of this thesis.

Chapter 2: Literature Review

Issues of Treatment Adherence

Introduction to Terms. According to the dictionary (Flexner, Stein, & Su, 1980), to adhere means to hold fast, or to hold firmly. In comparative outcome research, there have been several ways of answering the question of what holding fast and firmly to a treatment protocol means. Perhaps the most common use of the phrase "treatment adherence" in this research refers to the client's following of a treatment plan (for example, Vermilyea, Barlow, & O'Brien, 1984). There is however, a second use of the phrase which refers to the therapist's adherence to a specific treatment modality and it is this meaning that is expanded upon in this thesis.

The issue of appraising whether treatment was actually delivered as intended has been variously labeled *treatment integrity* (Butler, Fennell, Robson, & Gelder, 1991; Dobson & Shaw, 1988; Gresham, 1989; Kazdin, 1986, 1990, 1993; Shapiro, 1987; Szapnocznik et al., 1989; Waltz, Addis, Koerner & Jacobson; Yeaton & Sechrest, 1981), treatment fidelity (Boruch & Gomez, 1977; LeLaurin & Wolery, 1992; Moncher & Prinz, 1991; Startup & Shapiro, 1993b), *representativeness of treatment* (Kazdin, 1986; Kendall & Lipman, 1991; Shapiro & Shapiro, 1983), *treatment consistency* (Boruch & Gomez, 1977; Moncher & Prinz, 1991; Silove, Parker & Manicavasagar, 1990), *integrity of the independent variable* (LeLaurin & Wolery, 1992; Peterson et al., 1982; Sechrest, 1982), *procedural reliability* (Thomas, 1989), or treatment adherence (Elkin, Parloff, Hadley, & Autry, 1985; DeRubeis & Feeley, 1990; Hill et al., 1992; Koenigsberg, Kernberg, Appelbaum, & Smith, 1993).

Moncher and Prinz (1991) point out that a measure of treatment fidelity "is necessary to maintain internal validity and ensure a fair comparison of treatments" (p. 249). They add that fidelity also impacts on statistical power as well. LeLaurin and Wolery (1992) list four reasons for measuring treatment fidelity: to provide evidence that the intervention actually was implemented; to identify instances of deviation from the planned treatment; to help understand the generality and limits of interventions; and to identify the necessity of intervention characteristics. Sechrest (1982) says it is important to know precisely what treatment is used, for construct validity and for external validity. "If we do not know precisely what intervention we evaluated, we can hardly be on safe grounds in generalizing about its usefulness in other settings" (p. 73). Essentially, all of these terms (integrity, adherence, fidelity, etc.) refer to identifying whether the treatment was given as planned.

One reason that measuring treatment integrity is important is because research has shown a strong positive relationship between integrity and treatment effectiveness (DeRubeis & Feeley, 1990; Gresham, 1991; Luborsky, McLelland, Woody, O'Brien, & Auerbach, 1985). According to this research, when the actual treatment rigorously matches the planned treatment, outcome improves. Poor or ambiguous results may be due to the lack of integrity in treatment.

The simplest form of measuring treatment integrity would be a yes/no checklist, asking "Did the treatment include planned intervention A, B, C, etc.?"(Walz et al., 1993). This approach is in fact very similar to one of the validity checks used by Schefft and Kanfer (1987) where 5-minute audio-tape segments from three treatment groups were played and raters were asked to identify which treatment group each segment belonged, A, B, or C, based on the respective

manuals. However, to be able to thoroughly identify how the treatment as given adheres to the treatment as planned, more than a yes/no checklist is required.

There are a few other related concepts in the literature that are important to consider, some of which will be interpreted more fully in the definitions section. Essentially these terms are subdivisions of adherence as it has been defined so far: treatment as planned. *Extensiveness of treatment* considers how much of each intervention is used in a session (Hill et al., 1992). *Treatment strength* considers duration and intensity (Yeaton & Sechrest, 1981). *Treatment purity* examines the intrusion of non-planned interventions (Luborsky et al., 1985). *Therapist competency* (Kendall & Lipman, 1991; O'Malley et al., 1988; Shaw & Dobson, 1988; Waltz, et al., 1993) or *therapist skillfulness* (LeLaurin & Wolery, 1992; Schaffer, 1982) consider the quality of the therapist's work. *Therapeutic model emulation* is a new formulation aimed at estimating how close to the therapeutic model the therapist acts. *Therapist drift or deviation* looks for effects in therapist performance over time (Gresham, 1989; Hill et al, 1992; Svartberg, 1989b). Finally, *discrimination, differentiation* or *distinctiveness of treatments* considers the elements unique to or common to different treatment modalities (Dobson & Shaw, 1988).

For the purposes of this thesis, I propose a three stage analysis of what is now a rather broad definition of treatment adherence, including many of the above concepts: (a) extensiveness of treatment; (b) therapeutic model emulation; and (c) treatment discrimination.

Definitions. A measure of extensiveness of treatment records both the frequency of use of an intervention and measures treatment integrity. Treatment integrity (often referred to as adherence) refers to the degree to which a treatment

follows a specific, defined procedure (Hill et al., 1992; Moncher & Prinz, 1991; Peterson, 1977; Sechrest, 1982; Sechrest, West, Phillips, Redner & Yeaton, 1979; Shaw & Dobson, 1988; Szapocznik, Kurtines, Santisteban, & Rio, 1990; Yeaton & Sechrest, 1981). Dobson and Shaw (1988) suggest estimates of treatment integrity be based on "the extent to which the treatment is conducted according to the demands of the treatment manual" (p. 675). Luborsky and DeRubeis (1984) recommend calculating "the ratio of the use of the treatment techniques in the therapists' own manual over the use of all techniques (all techniques include those in other manuals as well as those in own manual)" (p. 6), a measure they define as treatment purity. A similar approach is used by Wills, Faitler and Snyder (1987) and Snyder, Wills, & Grady-Fletcher (1991a) when they classify therapist behaviors in three categories: "(a) consistent with and specific to the prescribed treatment modality, (b) consistent with but not specific to that treatment, or (c) inconsistent with that treatment" (p. 139). However, Hill et al. (1992) point out an important distinction between "how much therapists used techniques considered appropriate to their treatment approach" and "how adequately they apply the techniques" (p. 74). Waltz et al. (1993) also suggest that "successful tests of treatment integrity include both an assessment of therapist adherence to the treatment protocol and a determination that the interventions are being performed competently", (p. 620).

For the purposes of this study the extent to which the therapists used the techniques based on their own treatment manuals will be considered the basis for a measure of treatment integrity. Following Hill et al. (1992) this study assumes that therapists who rate higher in extensiveness on those items designed to tap the essential behaviors of their treatment approach (rather than on items designed

to tap the essential behavior of other treatment approaches) to show treatment integrity.

Strength of treatment refers to the intensity, frequency, amount, relevance, or focus of a therapist's intervention and its impact on a particular client, and grows out of a medical model of treatment (Sechrest, 1982; Sechrest et al., 1979; Yeaton & Sechrest, 1981), where therapy is seen as a sort of drug, to be given in measured doses. Stiles and Shapiro (1989) argue against the implicit assumptions of viewing therapy through a drug metaphor, where "the supposed `active ingredients' are process components -- therapeutic techniques such as interpretation, confrontation, reflection, self-disclosure, challenging assumptions, focusing on affect, efforts to give support, or (more abstractly), empathy, warmth or genuineness" (p. 522). They question whether such component names actually signify pure ingredients, whether it is really the "active ingredients" being measured, and whether the "active ingredients" of therapy are contained in the therapist's behavior. Additionally, estimating treatment strength requires value judgments on the part of a rater that may make agreement between raters difficult to achieve.

The term therapist competency refers to attempts to measure the quality of treatment given (Rounsaville, O'Malley, Foley, & Weissman, 1988; Shaw & Dobson, 1988; Svartberg, 1989a, 1989b; Vallis, Shaw, & Dobson, 1986). This is more than a measure of the time spent using a particular technique (extensiveness) but is instead a measure of the skill with which techniques and strategies are implemented, a very common consideration in the literature (Baucom, Sayers, & Sher, 1990; Jacobson, 1991c; O'Malley et al., 1988; Stiles, Shapiro, & Elliott, 1986; Strupp, Butler, & Rosser, 1988; Waltz et al., 1993).

Dobson and Shaw (1988) point out that the use of treatment manuals may emphasize fidelity (or integrity) over competency. They suggest that correct techniques may be delivered in a way that is awkward, poorly timed, inappropriate, or as part of a poor therapeutic relationship. Waltz et al. (1993) emphasize that measuring therapist competence must be context-sensitive, considering stage of therapy, client difficulty, and client presenting problems, and probably requires highly experienced raters to evaluate their performance.

Jacobson's (1991c) discussion of competency in behavioral marital therapy indicates how much different researchers may disagree about just what defines competency in a particular approach. Many authors, including Jacobson (1991c), Strupp (1986) and Snyder and Wills (1991), also argue that competency is much more difficult to specify. For example, Snyder and Wills (1991) argue that competency is best measured by documenting adherence to distinct treatments along with treatment efficacy or outcome and avoiding estimates of skillfulness. However, Goldfried, Greenberg and Marmar (1990) suggest that competency scores can be reliably obtained and are important to help therapists follow therapeutic guidelines. The following definition of competency gives a framework for evaluation:

> In summary, the competent therapist manifests (a) a theoretical or conceptual framework to guide interactions, (b) a memory of the patient's central issues, (c) the skillful use of intervention techniques to promote the desired changes in behavior or the conditions necessary for change, and (d) the knowledge of when to apply (and when not to apply) these interventions (Shaw & Dobson, 1988, p. 667)

However, like strength of treatment, competency requires a value judgment on the part of the rater, suggesting a competent therapist does the "right thing" at the right time, and that the "right thing" is whatever technique is being studied.

Adding to the conflict and confusion concerning the use, definition and measurement of the term competency, Koenigsberg et al. (1993) define the skillful conduct of psychotherapy quite differently. They refer to skill as the use of non-specific therapeutic factors like tactfulness, warmth or empathy, or the ability to tolerate affect (as opposed to the adherence to unique treatment-specific interventions). In the present study, these types of interventions are essentially measured by the Common subscale items both in terms of extensiveness and emulation. In yet another twist, Johnson and Greenberg (1985) operationally defined therapist skill in terms of the quality of the therapeutic alliance.

While both treatment strength and therapist competency attempt to address an important aspect of adherence, the literature review has made it clear that there is no general agreement about what they are or how exactly to measure them, other than that "experienced raters" would probably be needed. In the effort to address this topic, and to be able to use master's level raters, a more useful term than either strength or competency may be *therapeutic model emulation* (a term coined by Darryl Grigg). Emulation measures how close to the manualized ideal a therapist acts (Sechrest et al., 1979). Especially in systemic approaches, where collaborative alliances between client and therapist are emphasized (as opposed to the model of the therapist as a more or less skilled "objective" expert delivering a technique as a pill), emulation may be a more accurate and acceptable term. In other words, it requires less subjective judgment to rate a therapist's behavior and/or a therapist-client interaction in terms of how closely the behavior

or interaction matches the therapeutic model rather than trying to evaluate competency or skillfulness (i.e. just the right thing at just the right time). Jacobson (1991c), Kendall and Lipman (1991) and Shaw and Dobson (1988) suggest that competency may not be a consistent trait, but a state that "varies over time, across cases and perhaps even across phases of treatment" (Shaw & Dobson, 1988, p. 667), a view that supports a more descriptive approach to rating what treatment was done in the actual session compared to the manualized ideal. They also state that competence tends to be rated idiosyncratically, according to the rater's own personal standards, and that explicit standards of therapist operations according to the theory under study are necessary. Also, Pinsof (1989) suggests that descriptive, rather than evaluative ratings permit analysis of the impact of therapists' behavior rather than prejudging its value. Measures of emulation, as opposed to strength or competency, are more descriptive than evaluative, and based on explicit standards. For this study, emulation is therefore our second measure of adherence.

Treatment discrimination is a concept particularly relevant to outcome comparison research. When comparing the outcomes of supposedly different treatment modalities, it is important to have raters able to discriminate between the treatments as practiced (Baucom et al, 1990; Startup & Shapiro, 1993b). Ratings of sessions comparing different psychotherapies based on manuals "can reveal the ways in which psychotherapies are distinct from each other, as well as areas of overlap among therapeutic approaches" (Luborsky & DeRubeis, 1984, p. 5). Research has shown the importance of non-specific therapist behaviors and general therapeutic skills in affecting outcomes, so it is critical to be able to clarify the extent to which therapeutic approaches being compared overlap and what

specifically makes them unique (Dobson & Shaw, 1988; Jacobson, 1991c; Jacobson & Addis, 1993; Kazdin, 1986; Luborsky, Woody, McLellan, O'Brien, & Rosenzweig, 1982; Moncher & Prinz, 1991; Shapiro & Shapiro, 1987; Silove et al., 1990; Stiles, Shapiro, & Elliot, 1986; Snyder & Wills, 1991; Snyder, Wills, & Grady-Fletcher, 1991b; Wills et al, 1987). Rounsaville et al. (1988) for example, in a study of therapist competence, found that many general therapist factors like exploration, warmth and lower negative attitude (a "good therapist" factor), accounted for 40% of the variance in therapist skill ratings. Waltz et al. (1993) recommends four types of items to assess treatment distinctiveness: (a) therapist behaviors that are unique to that treatment modality and essential to it, (b) behaviors that are essential to the treatment but not unique to it, (c) behaviors that are compatible with the specified modality, and therefore not prohibited, but neither necessary nor unique, and (d) behaviors that are proscribed. Before research considers differences in outcome, sufficient differences in treatment as practiced must be demonstrated to merit direct comparison (DeRubeis, Hollon, Evans, & Bemis, 1982; Snyder & Wills, 1991). In a time when there are over 450 types of psychotherapy and there is a trend towards integrating ingredients from different therapies, it is important to attempt to record both the unique and overlapping ingredients of therapies being compared (Karasu, 1986). It may be the overlapping elements that represent unifying principles that are primarily responsible for therapeutic outcome. In this study, this concern is addressed by examining therapist behaviors in sessions using a rating scale that has items expected to be unique to each treatment modality as well as items expected to overlap both modalities.

So, out of the plethora of terms in the literature concerning treatment adherence, this study uses three components which appear to encompass the field of adherence: extensiveness, emulation and differentiation. The following section reviews the extensive literature describing how these three components have been previously measured.

Approaches to Measuring Adherence. The three components of treatment adherence proposed for this study have been researched using many different approaches for estimation. Sechrest et al. (1979) suggest several a priori forms of assessing the likely strength and integrity of treatment, including assessing the theoretical premises, the clarity of the treatment plan, the commitment of the staff involved, the supervisory plan, plans for the documentation of the service delivery, and the complexity and difficulty of treatment. Quay (1979) adds that the intended duration of treatment may also be a factor in treatment integrity (i.e. 6-8 weeks, or 24 sessions, etc.).

Moncher and Prinz (1991) provide an extensive review of the issue of treatment adherence in their analysis of 359 outcome studies, pointing out many of the key measures of adherence. For example, the therapist's amount of training, and years of experience with the treatment approach used in a study, provide very rough estimates of the treatment as practiced. Yet Shapiro and Shapiro (1983) report that even this rudimentary bit of information is missing from 36.5% of the 143 outcome studies examined in their meta-analysis. Elkin, Pilkonis, Docherty, & Sotsky (1988) stress the importance of equivalence in training and experience in outcome comparison research. They suggest different therapists be assigned to different treatment conditions, while ensuring they have comparable commitment, training, experience and competence in their respective

therapeutic approaches. Rounsaville, Chevron, Weissman, Prusoff, & Frank (1986) describe in detail the training program for the therapists involved in a larger study, which included reviewing the training manual, a didactic seminar and supervised casework. Startup and Shapiro (1993b) point out that careful training and screening of therapists must be combined with observer based adherence measures to confirm treatment integrity when the same therapists are providing two different types of therapy.

Self-report scales have also been used to determine therapist orientation, which is suggested to support claims of treatment integrity (Moncher & Prinz, 1991). For example, Stiles, Shapiro and Elliott (1986) report that "therapists who merely state their allegiance to alternative schools have shown appropriate, systematic differences" (p. 166). Shapiro (1987) suggests that "random checking of treatment integrity, or the use of a treatment implementation checklist on which the therapist monitors compliance with each step of the procedure" (p. 294) may be an effective way of keeping treatment consistent across therapists. Hahlweg et al. (1990) used self-reports by both the therapist and family members to evaluate the session. Newman, Kopta, McGovern, Howard, and McNeilly (1988) describe several self-report measures designed to reconstruct the therapist's in-session behavior. Therapists have completed self-report measures of adherence after sessions (Foley, O'Malley, Rounsaville, Prusoff, & Weissman, 1987; Greenberg & Goldman, 1988; Gresham, 1989; Rounsaville et al., 1987). However, Beutler and Hill (1992) point out that it is very common for there to be a discrepancy between how therapists describe what they do and what observers judge to have occurred. Post-session client-report scales have also been analyzed to identify the treatment

approach used with some degree of success(Hogg & Deffenbacher, 1988; Silove et al., 1990).

Avis and Sprenkle (1990) report on an empirical estimate of family therapy training that is based on essays written on four clinical case vignettes with differing presenting problems. Newman et al., (1988) also recommends using case vignettes rather than video tapes because greater levels of error variance are associated with video-taped interviews. On the other hand, Weiss, Marmar, and Horowitz (1988) found ratings based on video-taped sessions more reliable than audio. They also report a study indicating ratings based on verbatim transcripts were more reliable than audiotapes.

Continued individual or group supervision and feedback through the period of the study increases the likelihood of treatment integrity and strength (Baucom et al., 1990; Fonagy & Moran, 1990; Johnson & Greenberg, 1985; Rounsaville et al., 1988; Stiles, Shapiro, & Firth-Cozens, 1989; Wills et al, 1987; Zweben & Pearlman, 1983). Jacobson (1991c) suggests that the experience of the trainer and supervisor be considered in outcome studies. Supervisors or experts can make educated judgments of therapist adherence or skillfulness (for example, Elkin et al., 1985), although their judgments tend to be subjective. Supervisors may rate treatment procedures using a checklist to estimate fidelity and strength of treatment (Greenwood, Terry, Arreaga-Mayer & Finney, 1992). Supervisors may observe those carrying out the planned procedure and simply give constructive verbal feedback after sessions (Caplan, Vinokur, Price, & van Ryn, 1989).

The use of detailed training manuals and treatment manuals heralded a real advance in the assessment of treatment integrity (Luborsky & DeRubeis, 1984; Strupp et al., 1988), competency (O'Malley et al., 1988) and treatment

discrimination (Luborsky et al., 1982). The use of treatment manuals is rapidly spreading across many psychotherapeutic approaches (Glass & Freedman, 1986). Therapy tends to be more consistent, and more effective when based on manuals or when therapists are trained using manuals (Dobson & Shaw, 1988; Lambert & Ogles, 1988; Luborsky et al., 1982; Luborsky et al., 1985; Phillips, 1987; Rounsaville et al., 1988). In a study of training using manuals, Henry, Strupp, Butler, Schacht, and Binder (1993) add a note of caution that while trainees improved in their adherence to manualized technical interventions, they also unexpectedly deteriorated in certain interpersonal and interactional aspects of therapy. This led the authors to question whether "greater 'control' of the therapy variables is truly achieved via manuals and adherence scales. Attempts at changing or dictating specific therapist behaviors may alter other therapeutic variables in unexpected and even counterproductive ways" (p. 438).

It is easier to determine reliable and valid differences and similarities between treatments, when the manuals fully describe the treatments (Beutler, 1993; Dobson & Shaw, 1988; Kendall & Lipman, 1991; Woody, McLellan, Luborsky, & O'Brien, 1986). Manuals provide a basis for standardization and replication (Dobson & Shaw, 1988; Elkin et al., 1988; Kendall & Lipman, 1991; Szapocznik et al., 1990). Objective measures of treatment integrity, strength, and discrimination based on manuals are more reliable and explicit (Luborsky and DeRubeis, 1984). On the other hand, it may be unethical to create manuals that are not flexible enough to respond to unique client needs (Aradi & Piercy, 1985). LeLaurin and Wolery (1992) suggest that operational definitions of the independent variable (treatment) should be based on clearly defined and

appropriate theory in order to evaluate and integrate any findings into the existing literature.

Outcome comparison studies that include measures of adherence based on detailed manuals can provide a basis for rigorous debate over what should constitute a treatment format, like the debate between Wills et al (1987), Snyder et al. (1991a, 1991b), Snyder & Wills (1991), Gurman (1991), Baucom and Epstein (1991), Markman (1991), Johnson and Greenberg (1991) and Jacobson (1991a, 1991b, 1991c) over what constitutes behavioral and insight-oriented marital therapy. Startup and Shapiro (1993) use adherence measures to confirm factor analysis of two dimensions of cognitive therapy in an earlier study and were unable to show support for two other dimension structures in a second study.

Treatment manuals have often formed the basis for observer rating scales of therapist adherence or competency (Koenigsberg et al. (1993); Luborsky & DeRubeis, 1984; Newman et al., 1988; Piercy, Laird, & Mohammed, 1983; Startup & Shapiro, 1993b; Svartberg, 1989a, 1989b). When therapists are expected to perform different treatment modalities as part of the experimental design, it is especially important to document their adherence to the respective treatments (Snyder et al., 1991a; Startup and Shapiro, 1993b; Wills, et al., 1987). One unique way Johnson and Greenberg (1985) helped show equivalence in therapist skill, was by noting there were no significant differences between their two groups of therapists in the quality of the therapeutic alliance they were able to create. Goldfried et al. (1990) report that the use of guidelines like treatment manuals concerning treatment adherence is now required in research funded by NIMH.

Monitoring treatment integrity through observer ratings provides the most definitive assessment of therapeutic adherence available (Kazdin, 1986, 1987;

Phillips, 1987; Snyder & Wills, 1991), and has been accomplished in many different ways as well.

Hogg and Deffenbacher (1988) applied observer rating scales for assessing adherence to different forms of group therapy. Supervisors, as well as independent evaluators rated therapist performance using rating forms (Butler et al., 1991; Foley et al., 1987; Rounsaville et al., 1986). Greenberg and Goldman (1988) have student therapists rate videotaped psychotherapy sessions using scales that measure empathy, warmth and genuineness as part of their training to develop their perceptual discrimination of the conditions. Goldberg, Wool, Tull, and Boor (1984) use a similar training approach using a codable supervision format to analyze video-taped sessions as a group. Elkin et al. (1988) mention a "certification" program used as part of their attempt to ensure comparable skill and consistency across treatments. Henry et al. (1993) use pre- and posttraining adherence ratings to evaluate training. Moncher and Prinz (1991) report that ratings of student therapists have been used to aid in training, and to provide feedback in practice or in research. They note that role-played therapy has been observed and rated, to establish a consistent standard for therapeutic practice in a study. Responses to hypothetical therapeutic situations have also been rated. Sessions have been rated live, or using audio-tapes or video-tapes. Computers have also been used to help automate rating and data collection. Sessions have also been rated on the basis of verbatim transcripts (Weiss et al., 1988). Initial screening for competence, and a content analysis of systemic versus individual therapist statements has been used (Zweben & Pearlman, 1983). Therapist expertise has also been evaluated using a questionnaire based on a videotaped family therapy session reenacted by professional actors (Avis & Sprenkle, 1990).

Hahlweg et al. (1990) uses a coding system to describe the content of each session, and several rating scales to assess global and specific therapist competencies. A common approach to rating treatment adherence and/or discrimination of taped sessions is the examination of the verbal response mode usage of therapists (Elliott et al., 1987; Hill, 1982; Shapiro & Firth, 1987; Stiles et al., 1989). Another approach is categorizing therapist interventions from short, randomly selected session segments into general categories from an implementation checklist (Johnson & Greenberg, 1985) or dimensions of "core elements" (Schefft & Kanfer, 1987). Koenigsberg et al., (1993) working with borderline patients, developed yet another variation on measuring adherence and skill that involved comparing patient action resistance themes (like threats of suicide, transference, acting out in the session, etc.) in video-taped sessions with themes of therapists' responses. As well, they created distributions of indices of various categories of therapists interventions (like clarifications, expressive interventions, confrontations, etc.). In a very thoroughly designed study, Jacobson, Dobson, Fruzzetti, Schmaling and Salusky (1991) rated their therapists on two different scales, one measuring adherence and the other measuring competence. Rounsaville et al. (1987) used measures of both adherence and competence and also considered general therapeutic behaviors not specifically linked to manuals. The present study follows their example by measuring extensiveness (adherence) and emulation (competence) and adds to it by also considering treatment discrimination, which includes both general and unique therapeutic behaviors.

Various lengths of session excerpts have been rated, from whole sessions (e.g. DeRubeis et al., 1982), to 1- (Arnkoff, 1986), 2- (Greenberg & Goldman,

1988), 3- (Hill, 1982), 5- (Schefft & Kanfer, 1987; Snyder et al., 1991a; Wills et al. 1987), 10- (Johnson & Greenberg, 1985), 15- (Henry et al., 1993; Luborsky et al., 1982; Luborsky et al., 1985; Schefft & Kanfer, 1987) 20- (Hogg & Deffenbacher, 1988), or 30-minute excerpts (Avis & Sprenkle, 1990). Rating has been performed using Likert-type scales or checklists. Critical to the success of observer ratings is the quality of the rating scale developed and the characteristics of the raters themselves.

An interesting approach to assessing competency is proposed by Snyder and Wills (1991), when they suggest that observing treatment adherence within a study be complemented by comparing treatment efficacy between studies using similar treatment approaches. In other words, if two studies are evaluating behavioral marital therapy for example, the treatment outcomes should be comparable if the quality of treatment is similar in both studies.

One important factor in the development of a rating scale is finding a balance in process specificity or the degree of detail examined in the scale items (Pinsof, 1989).

> Like all methodological issues, process specificity involves trade-offs. The construction and use of highly specific coding systems and procedures is very time consuming, expensive, and complex. Global systems are easier to construct and use and have been able to predict outcome in individual studies with surprising consistency (p. 57).

Shaw and Dobson (1988) point out that competency measures vary widely in complexity from global ratings of therapist skillfulness, to rating specific sessions, to process-level ratings. In the present study, global ratings based on

observations of large (30-minute) segments of single sessions are analyzed on the level of overall ratings per therapist as well as overall ratings for each modality in terms of aggregated adherence and emulation measures. There are therefore important limitations inherent in the design of this study. The focus of the rating scale items is relatively global. It does not attempt to measure microskills or process variables like verbal response modes (Stiles et al., 1989), which is also important for a complete picture of adherence.

Hill et al. (1992) and Shaw and Dobson (1988) suggest it may be better to establish minimum cutoff points for therapists to exceed to be considered adherent to the approach. For example, in studies by Marmar, Gaston, Gallagher, and Thompson (1989) and Thompson, Gallagher & Breckenridge (1987), they rated therapist competency on one randomly selected tape per client. If the therapists' tape did not reach a preestablished cut-off score for competency, a second tape was evaluated, and if also found inadequate, that subject was dropped from the outcome analysis. While the therapists involved in the study were evaluated before being permitted to participate, which indicates some degree of competency, there are no cutoff points used in the present study. Henry et al. (1993) made a similar decision not to use a "red line" procedure for the following two reasons which also apply in this study: (a) the therapy emphasizes a variety of technical strategies, but does not specifically proscribe others, (b) the nature of their therapy does not lend itself to a priori determination of the optimal frequency of intervention techniques. In the present study, ExST in particular is designed to be flexible in terms of the inclusion of a wide variety of therapeutic techniques, as well as highly adaptive to client needs in terms of frequency of use of any intervention strategy. Items in the rating scale selected to be unique to SFT may

appear in an ExST session, but would be expected to be used much less frequently than in a SFT session. Also, specific items in either modality need not appear in any specific session.

Foley et al. (1987) suggest that "the influence of patient difficulty on therapist performance should be addressed whenever therapist performance is rated in therapeutic evaluation and outcome studies" (p. 215). The influence of characteristics of the clients (like patient hostility) may be an important factor in outcome studies. LeLaurin and Wolery (1992) recommend documenting the client's participation in the intervention as an important addition to adherence studies. The present study focuses on therapist behavior, and does not explore the impact of the client on therapy, a limitation offset by the random placement of clients into the respective treatment programs.

Pinsof (1989) points out several important concerns for family therapy research. He suggests that variables from both the therapist system (including the therapist's own family life, supervision, etc.) and the family system and subsystems, as well as relationships between the therapist and the family system and subsystems must be measured. He also suggests that change should be measured both within the session (traditionally the province of process research) and outside the session (traditionally the province of outcome research). This study of a systemic therapy attempts to include items that measure within-session client/therapist interactions, but the emphasis is on in-session therapist behaviors. This is an important limitation, and one that should be addressed in future research.

As the review above indicates, measuring adherence has become a complex arena, approached in many different ways. The following table is designed to provide an overview in outline form.

Table 1:

Summary of Methods for Estimating Adherence

A priori methods
- assess
  - -- theoretical premises
  - -- clarity of treatment plan
  - -- commitment of staff
  - -- supervisory plan
  - -- documentation plans
  - -- complexity & difficulty of treatment

Characteristics of the treatment program
- Adherence to expected duration of treatment
- Use of detailed training manuals
- Use of rating scales in training
- Rating of essays analyzing clinical vignettes

Characteristics of personnel
- Amount of training
- Years of experience
- Continued individual or group supervision
- Therapist orientation scales

Observer Ratings
- Judgment of therapist adherence by supervisors/experts
- Therapist self-report of adherence after sessions
- Client report scales of group or individual therapy
- Observer ratings of role-played or reenacted therapy in
  - whole sessions/specific situations
  - individual, couple's, or group therapy
- Observer ratings of sessions
  - -- live
  - -- audio-tapes
  - -- video-tapes
  - -- verbatim transcripts
  - -- automated data collection
- Rating whole sessions or 1-, 2-, 3-, 5-, 15-, 20-, or 30- minute excerpts
- Rating using scales or checklists
- Ratings of
  - -- verbal response modes
  - -- integrity/purity/adherence
  - -- competency/skillfulness/emulation
  - -- discrimination

Rating Scales and Raters. Newman and Scott (1988) point out that defining the construct of counselor performance is a complex task and difficult to reliably measure with a rating instrument. The use of training or therapeutic manuals as a basis for developing a rating scale is highly recommended (Luborsky et al., 1982). As well, it has been suggested that items in rating scales emphasize the concrete and observable, have high specificity, be easy to rate (Hill, O'Grady, & Price, 1988), and use extensive examples, especially with relatively naive raters. The more complex the treatment, the more difficult it is to assess integrity and emulation, and the more subjectivity intrudes (Gresham, 1989; Vallis, Shaw, & Dobson, 1986; Yeaton & Sechrest, 1981). Highly trained raters may be able to reliably rate these more subjective situations, but a rater-bias scale may also be necessary with such items (Moras & Hill, 1991).

Also, the detail of an item must reflect the intent of the study. A highly detailed interaction item may be appropriate for process research, where a broader, more general item may be more appropriate for measuring integrity (Gresham, 1989). It is important to include items expected to reflect unique elements of each therapy, as well as items expected to occur in each therapy being compared (including any control group). A common practice after rating sessions is to perform a factor analysis to verify these three categories or to identify any other associated items. In the process of developing a rating scale, questions from raters concerning any particular items may aid in developing items or their descriptions or examples.

Hill (1982) suggests three criteria for selecting raters: general intellectual functioning, motivation for the specific task, and ability to do a simulated task. Moras and Hill (1991) describe three variables that can influence the reliability

and validity of raters using process rating measures. These are also important in developing a high quality study of adherence. First, the therapeutic orientation, gender, experience and other characteristics of raters must be considered. Second, training may be required to use a scale consistently, and questions answered regarding specific items. The training process may include comparing the raters' scores to expert scores on standardized video tapes. Continued supervision of the raters and ongoing feedback from supervisors increases the reliability of their scores. Periodic comparisons with expert ratings also adds to reliability. At least two raters must rate each tape to provide a minimal level of reliability. Third, the nature of the inference or value judgment required in the rating scale may also effect the reliability and validity of rating data. Newman and Scott (1988) point out that differences in training, order of observation, personal values, biases and social pressure, as well as response sets (like leniency error or central tendency) may influence rater perception or evaluation. A study of rater bias (Hill, O'Grady & Price, 1988) suggests that when training is carefully done, individual rater characteristics and items requiring moderate levels of inference do not adversely effect reliability. The raters did however tend to rate facilitative items (communication style, warmth, rapport, etc.) in a way that was influenced by how much they liked and felt similar to the therapist and client. As these items are similar to the overlap items in The Adherence Rating Scale (TARS), it may be important to consider this possible source of error. Hill et al. (1988) also found that fatigue, waning interest or sensitivity and increased familiarity with the instrument as well as the length of their instrument (96 items) may have had some impact on the raters' reliability. For this study of the TARP project, a shorter

scale, shorter tape segments, and fewer sessions being rated will hopefully minimize these effects.

Research Questions

This study had two distinct phases of research, a preliminary phase concerned with establishing reliability of the rating instrument, and a second phase focused on adherence. The research question of the first phase was: Is The Adherence Rating Scale (TARS) and the rating process reliable and valid? In the second phase of research, the questions this study asks are: (a) To what extent do the therapists in The Alcohol Recovery Project (TARP) adhere to and emulate the respective treatments described in the manuals developed for individual and couples Experiential Systemic Therapy (ExST) (Friesen, Grigg, & Peel, 1988; Friesen et al., 1989) and Supported Feedback Therapy (SFT) (Grigg et al., 1989); and (b) Are the two forms of ExST similar to one another and distinct from SFT? These research questions can be partially answered by descriptive means, for example by reviewing therapist training, supervision, etc. However, a much more specific and objective analysis that answers both questions is possible through testing the following hypothesis using an observer rating scale.

Hypothesis

The clusters formed by similar scoring profiles will differentiate between treatment modalities, and the variance between these clusters of scores on the variables of extensiveness and emulation, in the subscales of ExST, SFT and Common items will be explained by treatment modality.

Chapter 3: Methodology

Research Design

Purpose of the study. The purpose of this study is to measure the therapists' treatment adherence of individual Experiential Systemic Therapy (ExST), couple's ExST and Supported Feedback Therapy (SFT) in the Alcohol Recovery Project (TARP). Based on the literature review, both a descriptive and quantitative analysis was undertaken. As a preliminary step, the observer rating scale, the rating process and rater bias were examined. Then, adherence was examined in terms of extensiveness, emulation and differentiation. This study is intended to support and further the comparative outcome research that was the central goal of TARP.

History of TARP. The Alcohol Recovery Project was a large-scale, comparative outcome, alcohol treatment study begun in 1989 and completed in 1994. The detailed design and results of this study are reported in the dissertation by Grigg, 1994, and briefly outlined here. Clients selected for treatment were distressed married alcoholic men with children. Two drug and alcohol treatment clinics were chosen in which to conduct the study, one in Surrey, a large suburb of metropolitan Vancouver, and the second in Duncan, a rural community on Vancouver island. The original design of TARP involved comparing the outcomes of both individual and couple's ExST treatment with a wait-list control group. SFT was developed during the initial stages of the project to replace a wait-list control group, because of clinic discomfort with the wait-list alternative, and because of anticipated client reluctance to accept this treatment option. Therefore when TARP was implemented, there were three treatment modalities being compared: individual ExST, couple's ExST, and SFT. Outcomes

were examined systemically on multiple levels: within the individual, within the couple, within the family, at the level of friends and work relationships.

Alcohol Recovery Project Treatment Modalities . Experiential Systemic therapy is a relatively new integration of individual and systemic therapeutic formats (Friesen et al., 1988; Friesen et al., 1989). This approach employs a wide variety of therapeutic techniques and uses a systemic, symbolic, and experiential framework. The main focus is on the here and now experiencing of relationships for an individual, couple or family. The purpose of ExST therapy is to perturb clients from rigid, dysfunctional patterns of interaction, to healthier, more adaptive styles of relationship. There are two experimental groups within ExST: individual and couple's therapy. Therapists from both experimental groups were expected to adhere to the same ExST manuals, and one purpose of this study is to see whether adherence ratings for both groups are similar or distinct.

On the other hand, Supported Feedback Therapy was developed to contrast with ExST (Grigg et al., 1989). In SFT, the emphasis is on using charts to monitor behavior on several dimensions: alcohol, abstinence, the self, marriage, family, friends and work. The charts are the source of feedback for the client (not the therapist), and the emphasis is on the client's ability to derive meaning from them. But beyond the charts and questionnaires, the therapist is responsible for developing a therapeutic relationship that is caring, warm and accepting and provides validation of the client's experience and it is in the use of these skills that the two approaches overlap.

Treatment Adherence and The Adherence Rating Scale. In order to support the conclusions of the comparative outcome study of TARP, undertaken

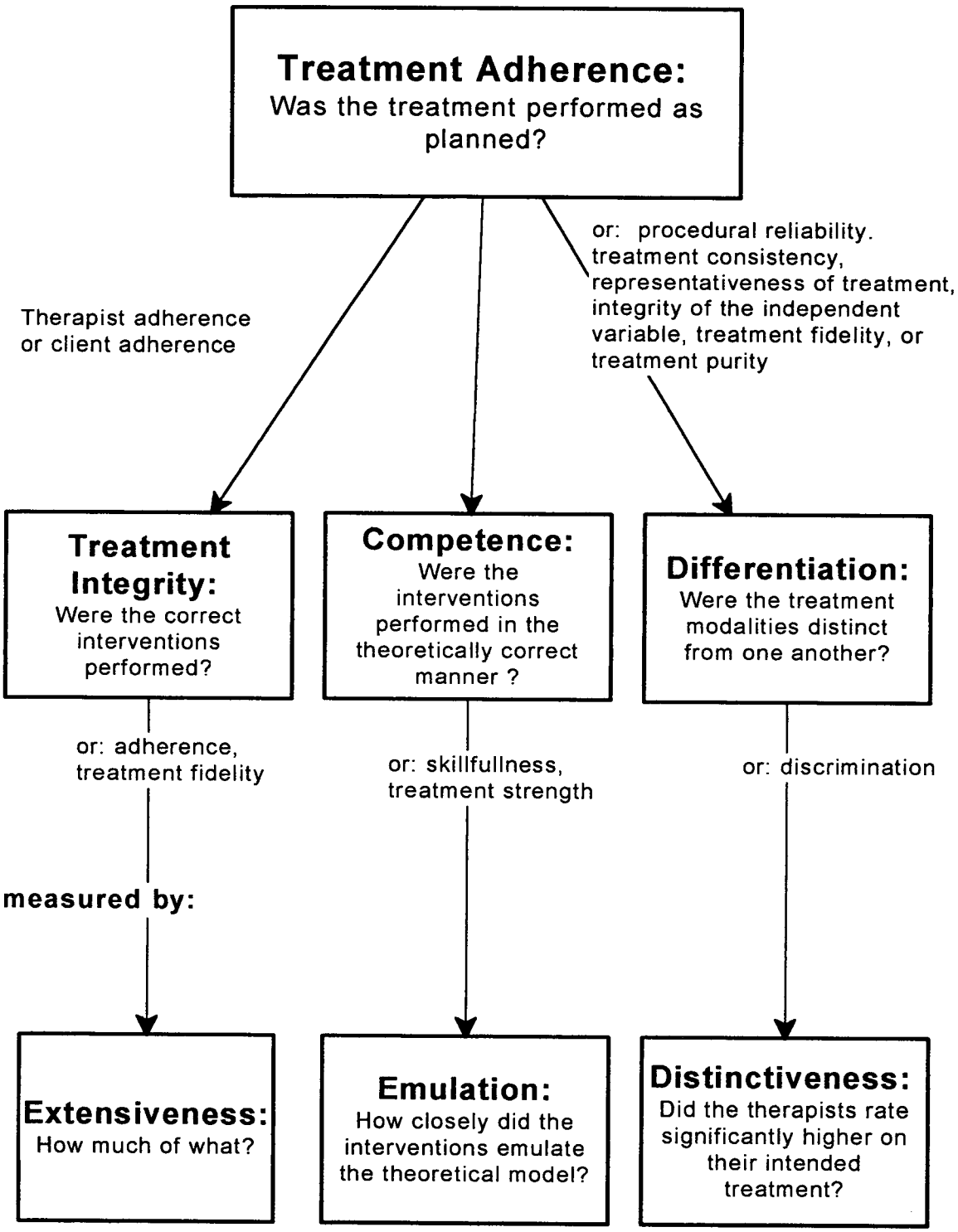


Figure 1. Treatment Adherence Components

by Grigg (1994), treatment adherence must also be examined, and so the present study was begun. Three components of treatment adherence were identified and measured by extensiveness, emulation and distinctiveness as shown in Figure 1.

There are essentially two phases in this study of treatment integrity in the TARP project: descriptive, and quantitative. In the first phase, the amount of therapist and rater training, the type and extent of supervision, the use of manuals, checklists, and ongoing training is reviewed as elements contributing to our confidence in treatment adherence.

In the second phase, an observer's rating scale is used. Rating scales based on the treatment manuals are considered the most rigorous measure of adherence available (Kazdin, 1986, 1987; Luborsky & DeRubeis, 1984; Startup and Shapiro, 1993b). In order to measure adherence in TARP, it was necessary to create The Adherence Rating Scale (TARS). TARS is a new rating scale, consisting of a manual and rating form that fit the unique characteristics of the ExST and SFT treatment modalities. The TARS rating form is presented in Appendix A, the TARS manual is presented in Appendix B, and lists of the items in each subscale is presented in Appendix C.

TARS lists interventions unique to ExST or SFT, as well as nonspecific or general therapeutic traits common to both, drawn from the two training manuals and various existing general family therapist skill measures. Measuring adherence to both unique and overlapping therapeutic interventions is crucial according to Luborsky and DeRubeis (1984). Many of the items and the basic scale design were modeled on the Collaborative Study Psychotherapy Rating Scale (CSPRS), developed by Hollon, Waskow, Evans, and Lowery (1984), a strategy also adopted by Startup and Shapiro, (1993b). Each item uses illustrative examples

and definitions emphasizing the concrete and observable. This approach is characteristic of a good adherence rating scale according to Hill, O'Grady and Price (1988).

Construction of TARS was a lengthy process. The author of this thesis gathered lists of key interventions for ExST and SFT from John Friesen, Darryl Grigg, Warren Weir and Dan Mitchell, each experts in the models. Then, using the manuals as a guide, a composite list was constructed, revised repeatedly and basic item definition was sketched out. A final definition version of this list and set of definitions became the basis for TARS. Jennifer Newman and Cheryl Bates, who were therapists experienced in the models, provided some additional feedback. Robert Conry, methodologist for the TARP project, helped define the structure of the rating form and the wording of the choice points on each item. TARS has 15 items in each of three categories: (a) ExST, (b) SFT, and (c) overlap. These final 45 items were then randomly ordered to avoid rater biasing. Each item requires completing two scales: one measuring extensiveness, the other measuring emulation. The scale design of examining each type of therapist intervention in terms of extensiveness and then in terms of emulation is novel to this study. Other research has indicated the value of measuring both frequency and competence (for example, Jacobson, Dobson, Fruzzetti, Schmaling and Salusky, 1991), but this study is the first to combine both in the analysis of each item. In order to establish the reliability and validity of TARS, a preliminary analysis of the rating data was conducted.

Rater Characteristics. There were two primary raters who rated the full set of 120 tapes, one of whom was the author of this thesis. The third rater rated a smaller subset of 30 tapes. All three raters were in their early 40's, were

Master's students in the Counselling Psychology program, and had completed courses in basic systemic counselling theory.

Hill et al. (1988) provide a good model for training raters, including reading the training and rating manuals, discussing each item, and rating and comparing ratings of several tapes during the training phase, and throughout the study. In this study, each rater spent about 20 hrs. in training using the methods outlined by Hill et al. (1988). Beutler and Hill (1992) recommend using at least two raters for every session to increase the reliability of the results on a rating scale, a suggestion followed in this study. Additionally, in the present study, a third rater was used to further test inter-rater reliability and to test for possible rater bias because of the inclusion of the author of this thesis as a rater. Beutler and Hill suggest that training and continued supervision of the raters is critical to providing reliable measures of adherence, particularly on the emulation scale, again part of our procedure. Ratings of six video tapes in each modality of clients who did not complete the treatment were used as training models and for rater standardization during the study.

Video Tape Library and Sampling Procedures. At the time this study was begun there were 60 clients who had completed the TARP project: 20 individual ExST, 20 couples ExST, and 20 SFT, involving 8 therapists on two sites over two years. Of the 90 clients who began treatment, dropouts averaged about 25%. As part of the procedures for the TARP project, all sessions were video-taped in the two locations, Surrey and Duncan. In total, for these 60 completing clients, there was a library of about 740 videotapes. These videotapes were stored in the TARP office at the University of British Columbia.

Thirty-minute segments of two videotapes from every completed client(s) were each rated by thetwo primary raters using the developed scale. The segment began 10 minutes from the beginning of the session, and raters were instructed to watch the entire 30-minute segment through, taking notes as watching, and reviewing the tape later as necessary. Probably the most common approach to rating global therapist behaviors in the literature is to watch the full session (ex. DeRubeis et al., 1982; Startup & Shapiro, 1993, etc.), although much shorter segments have also reliably been used (ex. Henry et al., 1993; Snyder et al., 1991a). Given the wide range of segment lengths studied, a judgment call was necessary in this study. A 30-minute segment from the middle portion of the session provides a very significant portion of a 50-minute session, and allows the design to almost double the number of sessions viewed over rating full-sessions. Henry et al. (1993) made a very similar argument for rating 15 minute segments from the middle of sessions, and produced very reliable results.

Raters viewed the third session (week 3-6), and the second or third session from the end of treatment (week 10-16), which allowed us to avoid rating the unique therapist behaviors associated with forming a therapeutic relationship or with termination. This is again a similar strategy to Henry et al, (1993). As well, any differences in treatment adherence over time may still be observed, allowing us to look for therapist drift (Hill et al, 1992). This sampling procedure means that up to 25 tape segments from sessions of each therapist over the course of the Project were selected. The three raters did not see the tapes in the same order, and did not see two sessions with the same client back to back to help minimize biasing effects of presentation. The most common procedure for this tape selection is a *balanced incomplete block design* (Fleiss, 1981; Hill et al., 1992),

however since the two main raters being used in this study saw every tape, the incomplete block design was not necessary.

Reducing Rater Bias. It is also important to reduce the risk of rater bias in a study relying on raters. One issue concerning rater bias addressed by the study was one raised by Newman and Scott (1988), order of presentation. Three groups of cases were randomly drawn for each of the two primary raters from the first 60 completed cases. The first group of 21 cases was randomized for early/late session time, with 7 SFT, 7 individual ExST and 7 couple's ExST cases, resulting in a group of 42 video tapes. The same procedure was repeated for the second group of 42 tapes. The third group was randomized in the same fashion, but used a smaller number of tapes (36) to finish off the first 60 cases. The tapes the third rater watched were drawn from the full set of 120 tapes. Thirty cases were selected, 10 from each of the three groups of clients, with half early and half late sessions, and equal numbers of the three treatment modalities.

Newman and Scott (1988) also comment on differences in training as a potential source of rater bias. They also suggest personal values, biases and social pressure or response sets like leniency or central tendency may contribute to rater bias. Hill et al. (1988) note raters may bias their rating when they like a therapist or feel similar to them. As well, fatigue, waning interest or sensitivity, familiarity with the instrument, complexity or length of the rating scale may produce rater bias. However, they also suggest that effective rater training can eliminate many of these sources of bias.

In this study, a similar procedure was used to train each of the three raters. Each rater was given a copy of the rating scale to read and study and questions regarding the rating scale were answered. The author of this thesis and

one rater then watched a video tape taken from the TARP study of a case that did not complete the sessions. Each rater then rated the session. Finally each item was discussed in detail and a final score was agreed on. This same process was repeated with each of the raters on a total of 6 tapes. At this point, a great deal of agreement was observed on the scoring, and each rater received their first set of tapes to rate for the study. Special note was made during and after training to be careful to rate only what was observed and to try to avoid using personal values to rate the sessions. Each rater noted tendencies to prefer certain therapists and were asked to be especially careful rating those sessions. As the sample of videotapes was fairly large, the raters were asked to limit the amount of tapes they rated at a sitting, to avoid fatigue. Interest remained high throughout the rating process. The length of the instrument was moderate, also reducing fatigue.

One possible source of rater bias was the distinctiveness of the treatments. Raters could very quickly identify two very different approaches, simply because one treatment (SFT) had large charts on the wall. The raters were cautioned to focus on each item as it actually appeared in the session and not to assume certain behaviors appeared automatically. The scale items were presented in random order, with no indication of which subscale they belonged to, or even how many subscales there were. This attention to the smaller detail, and the construction of the scale to mix SFT, ExST and common items was intended to reduce this bias.

Data Analysis

There were five stages in the data analysis in this study. First, descriptive information indicating adherence (like the use of manuals, therapist training and supervision, etc.) was gathered. Second, a preliminary analysis was conducted on

the rating scale to establish reliability for the rating scale. Third, an analysis of rater bias was conducted. Fourth, the reliability of the rating process was examined. Fifth, a hypothesis-testing analysis was performed.

Preliminary Analysis. A preliminary phase in the data analysis was conducted, examining the reliability of the whole scale, and the individual subscales. Item-total correlations were produced based on the whole scale. Then, the internal consistency reliability of the whole scale was examined. Item-total correlations were then performed on the subscales. Internal consistency of the subscales was then measured. Finally, an ANOVA of items broken down by treatment modality was performed.

Analysis of Rater Bias. One significant potential source of rater bias arose through the choice of raters. The author of this thesis was both highly involved in the construction of the rating scale, and was one of the two primary raters. In order to evaluate the bias this familiarity with the format and intentions of the rating scale might have, a third rater was asked to rate a smaller group of tapes. A comparison of the three raters was made to discover if there were any systematic bias on the part of the author. Subscale internal consistency scores were compared, as were the inter-rater correlations on the subscales. Also, an ANOVA of the subscale scores broken down by raters was performed.

The general level of adherence (extensiveness and emulation) for each modality was calculated then used to determine that the data distinguishes between treatments (discrimination). There are of course many statistical approaches to measuring each of these constructs.

Reliability of the Rating Process. The correlations of subscale mean scores between the two main raters were examined. As well, subscale correlations of the averaged scores of the two raters were analyzed.

Hypothesis-Testing Analysis. Treatment adherence and differentiation were estimated in three steps. First a session-wise cluster analysis was performed. This is done by creating a 6-point profile for each session on both extensiveness and emulation in each of the three subscales: SFT, ExST, and overlap. Second, similar 6-point profiles were clustered, and membership in the resulting clusters is then cross tabulated with the treatment given. Integrity is indicated when the clusters match the treatment given. Statistically, treatment integrity is estimated in the third step by measuring the strength of association between profile cluster and treatment modality using a chi-squared test. Discrimination can also be estimated at this point by the lack of similarity in profile clusters across treatments. If the treatments are distinct from one another, there should be clear matches between clusters and treatment type.

Simple overall averaged tallies of item ratings were examined to look for unique item effects, like overuse or underuse. The items were also be tallied by treatment type to indicate unusual patterns of use within a modality. This was done to give some preliminary indication of the success of training for specific items, as well as some of the specifics of adherence in the project.

Finally, comparison of averaged score profiles by therapist for early and late sessions provided a measure of therapist drift.

Chapter 4: Results

There are three sections in the results chapter. The first section describes the descriptive estimates of adherence based on the design and implementation aspects of TARP that relate to adherence. The second section reports results concerning the reliability of the rating scale and the rating process, as well as data used to check for rater bias. The third section reports the findings from the hypothesis testing relating to adherence, including extensiveness, emulation and differentiation, and relative levels of adherence.

Descriptive Estimates of Adherence.

A Priori Methods. Sechrest et al. (1979) suggest examining the original design of a research project to begin to get an idea of treatment integrity. This section reports the various design characteristics of TARP related to treatment adherence.

The theoretical premises and treatment plan of ExST were clearly articulated in manual form well before the project began (Friesen, et al., 1988; Friesen, et al., 1989). ExST is a complex therapy requiring considerable clinical skill, and careful selection of the therapists participating inthe project, as well as their extensive, ongoing training and supervision was a basic part of the design.

SFT was developed during the initial stages of the project to replace a wait-list control group, because of clinic discomfort with the wait-list alternative, and anticipated client reluctance to accept this treatment option (Grigg, 1994). SFT was generated and manualized by Grigg et al. (1989) and intended to be an elaborate quasi-control group. The theoretical basis of SFT is therefore very simple, based on a warm, caring, non-judgmental therapeutic relationship, where supportive feedback can be explored using charts of the client's previous week(s).

The manual clearly articulates both the theoretical basis, and the particular tasks for the SFT therapist.

The present adherence study was also part of the overall Project design, so all sessions were videotaped, in part so a random selection of tapes could be rated in terms of adherence.

This section has summarized the elements of TARP's planned procedure. The following sections examine the procedures actually carried out in TARP that relate to adherence.

Therapist Experience. Moncher and Prinz (1991) point out the significance of therapist experience and amount of training as an indication of adherence. There were a total of 12 therapists involved in the delivery of treatment in TARP, 5 involved in providing the individual and marital forms of ExST, and 7 delivering SFT. All therapists who participated in the study were required to have a minimum of 3 years direct work experience involving alcohol dependent individuals and their families. They also had a Master's degree or higher in Psychology, Social Work or a related field. The average experience of therapists was considerable ($\bar{X}$ = 8.17 yrs, SD = 4.37 yrs) and the average number of years specifically working with alcoholics indicates a knowledgeable group of therapists ($\bar{X}$ = 5.70 yrs, SD = 3.25).

Manuals. Manualization of treatment, training and supervision using manuals and observer rating forms based on treatment manuals are all considered good indicators of treatment integrity (Luborsky & DeRubeis, 1984; Dobson & Shaw, 1988; Newman et al., 1988). In this study, both SFT (Grigg et al., 1989) and ExST manuals (Friesen et al, 1988; Friesen et al., 1989) were used in training and supervision and formed the basis for the rating scale.

SFT Therapist Training and Selection. Nine therapists began to train in this approach and studied the SFT treatment manual before participating in a series of training sessions. These sessions were conducted over a 6 week period, lasting a total of 20 hours. SFT training focused on mastering the manualized treatment, using didactic presentations, discussion, role-play, video tape presentations and rehearsals. Trainers and developers then selected 6 of the therapists who demonstrated competent treatment implementation in role play situations.

SFT Supervision. Regular biweekly meetings of the SFT therapists with the developers of the treatment approach was used as an opportunity for group supervision. These sessions included discussion of case management and planning, as well as reviews of videotaped sessions, or didactic presentations regarding the principles and implementation of the therapy. In addition to the group sessions, individual supervision was available on a case by case basis when a therapist requested specific supervision, or when a supervisor requested further in-depth supervision in order to maintain a high level of treatment adherence.

ExST Therapist Training and Selection. A total of 14 therapists began to train in this approach and studied the ExST training manual before and during the training sessions. Training began with three 4-day workshops conducted over 3 months spaced one month apart. In addition, each therapist received 10 hrs. of direct or videotaped supervision. These individual supervision sessions focused on the implementation of the relatively complex techniques used in individual and couple's ExST.

Each therapist was then required to collect 5 videotapes from both individual and couple's therapy sessions. Two trained adjudicators then randomly

selected 2 individual and 2 couple's therapy sessions from the total of 10 videotapes submitted. These sessions were reviewed by the adjudicators to determine the therapist's capacity to implement the treatment in either condition, in terms of integrity and competency. There were 14 therapists who received training and 5 were selected through this procedure who adequately delivered the individual and marital forms of ExST treatment.

ExST Supervision. Supervision occurred on a weekly basis throughout the course of the research. The 2 trained supervisors used both group and individual supervision formats. The ExST therapists were provided both live and videotape supervision of individual and marital therapy treatment. Supervision included review of therapy and technique intended to insure emulation of the ExST model and minimize therapist drift.

The ExST model of supervision (Newman, Friesen, & Grigg, 1991) informed the focus and intent of the consultation meetings. According to this model, there are three key domains of the therapist's development: (a) theoretical development; (b) technical refinement; and (c) personal growth. Experiential activities typical of ExST therapy (e.g. role play, role reversal, enactment, artwork, sculpting, use of symbols and ritual, as well as process recall) were used in supervision in order to address these three domains. An emphasis on the "here and now" experiencing of issues with clients was used as a means to communicate and work through problems. In other words, supervision was an opportunity for both therapists and supervisors to directly experience the ExST model in order to deepen appreciation and understanding of the practical implementation and integrity of the approach.

Preliminary Analysis: Scale and Rater Reliability

In order to quantitatively measure adherence, the measuring device must be valid and reliable, and the raters must be unbiased and reliable. Only then can the results of such a measure be considered. Reliability statistics on TARS have not previously been reported, so they are examined in some detail in this thesis. Rater bias is examined later in the chapter and validity issues are discussed in the final chapter. The results of the adherence study itself are presented at the end of this chapter.

Reliability of TARS. Reliability in TARS can be studied on both the overall scale and the subscale level by examining internal consistency. Item-total correlations of the three raters on the scale as a whole were generally low, varying from -.47 to .73 as shown in Tables 2-4. However the standardized item alpha for the extensiveness items on the whole scale averaged over three raters was .71. For the emulation items the standardized item alpha averaged over three raters was .78. The average reliability of the two performance subscales (and therefore the whole scale) is .75.

Reliability of TARS Subscales. Item-total correlations of the three raters within a subscale were generally much higher than on the scale as a whole, ranging from -.46 to .96 as shown in Tables 5-7. (Items from the subscale with low or negative item-total correlations, or low inter-rater correlations will be discussed in Chapter 5.)

Table 2

Item-Total Correlation of ExST Items on Whole Scale

	Extensiveness			Emulation		
Item#	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
3	.12	.02	.38	.23	.14	.32
4	.12	.03	.30	.25	.17	.52
6	.07	-.20	.35	.24	.04	.35
9	.26	-.23	.27	.33	.01	.50
16	.13	-.09	.01	.18	.01	.29
18	.48	-.17	.23	.48	.13	.51
21	.25	-.34	.28	.37	-.04	.39
24	.17	-.11	.13	.26	.13	.40
26	.23	-.20	.22	.31	.04	.32
28	.38	.10	.49	.37	.22	.50
30	.10	-.19	-.02	.15	.10	.25
36	.04	-.18	-.06	.12	-.07	.25
37	.14	-.01	.25	.24	.26	.40
39	.16	-.27	.05	.25	-.01	.31
43	.16	.04	.20	.23	.10	.36

Note. n=120 for Raters 1 & 2. n=30 for Rater 3.

Table 3

Item-Total Correlation of SFT Items on Whole Scale

	Extensiveness			Emulation		
Item#	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
2	.16	.21	.04	.08	.03	-.12
7	.17	.41	-.11	.12	.19	-.32
13	.20	.43	.02	.13	.15	-.03
14	.26	.35	-.34	.23	.36	-.09
15	-.11	.17	.37	-.11	-.02	.21
17	.51	.46	.41	.43	.28	.20
19	.21	.37	.05	.14	.22	.00
22	-.01	.27	-.05	.01	.09	-.03
23	.19	.20	.31	.17	.06	.25
27	-.21	.20	-.42	-.24	.05	-.47
32	.20	.22	.31	.22	.08	.35
34	.07	.43	-.11	.03	.17	-.18
35	.20	.45	-.09	.13	.21	-.16
38	.21	.45	.02	.13	.26	-.12
42	.26	.42	.02	.15	.19	-.17

Note. n=120 for Raters 1 & 2. n=30 for Rater 3.

Table 4

Item-Total Correlation of Common Items on Whole Scale

Item#	Extensiveness			Emulation		
	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
1	.38	.28	.38	.37	.42	.43
5	.48	.18	.56	.57	.34	.66
8	.35	.10	.27	.37	.25	.24
10	.66	.26	.34	.73	.45	.50
11	.53	.22	.18	.54	.43	.41
12	.12	.11	.31	.19	.13	.51
20	.39	.06	.10	.49	.21	.41
25	.31	.30	.55	.37	.44	.47
29	.34	-.11	.26	.41	.29	.29
31	.64	.36	.36	.67	.52	.43
33	.34	.09	.27	.40	.28	.38
40	.24	.38	-.12	.23	.18	-.07
41	.35	.01	.23	.45	.36	.34
44	.60	.33	.35	.54	.31	.44
45	.12	.02	.28	.13	.06	.36

Note. n=120 for Raters 1 & 2. n=30 for Rater 3.

Table 5

Item-Total Correlation of the ExST Subscale

Item#	Extensiveness			Emulation		
	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
3	.36	.50	.25	.37	.56	.38
4	.64	.64	.72	.73	.70	.76
6	.82	.90	.87	.86	.90	.87
9	.79	.81	.68	.82	.83	.75
16	.80	.80	.91	.81	.84	.90
18	.75	.83	.73	.75	.82	.76
21	.87	.90	.90	.90	.90	.88
24	.70	.47	.06	.74	.50	.46
26	.83	.85	.71	.87	.86	.79
28	.12	.10	-.02	.11	.19	.12
30	.83	.87	.78	.83	.87	.83
36	.70	.69	.82	.74	.72	.86
37	.84	.72	.84	.86	.75	.84
39	.85	.90	.89	.87	.85	.92
43	.75	.56	.87	.77	.55	.87

Note. n=120 for Raters 1 & 2. n=30 for Rater 3.

Table 6

Item-Total Correlation of the SFT Subscale

	Extensiveness			Emulation		
Item#	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
2	.64	.85	.89	.63	.65	.84
7	.96	.96	.92	.96	.96	.93
13	.94	.88	.88	.94	.90	.80
14	.35	.37	-.03	.38	.23	-.04
15	.73	.83	.35	.76	.82	.38
17	.71	.87	.64	.70	.83	.62
19	.95	.93	.87	.95	.95	.87
22	.69	.91	.85	.69	.90	.85
23	.69	.90	.67	.73	.87	.64
27	.74	.89	.68	.76	.81	.68
32	.21	.59	.18	.18	.60	.07
34	.59	.88	.61	.57	.90	.58
35	.94	.90	.92	.94	.91	.91
38	.95	.88	.93	.94	.90	.92
42	.91	.91	.92	.91	.90	.90

Note. n=120 for Raters 1 & 2. n=30 for Rater 3.

Table 7

Item-Total Correlation of the Common Subscale

	Extensiveness			Emulation		
Item#	Rater 1	Rater 2	Rater 3	Rater 1	Rater 2	Rater 3
1	.40	.43	.24	.41	.52	.43
5	.52	.44	.63	.56	.48	.65
8	.38	.32	.10	.42	.38	.08
10	.65	.56	.51	.68	.56	.63
11	.56	.55	.46	.54	.58	.54
12	.16	.18	.39	.23	.25	.49
20	.46	.30	.36	.49	.43	.49
25	.41	.37	.44	.46	.46	.44
29	.38	.30	.50	.43	.50	.41
31	.64	.52	.56	.63	.55	.38
33	.45	.35	.49	.46	.37	.53
40	-.04	-.46	-.30	.05	-.40	-.25
41	.44	.40	.37	.47	.55	.45
44	.32	.23	.09	.34	.37	.23
45	.21	.08	.37	.20	.12	.37

Note. n=120 for Raters 1 & 2. n=30 for Rater 3.

When the internal consistency of the subscales was measured, the standardized item alpha scores, averaged across three raters, were .94 for ExST Extensiveness (EExt), .95 for ExST Emulation (EEmu), .95 for SFT Extensiveness (SExt), .95 for SFT Emulation (SEmu). The common items were less internally consistent, but still showed moderately high alpha's: .76 for Common Extensiveness (CExt) and .80 for Common Emulation (CEmu).

An ANOVA of the items broken down by treatment modality was highly consistent in its results. Much of the variance in mean ExST and SFT items scores was explained by treatment modality. Very little of the variance in mean Common item scores was explained by treatment modality. Tables 8-13 show the F score and correlation ratio (eta^2or shared variance) results of the ANOVA for the three raters on each item. The tables of the mean scores and standard deviations for each item that form the basis of the ANOVA are presented in Appendix D.

Table 8

ANOVA: Effect of Treatment on ExST Extensiveness Item Means

	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	F	eta^2	F	eta^2	F	eta^2
3	8.26**	.12	14.31**	.20	.36	.03
4	31.75**	.35	21.87**	.27	12.63**	.48
6	102.08**	.64	135.94**	.70	25.42**	.65
9	74.06**	.56	113.57**	.66	17.98**	.57
16	64.92**	.53	63.03**	.52	90.05**	.87
18	25.39**	.30	110.29**	.65	12.79**	.49
21	64.54**	.52	268.62**	.82	34.50**	.72
24	36.76**	.39	23.56**	.29	.11	.01
26	62.89**	.52	123.88**	.68	17.21**	.56
28	3.30*	.05	.59	.01	.66	.05
30	72.54**	.55	148.68**	.72	23.00**	.63
36	39.56**	.40	45.06**	.44	63.53**	.82
37	64.56**	.52	36.99**	.39	22.12**	.62
39	64.89**	.53	181.81**	.76	42.72**	.76
43	35.24**	.38	19.52**	.25	25.83**	.66

Note. *p < .05. **p < .001.

Table 9

ANOVA: Effect of Treatment on ExST Emulation Item Means

Item#	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
	F	eta^2	F	eta^2	F	eta^2
3	6.02*	.09	18.69**	.24	1.54	.10
4	37.46**	.39	31.44**	.35	12.74**	.49
6	86.94**	.60	110.51**	.65	26.98**	.67
9	64.67**	.53	134.26**	.70	20.78**	.61
16	74.03**	.56	113.56**	.66	80.19**	.86
18	32.18**	.35	88.61**	.60	10.83**	.45
21	74.78**	.56	194.12**	.77	35.74**	.73
24	29.47**	.40	16.41**	.22	2.21	.14
26	71.59**	.55	115.45**	.66	16.17**	.55
28	3.43*	.06	1.53	.03	.22	.02
30	76.08**	.57	126.77**	.68	32.44**	.71
36	42.85**	.42	60.89**	.51	66.21**	.83
37	70.22**	.55	42.12**	.42	19.20**	.59
39	72.38**	.55	137.86**	.70	65.68**	.83
43	41.65**	.42	20.82**	.26	28.09**	.68

Note. *p < .05. **p < .001.

Table 10

ANOVA: Effect of Treatment on SFT Extensiveness Item Means

Item#	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
	F	eta^2	F	eta^2	F	eta^2
2	36.99**	.39	222.66**	.79	59.02**	.81
7	226.68**	.79	219.32**	.79	140.37**	.91
13	168.21**	.74	89.83**	.61	41.24**	.75
14	5.80*	.09	6.65*	.10	.58	.04
15	102.45**	.64	200.52**	.77	1.18	.08
17	27.35**	.32	114.67**	.66	4.89*	.27
19	174.23**	.75	359.86**	.86	43.19**	.76
22	61.15**	.51	419.61**	.88	30.37**	.69
23	37.85**	.39	444.00**	.88	6.53*	.33
27	112.58**	.66	317.93**	.84	14.41**	.52
32	2.51	.04	37.07**	.39	.92	.06
34	26.69**	.31	100.56**	.63	10.29**	.43
35	167.33**	.74	100.56**	.63	55.72**	.81
38	183.23**	.76	108.02**	.65	54.22**	.80
42	149.39**	.72	138.68**	.70	86.54**	.87

Note. *p < .05. **p < .001.

Table 11

ANOVA: Effect of Treatment on SFT Emulation Item Means

	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	F	eta^2	F	eta^2	F	eta^2
2	38.14**	.39	61.04**	.5106	43.76**	.76
7	226.68**	.79	242.71**	.8058	163.51**	.92
13	181.73**	.76	140.17**	.7055	20.67**	.60
14	6.77*	.10	2.44	.0400	.16	.01
15	114.20**	.66	160.90**	.7334	1.60	.11
17	26.43**	.31	92.08**	.6115	3.76*	.22
19	170.03**	.74	392.71**	.8703	23.43**	.63
22	50.74**	.46	323.79**	.8470	26.45**	.66
23	45.72**	.44	284.39**	.8294	4.79*	.26
27	125.08**	.68	154.03**	.7247	18.22**	.57
32	1.82	.03	35.03**	.3745	.09	.01
34	25.48**	.30	161.85**	.7345	10.57**	.44
35	192.45**	.77	147.67**	.7162	38.05**	.74
38	194.22**	.77	134.86**	.6975	51.88**	.79
42	184.89**	.76	159.55**	.7317	53.57**	.80

Note. *p < .05. **p < .001.

Table 12

ANOVA: Effect of Treatment on Common Extensiveness Item Means

	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	F	eta^2	F	eta^2	F	eta^2
1	.62	.01	.55	.01	1.76	.12
5	5.86*	.09	6.19*	.10	.22	.02
8	1.15	.02	4.57*	.07	2.67	.17
10	3.55*	.06	11.07**	.16	16.56**	.55
11	5.29*	.08	10.05**	.15	5.75*	.30
12	3.86*	.06	7.92**	.12	1.79	.12
20	8.67**	.13	13.25**	.18	16.40**	.55
25	1.07	.02	.86	.01	.02	.00
29	16.18**	.22	38.68**	.40	14.96**	.53
31	4.13*	.07	1.85	.03	4.82*	.26
33	8.80**	.13	2.05	.03	2.08	.13
40	31.26**	.35	158.75**	.73	18.83**	.58
41	23.16**	.28	29.39**	.33	2.92	.18
44	15.35**	.21	2.88	.05	.70	.05
45	4.22**	.07	2.43	.04	1.33	.09

Note. *p < .05. **p < .001.

Table 13

ANOVA: Effect of Treatment on Common Emulation Item Means

	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	F	eta^2	F	eta^2	F	eta^2
1	1.77	.03	1.46	.02	.21	.02
5	6.15*	.10	8.55**	.13	.36	.03
8	1.60	.03	4.27*	.07	.50	.04
10	3.25*	.05	9.72**	.14	10.31**	.43
11	5.72*	.09	13.72**	.19	3.77*	.22
12	2.76	.05	15.14**	.21	4.26*	.24
20	7.86**	.12	13.51**	.19	9.64**	.42
25	.93	.02	.52	.01	.80	.06
29	20.95**	.26	29.02**	.33	14.00**	.51
31	3.77*	.06	3.68*	.06	2.38	.15
33	8.80**	.13	1.48	.02	13.00**	.49
40	17.64**	.23	189.91**	.76	16.24**	.55
41	15.16**	.21	23.79**	.29	2.95	.18
44	12.30**	.17	6.50*	.10	.40	.03
45	3.05	.05	2.89	.05	3.42*	.20

Note. *p < .05. **p < .001.

The ExST and SFT items tended to show more of an effect of treatment modality. As a way of simplifying to clarify the pattern in this data, the correlation ratio (eta^2or shared variance(of item by treatment can be averaged over the three raters for all the items from each subscale. As Table 14 shows, the average Common subscale score is less than half of the average ExST or SFT scores.

Table 14

Averaged Correlation Ratio of Subscales (ANOVA: Effect of Treatment on Item

	EExt	EEmu	SExt	SEmu	CExt	CEmu
M eta^2	.48	.49	.58	.57	.17	.17

Note. Correlation Ratios are averaged over three raters and over 15 items on each subscale.

Low variance in the ratings of several items produced low inter-rater correlation scores, particularly from the Common subscale. For example, on Item 33, Respectful, all three raters rated every session with a 3 or 4. Rater 1 rated 88% of the sessions with a 4, and rater 2 rated 98% of sessions with a 4. Despite the high degree of similarity in the scores, because of the limited variation in the

scoring, the reliability is reduced. The relatively low variance of the common subscale items overall is readily apparent in the standard deviation scores for subscale means, less than half the size of the SD for the ExST and SFT subscale means as shown in Table 15.

Table 15

Mean & Standard Deviation Scores for Subscales of Whole Sample

Scale	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
	M	SD	M	SD	M	SD
EExt	32.11	14.09	31.57	13.69	36.23	11.35
EEmu	35.55	15.96	37.48	16.38	39.60	14.04
SExt	20.83	15.33	17.61	19.07	23.93	12.56
SEmu	21.21	15.87	20.70	20.26	24.93	12.91
CExt	34.50	6.58	34.55	4.78	38.80	5.10
CEmu	37.17	7.73	38.68	7.42	40.47	5.86

Consistency of Rating and Rater Bias. Subscale internal consistency scores for the three raters were virtually identical as shown in Table 16.

Table 16

Reliability of Subscale Internal Consistency

Rater	EExt	EEmu	SExt	SEmu	CExt	CEmu
1 (n=120)	.94	.95	.95	.95	.80	.83
2 (n=120)	.94	.95	.97	.97	.72	.80
3 (n=30)	.93	.95	.93	.93	.76	.79

A comparison of subscale means and SDs has already been presented in Table 15. When the inter-rater correlations are examined for the items on each subscale, a more complex relationship emerged as shown in Tables 17-19.

Table 17

Inter-Rater Correlations on ExST Extensiveness and Emulation Items

	Rater 1:2 (n=120)		Rater 1:3 (n=30)		Rater 2:3 (n=30)	
Item#	Ext	Emu	Ext	Emu	Ext	Emu
3	.25**	.26**	.21	.24	.21	.34*
4	.48***	.51***	.65***	.69***	.43**	.58***
6	.73***	.73***	.73***	.68***	.75***	.73***
9	.68***	.71***	.81***	.75***	.82***	.74***
16	.61***	.70***	.67***	.72***	.81***	.87***
18	.61***	.62***	.70***	.58***	.69***	.65***
21	.71***	.74***	.71***	.68***	.78***	.84***
24	.33***	.38***	.31*	.46**	-.05	.15
26	.70***	.68***	.59***	.60***	.67***	.61***
28	.30***	.35***	.16	.18	.30	.33*
30	.73***	.69***	.723***	.72***	.71***	.72***
36	.76***	.74***	.87***	.90***	.81***	.86***
37	.50***	.59***	.95***	.82***	.84***	.76***
39	.68***	.66***	.79***	.79***	.83***	.81***
43	.41***	.40***	.80***	.81***	.26	.20
median	.61***	.66***	.71***	.72***	.71***	.72***
mean	.57***	.58***	.63***	.64***	.59***	.61***

Note. *p < .05, one-tailed. **p < .01 one-tailed. ***p < .001 one-tailed.

Table 18

Inter-Rater Correlations on SFT Extensiveness and Emulation Items

	Rater 1:2 (n=120)		Rater 1:3 (n=30)		Rater 2:3 (n=30)	
Item#	Ext	Emu	Ext	Emu	Ext	Emu
2	.59***	.47***	.53**	.57***	.83***	.58***
7	.90***	.91***	.82***	.84***	.90***	.91***
13	.87***	.87***	.81***	.77***	.72***	.74***
14	.67***	.63***	.70***	.56**	.70***	.60***
15	.80***	.77***	.35*	.49**	.30	.32*
17	.68***	.67***	.68***	.56**	.64***	.53**
19	.86***	.89***	.68***	.66***	.88***	.85***
22	.64***	.65***	.71***	.77***	.83***	.77***
23	.58***	.62***	.60***	.59***	.46**	.42*
27	.74***	.69***	.76***	.77***	.78***	.71***
32	.14	.14	.14	.06	.20	.13
34	.49***	.55***	.90***	.85***	.61***	.67***
35	.854***	.90***	.81***	.83***	.82***	.84***
38	.88***	.89***	.95***	.94***	.82***	.85***
42	.86***	.89***	.79***	.80***	.87***	.86***
median	.74***	.77***	.71***	.77***	.82***	.71***
mean	.70***	.70***	.68***	.67***	.69***	.65***

Note. *p < .05, one-tailed. **p < .01 one-tailed. ***p < .001 one-tailed.

Table 19

Inter-Rater Correlations: Common Extensiveness and Emulation Items

	Rater 1:2 (n=120)		Rater 1:3 (n=30)		Rater 2:3 (n=30)	
Item#	Ext	Emu	Ext	Emu	Ext	Emu
1	.50***	.48***	.38*	.13	.29	.35*
5	.20*	.13	.08	.11	.15	.03
8	.21*	.20*	.17	.26	-.12	.05
10	.27**	.26**	.52**	.49**	.45**	.51**
11	.35***	.22**	.44**	.49**	.44**	.23
12	.41***	.30***	.38*	.21	.51**	.39*
20	.34***	.37***	.58***	.51**	.52**	.59***
25	.36***	.44***	.63***	.44**	.20	.28
29	.42***	.41***	.61***	.58***	.63***	.58***
31	.38***	.48***	.32*	.10	.45**	.42*
33	-.05	-.06	.07	.14	.24	.31*
40	.64***	.54***	.56**	.59***	.60***	.64***
41	.54***	.33***	.60***	.53**	.56**	.22
44	.26**	.09	.44**	.43**	.313*	.26
45	.37***	.33***	-.02	-.13	-.06	-.07
median	.37***	.43***	.44**	.43**	.45**	.31*
mean	.35***	.32***	.38*	.32*	.35*	.32*

Note. *p < .05, one-tailed. **p < .01 one-tailed. ***p < .001 one-tailed.

An ANOVA of the subscale scores broken down by raters indicated the variance in the subscale scores was not explained by raters. The correlation ratio scores on the six subscales ranged from .01 to .05 as shown in Table 20. The one subscale that approached significance was CExt and rater only accounted for 5% of the variance in scores. As well, it was rater 3 that had the slightly elevated score on that subscale, with rater 1 and 2 almost identical in their mean scores. The ratings of Rater 1 did not stand out on any subscale.

Table 20

ANOVA: Subscale X Rater

Subscale	F	sig.	eta^2
EExt	1.43	.2405	.01
EEmu	0.94	.3924	.01
SExt	2.14	.1199	.02
SEmu	0.70	.4983	.01
CExt	7.55	.0006*	.05
CEmu	2.83	.0611	.02

Note. *p < .001

Reliability of the Rating Process. When looking at the correlation between the extensiveness and emulation subscale scores averaged over the two main raters on the full sample, there are extremely high correlations (.96 and higher) for each treatment subscale as shown in Table 21. Even correlations across extensiveness and emulation between raters was very high although only moderate on the common subscales (above .82 for ExST, above .92 for SFT, around .50 for Common, as shown in Table 22).

Table 21

Subscale Correlations Using Averages Over Raters 1 & 2

Subscale	EExt	EEmu	SExt	SEmu	CExt	CEmu
EExt	--	.99**	-.86**	-.85**	.45**	.49**
EEmu		--	-.86**	-.84**	.47**	.52**
SExt			--	.99**	-.10	-.17*
SEmu				--	-.08	-.14
CExt					--	.96**
CEmu						--

Note. n=120 for each subscale. *p < .05. **p < .001

Table 22

Inter-Rater Correlations: Subscale Scores

	EExt1	EEmu1	SExt1	SEmu1	CExt1	CEmu1	EExt2	EEmu2	SExt2	SEmu2	CExt2	CEmu2
EExt1	--	.99**	-.77**	-.76**	.45**	.44**	.85**	.83**	-.77**	-.76**	.47**	.50**
EEmu1		--	-.76	-.75**	.46**	.46**	.84**	.83**	-.76**	-.75**	.48**	.50**
SExt1			--	.99**	.03	.02	-.85**	-.82**	.93**	.93**	-.19*	-.26*
SEmu1				--	.05	.04	-.84**	-.82**	.93**	.93**	-.18*	-.25*
CExt1					--	.97	.19*	.20*	-.04	-.02**	.54**	.50**
CEmu1						--	.19*	.21*	-.05	-.04	.53**	.48**
EExt2							--	.97**	-.88**	-.86**	.47**	.49**
EEmu2								--	-.87**	-.85**	.50**	.54**
SExt2									--	.99**	-.21*	-.28**
SEmu2										--	-.17*	-.23*
CExt2											--	.91**
CEmu2												--

Note. *p < .05. **p < .001

In order to more easily visualize the pattern in the correlation scores between the two raters, Figure 2 is based on only the data from Table 22 that refers to inter-rater correlations.

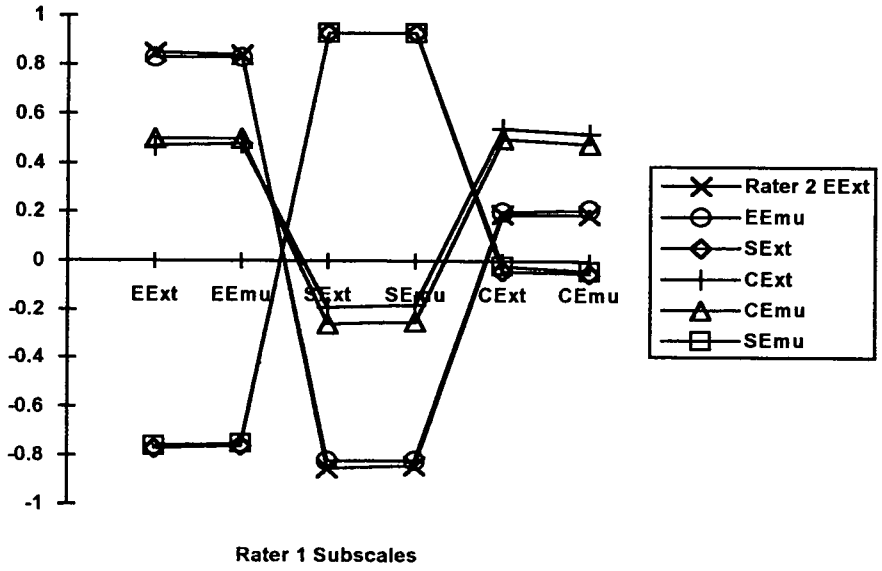


Figure 2. Inter-Rater Subscale Correlations.

Note: Ordinate variable is Pearson correlation score.

Hypothesis Testing Results

Adherence and Differentiation. Initial measurement of adherence was begun with a simple ANOVA of the subscale scores broken down by treatment as shown in Table 23.

Table 23

ANOVA: Subscale X Treatment

	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)		mean
Scale	F	eta^2	F	eta^2	F	eta^2	eta^2
EExt	124.18***	.68	272.26***	.82	75.47***	.85	.78
EEmu	119.58***	.67	221.50***	.79	71.88***	.84	.77
SExt	255.28***	.81	494.14***	.89	108.73***	.89	.87
SEmu	262.94***	.82	529.59***	.90	71.53***	.84	.85
CExt	3.67*	.06	7.77**	.12	2.17	.14	.10
CEmu	3.19*	.05	9.86**	.14	3.99*	.23	.14

Note. *p < .05. **p < .001. ***p < .0001

Composite subscale scores were calculated for each session then averaged over the two raters, generating a 6-point profile for each session. These points were labeled ExST extensiveness (EExt), ExST emulation (EEmu), SFT extensiveness (SExt), SFT emulation (SEmu), Common extensiveness (CExt), and

Common emulation (CEmu). Correlations scores across subscales indicate the ExST and SFT subscales are highly negatively correlated, therefore measuring very different behaviors.

A cluster analysis was performed on the 6-point profile scores and the 6, 5, 4, and 3-cluster results were carefully examined. Error differences were calculated and the 6-cluster group showed a relatively low increase in error as shown in Figure 3, with some interesting information to explore, so the following results are based on a 6-cluster analysis. However, the 3 cluster analysis is so clear-cut in its results that brief mention will be made of it as well.

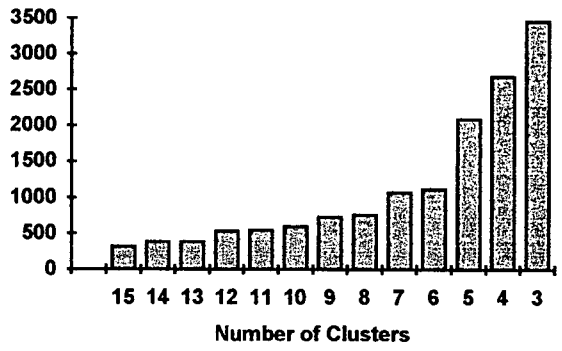


Figure 3. Accumulated Error Differences of Clusters.

Note: Ordinate is Error Coefficient Difference

The six clusters were descriptively labeled, based on their composition and score profiles, as follows: a) Cluster 1, moderate ExST (n=35); b) Cluster 2, moderate SFT (n=10); c) Cluster 3, high SFT (n=17); d) Cluster 4, high ExST (n=43); e) Cluster 5, blended SFT (n=11); and f) Cluster 6, low mixed (n=4). Subscale means and standard deviations for the six clusters provide distinctive profiles as shown in Table 24.

Table 24

6-Cluster Means & Standard Deviations on Subscales

Cluster	EExt		EEmu		SExt		SEmu		CExt		CEmu	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1 (n=35)	36.29	4.18	41.53	5.08	7.96	2.31	9.13	2.56	32.10	2.94	34.74	4.18
2 (n=10)	11.60	5.41	13.40	6.41	33.75	4.61	36.15	5.87	28.30	6.42	30.40	7.65
3 (n=17)	12.71	1.61	13.68	2.31	49.15	3.51	51.56	3.85	34.12	3.95	36.44	5.14
4 (n=43)	44.29	3.59	51.01	3.21	8.08	1.75	9.26	1.86	38.40	3.06	43.38	3.85
5 (n=11)	21.18	5.12	25.18	5.75	41.41	6.38	45.86	6.19	36.64	2.40	39.73	3.59
6 (n=4)	20.25	2.87	22.88	2.50	12.88	6.93	13.63	8.43	25.63	2.06	27.25	2.90

A visual comparison of the clusters mean score profiles is presented in Figure 4.

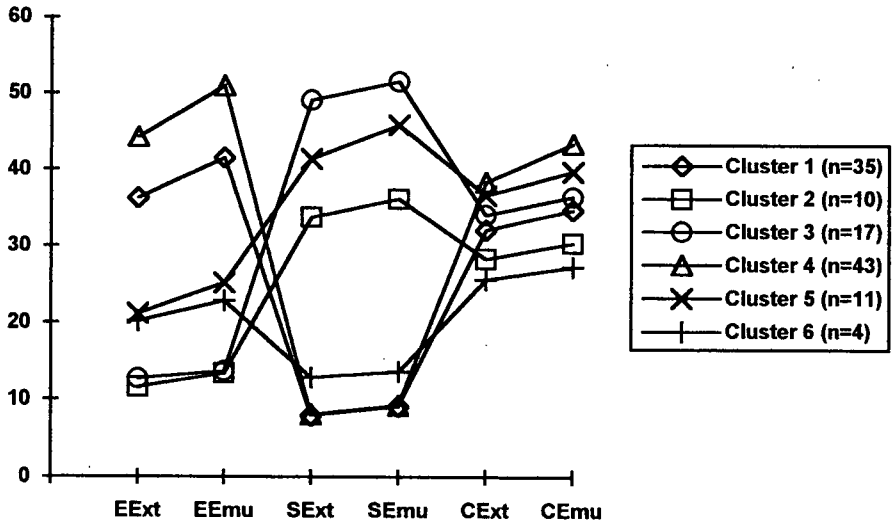


Figure 4. 6-Cluster Mean Score Profiles for Subscales.

Note: Ordinate is accumulated mean score.

When the subscale means are broken down by an ANOVA on the 6 clusters, high F scores and correlation ratio scores are produced as shown in Table 25. There is a possibility of some dependency effects in this analysis because the 120 sessions were composed of early and late sessions with the same clients.

Table 25

ANOVA scores: Subscale X 6 Clusters

Subscale	F	eta^2
EExt	257.19	.92
EEmu	283.35	.93
SExt	596.12	.96
SEmu	543.61	.96
CExt	26.61	.54
CEmu	26.85	.54

Note. $p < .001$ for all F values. Degrees of freedom = (5,114).

The results of cross-tabulating the 6 clusters with treatment type are shown in Table 26. A chi squared test on the table produced these results: χ^2 (10, N=120) = 124.69, $p < 0.001$.

Table 26

Cross-Tabulation of 6 Clusters with Treatment Modality

Cluster	ExST Individual	ExST Couples	SFT
1 (n=35)	22	13	
2 (n=10)			10
3 (n=17)			17
4 (n=43)	16	27	
5 (n=11)			11
6 (n=4)	2		2

A brief examination of the 3-Cluster analysis reveals equally striking results. The ANOVA of subscales by cluster produced F scores ranging from 317 to 655 and correlation ratio scores ranging from .84 to .92 on the SFT and ExST subscales, all with significance less than 0.001. The common subscales had lower, but still significant scores (for CExt F=32.82 Eta^2 = .3594.; for CEmu F=40.35, Eta^2 = .41). In the 3-Cluster analysis the low mixed sessions are grouped in with a moderate ExST cluster (n=39), leaving the high ExST cluster the same (n=43) and grouping the three SFT clusters together in one SFT cluster (n=38). There are only 2 SFT sessions that are included in the moderate ExST cluster (a 2% error). Refer to Table 27 and Figure 5 for the complete results. The

same convincing results appear in the smaller number of clusters, but some of the detail is lost that makes the 6-cluster analysis so interesting.

Table 27

3-Cluster Means & Standard Deviations on Subscales

Cluster	EExt		EEmu		SExt		SEmu		CExt		CEmu	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1 (n=39)	34.64	6.37	39.62	7.51	8.46	3.30	9.59	3.66	31.44	3.46	33.97	4.64
2 (n=38)	14.87	5.68	16.93	7.05	42.86	7.92	45.86	8.10	33.32	5.36	35.80	6.47
3 (n=43)	44.29	3.59	51.01	3.21	8.08	1.75	9.26	1.86	38.40	3.06	43.38	3.85

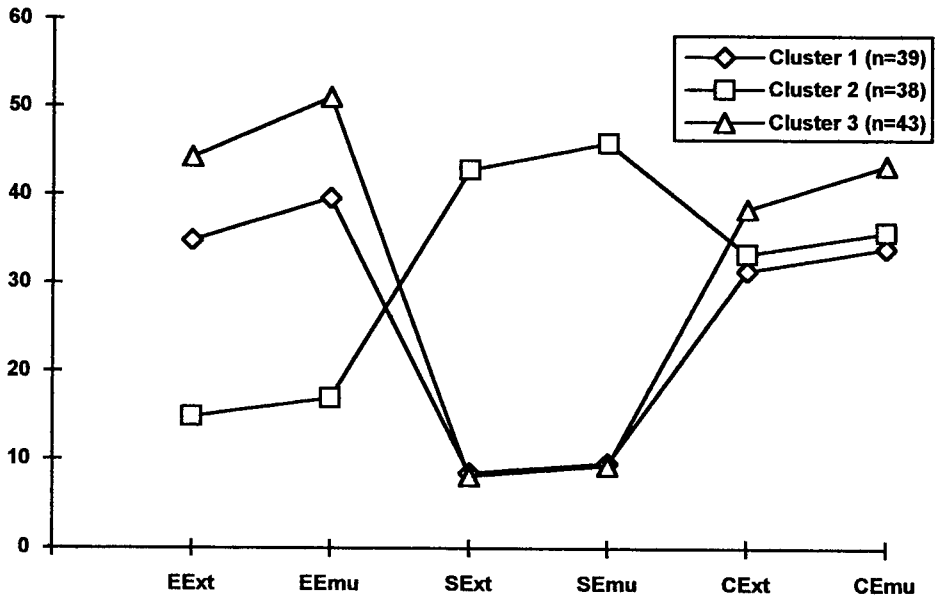


Figure 5. 3-Cluster Mean Score Profiles for Subscales.

Note: Ordinate is Accumulated mean score

Relative Average Adherence. Another consideration impacting adherence is examining whether the treatments were performed relatively fully. In other words, were the ExST and SFT therapists performing their respective treatments to the same degree? While analyzing this issue was not a part of the planned hypothesis testing procedure, a simple manipulation of the data generated by the 6-cluster analysis was done to indicate an area worthy of further analysis. The

data was used to create a Relative Adherence chart (Figure 6). Essentially, the scores of the intended treatment for the clusters were compared. For example, a comparison can be made of scores for ExST therapists performing ExST with SFT therapists performing SFT. This simply involved rearranging the data from the 6-Cluster Subscale Profile graph in Figure 4.

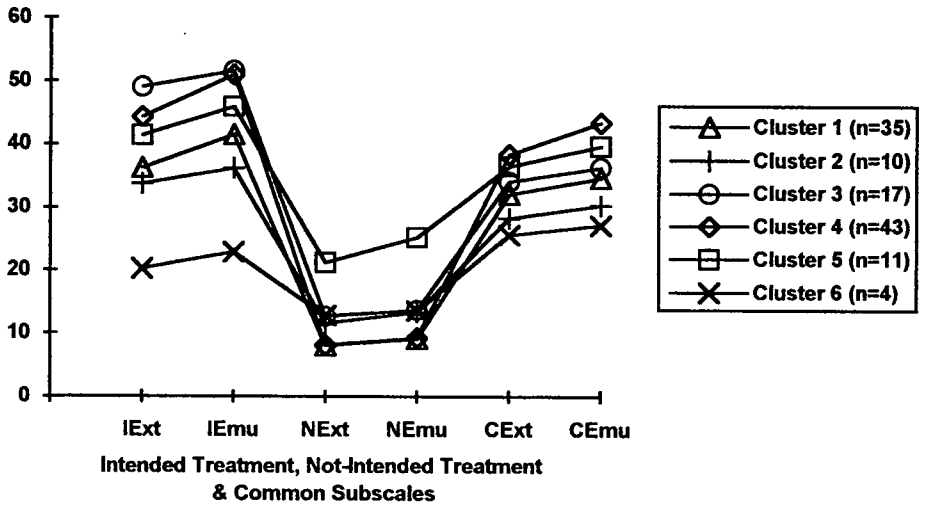


Figure 6. Relative Adherence of Treatments.

Note: Ordinate is accumulated mean score

Session Time. The results of cross-tabulating the 6 clusters with session time are shown in Table 28. A chi squared test on the table produced these results: $\chi^2(5, N=120) = 10.90$, $p = 0.05$.

Table 28

Cross-Tabulation of 6 Clusters with Session Time

Cluster	Early Session	Late Session
1 (n=35)	19	16
2 (n=10)	7	3
3 (n=17)	11	6
4 (n=43)	20	23
5 (n=11)	1	10
6 (n=4)	2	2

Chapter 5: Discussion

This chapter opens with a discussion of the results of the descriptive data analysis. Next, discussion of the preliminary quantitative analysis will begin by exploring in some depth the reliability and validity of the rating scale and rating process used to measure treatment adherence before examining the conclusions of the study regarding adherence. This will be followed by an exploration of the limitations of the study, and will end by considering possible directions for future adherence research.

Discussion of Results

Descriptive Analysis of Adherence. TARP was a well-designed study carried out in a thorough and careful manner. In terms of preparation for the project, there is clear support for the claim that SFT and ExST were distinct treatments likely to be performed with a high degree of adherence. Manuals were developed to outline theory and practice. Selection, training, and supervision of the therapists was carefully planned to address issues of adherence. All sessions were planned to be video-taped in order that in-session adherence could be rated. All of these a priori aspects of the project were then followed up in the actual running of the project.

Training and supervision were extensive, based on well-developed manuals. Therapists were well experienced in the alcohol recovery field and carefully selected for their performance of the selected treatment. It was clear in reviewing the planning and implementation of TARP that the literature describing the importance of treatment adherence in outcome research had been consulted in the study design. It is only through such careful elimination

of possible research error that the central determinants of therapy can be identified.

Preliminary Analysis. First, the reliability of TARS is discussed as a whole scale, then on the subscale level, and finally on the item level. Both content and construct validity of TARS are then explored. Next, rater bias and the rating process are examined.

Reliability of TARS. In order to measure the actual performance of therapist behaviors, the TARS must be both reliable and valid. The internal consistency of TARS as a whole was quite high, with a standardized item alpha on average of .75. This suggests TARS is reliably measuring something (like "systemic therapy").

Item-total correlations on the scale as a whole varied widely, with many low scoring items. When the three raters' item-total correlations over the whole scale of the extensiveness and emulation items from each subscale are averaged, a clear pattern emerges. The Common items on average show a .33 correlation indicating they relate most strongly to the scale as a whole. The ExST items on average show a much weaker correlation of .17, and SFT items on average have the lowest correlation with the scale as a whole, only .12. These figures suggest that the Common items reflect a construct that is to some degree common to each of these forms of therapy, while the ExST and SFT items reflect constructs that are much less general.

Reliability of TARS subscales. The internal consistency of the subscales is, of course, more important in terms of both reliability and validity, than that of the scale as a whole. The ExST and SFT treatment subscales showed extremely high internal consistency (.94 to .95). As these were the

subscales used to identify treatment performance, it was critical that they be highly reliable. The Common subscales had lower internal consistency scores (.76 to .80) which suggests they did indeed overlap with the other scales to a greater degree. In fact, internal consistency of the Common items was only slightly higher than the internal consistency of the scale as a whole (.75). While this indicates lower reliability on the common subscales, it is consistent with the intentions of the common subscale: to identify and measure therapist behaviors common to any of the treatment modalities.

A second way to measure subscale reliability is by examining the item-total correlations within each subscale. Most items on the two treatment subscales (ExST and SFT) had high item-total correlation scores (around 90% above .50), indicating good reliability. However, a few low-scoring items raised concerns that will be addressed in the section on item-analysis.

The third form of addressing reliability is through an ANOVA of items broken down by treatment. As reported in the results chapter, the variance in mean scores on the items from the ExST or SFT subscales is largely explained by treatment modality (on average 50% or more). On the other hand, very little of the variance in mean scores on the items from the Common subscale was explained by treatment (on average only 17%). In other words the ExST and SFT items were reliably related to treatment , while the Common items were not related to a specific treatment. Again, as with the item-total correlations, a few items from each subscale did not fit the pattern to varying degrees. These items will be discussed in the next section.

In summary, when examining subscale internal consistency, item-total correlations of the subscales, and the ANOVA of item by treatment, TARS has

demonstrated very high reliability overall. Items with anomalous scores on each of these measures will be examined in the next section.

Issues with Individual Items. Low variance on the item scores produced low inter-rater correlation scores on several items, particularly from the Common subscale. The relatively low variance of the Common subscale items overall is readily apparent in the SD scores for subscale means, which are less than half the size of the SD for the ExST and SFT subscale means as shown in Table 15. Specifically, item 5 (Accepting and Validating Client's Experience), item 11 (Warmth and Caring), item 31 (Therapeutic Rapport), and item 33 (Respectful) all had very high mean scores and little variation in the scores which lowered inter-rater reliability. Therapists were trained in both ExST and SFT to consistently perform the behaviors measured by the Common subscale. While producing low reliability scores, this does not necessarily reflect badly on the actual reliability or validity of the scale. In order to achieve more variation in the scoring, either more divergence in the sample (therapists' performance) or more subtle distinctions in the definition and/or rating of these items may be required. Either solution to this issue would increase reliability on these items.

Some items stood out because of low correlation ratio (eta^2) scores in an ANOVA of item broken down by treatment modality as shown in Tables 8-13. Items with low item-total correlation on the subscale level (see Tables 5-7), or items with low inter-rater correlation were also an issue (see Tables 17-19).

Item 44 (Ecosystemic Focus) was another Common subscale item. It had a high mean score with relatively low item-total correlations and somewhat low inter-rater correlations. As limited variation was not an issue with this

item, the low correlation scores suggest some problem with either item definition or rater training on this item. Ecosystemic Focus could look quite different in the different treatment modalities, and this may have contributed to the error.

Item 24 (Clients Define Their Own Experience) and item 3 (Developmental Focus) had fairly high mean scores, with somewhat low inter-rater correlations and one low inter-item correlation. These weak scores probably reflected an issue with item definition and/or rating. These were items that stood out in training as difficult to consistently rate.

Item 1 (Supportive Encouragement), item 8 (Tracking Learning), item 12 (Changes Desired in Relationships), item 28 (Relating Interpersonal Change to Therapy), item 32 (Encourages Independence), item 40 (Setting and Following Agenda) and item 45 (Discussing Ways of Changing Relationships) all had low mean scores and also had low variation in the scores, again lowering inter-rater reliability. Most were Common subscale items (except for items 28 and 32) and the low mean scores probably reflect either poor item definition and/or rating or therapist training issues. Another interpretation of these results is that these items reflect common therapeutic behaviors but ones which are infrequently used. Supporting this interpretation, items 1, 8, 12 and 28 showed little differences over treatment modality.

Item 45 (Discussing Ways of Changing Relationships) was infrequently rated, but showed a higher mean score on the ExST Couples treatment. This is the one of the few items that slightly distinguished the two ExST treatment modalities although couple's ExST tended to score very slightly higher than individual ExST.

Item 40 (Setting and Following Agenda) had very low item-total correlations with high inter-rater correlations. This is understandable because it showed a clearly higher mean score in the SFT treatment modality where a clear agenda is a basic part of the treatment. In ExST in contrast, spontaneity rather than an agenda is emphasized, although setting and following agendas are considered a part of ExST. Another interpretation of the low score could be that in ExST treatment, agenda may have been discussed in the first 10 minutes of the session, i.e. before the rated segment. Therefore, although this item was expected to be seen in both ExST and SFT, (and in fact was, though minimally for ExST) there was a basic difference between the modalities in terms of expected frequency for this item. Again, there may be some rater error for this item or the low mean score in ExST may reflect a therapist training issue, but this item does seem to reflect the intended differences in treatment.

Item 32 (Encourages Independence) was a SFT item but showed a low mean in all treatment modalities. This result may reflect a therapist training issue or may simply reflect the relative novelty of the intervention, which would suggest it is not a central technique of SFT.

Finally SFT item 14 (Primary Emphasis on Relationship Between Alcohol and Client System) showed low item-total correlations scores with one rater and moderate scores with the other two. Although this was an SFT item, the mean scores of the low scoring rater were very similar in all modalities. There was also some similarity in mean scores across modalities on this item for the other two raters. Since the inter-rater correlations for this item were high, these results probably reflect primarily rater training error.

When an ANOVA was performed on the items broken down by treatment, two different kinds of results stood out. First, there were ExST and SFT items whose mean scores did not seem to be explained by treatment. ExST Item 3 (Developmental Focus) showed clearly higher means on the ExST modalities, but the SFT means were also somewhat elevated, especially for Rater 3. This suggests the item definition was overlapping the two treatments to some degree, and/or that rater training could have been improved on that item. ExST item 28 (Relating Interpersonal Change to Therapy) had a very low mean for each treatment modality. This item appeared very infrequently and may reflect poor item choice. SFT item 14 (Primary Emphasis on Relationship Between Alcohol & Client System) and SFT item 32 (Encourages Independence) showed clearly higher scores on SFT than on either ExST modality, but the mean was fairly low, and there was some (even lower) ExST score. This result probably reflects poor item definition or poor item choice due to the low frequency of use.

The second kind of result was when the mean scores of Common subscale items could be partially explained by treatment. Most common items were not associated with treatment modality, which was expected and in fact supports the claim that basic therapeutic behaviors were performed equally fully across treatment. Common item 20 (Appropriate Expression of Therapist's Feeling), item 29 (Focus on the Therapist and Client Relationship) and item 41 (Primary Empathy) all showed slightly elevated ExST scores, and probably reflects genuine intended differences in the levels these items were performed in the different modalities. These items are still common to both, but more common to ExST. Item 40 (Setting and Following Agenda) on the

other hand, showed clearly higher scores in the SFT treatment modality. As discussed above, this is partly due to the clear, even central part it played in SFT treatment, and the less frequent part it played in ExST treatment. The variation in rater scores, however suggests there is also an issue of rater training on this item. This item looks very different in the different modalities, and special care must be made to identify the different versions of it.

In summary, there were 18 items with scores that raised questions on eta squared scores, means, inter-rater correlations, or item-total correlations. The vast majority of these items were from the common subscale (13 items), with 3 from the ExST subscale and 2 items from SFT. 11 of the items (9 common, 1 ExST and 1 SFT) had low score variance, which lowers reliability. Error on 12 of the items may be influenced by some combination of problems with item definition, rater training or therapist training. 7 items may have questionable scores simply because they appear with low frequency. 2 items stood out because of apparent modality-specific variation in therapist behaviors in SFT or ExST treatment. Three Common subscale items showed slightly more variation across treatment than could be explained by chance, and probably reflect differences in degree for intended treatment. This summary suggests the rating scale is a robust measuring device, with room for improvement on a few items.

Validity of TARS. Whenever a new measuring device is developed and found to be reliable, the next question must be whether it is a valid measure. TARS has strong support for its validity in two ways. First, content validity can be shown as TARS was developed based on training manuals for the treatments used in the study, and the scale was developed by the author in close

collaboration with the developers of the treatment models. Each item and the examples were reviewed by four of the people instrumental in the practical and theoretical development of the treatment modalities. Therapist behaviors expected only in unique sessions like beginnings or endings were deleted from the scale. Consequently TARS represents the full range of mid-session content for typical sessions of each treatment type.

Second, construct validity was indicated in two ways. Cluster analysis shows very clear groupings of score profiles with a very high relationship between treatment type and the SFT or ExST subscales (χ^2 (10, N=120) = 124.69, $p < 0.001$). Secondly, validity is supported when the results showed very strong internal consistency (i.e. high correlation) within the SFT and ExST subscales. The standardized item alpha scores, averaged across three raters, were .94 for ExST Extensiveness (EExt), .95 for ExST Emulation (EEmu), .95 for SFT Extensiveness (SExt), .95 for SFT Emulation (SEmu). These scores suggest that the items from each treatment subscale measure different aspects of the same basic constructs: ExST and SFT.

As well, there were strong negative correlations across the SFT and ExST subscales: Pearson correlation coefficient scores ranged from -.84 to -.86. This shows that the items on the SFT and ExST subscales were measuring different constructs from one another, each with clear boundaries and distinct contents.

The items expected to be common to both treatments were less internally consistent, but still showed moderately high alpha's: .76 for Common Extensiveness (CExt) and .80 for Common Emulation (CEmu). The Common item subscales were however not statistically distinct from the SFT and ExST subscales: Pearson correlation coefficients ranged from .45 to .52 between the

ExST and the Common subscales which indicates a weak positive relationship, and from -.17 to -.08 between the SFT and Common subscales, which indicates no relationship. These results are consistent with the expectations of the developers of the treatments. While both treatments were expected to use the common items, SFT was intended to be a more minimal treatment. The Common items show up in both treatment modes, but with somewhat more extensiveness and/or more emulation in ExST.

To sum up, TARS has been shown to have both content and construct validity in terms of the ExST, SFT and Common subscales.

Rater Bias. Rater bias may show up in several forms. While every effort was made to reduce rater bias in the training and implementation of this study, it is still a topic worth examining carefully. The tapes were viewed in a randomized order which should prevent any bias in rating due to presentation order.

All three raters went through the same specific training process, watching the same training tapes. The author of this study was involved both in the construction of the rating scale and in the rater training, as well as the actual rating of tapes. Despite this lack of blindness, the three raters produced a high degree of inter-rater reliability, with no evidence of rater bias. This lack of bias is partially explained by the scale instructions focusing on looking for the presence of items without overtly ascribing them to a treatment modality. The rater training process was also collaborative, as all the raters, including the author of this study worked together to understand and apply the rating of items to actual sessions.

Special care was taken to examine potential rater bias in the case of the author of this thesis, by comparing his scoring profile with two other raters. Subscale reliability scores were extremely similar. Subscale means were also very similar, and showed a slight tendency for rater 3 to rate higher overall as shown in Table 15. Means for Rater 1 and 2 were virtually identical.

When the inter-rater correlations are examined for the items on each subscale, a more complex relationship emerged as shown in Tables 17-19. The most important result from these charts is that no consistent pattern emerged. There is no indication that Rater 1 (the author) was systematically biased. Each pair of raters showed higher correlation scores on different items, but no pair emerged as consistently more similar. Correlation scores tended to be very similar on most items, with only a few items showing a wide disparity in correlations. The Common items showed the most variation in correlation scores, but again showed no pattern of systematic differences between the raters.

Finally, an ANOVA of the subscale scores broken down by the three raters indicated the variance in the subscale scores was not explained by raters. The only subscale with an F score approaching significance was the Common ExST, and rater accounted for only 5% of the variance in the subscale scores. Given the above results, there is strong evidence supporting the claim that Rater 1 was not systematically biased because of his involvement in developing the rating scale. The main portion of the study then, is based on ratings by the two primary raters, 1 and 2.

Reliability of the Rating Process. While this study does not examine every session of therapy, it does provide a extremely good sample of therapist

behavior in TARP. The results are so consistent that there is good reason to be very confident that these results do indeed reflect the standard of therapist performance throughout TARP.

There was a great deal of evidence supporting claims of inter-rater reliability. Inter-rater correlations on the individual items was generally high as shown in Tables 17-19, and internal consistency scores on the subscales for each rater were high and virtually identical as shown in Table 16. Mean scores and standard deviations for the subscales were very similar across the raters (see Table 15). Inter-Rater correlations for the two primary raters on the subscales was very high (EExt .85; EEmu .83; SExt .93; SEmu .93; CExt .54; CEmu .48) and all highly significant (p= 0.000) as shown in Table 22. Correlations between raters across extensiveness and emulation subscales for the same treatment subscale were slightly lower but highly consistent with results within each rater. For example, Extensiveness and Emulation within each subscale were highly correlated across raters: from .83 to .84 on the ExST subscales; from .93 to .93 on the SFT subscales; and .52 to .50 on the Common subscales. In other words, the raters were highly reliable in their rating of the ExST and SFT treatment items. As noted above, a large part of the low inter-rater reliability scores on the Common items was due to low variability in the rating scores, high or low, and may in fact not reflect inconsistent rating.

When correlations on the extensiveness and emulation subscales are averaged for each treatment subscale, there is a very clear correspondence between subscale correlations within each rater and across raters. Both within and across raters the ExST and SFT treatment subscales show high positive

correlation within subscales and high negative correlation across subscales. When the common subscales enter the picture, between raters correlation is only moderately high (.51) though still highly significant (p= 0.000).

In this section, inter-rater correlations have been examined on the individual item level, the subscale level and the averaged treatment subscale level. As well, internal consistency, means and standard deviations have been compared. Altogether, these results indicate the rating process was highly reliable.

Discussion of Hypothesis Testing Results. The following discussion of the hypothesis testing results begins with adherence and differentiation, focusing on a 6 cluster analysis of the rated sessions, and finishes with a discussion of the relative average treatment adherence. A brief discussion of the impact of early or late session time completes the section.

Hypothesis-Testing: Adherence and Differentiation. The central aim of this study was to determine whether actual treatment as performed scored most highly in terms of extensiveness and emulation on the related subscale items based on the intended treatment. Examining the clusters formed by similar scoring profiles will show if the clusters differentiate between treatment modalities, and studying how much of the variance between these clusters of scores on the variables of extensiveness and emulation, in the ExST, SFT and Common subscales is explained by treatment modality will give a measure of the degree of treatment adherence.

An ANOVA of the subscale scores broken down by treatment gave preliminary indication that treatment was adhered to, and that ExST and SFT treatments were distinct from one another as shown in Table 23. The variance

in subscale scores was largely explained by treatment. Averaged over the three raters, the correlation ratio of the items on the ExST extensiveness and emulation subscales was .78. The correlation ratio of the items on the SFT extensiveness and emulation subscales, averaged over three raters was .86. These scores very strongly suggest that ExST and SFT are distinct from one another and that the ExST items were seen to be performed in the ExST sessions, while the SFT items were seen to be performed in the SFT sessions. The correlation ratio of the items on the Common extensiveness and emulation subscales, when averaged over the three raters was, on the other hand very low, only .12. In other words, very little of the variance in the Common subscale scores was explained by treatment type. This fits the intentions of the Common items, namely to reflect therapist behaviors common to any treatment modality.

In order to develop a detailed discussion of the treatment adherence in TARP, it is important to consider the results of the cluster analysis performed on the 6-point profiles of each session. The high inter-rater reliability scores supports the decision to average the two primary raters scores. The high internal consistency reliability of the subscales supports the decision to create 6 average subscale scores, rather than trying to deal with a total of 90 individual items. The results of the cluster analysis indicate extremely strong evidence that the variance in the data can be explained almost fully by the clusters.

As mentioned above, the six clusters are descriptively labeled as follows: a) Cluster 1, moderate ExST (n=35); b) Cluster 2, moderate SFT (n=10); c) Cluster 3, high SFT (n=17); d) Cluster 4, high ExST (n=43); e) Cluster 5,

blended SFT (n=11); and f) Cluster 6, low mixed (n=4). When the profiles of mean subscale scores for each cluster are examined in Figure 3, the shape of the profile of the two ExST clusters are clearly very similar. The profiles of the three SFT clusters are also clearly very similar to one another. Only the low mixed cluster (n=4) stands out as being unique.

The two ExST treatment modalities (individual and couples) were split nearly evenly into the two ExST clusters as shown in Table 26. Individual ExST sessions were split in half (22 sessions in Cluster 1 and 16 sessions in Cluster 2). The two anomalous ExST sessions in the low mixed cluster were from the same individual case. Two thirds of the Couple's ExST sessions were in the high ExST cluster (13 sessions in Cluster 1 and 27 sessions in Cluster 4). However, early and late sessions of both individual and couples cases were evenly split in Clusters 1 and 4 as shown in Table 28. While mean scores for the couple's ExST sessions were slightly higher than the individual ExST sessions, this difference was not significant. ExST appears to be performed in a very similar fashion with individuals or couples. The high degree of overlap of the individual and couples ExST modalities lends itself to treating the two modalities as one unit (ExST), while recognizing a general tendency to see slightly higher ExST scores in the Couple's sessions.

Both the moderate and high ExST clusters were composed entirely of ExST treatment modality cases. Mean ExST scores were very high on both the extensiveness and emulation subscales (Cluster 1 EExt 36.29, EEmu 41.53; Cluster 4 EExt 44.29, EEmu 51.01). Mean SFT scores for these clusters were very low on both extensiveness and emulation subscales (Cluster 1 SExt 7.96, SEmu 9.13; Cluster 4 SExt 8.08, SEmu 9.26). Mean Common scores for these

clusters were moderate to high on both the extensiveness and emulation subscales (Cluster 1 CExt 32.10, CEmu 34.74; Cluster 4 CExt 38.40, CEmu 43.38). These results show that the therapists in the ExST sessions adhered to the ExST and Common items, and did not adhere to the SFT items.

The moderate, high and blended SFT clusters showed an almost opposite profile to the ExST clusters, except for the Common subscales. These clusters were composed entirely of SFT treatment modality cases. Mean SFT scores were very high on both the extensiveness and emulation subscales (Cluster 2 SExt 33.75, SEmu 36.15; Cluster 3 SExt 49.15, SEmu 51.56; Cluster 5 SExt 41.41, SEmu 45.86). Mean ExST scores for Clusters 2 and 3 were very low on both extensiveness and emulation subscales (Cluster 2 EExt 11.60, EEmu 13.40; Cluster 3 EExt 12.71, EEmu 13.68). Cluster 5 was made unique in the SFT clusters by the somewhat elevated ExST scores (EExt 21.18, EEmu 25.18).

Careful study of these session ratings revealed that three items "leaked" in from ExST: item 4 -- Advanced Empathy, item 24 --Clients Define Their Experience, and item 37 -- Therapist Involvement. There was also a very high proportion of late sessions in this cluster (10 out of 11), the only cluster to show this kind of bias concerning session time. This "leaking" seems likely to result from deeper client-therapist connection as time goes on.

Other items, like Developmental Focus, Spontaneity, Systemic/Relational Rationale, Collaboration Or Mutuality, Relating Interpersonal Change to Therapy, and Emphasis on Awareness were more sporadically rated highly and probably reflect some therapist drift from the original training. Alternatively there could be some problem with item

definition, rater or therapist training for some of these items, especially Developmental Focus, Clients Define Their Own Experience and Relating Interpersonal Change to Therapy (as discussed in the section "Issues with Individual Items").

Mean Common scores for all 3 SFT clusters were moderate to high on both the extensiveness and emulation subscales (Cluster 2 CExt 28.30, CEmu 30.40; Cluster 3 CExt 34.12, CEmu 36.44; Cluster 5 CExt 36.64, CEmu 39.73). The high common scores for cluster 5 and the moderate positive correlation between the Common and ExST subscales may suggest these sessions reflect some overlap between ExST and Common therapeutic treatment. These results show that in all cases the therapists in these SFT sessions adhered to the SFT and Common items, and most did not adhere to the ExST items. There was a limited amount of "leaking" on a few ExST items in 11 of the SFT sessions.

The low mixed cluster was the odd group out. Composed of 2 ExST and 2 SFT sessions, these sessions showed low ExST, SFT and Common subscale means (EExt 20.25, EEmu 22.88, SExt 12.88, SEmu 13.63, CExt 25.63, CEmu 27.25). These very flat profiles showed a generally low score for the 2 sessions from the same ExST case. Notes in the margin of the rating forms of these ExST sessions indicate this was a non-responsive client. For the 2 SFT sessions relatively high scores on Advanced Empathy, Spontaneity, Systemic/Relational Rationale, Appropriate Intensification of Experience, Symbolic Orientation or Therapist Involvement slightly elevated the ExST scores for the SFT sessions and these sessions were extremely flat on the SFT items. These sessions seem to reflect very low use of unique therapist

interventions from either treatment modality, with the common items reaching only moderate scores (the lowest of any cluster). As these sessions compose only 3% of the total sample, these sessions appear to be exceptions in the generally high performance of the therapies.

The results of the cluster analysis overwhelmingly argue for both treatment adherence and treatment differentiation. A chi square test of treatment modality and cluster produces these results: χ^2 (10, N=120) = 124.69, $p < 0.001$. Treatment modality is extremely highly significantly related to cluster.

Given the strong indication of treatment adherence based on TARP's design and procedural implementation, the quantitative study of actual sessions was likely only to add details to the picture, not disconfirm expectations. However, the results of this study are remarkably unequivocal: intended treatment was seen extensively; it closely emulated the planned treatment; and it was distinctive.

Relative Average Adherence. There is strong evidence supporting the claim that the therapists did adhere to the treatment modalities assigned, and that the treatment modalities were distinct from one another. There was also some indication in the results concerning adherence that addressed the issue of whether averaged measures of extensiveness and emulation for the intended treatment items, not-intended treatment items and common items were consistent across treatments. While not statistically analyzed, this information is worth mentioning briefly, and merits further statistical analysis in future research. Were the ExST and SFT therapists performing their respective treatments to the same degree? In order to answer this question, the data

generated by the 6-cluster analysis was used to create a Relative Adherence chart (Figure 5). In this case, the intended and not-intended treatment scores for the clusters were compared. For example, a comparison was made of scores for ExST therapists performing ExST with SFT therapists performing SFT. Essentially this simply involves rearranging the data from the 6-Cluster Subscale Profile graph in Figure 3.

Looked at in this way it is clear that cluster 4 (high ExST) and cluster 3 (high SFT) have a virtually identical profile. Cluster 1 (moderate ExST) and Cluster 2 (moderate SFT) have very similar profiles. The common subscale scores for these four clusters were all quite high and similar.

Cluster 5 (blended SFT) shows a close relationship to all four clusters on the intended treatment subscale, but there is a definite elevated score on the not-intended treatment subscale, again indicating some limited blending of modalities. Also, the Common subscale scores were quite high for cluster 5, in fact the highest of the SFT clusters.

Cluster 6 (Low mix) showed low scores on the "intended" treatment modality (ExST) and even lower ones on the "not-intended" treatment modality (SFT). As discussed above, 2 sessions were from the same ExST case, and may have been simply weak sessions. However the 2 SFT sessions revealed the odd pattern of slightly elevated ExST scores and very flat SFT scores. These sessions may be the only examples of poor relative adherence in the sample (less than 2%). Cluster 6 also showed the lowest Common subscale scores, although they were fairly close to cluster 2 on the Common subscales.

In general, the Common subscale means for each cluster were moderate to high for all six of the clusters, suggesting the therapists from each modality

performed the common therapeutic interventions to a high and similar degree. When the Common extensiveness and emulation subscale means for the clusters are examined in Table 24, it is apparent that the common items had consistently high means. This does suggest the therapists were performing the common items to a similar degree.

Obviously, a statistical analysis examining the significance of the differences between the cluster profiles, or a MANOVA of the profiles of intended/not-intended/common scores on extensiveness and emulation, broken down by treatment modality would be a valuable addition to this research.

In this study however, only a visual comparison of profiles suggests the relative levels of adherence were very similar across the different treatment modalities, particularly on the Common subscale items. This indicates that the therapists in almost all cases performed the different forms of treatment to a similar degree. This point is important because if both treatments were adhering, but one was consistently performed less frequently, or with a lower degree of emulation, then treatment adherence could still be a factor contaminating any outcome research.

Extensiveness and Emulation. Some, though not all, of the literature on adherence suggests there is a critical distinction to be made between extensiveness of treatment and skill or quality of treatment. Emulation was introduced in this study as an attempt to measure quality in the most descriptive, objectively based way possible.

Each scale item was rated both in terms of extensiveness and emulation. Every effort was made to clarify the distinction between these two subscales, and the raters commented they used different criteria for their scoring as suggested in the rating manual. However, in this study, extensiveness and emulation were so highly correlated, they seem to be measuring essentially the same construct. This was shown most clearly in Tables 21 and 22. In Table 21, subscale correlations were based on the averaged scores of Raters 1 and 2. For both the ExST and the SFT subscales, extensiveness and emulation were correlated .99. The Common extensiveness and emulation subscales were correlated .96. It seems that the extensiveness and emulation subscales are essentially measuring the same construct.

This evidence strongly suggests that measuring extensiveness alone does provide an adequate measure of adherence both in terms of frequency and skill. However, another possible explanation is that more rigorous item definition of the distinction between extensiveness and emulation and/or rater training is required. Also, another possibility is that emulation as defined in this study is too specific, and that "skill" in the sense of the planning, timing and performance of therapist interventions can only be judged on an even more global level, by highly trained experts. However, given the high degree of agreement by the raters in this study, this argument lacks a lot of support. There appears to be a great deal of correlation between frequency and quality. While this argues for scale reliability, it raises questions about the validity of the distinctiveness of the extensiveness and emulation scores. While the mean extensiveness scores were slightly lower than the emulation scores, this difference was not statistically significant. These subscales may be measuring

essentially the same thing and there may be no valid difference between them. However, this lack of distinction between extensiveness and emulation does not detract from the clear evidence supporting adherence.

Session Time. The sample of video taped sessions was composed of two sessions from each case, one early session and one late session. When early or late session time was cross tabulated with cluster, these results were produced: $\chi^2(5, N=120) = 10.90$, $p = 0.05$. This suggests session time had a slight but still significant impact on the rating. This appears to accounted for primarily by the SFT clusters (Cluster 2,3, & 5). When half of the late SFT sessions are in the blended SFT cluster, there is some indication that SFT was affected by time. Whether this is a result of problems with item definition, rater or therapist training, or therapist drift is not entirely clear. It does indicate that some items expected to be performed only in ExST sessions do in fact show up minimally in later SFT sessions.

Limitations of the Study.

One limitation of this study is in the use of TARS as a new measuring instrument. Although the reliability and validity is quite strong, because the scale was developed and tested on the study sample, there are four issues regarding scale development or use that could be improved.

First, as discussed above, a few items stood out as questionable. This was partly because of the extremely consistent behavior of the therapists on several Common subscale items. One possible way to improve the reliability of these items would be to use the scale on wider sample of therapists. Improving item definition is also another possibility. There were also a few items with low internal consistency scores on the subscales. The scale could be improved by

reexamining these items. The experts involved in the scale construction now have more information in order to improve item construction, definition or selection. It is apparent that some items need to be questioned in terms of how well they fit the treatment modality, or how central they are to the therapy, or whether the item definitions captured the key characteristics as well as they could. Low scores of one or two items also indicated that a particular aspect of the theory could have been emphasized more in the therapists' training. These improvements would only have a marginal impact on the reliability of the scale however.

A second limitation of the study due to the use of this scale is the extent to which it is specific to TARP. Both ExST and SFT subscales represent unique forms of therapy. While the scale needs to closely follow the particular treatment methodology in order to evaluate adherence, it may make comparisons with other treatment approaches difficult. While the Common subscale builds on ideas from other rating scales which address "general therapeutic interventions", it cannot pretend to represent a consensus of what the field of counselling considers general skills. In essence then, two questions arise. First, to what extent is this research generalizable? To what extent do the subscales or their individual items reflect counselling skills in other forms of therapy? Second, under what circumstances can this scale, in whole or in part, be used again? This question may an inevitable part of the process of developing research in counselling as it searches for just what characterizes good therapy.

A third limitation to the study related to the development of the rating scale is the use of the concept of therapeutic emulation. Emulation as such, is

a new concept, although it attempted to address issues already raised in the literature regarding therapeutic quality or skill level. The high correlation between extensiveness and emulation in this study indicates, but does not prove, that measures of extensiveness may in fact be highly indicative of therapeutic quality or skill level. However, because this study contains the only research regarding this term, the use of emulation measures remains a limitation regarding the generalizability of this study. Waltz et al. (1993) suggest that certain items on the Collaborative Study Rating Scale require the rater's inference about the quality of the intervention, not just the frequency, in effect overlapping the constructs of adherence and competency. Also, Koenigsberg et al. (1993) define skill as general or non-specific therapeutic behaviors. In the present study, TARS includes items that require inferences about appropriate levels of therapist's behavior (for example Item 21: Appropriate Intensification of Experience), and clearly includes measures of general therapeutic behaviors like warmth and empathy in the Common subscale. Perhaps the high degree of correlation between extensiveness and emulation reflects an overlap of constructs in the construction of the scale items. However, there is no clear agreement in the literature at this time as to exactly what "skill" is and is not.

Fourth is the issue of the level of focus of the rating scale items. TARS emphasizes a more global level of rating, and more detailed, process-level rating may have revealed more or different information.

Another area of limitation of this study is the lack of statistically exploring the link between adherence and outcome. Ideally this would be done

in terms of extensiveness and emulation on both the level of the treatment subscales and the individual items.

Another significant limitation of the study was the lack of blindness of the author of the study as one of the primary raters. While the data supports the claim of a lack of bias, it would of course be preferable to use raters who were completely blind to the intentions and design of the study.

This study also makes no attempt to measure client difficulty or degree of participation. Waltz et al. (1993) suggest that client difficulty must be taken into account in order to properly interpret adherence or competency ratings. They suggest that client "difficulty ratings could either be used as covariates or as a basis for the selection of cases" (p. 628). LeLaurin & Wolery (1992) recommend documenting the participation or involvement of clients. A client's reaction, involvement or attention during an intervention may have as much, or more, impact on outcome as treatment adherence. While clients in this sample were selected to have a similar degree of distress with alcohol as the primary problem, neither client difficulty nor participation was monitored during the period of therapy, and was not considered in this study. Perhaps this information is more appropriate in outcome research, but it clearly may have some impact on adherence as well.

A critical issue in outcome research in general is in the selection and definition of the independent variables. What is the range of constructs being measured? What are the key therapeutic elements? While this rating scale adds to the research addressing these questions, it cannot claim to conclusively answer any of them. The process of item selection and definition was based on careful study of the treatment training manuals, as well as the

study of other rating scales, and involved careful discussion and agreement of several experts in these models. However, these items are in essence only educated guesses as to what really are the key therapeutic factors. As research progresses, it may be found that some apparently incidental therapeutic interventions were in fact key to therapy, while apparently significant (and cherished) interventions had little impact in truth. This limitation is of course not unique to this study. All attempts to define and measure therapy have this issue in common.

Conclusions.

In this study of treatment adherence, a cluster analysis produced extremely clear results indicating that treatment was delivered as planned in the respective manuals. Both individual and couple's ExST therapists rated highly on ExST and Common subscale items and low on SFT items. Conversely, SFT therapists rated highly on SFT and Common subscale items and low on ExST items. There were four sessions (3%) that proved the exception to the general rule, and were relatively flat sessions on every subscale. Early or late session time was not shown to be significant in the cluster analysis, suggesting that therapist drift was not a significant factor and that adherence was maintained throughout the study. There was some limited blending of treatments in some of the later SFT sessions, but there was still a clear distinction between SFT and ExST treatment. These results indicate a high degree of treatment adherence as well as clear differentiation between treatments. There is some indication that the treatments were performed to very similar relative levels of extensiveness and emulation.

Further Research Directions.

As indicated earlier, one obvious piece of future research would involve statistically measuring the relative levels of adherence across treatment modalities.

Another important future research direction involves exploring the construct of emulation. In this study emulation and extensiveness were highly correlated and seem to be measuring essentially the same construct. Further research is needed to examine this result. One approach to studying the relationship between extensiveness and emulation would be to have more experienced therapists rating sessions on these two dimensions and determine whether they observe more differences between the measures. A second approach could be to have raters rate extensiveness or emulation but not both, and then examine what the correlation is between the two scales. The inclusion of both dimensions on each item may have biased the degree of correlation. A third approach, which some research has taken, would be to define extensiveness and emulation items in a way that are not obviously related. The research using terms like therapist skill or quality has taken this route, with a distinctly different (more global) level of analysis of "quality". One possible problem with this approach is that it assumes extensiveness and quality are not measuring the same thing, and may create a false distinction between them. More research is required in order to create agreement in the literature as to just what constitutes skill, competency or emulation, in contrast to treatment fidelity in terms of simple extensiveness.

One obvious area for future exploration would be to compare adherence ratings with outcome, both on the level of the treatment subscales, but also on

the level of individual items. In this study adherence is so consistent that outcome cannot fruitfully be compared with variation in adherence. Deliberate, systematic manipulation of specific elements of treatment adherence may become a profitable alternative to control-group studies. In this manner the keys to effective outcome could possibly be identified.

Another potential area of research involves exploring the covariation of client difficulty and treatment adherence and their combined effect on outcome. Especially given the systemic theoretical foundation of ExST, the impact of both therapist variables and client variables may reveal much about the process and outcome of therapy. Also significant may be the actual effect of client difficulty on the therapists' treatment adherence itself. Waltz et al., (1993) point out that competent therapist behaviors will vary with the difficulty of issues a client is dealing with. Certainly therapist performance of some items in this study could be influenced by the level of client difficulty or participation.

The use of adherence rating in this study has focused on verifying therapist performance for comparative outcome research. However, there appear to be many other potential uses for adherence ratings, some of which have been mentioned in the literature. Referring to ratings of treatment adherence can be used to train and/or supervise therapists. The ratings may also be used to clarify the therapeutic theory or method a rating scale is based on. When coupled with outcome research, adherence ratings may help determine what really makes therapy work. As computers improve, the potential for a recursive, developmental use of adherence ratings is more feasible. In other words, adherence ratings may influence our definition of

therapy, our training as therapists, and our understanding of what is central to therapy, and these new understandings will in turn change our approach to adherence rating. As feedback concerning treatment adherence becomes faster and easier to implement, counselors have a much greater potential for identifying what specifically works in therapy, and of consistently performing it. However, caution must be used in this pursuit as Aradi and Piercy (1985) point out. Its is important not to put adherence to treatment regime above our ethical duty to respond to the client's needs.

Another avenue of further research would be to compare ratings of the relatively inexperienced Master's students used in this study with the ratings of more experienced therapists. This avenue of research may indicate whether there are systematic or significant differences in the ratings. This may lead to clearer item definition, or help define essential rater characteristics.

Given the slight, but significant impact of session time on clustering, further exploration of therapist drift, particularly in SFT is also warranted.

TARS has been shown to be effective as a measuring instrument in this study. Despite the instrument's strength, improvements should be made in item construction and selection before it is used in further research. Potential areas for further research include using the scale as a whole, individual subscales or selected individual scale items.

As the library of rating scales continues to grow in adherence literature, the possibility of developing an outline of therapeutic behaviors that is capable of rating adherence across different treatments also grows. The literature has shown that rating scales developed to be used in one setting may be profitably applied, adapted, or modeled in another. Whether used in rating adherence for

the purposes of training, supervision, or outcome research, TARS is now part of that library.

This study has been an important part of supporting the research of The Alcohol Recovery Project. Hopefully this research can be used and furthered by other studies conducted under its auspices.

References

Aradi, N. S., & Piercy, F. P. (1985). Ethical and legal guidelines related to adherence to treatment protocols in family therapy outcome research. The American Journal of Family Therapy, 13, 60-64.

Arnkoff, D. B. (1986). A comparison of the coping and restructuring components of cognitive restructuring. Cognitive Therapy and Research, 10, 147-158.

Avis, J. M., & Sprenkle, D. H. (1991). Outcome research on family therapy training: A substantive and methodological review. Journal of Marital and Family Therapy, 16, 241-264.

Beutler, L. E. (1993). Designing outcome studies: Treatment of adult victims of childhood sexual abuse. Journal of Interpersonal Violence, 8, 402-414.

Beutler, L. E., & Hill, C. E. (1992). Process and outcome research in the treatment of adult victims of childhood sexual abuse: Methodological issues. Journal of Consulting and Clinical Psychology, 60, 204-212.

Boruch, R. F., & Gomez, H. (1977). Sensitivity, bias, and theory in impact evaluations. Professional Psychology, 8, 411-434.

Butler, G., Fennell, M., Robson, P., & Gelder, M. (1991). Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety disorder. Journal of Consulting and Clinical Psychology, 59, 167-175.

Caplan, R. D., Vinokur, A. D., Price, R. H., & van Ryn, M. (1989). Job seeking, reemployment, and mental health: A randomized field experiment in coping with job loss. Journal of Applied Psychology, 74, 759-769.

DeRubeis, R. J., & Feeley, M. (1990). Determinants of change in cognitive therapy for depression. Cognitive Therapy and Research, 14, 469-482.

DeRubeis, R. J., Hollon, S. D., Evans, M. D., & Bemis, K. M. (1982). Can psychotherapies for depression be discriminated? A systematic investigation of

cognitive therapy and interpersonal therapy. Journal of Consulting and Clinical Psychology, 50, 744-756.

Dobson, K. S., & Shaw, B. F. (1988). The use of treatment manuals in cognitive therapy: Experience and issues. Journal of Consulting and Clinical Psychology, 56, 673-680.

Elliott, R., Hill, C. E., Stiles, W. B., Friedlander, M. L., Mahrer, A. R., & Margison, F. R. (1987). Primary therapist response modes: Comparison of six rating systems. Journal of Consulting and Clinical Psychology, 55, 218-223.

Elkin, I., Parloff, M., Hadley, S. W., & Autry, J. H. (1985). NIMH treatment of depression collaborative research program: Background and research plan. Archives of General Psychiatry, 42, 305-316.

Elkin, I., Pilkonis, P. A., Docherty, J. P., & Sotsky, S. M. (1988). Conceptual and methodological issues in comparative studies of psychotherapy and pharmacotherapy, I: Active ingredients and mechanisms of change. The American Journal of Psychiatry, 145, 909-917.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pikonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J., & Parloff, M. B. (1989). National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. Archives of General Psychiatry, 46, 971-982.

Fleiss, J. L. (1981). Balanced incomplete block designs for interrater reliability studies. Applied Psychological Measurement, 5, 105-112.

Flexner, S. B., Stein, J., & Su, P. Y. (Eds.). (1980). The Random House dictionary: Concise edition. New York: Random House.

Foley, S. H., O'Malley, S. H., Rounsaville, B., Prusoff, B. A., & Weissman, M. M. (1987). The relationship of patient difficulty to therapist performance in

interpersonal psychotherapy of depression. Journal of Affective Disorders, 12, 207-217.

Fonagy, P., & Moran, G. S. (1990). Studies on the efficacy of child psychoanalysis. Journal of Consulting and Clinical Psychology, 58, 684-695.

Friesen, J. D., Grigg, D. N., & Peel, C. P. (1988, December). Therapy manual: Experiential systemic therapy for individuals, couples and families. Unpublished manuscript.

Friesen, J. D., Grigg, D. N., Peel, C. P., & Newman, J. A. (1989, December). Systemic experiential therapy manual for individuals and couples. Unpublished manuscript.

Goldberg, R. J., Wool, M., Tull, R., & Boor, M. (1984). Teaching brief psychotherapy for spouses of cancer patients: Use of a codable supervision format. Psychotherapy and Psychosomatics, 41, 12-19.

Goldfried, M. R. (1990). Individual psychotherapy: Process and outcome. Annual Review of Psychology, 41, 659-688.

Grigg, D. N. (1994). An ecological assessment of the efficacy of individual and couples treatment formats of Experiential Systemic Therapy for alcohol dependency. Unpublished doctoral dissertation, University of British Columbia, Vancouver.

Greenberg, L. S., & Goldman, R. L. (1988). Training in experiential therapy. Journal of Consulting and Clinical Psychology, 56, 696-702.

Greenwood, C. R., Terry, B., Arreaga-Mayer, C., & Finney, R. (1992). The classwide peer tutoring program: Implementation factors moderating student's achievement. Journal of Applied Behavior Analysis, 25, 101-116.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. School Psychology Review, 18, 37-50.

Gresham, F. M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. School Psychology Review, 20, 23-36.

Grigg, D. N., Friesen, J. D., Weir, W., & Bate, C. (1989). Supported feedback therapy manual. Unpublished manuscript.

Gurman, A. S. (1991). Back to the future, ahead to the past: Is marital therapy going in circles? Journal of Family Psychology, 4, 402-406.

Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. Journal of Consulting and Clinical Psychology, 61, 434-440.

Hill, C. E., O'Grady, K. E., & Elkin I. (1992). Applying the collaborative study psychotherapy rating scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. Journal of Consulting and Clinical Psychology, 60, 73-79.

Hill, C. E., O'Grady, K. E., & Price, P. (1988). A method for investigating sources of rater bias. Journal of Counseling Psychology, 35, 346-350.

Hogg, J. A., & Deffenbacher, J. L. (1988). A comparison of cognitive and interpersonal-process group therapies in the treatment of depression among college students. Journal of Counseling Psychology, 35, 304-310.

Hollon, S. D., Waskow, I. E., Evans, M., & Lowery, H. A. (1984, May). Systems for rating therapies for depression. Paper presented at the annual convention of the American Psychiatric Association, Los Angeles, CA.

Jacobson, N. S. (1991a). Behavioral versus insight-oriented marital therapy: Labels can be misleading. Journal of Consulting and Clinical Psychology, 59, 142-145.

Jacobson, N. S. (1991b). To be or not to be behavioral when working with couples: What does it mean? Journal of Family Psychology, 4, 436-445.

Jacobson, N. S. (1991c). Toward enhancing the efficacy of marital therapy and marital therapy research. Journal of Family Psychology, 4, 373-393.

Jacobson, N. S., & Addis, M. E. (1993). Research on couples and couple therapy: What do we know? Where are we going? Journal of Consulting and Clinical Psychology, 61, 85-93.

Jacobson, N. S., Dobson, K., Fruzzetti, A. E., Schmaling, K. B., & Salusky, S. (1991). Marital therapy as a treatment for depression. Journal of Consulting and Clinical Psychology, 59, 547-557.

Johnson, S. M., & Greenberg, L. S. (1985). Differential effects of experiential and problem-solving interventions in resolving marital conflict. Journal of Consulting and Clinical Psychology, 53, 175-184.

Johnson, S. M. & Greenberg, L. S. (1991). There are more things in heaven and earth than are dreamed of in BMT: A response to Jacobson. Journal of Family Psychology, 4, 407-415.

Karasu, T. B. (1986). The psychotherapies: Benefits and limitations. American Journal of Psychotherapy, 40, 324-342.

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. Journal of Consulting and Clinical Psychology, 54, 95-105.

Kazdin, A. E. (1987). Treatment of antisocial behavior in children: Current status and future directions. Psychological Bulletin, 102, 187-203.

Kazdin, A. E. (1990). Psychotherapy for children and adolescents. Annual Review of Psychology, 41, 21-54.

Kazdin, A. E. (1993). Evaluation in clinical practise: Clinically sensitive and systematic methods of treatment delivery. Behavior Therapy, 24, 11-45.

Kendall, P. C., & Lipman, A. J. (1991). Psychological and pharmacological therapy: Methods and modes for comparative outcome research. Journal of Consulting and Clinical Psychology, 59, 78-87.

Koenigsberg, H. W., Kernberg, O. F., Appelbaum, A. H., & Smith T. (1993). A method for analyzing therapist interventions in the psychotherapy of borderline patients. Journal of Psychotherapy Practice and Research, 2, 119-134.

Lambert, M. J. (1989). The individual therapist's contribution to psychotherapy process and outcome. Clinical Psychology Review, 9, 469-485.

Lambert, M. J., & Ogles, B. M. (1988). Treatment manuals: Problems and promise. Journal of Integrative and Eclectic Psychotherapy, 7, 187-204.

LeLaurin, K. & Wolery, M. (1992). Research standards in early intervention: Defining, describing, and measuring the independent variable. Journal of Early Intervention, 16, 275-287.

Luborsky, L., & DeRubeis, R. J. (1984). The use of psychotherapy treatment manuals: A small revolution in psychotherapy research style. Clinical Psychology Review, 4, 5-14.

Luborsky, L., McLellan, A. T., Woody, G. E., O'Brien, C. P., & Auerbach, A. (1985). Therapist success and its determinants. Archives of General Psychiatry, 42, 602-611.

Luborsky, L., Woody, G. E., McLellan, A. T., O'Brien, C. P., & Rosenzweig, J. (1982). Can independent judges recognize different psychotherapies? An experience with manual-guided therapies. Journal of Consulting and Clinical Psychology, 50, 49-62.

Marmar, C. R., Gaston, L., Gallagher, D., & Thompson, L. W. (1989). Alliance and outcome in late-life depression. Journal of Nervous and Mental Disease, 177, 464-472.

Markman, H. J. (1991). Backwards into the future of couples therapy and couples therapy research: A comment on Jacobson. Journal of Family Psychology, 4, 416-425.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. Clinical Psychology Review, 11, 247-266.

Moras, K., & Hill, C. E. (1991). Rater selection for psychotherapy process research: An evaluation of the state of the art. Psychotherapy Research, 1, 113-123.

Newman, F. L., Kopta, S. M., McGovern, M. P., Howard, K. I., & McNeilly, C. L. (1988). Evaluating trainees relative to their supervisors during the psychology internship. Journal of Consulting and Clinical Psychology, 56, 659-665.

Newman, J. L., & Scott, T. B. (1988). The construct problem in measuring counseling performance. Counselor Education and Supervision, 28, 71-79.

O'Malley, S. S., Foley, S. H., Rounsaville, B. J., Watkins, J. T., Sotsky, S. M., Imber, S. D., & Elkin, I. (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression. Journal of Consulting and Clinical Psychology, 56, 496-501.

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. Journal of Applied Behavior Analysis, 15, 477-492.

Peterson, P. L. (1977). Interactive effects of student anxiety, achievement orientation, and teacher behavior on student achievement and attitude. Journal of Educational Psychology, 69, 779-792.

Phillips, R. D. (1987). A primer for conduction child psychotherapy outcome research. Psychotherapy, 24, 178-185.

Piercy, F. P., Laird, R. A., & Mohammed, Z. (1983). A family therapist rating scale. Journal of Marital and Family Therapy, 9, 49-59.

Pinsof, W. M. (1989). A conceptual framework and methodological criteria for family therapy process research. Journal of Consulting and Clinical Psychology, 57, 53-59.

Quay, H. C. (1979). The three faces of evaluation: What can be expected to work. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), Evaluation studies review annual: Vol. 4 (pp. 95-109). Beverly Hills, CA.: Sage Publications

Rounsaville, B. J., Chevron, E. S., Prusoff, B. A., Elkin, I., Imber, S., Sotsky, S., & Watkins, J. (1987). The relation between specific and general dimensions of the psychotherapy process in interpersonal psychotherapy of depression. Journal of Consulting and Clinical Psychology, 55, 379-384.

Rounsaville, B. J., Chevron, E. S., Weissman, M. M., Prusoff, B. A., & Frank, E. (1986). Training therapists to perform interpersonal psychotherapy in clinical trials. Comprehensive Psychiatry, 27, 364-371.

Rounsaville, B. J., O'Malley, S., Foley, S., & Weissman, M. M. (1988) Role of manual-guided training in the conduct and efficiency of interpersonal psychotherapy for depression. Journal of Consulting and Clinical Psychology, 56, 681-688.

Schaffer, N. D. (1982). Multidimensional measures of therapist behavior as predictors of outcome. Psychological Bulletin, 92, 667-681.

Sechrest, L. (1982). Program evaluation: The independent and dependent variables. The Counseling Psychologist, 10 (4), 73-74.

Sechrest, L., West, S. G., Phillips, M. A., Redner, R., & Yeaton, W. (1979). Some neglected problems in evaluation research: strength and integrity of treatments. In L. Sechrest, S. G. West, M. A. Phillips, R. Redner, & W. Yeaton (Eds.), Evaluation studies review annual: Vol. 4 (pp. 15-35). Beverly Hills, CA.: Sage.

Shapiro, D., & Shapiro, D. A. (1987). Change processes in psychotherapy. British Journal of Addiction, 82, 431-444.

Shapiro, D. A., & Firth, J. (1987). Prescriptive v. exploratory psychotherapy: Outcomes of the Sheffield psychotherapy project. British Journal of Psychiatry, 151, 790-799.

Shapiro, D. A. & Shapiro, D. (1983). Comparative therapy outcome research: Methodological implications of meta-analysis. Journal of Consulting and Clinical Psychology, 51, 42-53.

Shapiro, E. S. (1987). Intervention research methodology in school psychology. School Psychology Review, 16, 290-305.

Shaw, B. F., & Dobson, K. S. (1988). Competency judgements in the training and evaluation of psychotherapists. Journal of Consulting and Clinical Psychology, 56, 666-672.

Silove, D., Parker, G., & Manicavasagar, V. (1990). Perceptions of general and specific therapist behaviors. The Journal of Nervous and Mental Disease, 178, 292-299.

Snyder, D. K., & Wills, R. M. (1991). Facilitating change in marital therapy and research. Journal of Family Psychology, 4, 426-435.

Snyder, D. K., Wills, R. M., & Grady-Fletcher, A. (1991a). Long-term effectiveness of behavioral versus insight-oriented marital therapy: A 4-year follow-up study. Journal of Consulting and Clinical Psychology, 59, 138-141.

Snyder, D. K., Wills, R. M., & Grady-Fletcher, A. (1991b). Risks and challenges of long-term psychotherapy outcome research: Reply to Jacobson. Journal of Consulting and Clinical Psychology, 59, 146-149.

Startup, M., & Shapiro, D. A. (1993a). Dimensions of cognitive therapy for depression: A confirmatory analysis of session ratings. Cognitive Therapy and Research, 17, 139-151.

Startup, M., & Shapiro, D. A. (1993b). Therapist treatment fidelity in prescriptive vs. exploratory psychotherapy. British Journal of Clinical Psychology, 32, 443-456.

Stiles, W. B., & Shapiro, D. A. (1989). Abuse of the drug metaphor in psychotherapy process-outcome research. Clinical Psychology Review, 9, 521-543.

Stiles, W. B., Shapiro, D. A., & Elliott, R. (1986). Are all psychotherapies equivalent? American Psychologist, 41, 165-180.

Stiles, W. B., Shapiro, D. A., & Firth-Cozens, J. A. (1989). Therapist differences in the use of verbal response mode forms and intents. Psychotherapy, 26, 314-322.

Strupp, H. H. (1986). Research, practise and public policy (How to avoid dead ends). American Psychologist, 41, 120-130.

Strupp, H. H., Butler, S. F. & Rosser, C. L. (1988). Training in psychodynamic therapy. Journal of Consulting and Clinical Psychology, 56, 689-695.

Svartberg, O. M. (1989a). Manualization and competence monitoring of short-term anxiety-provoking psychotherapy. Psychotherapy, 26, 564-571.

Svartberg, O. M. (1989b). Monitoring the competence of dynamic psychotherapists: A practical demonstration. Acta Psychiatrica Scandinavica, 80, 292-296.

Szapocznic, J., Kurtines, W., Santisteban, D. A., & Rio, A. T. (1990). Interplay of advances between theory, research, and application in treatment interventions aimed at behavior problem children and adolescents. Journal of Consulting and Clinical Psychology, 58, 696-703.

Szapocznic, J., Rio, A., Murray, E., Cohen, R., Scopetta, M., Rivas-Vazquez, A., Hervis, O., Posada, V., & Kurtines, W. (1989). Structural family versus psychodynamic child therapy for problematic Hispanic boys. Journal of Consulting and Clinical Psychology, 57, 571-578.

Thomas, E. J. (1989). Advances in developmental research. Social Service Review, 63, 586-597.

Thompson, L. W., Gallagher, D., & Breckenridge, J. (1987). Comparative effectiveness of psychotherapies for depressed elders. Journal of Consulting and Clinical Psychology, 55, 385-390.

Vallis, T. M., Shaw, B. F., & Dobson, K. S. (1986). The cognitive therapy scale: Psychometric properties. Journal of Consulting and Clinical Psychology, 54, 381-385.

Vermilyea, B. B., Barlow, D. H., & O'Brien, G. T. (1984). The importance of assessing treatment integrity: An example in the anxiety disorders. Journal of Behavioral Assessment, 6, 1-11.

Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. Journal of Consulting and Clinical Psychology, 61, 620-630.

Weiss, D. S., Marmar, C. R., & Horowitz, M. J. (1988). Do the ways in which psychotherapy process ratings are made make a difference? The effects of mode of presentation, segment, and rating format on interrater reliability. Psychotherapy, 25, 44-50.

Wills, R. M., Faitler, S. L., & Snyder, D. K. (1987). Distinctiveness of behavioral versus insight oriented marital therapy: An empirical analysis. Journal of Consulting and Clinical Psychology, 55, 685-690.

Woody, G. E., McLellan, A. T., Luborsky, L., & O'Brien, C. P. (1986). Psychotherapy for substance abuse. Psychiatric Clinics of North America, 9, 347-362.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. Journal of Consulting and Clinical Psychology, 49, 156-167.

Zweben, A., & Pearlman, S. (1983). Evaluating the effectiveness of conjoint treatment of alcohol-complicated marriages: Clinical and methodological issues. Journal of Marital and Family Therapy, 9, 61-72.

APPENDIX A

The Alcohol Recovery Project Adherence Rating Form

THE ADHERENCE RATING SCALE – RATING FORM

Tape #_____ Date Rated_____ Rater_____

Tape ID#_____ Session # _____ Date Coded _____

	EXTENSIVENESS 0 = not at all 1 = little 2 = some 3 = considerably 4 = extensively	EMULATION 0 = not at all 1 = slightly 2 = some 3 = close 4 = very close
1. SUPPORTIVE ENCOURAGEMENT:	0 // 1 2 3 4	0 // 1 2 3 4
2. CONCRETE LANGUAGE:	0 // 1 2 3 4	0 // 1 2 3 4
3. DEVELOPMENTAL FOCUS:	0 // 1 2 3 4	0 // 1 2 3 4
4. ADVANCED EMPATHY:	0 // 1 2 3 4	0 // 1 2 3 4
5. ACCEPTING AND VALIDATING CLIENT'S EXPERIENCE:	0 // 1 2 3 4	0 // 1 2 3 4
6. SPONTANEITY:	0 // 1 2 3 4	0 // 1 2 3 4
7. SEMI-STRUCTURED SESSION FORMAT:	0 // 1 2 3 4	0 // 1 2 3 4
8. TRACKING LEARNING:	0 // 1 2 3 4	0 // 1 2 3 4
9. SYSTEMIC/RELATIONAL RATIONALE:	0 // 1 2 3 4	0 // 1 2 3 4
10. TIMING:	0 // 1 2 3 4	0 // 1 2 3 4
11. WARMTH AND CARING:	0 // 1 2 3 4	0 // 1 2 3 4
12. CHANGES DESIRED IN RELATIONSHIPS:	0 // 1 2 3 4	0 // 1 2 3 4
13. TASK ORIENTED:	0 // 1 2 3 4	0 // 1 2 3 4
14. PRIMARY EMPHASIS ON RELATIONSHIP BETWEEN ALCOHOL & CLIENT SYSTEM:	0 // 1 2 3 4	0 // 1 2 3 4
15. RECOVERY DIRECTED BY CLIENT:	0 // 1 2 3 4	0 // 1 2 3 4
16. GENERATING RELATIONAL NOVELTY:	0 // 1 2 3 4	0 // 1 2 3 4
17. ASSESSING GENERAL FUNCTIONING:	0 // 1 2 3 4	0 // 1 2 3 4
18. WHOLISTIC APPRECIATION:	0 // 1 2 3 4	0 // 1 2 3 4
19. CUMULATIVE LEARNING FORMAT:	0 // 1 2 3 4	0 // 1 2 3 4
20. APPROPRIATE EXPRESSION OF THERAPIST'S FEELINGS:	0 // 1 2 3 4	0 // 1 2 3 4
21. APPROPRIATE INTENSIFICATION OF EXPERIENCE:	0 // 1 2 3 4	0 // 1 2 3 4

	EXTENSIVENESS 0 = not at all 1 = little 2 = some 3 = considerably 4 = extensively	EMULATION 0 = not at all 1 = slightly 2 = some 3 = close 4 = very close
22. COGNITIVE ORIENTATION:	0 // 1 2 3 4	0 // 1 2 3 4
23. RECENT PAST ORIENTATION:	0 // 1 2 3 4	0 // 1 2 3 4
24. CLIENTS DEFINE THEIR OWN EXPERIENCE:	0 // 1 2 3 4	0 // 1 2 3 4
25. CLIENT CAPABILITY AND RESOURCES:	0 // 1 2 3 4	0 // 1 2 3 4
26. COLLABORATION OR MUTUALITY:	0 // 1 2 3 4	0 // 1 2 3 4
27. LOW TO MODERATE INTENSITY:	0 // 1 2 3 4	0 // 1 2 3 4
28. RELATING INTERPERSONAL CHANGE TO THERAPY:	0 // 1 2 3 4	0 // 1 2 3 4
29. FOCUS ON THE THERAPIST AND CLIENT RELATIONSHIP:	0 // 1 2 3 4	0 // 1 2 3 4
30. EMPHASIS ON AWARENESS:	0 // 1 2 3 4	0 // 1 2 3 4
31. THERAPEUTIC RAPPORT:	0 // 1 2 3 4	0 // 1 2 3 4
32. ENCOURAGES INDEPENDENCE:	0 // 1 2 3 4	0 // 1 2 3 4
33. RESPECTFUL:	0 // 1 2 3 4	0 // 1 2 3 4
34. THERAPIST ENSURES CLIENT COMPLETES WSD:	0 // 1 2 3 4	0 // 1 2 3 4
35. THERAPIST GATHERS QUANTITATIVE INFORMATION:	0 // 1 2 3 4	0 // 1 2 3 4
36. SYMBOLIC ORIENTATION:	0 // 1 2 3 4	0 // 1 2 3 4
37. THERAPIST INVOLVEMENT:	0 // 1 2 3 4	0 // 1 2 3 4
38. ACCENTS FEEDBACK FROM CHARTS:	0 // 1 2 3 4	0 // 1 2 3 4
39. HERE AND NOW EXPERIENTIAL FOCUS:	0 // 1 2 3 4	0 // 1 2 3 4
40. SETTING AND FOLLOWING AGENDA:	0 // 1 2 3 4	0 // 1 2 3 4
41. PRIMARY EMPATHY:	0 // 1 2 3 4	0 // 1 2 3 4
42. CLIENTS DERIVE OWN MEANING FROM CHARTS:	0 // 1 2 3 4	0 // 1 2 3 4
43. EXPRESSIVE/CEREMONIAL ACTIVITIES:	0 // 1 2 3 4	0 // 1 2 3 4
44. ECOSYSTEMIC FOCUS:	0 // 1 2 3 4	0 // 1 2 3 4
45. DISCUSSING WAYS OF CHANGING RELATIONSHIPS:	0 // 1 2 3 4	0 // 1 2 3 4

APPENDIX B

The Alcohol Recovery Project Adherence Rating Scale: Rater's Manual

THE ALCOHOL RECOVERY PROJECT

ADHERENCE RATING SCALE

RATER'S MANUAL

THE ALCOHOL RECOVERY PROJECT ADHERENCE RATING SCALE

RATER'S MANUAL

I. Introduction

This Rater's Manual is intended to accompany the Alcohol Recovery Project's Adherence Rating Scale (TARS). In the manual, we attempt to more fully explain the basis for rating items more thoroughly than was possible in the scale itself. The Manual contains information on every item in the TARS. It is essential that the rater be familiar with the material in the Rater's Manual before he/she makes ratings on the TARS.

The Manual begins with General Comments and Instructions to Raters which are important in rating the TARS. The remainder of the Manual is organized according to item number. Each item contains (when applicable):

1) The exact wording and expanded format of the item as it appears in the scale.
2) A restatement of or elaboration on the item's purpose.
3) Definitions of terms used in the item.
4) General guidelines for rating the item.
5) Examples of therapist behavior which should and should not be considered in rating the item.
6) Specific rules for rating the item.
7) Important distinctions to be made between items.

the ideal of each item in the training period, and frequent careful referral to the Rating Manual are vital.

EXAMPLES

Because the distinction between extensiveness and emulation are key concepts in this rating scale, the following examples are included to help illustrate the differences.

a. High Extensiveness, Low Emulation.
In this example the therapist is being rated on Item #4, *Advanced Empathy.*

C: "I just can't stand his threats any more. Every time he's been out drinking with his buddies, he comes home threatening to leave me and the kids."

T: "You're feeling angry and scared that he's setting himself up to leave you. I've got a hunch someone else has abandoned you before this."

C: "Well, yes, my dad left my mom and me when I was turning twelve. He'd been drinking too."

T: "And you begin to question your own worth, when the men you love decide to leave you."

C: "I can't help thinking it must be my fault."

T: "At the same time you feel furious that he might consider leaving."

C: "Uh huh."

T: "In fact, I'm wondering if you think you might be pushing him away as a way to assert your independence. Is it possible you really want to be free of him?"

C: Uh, well..."

T: "I'm sensing your hesitation, and have a hunch that you aren't sure whether to trust me either, and that you might see me as just another unreliable man."

This example shows the therapist using advanced empathy very frequently, but not matching the description in the manual in terms of timing or sensitivity to the client.

b. Low Extensiveness, High Emulation.
In this example, the therapist is being rated on Item #20, *Appropriate Expression of Therapist's Feelings.*

C: "I did it! I got the job!"

T: "Wow! Congratulations! You sound incredibly excited."

C: "I am excited. Man, all this work we've done has really paid off. I wouldn't even have applied without our talks."

T: "Hey, I'm really pleased I could help. And I want to tell you how excited I feel about what you have accomplished."

If this were the only expression of the therapist's feelings in the session segment viewed, then this item would rate "1" or "little" on extensiveness, but would still rate "3" or "close" on emulation.

II. General Comments

1. Rating Therapist Behaviors: The TARS is designed to rate therapist behavior. In rating the scale items it is important to distinguish the therapist behavior (as much as possible) from the client behavior in response to the therapist. This is not possible when rating items such as Item #31, *Rapport*, in which client behavior must be taken account. For the vast majority of items, however, the rater should attempt to rate the therapist behavior, not the client response to that behavior.

   In rating therapist behavior, the rater should consider what the therapist attempted to do, not whether those attempts met with success or failure. For example, in rating Item #1, *Supportive Encouragement*, the rater must determine how supportive the therapist was of the client's gains during therapy or how much the therapist reassured the client that gains would be forthcoming. This item should receive a high rating if the therapist tried very hard to be supportive and/or reassuring, whether or not the client could be encouraged or reassured.

2. Rating Therapist Facilitation: One difficulty that arises in attempting to rate therapist behavior (which is the goal of the TARS) is that sometimes the client initiates a behavior which is measured in a TARS item. Similarly, in other cases the client may actually engage in a behavior being measured in a TARS item with limited therapist involvement. An item should not be necessarily receive a lower rating in either of these instances. In these cases, ratings should reflect the degree to which the therapist facilitates the behavior being measured. In the TARS, facilitation refers to more than a passive acceptance on the part of the therapist of the client's behavior. The therapist must actively encourage, prompt, or work with the client.

   For example, Item #42, *Clients Derive Own Meaning From Charts*, would receive a rating of greater than "0" if: sometime during the session the client discusses the meaning developed from considering her/his charts which record drinking habits, relationships, feelings, etc., AND the therapist assists or actively encourages the client to do so.

3. Prerequisite Knowledge to Rate the TARS: Raters are not required to have extensive knowledge of the behaviors being measured in the TARS in order to rate the items in this scale. The TARS was specifically designed so that raters with little or no previous exposure to the therapeutic modalities represented in the scale could reliably and validly rate therapist behaviors which occur in those modalities. The Rater's Manual has been designed to provide the rater with the specific background she/he needs in order to rate the TARS items.

   However, when using the TARS, the rater must be careful and conscientious in listening to and rating therapy sessions. Because rating the TARS is a complex task, it requires that the rater be thoughtful and exercise good judgement.

4. Rating Extensiveness and Emulation: Each item on the TARS is designed to measure first the extent to which the therapist engages in the behaviors being measured, rather than the quality with which those behaviors are performed. Extensiveness refers mainly to quantity, frequency, consistency, thoroughness or comprehensive-ness. Although extensiveness is not totally independent of the quality of therapist behavior, the rater should not consider the quality of the therapist behavior per se when rating the extensiveness scales on TARS.

   Secondly, each item on the scale also measures emulation. The rater should judge how close to the ideal as indicated in the manual the therapist performs the behavior. This is not a measure of how effective the behavior was, nor is it directly a measure of quality, but is intended to be descriptive of the therapist's behavior. Emulation often requires more personal judgement on the part of the rater, so a clear understanding of

c. High Extensiveness, High Emulation.
In this example the therapist is being rated on Item #2, *Concrete Language.*

T: "You said you felt `freaked out' when you were alone in the house. Could you tell me more about that?"

C: "I knew I could drink and no one was there to stop me. I was scared of going out of control."

T: "So when you became aware of being alone, you felt scared, and thought about drinking. When you say `out of control', do you mean you imagined drinking enough to black out again?"

C: "No, I just thought I'd be drunk when the family came back, then we'd argue, and I'd feel bad, like I'd screwed up again."

T: "Sounds like a whole series of images came into your mind. First you imagined getting drunk, then imagined arguing with your family, then you imagined telling yourself you'd screwed up. What really happened?"

C: "I paced and kept imagining drinking and the arguments it would cause, till finally they came home."

T: "So you used the tension created by your nervousness to begin pacing, and you used your imagination to consider your options until the family returned."

If the therapist's use of concrete language to explore and express the client's ideas or feelings continued consistently through the whole session, and the therapist limited the use of metaphors, analogies, or symbolism in responding to the client, the therapist would receive a "4" or "extensively" on extensiveness and a "4" or "very close" on emulation.

d. Low Extensiveness, Low Emulation.
In this example, the therapist is being rated on Item #23, Recent Past Orientation.

C: "I feel quite lonely at work these days. Since Bob moved, there's no one at work who shows the interest in me that he did."

T: "You really miss that mentoring or fathering Bob gave you."

C: "Hmm, fathering... Yes, Bob was kind of like the dad I never had. And now he's gone too."

T: "It sounds like you feel abandonded by Bob."

C: "Yeah, just the way I feel about my dad."

T: "You feel abandonded by your dad?"

C: "My dad just wasn't there for me to talk to when I was growing up. He was always off on trips. When he did come home, he was too tired to talk."

T: "I notice when you talk about your childhood with your dad that your shoulders seem to tighten up. Could you exaggerate that tension in your shoulders and tell me again about your dad's absence?"

In this example, the therapist spends little or no time focusing on the recent past, instead talking about the more distant past or the client's feelings in the immediate present. This example would recieve a rating of "0" or "not at all" on extensiveness and emulation for Item #23.

However, this item would receive a very high rating on extensiveness ("4" or "extensively") and emulation ("4" or "very close") on Item #39, *Here and Now Experiential Focus*.

5. Frequency, Intensity and Appropriateness: Most of the TARS items require the rater to rate how "extensively" (or "thoroughly") the therapist behavior occurred, as well as how close to the ideal the behavior was (or how well it emulated the model). In order to determine the "extensiveness" or "degree of emulation" of a therapist's behavior, the rater must consider BOTH the frequency with which that behavior occurred during the session and the intensity with which that behavior was engaged in when it did occur. For certain items, the appropriateness of the behavior also must be considered. Judgements of appropriateness should be based on the descriptions of items in the rating manual.

Items vary with regard to how relevant each of the aspects, frequency and intensity, are in determining how that item should be rated. For example, with regard to Item #40, *Setting and Following Agenda*, the therapist would not be expected to take up a large part of the session setting an agenda. Thus the amount of time setting an agenda (i.e., frequency) is not as relevant to rating this item as is the completeness with which the agenda is developed (i.e., intensity). (The manual entry for Item #40 describes what the rater should consider in rating this Item.)

For Item #7, *Semi-Structured Session Format*, the amount of time spent in the session using the session structure described is more relevant. It is important to consider the structure of the whole session in rating this item.

Appropriateness particularly refers to the items #10, *Timing*, #20, *Appropriate Expression of Therapist's Feelings* and #21, *Appropriate Intensification of Experience*. However, appropriateness is an important element in many other items as well and should be considered in rating. For example, with item #43 *Expressive/Ceremonial Activities*, the example in the manual should be preceded by the therapist setting a context for the activity, using empathy, and providing support, etc. A ceremony could only be done once in a session and still could receive a high rating. For example a ceremonial toast, celebrating a success for the client may only take a few minutes, but this could still be a very intense experience for the client, and would very appropriately be done only once. Repeating the ceremony, or adding other similar ones would actually reduce the impact of the intervention. This example would receive a high rating on extensiveness if the ceremony held clear significance for the client. It would also receive a high rating on emulation if the therapist prepared the client for the activity, imbued the ceremony with significance and was able to capture a strong sense of the client's accomplishment in the toast.

In rating an item, there are no fixed rules for determining the weights assigned to the concept's frequency and intensity or appropriateness. The relative weighting of these concepts depends not only on which item is being rated, but also on which specific techniques the therapist uses to accomplish the strategy or goal stated in the item. For example, with respect to Item #3, *Developmental Focus*, the therapist might help the client consider the problem in terms of several possible developmental issues:

1) The therapist could work with goal setting, or teleological concerns.
2) The therapist could frame the problem in terms of the life cycle, or the stages of development in a family, or relationship, etc.

3) A developmental perspective includes framing all therapeutic experiences, even lapses, as positive for learning and growth, contributing to the therapeutic process (i.e. visit to an old pattern with new eyes).
4) The therapist may help the client prepare for difficult moments, times of crisis, or explore losses and gains involved in behavior change.
5) Life themes, or underlying assumptions about life, may be explored from a developmental perspective.
6) Finally, the therapist may help the client review past difficulties as adaptive responses to interpersonal relationships.
7) A developmental focus emphasizes the stages a client goes through both in the short term and the long term.

Not every possible approach would be expected to be used in any one session. Thus, the amount of time spent is more important for thoroughly executing some strategies than it is for others.

There are no fixed rules for determining the equivalence of doing something intensively for a short period of time versus doing something not very intensively for a long period of time. Because the rules for combining frequency and intensity would be very complex and might not always lead to valid ratings, we have left it up to the rater to appropriately weight these concepts when rating TARS items.

6. Avoiding Haloed Ratings: The TARS was designed for the purpose of describing the therapist's behavior in the session. In order to use the TARS correctly, it is essential that the rater rates what she/he hears, NOT what she/he thinks ought to have occurred.

The rater must be sure to apply the same standards for rating an item regardless of:

(1) what type of therapy the rater thinks she/he is rating;

(2) what other behaviors the therapist is engaged in during the session;

(3) what ratings were given to other TARS items;

(4) how skilled the rater believes the therapist to be in a particular modality;

(5) how much the rater likes the therapist;

(6) whether the rater thinks the behavior being rated is a good thing to do or a bad thing to do.

Example of rater halo resulting from rater's judgment of therapy modality (#1 above):

The rater assumes that the item being rated is meant to measure an aspect of Modality A. This item might be rated higher than it should be as a result of the rater also assuming that the therapist was practicing Modality A. Conversely, this item might be rated lower than it should be as a result of the rater assuming that the therapist was not practicing Modality A.

Example of rater halo resulting from a consideration of other behaviors the therapist engaged in during the session (#2 above):

In deciding what rating to assign an item, the rater might erroneously base her/his ratings on behaviors which are similar to or which are likely to covary with the behaviors which are supposed to be considered in rating the item. (General Comment #10, *Making Distinctions*, discusses this further).

Example of rater halo resulting from ratings given to other items (#3 above):

In deciding what rating to assign to an item, the rater might erroneously base her/his rating on ratings given to other TARS items. This is likely to occur when the rater believes that the rating given to another item affects the rating given to the item currently being rated. For example, the rater might assign a high rating to the item *Empathy* because she/he rated the therapist highly on the item *Warmth and Caring*. Each of the TARS items should be rated independently.

Example of rater halo resulting from rater's judgement of the therapist's level of skill (#4 above):

The rater assumes that the therapist is practicing Modality A. Furthermore the rater assumes that the item being rated is meant to measure an aspect of Modality A. Based on these assumptions, the item might be rated lower than it should be if the rater judges that the therapist is not skilled in practicing Modality A; and higher than it should be if the rater judges the therapist to be skilled in practicing Modality A.

Example of rater halo resulting from how much the rater likes the therapist (#5 above):

In deciding what rating to assign an item such as *Warmth and Caring*, the rater might rate this item higher than it should be because she/he has a positive affective reaction to the therapist.

Example of rater halo resulting from rater's judgement of whether the behavior is a good or bad thing to do (#6 above):

The rater might assign a lower rating to an item than is warranted because she/he thinks the therapist is a good therapist and the behavior measured is undesirable. Similarly, the rater might assign a higher rating than is warranted because the rater believes the therapist is a good therapist and the behavior being measured is desirable.

7. Rating Conjunctive Relationships: Instances of AND and OR which are particularly important to note have been capitalized. When two aspects of a behavior specified in an item are joined by "AND", both must be present in order for the item to be rated highly. When two aspects are joined by "OR", the item can be rated highly if either aspect is present.

8. Use of Guidelines: The descriptions and definitions of items in this Manual are intended to be guidelines for use in rating The Adherence Rating Scale (TARS). In some cases, there are specific rules which the rater should use in assigning a particular rating to an item. These rules are referenced in the scale as "//" and are clearly noted in the Rater's Manual as "NOTES." In most cases, however this manual contains guidelines. We expect the rater to exercise her/his judgement in applying these guidelines as well as in rating situations for which the guidelines do not apply.

9. Use of Examples: For many of the items in this Manual, we have given examples of therapeutic exchanges which provide guidelines for how to rate the therapist behavior. Examples have been given when, in our experience with training raters, they have proven to be helpful. The examples in this manual are nevertheless only guidelines for how to rate an item. Often, in fact, examples state only that they should result in a rating of greater than "0" on an item. This is because the examples are only of brief interchanges that might occur in the midst of a session when rating an item. Thus, as an aid to rating items, the examples are a better guide to the kinds of behavior and the intensity with which they should occur, than they are to the frequency with which the behaviors should occur.

Examples in the manual can occur in three different forms:

(1) a list of relevant aspects of the behavior which should be considered in rating an item;

(2) a synopsis of a therapy exchange which should (or should not) result in a rating of greater than "0" on an item; or

(3) dialogue between the therapist and client which should (or should not) result in a rating of greater than "0".

When dialogue is given in an example, it is italicized and the letter "T" is used to indicate what the therapist said, and the "C" is used to indicate what the client said. All names which appear in these examples are fictitious as are most of the situations which are depicted.

In the Manual, reference is often made to a "low rating", a "medium rating", or a "high rating" in discussions regarding how examples should be rated. Because the rater must consider the entire session and not just a discrete incident or period of time (as is represented in an example) in deciding the exact rating given to an item, these suggested ratings should not be considered to be fixed. In general, however, a "low" rating corresponds to a "1"; a "medium" rating, a "2"; a "high" rating, a "3"; and a "very high" rating, a "4". The Manual explicitly states when the rater should assign a rating of "0". A "low rating" does not refer to a "0".

10. Making Distinctions: Because the TARS items vary in terms of breadth of coverage, the same therapist behaviors which are appropriately rated in one item may also rated in another item. For example, if the therapist worked with a couple, asking them to describe and act out metaphors that represent themselves and their partners, this should be considered in rating Item #9, *Systemic/Relational Rationale*, Item #36, *Symbolic Orientation*, as well as Item #39, *Here and Now Experiential Focus*.

Conversely, the rater is often required to make fine distinctions between therapist behaviors which are similar yet should be rated distinctly. For example, in Item #11, *Changes Desired in Relationships*, the rater rates the extent to which the therapist explored changes the client would like to occur in her/his relationship. In Item #45, *Discussing Ways of Changing Relationships*, the rater rates the extent to which the therapist helped the client consider ways to bring about change in her/his relationship. In Item #16, *Generating Relational Novelty*, the rater rates the extent to which the therapist helps bring about changes by generating new experiences in the relationship. These three items measure therapist behaviors which are similar and which may covary, but yet are distinct. The rater should be careful to rate them distinctly (i.e., in rating each item, the rater should consider the extent to which the behavior specified in that item occurred and should not consider other similar behaviors).

When possible, similar TARS items have been compared to one another to help the rater make these distinctions. The rater should bear in mind the subtle differences between some items, and not use the same exact behavior to substantiate ratings given to different items unless it is appropriate to do so.

The Rater's Manual also contains an "Important Distinctions" section within the Manual entry for many of the TARS items. This section contains information regarding how a "target" item is similar to and/or different from other "comparison" items in the TARS. The "comparison" items contain a cross-reference to refer the rater to a discussion of how that item is similar to or different from the "target" item.

The rater should not infer that the existence of these "Important Distinctions" means that they are the only important similarities or differences that need to be made. Thus, the rater should not rely on "Important Distinctions" to point out all of the important similarities or differences which exist.

11. Distinguishing Between Thoughts, Feelings and Behavior: Several items in the TARS refer to thoughts and/or feelings and/or behavior. In order to rate these items as they are intended, it is important to interpret each of these terms literally; i.e., "feelings" refers to emotions or affective states; "thoughts" refers to cognition; "behavior" refers to gestures, body sensations, or actions. Therefore, if a client says, "I'm feeling really scared about that," she/he is indeed talking about an affective state. If, on the other hand, she/he says, "I feel that I couldn't succeed at that;" she/he is referring to the thought or belief, "I don't believe that I could succeed at that." In this latter example, the client may also be attempting to convey a feeling of sadness or hopelessness. If a client says, "It made me feel sick"; she/he may be referring to a physical sensation, an emotional response, or a cognition. It is sometimes difficult to untangle whether statements refer to feelings, cognitions, or behaviors. Therefore, the rater must carefully distinguish whether thoughts, feelings or behaviors are the object of the therapist behavior.

12. Charts and Weekly Situation Diaries (WSD): Some sessions involve the use of charts to record weekly alcohol use, the client's perceptions of relationships with self, spouse, family, friends, etc. The information for the charts comes from a record kept by the client in a weekly format. This record is called the Weekly Situation Diary. The rater must attempt to rate every item carefully when charts are being used. DO NOT assume that certain items will not be seen or that other items must be there in these sessions. There are a couple of items that refer ONLY to the sessions where charts are used, but this does not necessarily mean that when a chart is used that the particular item will be seen.

13. Defining "Important Other": An important other is a person with whom the client has a relationship that is more involved than a casual acquaintance relationship, (i.e., involves some degree of intimacy or significant contact). This can include friends, family, boss, work associates, etc.

14. Specific Instances Required for Rating: In order to rate an item greater than "0", the rater must hear a specific example of the therapist behavior being rated. The rater should be careful not to rate behavior as having occurred if she/he thinks it probably occurred but cannot think of a specific example.

15. Substantiating Ratings: The starting point for rating each item in the scale is "0". The rater should assign a rating of greater than "0" only if she/he hears examples of the behavior specified in the item. This is particularly difficult to do when rating items such as Item #11, *Warmth and Caring*, or Item #10, *Timing*, in which the rater may be tempted to assign an average rating (e.g., "2" or "some") unless the therapist's behavior was remarkably "warm" or lacking in "warmth.". DO NOT DO THIS. The rater must be able to substantiate the rating she/he assigns to every item.

16. Overlap Between Current Versus Prior Sessions: Often an issue that was discussed in an earlier session is implicitly or explicitly referred to in the session being rated, for example, when the client seems to know what the therapist means when the latter request thoughts (because the distinction has obviously been made between thoughts and feelings in an earlier session). Discussions which took place in an earlier session should not be considered in determining a rating given to the current session.

17. Couples Versus Individual Therapy: Some sessions involve working with the couple, while others are purely individual sessions. When rating the couples sessions, some scale items may need to be adapted slightly to apply to working with two clients. For example, in Item #31, *Therapeutic Rapport*, there may be a very different quality to the rapport with either member of the couple than might occur in an idividual session. In this example, the rater should look for rapport to be developed with both members of the couple. Generally, the rater should rate items in couple's sessions taking the different dynamics of three people interacting into account.

III. INSTRUCTIONS TO RATERS

1. Rate Every Item. This scale is designed so that every item can be rated on a scale from zero to four for every therapy session. In other words, DO NOT LEAVE ANY ITEMS BLANK.

2. Read Items Each Time They Are Rated. We recommend that the rater read each item entirely every time it is rated. Careless errors may result when raters rate an item from reading only the item name on the answer sheet and not the item as defined in the scale. Because of the complexity of the TARS item, it is also essential that the rater be completely familiar with the information in the Manual for each item before rating it. It is important that the rater continually refer to the manual, even after she/he has become familiar with it, in order to prevent subsequent rater drift.

3. Attend To Manual Notes. Breaks (//) which appear in the lines above the scale points in the TARS indicate that the Rater's Manual contain a NOTE for this item which specifies what conditions are necessary in order for the item to be given a rating to the right of the "//".

4. Watch Before Rating. Do not rate any items on the scale until the entire 30 minute session segment has been viewed. Remember to forward the tape ten minutes past the beginning before viewing.

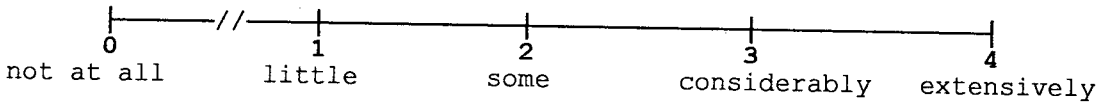
5. Take Notes. We recommend that the rater take notes while viewing to the session. We have found that this procedure enhances the accuracy of ratings both because it helps remind raters of information which is relevant to rating TARS items, and because it helps keep the rater focused on what is occurring in the session. Because the TARS requires the rater to make many fine distinctions, it is essential that the rater attend to the session carefully. The rater should not attempt to do other tasks while listening to tapes of therapy sessions which are to be rated on the TARS.

6. Use Answer Sheet Correctly. We have developed an answer sheet that can be easily read by keypunchers. When using this answer sheet it is important to clearly circle the desired response in pencil and to avoid making any stray marks on the answer sheet. It is crucial that the raters review their answer sheet to ensure that the ┼—necessary identifying information has been filled in, that every item is rated and that no item is assigned more than one response. Be sure to pick a number, not the space between the numbers.

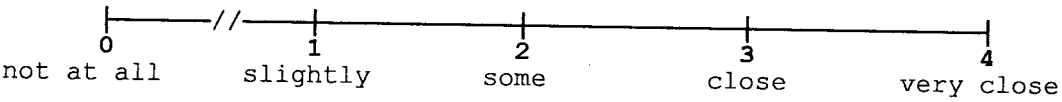
IV. SPECIFIC GUIDELINES FOR RATING ITEMS

1. SUPPORTIVE ENCOURAGEMENT:

1a. Extensiveness: How much was the therapist supportive of the client by acknowledging the client's gains during therapy OR by reassuring the client that gains will be forthcoming?



1b. Emulation: How close to the ideal in terms of the description did the therapist support client gains OR encourage and reassure the client?



The purpose of this item is to measure how supportive the therapist is of the client's efforts in therapy. The therapist might have accomplished this by:

1) Pointing out or acknowledging positive changes the client has made during the time she/he has been in therapy.
2) Reinforcing the client for accomplishments she/he has made thus far (just recently or earlier in therapy).
3) Encouraging the client to continue with the work of therapy because (more) accomplishments will be forthcoming.

The therapist need not have done all of these in order for this item to be rated highly. The gains which the therapist pointed to may range from a client continuing to come to therapy, to a client making major positive life changes.

EXAMPLES

The following are examples of behaviors which should result in a rating of greater than "0" on this item:

T: "I'm remembering when we first started working together. Your depression has eased since then and you seem to be feeling much more independent lately."

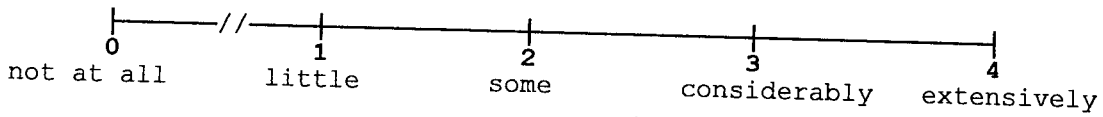
T: "The fact that you are making better eye contact with me is a sign that you are improving."

T: "That's good that you were able to let your boss know that you were angry with her when she did something on your project without your knowledge. That is different from the way you would have reacted before, which would have been to sit and steam over it without letting her know."

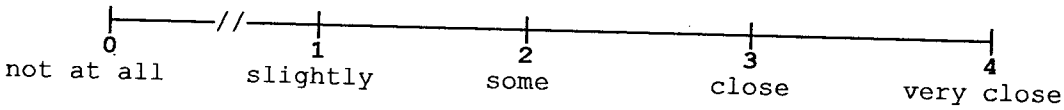
---

2. CONCRETE LANGUAGE:

2a. Extensiveness: How often did the therapist use concrete language?



2b. Emulation: How well did the therapist emulate using concrete language?



This item measures the use of specific, concrete, sensory-based language with the client. To receive a high rating on this item, the therapist should avoid general, abstract, metaphorical, analogic, symbolical or image-laden language. A high rating on this item indicates the strong emphasis on the concrete and observable.

EXAMPLE

This example involves the therapist using concrete language.

T: "You said you felt `freaked out' when you were alone in the house. Could you tell me more about that?"

C: "I knew I could drink and no one was there to stop me. I was scared of going out of control."

T: "So when you became aware of being alone, you felt scared, and thought about drinking. When you say `out of control', do you mean you imagined drinking enough to black out again?"

C: "No, I just thought I'd be drunk when the family came back, then we'd argue, and I'd feel bad, like I'd screwed up again."

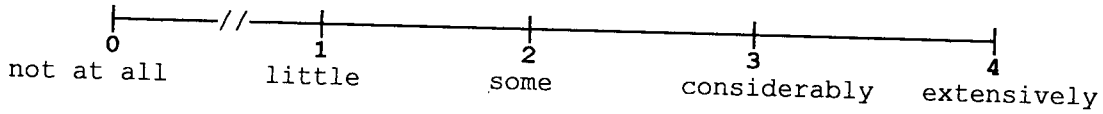
T: "Sounds like a whole series of images came into your mind. First you imagined getting drunk, then imagined arguing with your family, then you imagined telling yourself you'd screwed up. What really happened?"

C: "I paced and kept imagining drinking and the arguments it would cause, till finally they came home."

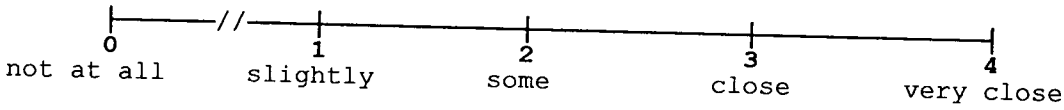
T: "So you used the tension created by your nervousness to begin pacing, and you used your imagination to consider your options until the family returned."

3. DEVELOPMENTAL FOCUS:

3a. Extensiveness: How frequently did the therapist focus on developmental issues?



3b. Emulation: How well did the therapist emulate focusing on developmental issues?



This item measures the focus on developmental issues by the therapist. The therapist might help the client consider the problem in terms of several possible developmental issues:

1) The therapist could work with goal setting, or teleological concerns.
2) The therapist could frame the problem in terms of the life cycle, or the stages of development in a family, or relationship, etc.
3) A developmental perspective includes framing all therapeutic experiences, even lapses, as positive for learning and growth, contributing to the therapeutic process (i.e. visit to an old pattern with new eyes).
4) The therapist may help the client prepare for difficult moments, times of crisis, or explore losses and gains involved in behavior change.
5) Life themes, or underlying assumptions about life, may be explored from a developmental perspective.
6) Finally, the therapist may help the client review past difficulties as adaptive responses to interpersonal relationships.
7) A developmental focus emphasizes the stages a client goes through both in the short term and the long term. On the short term, this includes development in the therapeutic process, or in immediate relationships. In the long term, this would include considering stages of life or patterns in relationship over extended periods of time.

EXAMPLE

The following example would indicate a rating over "0" depending on how it was followed up in the rest of the session.

C: I haven't had a drink for three months until last night watching the hockey game, I just went right back to drinking. I got totally smashed and my wife was furious when I got home at 2:00 am."

T: "You found it easy to start drinking at your buddy's house. I wonder if you have a new perspective on what happened?"

C: "I can see how easy it is to drink there. All the guys are drinking, and we always have. I'd feel out of place not drinking."

T: "So you see yourself as trying to fit in and drinking with your buddies is one way of being part of them?"

C: "That's right. I was thinking of it as being like them. Doing what they do. I don't think of myself as a hard-drinking man anymore."

T: "You view yourself differently now, and you might drink to fit in, but its not really you."

C: "I never thought about that. You're right, it's been since I was accepted that I've been so edgy."

T: "So you are feeling some anxiety about your performance in this new position?"

C: "No, its not my performance I'm worried about, it's whether the people there are going to accept me, since I just got my degree a few months ago. They've all been working in this field for years."

T: "You're worried about being accepted."

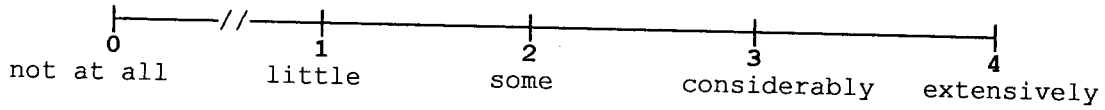
C: "Yeah, and I've been acting out that fear at home. I don't need to do that. I can talk about it with my husband and kids. I know I've got their support."

Important Distinction for Item #24
with Item #4, *Advanced Empathy*

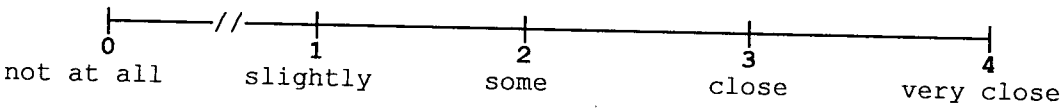
It is possible to confuse *Advanced Empathy* and a focus on underlying feelings and issues with the therapist imposing a definition on the client. The rater should look for the therapist's sensitivity to accepting the client's interpretation rather than forcing the therapist's perspective, even when the therapist is accurately intuitive and empathic. Offering another viewpoint is quite different from imposing one.

4. ADVANCED EMPATHY:

4a. Extensiveness: To what extent did the therapist express deeper meanings and feelings of the client or make intuitive connections?



4b. Emulation: How close to the model of advanced empathy was the therapist, in terms of expressing deeper feelings and meanings or making intuitive connections?



In advanced empathy the therapist responds not only to the overtly expressed feelings of the client but also to thoughts and feelings covertly expressed or implicit. The therapist, in adding deeper meaning and feeling to the statement of the client often expresses feelings that the client was unable to express.

An advanced empathic statement can be stated in a definite manner: "You feel..." or more tentatively using phrases such as: "Could it be that..." or "I'm wondering if..." or "I have a hunch that...". The therapist allows the client to confirm, correct or deny the statement.

A skillful use of advanced empathy is accurate, yet may not be confirmed by the client at the time, and still receive a high rating. There is also an element of timing important to advanced empathy (see Item #04 *Timing*), so the rater should look for a sense that the client was ready for the insight or intuitive connection.

Item #4 and Item #41 *Primary Empathy* are to be considered the therapist's "home base" to which he/she can return to maintain or re-establish contact with the family member while employing another technique. Skillful use of these interventions should be seen frequently to receive a rating greater than "0".

EXAMPLE

The following is an example of one instance of advanced empathy, and should receive a rating greater than "0" if seen throughout the session.

C: "I don't think I need to have my husband at these meetings, it's really just my problem."

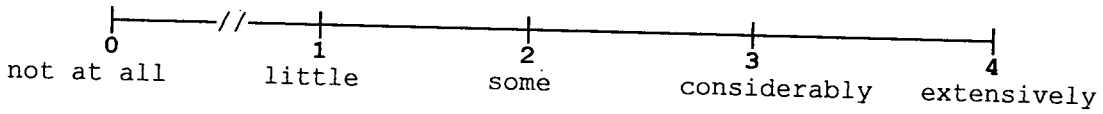
T: "You're feeling hesitant about inviting your husband to these meetings because you think this is not his problem. But, could it be that you're also worried that he might refuse to come if you do invite him, and what that might mean to your marriage?"

C: "I just seem to be irritable all the time these days. The kids can do the littlest things, and I'm snapping at them. And I'm tired. I can't seem to get enough rest."

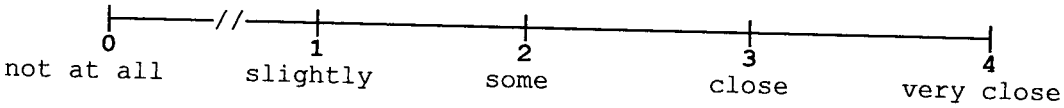
T: "You know, I've got a hunch you're feeling worried about your new job, and your fatigue and irritation at home is about that. Does that fit for you at all?"

5. ACCEPTING AND VALIDATING CLIENT'S EXPERIENCE:

5a. Extensiveness: To what extent did the therapist accept and validate the client's experience?



5b. Emulation: How well did the therapist emulate accepting and validating the client's experience?



This item considers how accepting and validating of the client's experience the therapist can be. A high rating would involve the therapist frequently responding with a sense of understanding and openness to the feelings, thoughts and physical presence of the client, or showing supportive appreciation for their experience. The client's experience is taken as meaningful and important for the client at that time, and the therapist expresses that attitude verbally or non-verbally.

A low rating on this item would occur when the therapist is critical, judgmental or intolerant of a client's experience. However, the therapist may challenge the client, or express a negative personal reaction, without receiving a low rating if there is a strong sense of accepting the client's experience at the same time.

EXAMPLE

This item should receive a rating greater than "0" because the therapist accepted and validated the client's experience (while challenging them).

C: "I just hate her when she keeps quiet, and won't talk to me. She's always withdrawn. I feel like leaving her."

T: "You're feeling furious right now. You really want her to talk to you."

C: "Yes, that's right, but then I feel guilty for being so demanding."

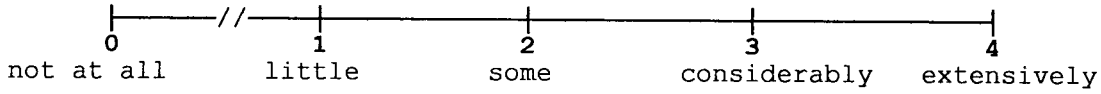
T: "It seems that when you begin to feel angry, you think there is something wrong with you, and that stops you from expressing your anger. I'm wondering what you think might happen if you expressed your anger to her more directly?"

C: "I don't know. I couldn't..."

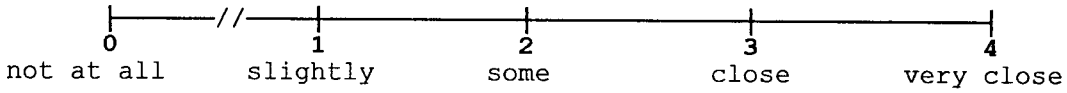
T: "When I suggest you express your anger, I can see how frightened you begin to feel. It's important to me that you feel safe with me, and at the same time, I want to challenge you to explore these ideas with me."

6. SPONTANEITY:

6a. Extensiveness: How frequently did the therapist behave with spontaneity and engender spontaneity in the client's behavior?



6b. Emulation: How well was the therapist able to behave with spontaneity and able to engender spontaneity in the client's behavior.



The rater should see the therapist behave with spontaneity and engender spontaneity in the client's behavior. The therapeutic space should be safe enough to explore and trust intuition. The therapist should take intuitive leaps or initiate creative processes, rather than be rigidly committed to an agenda. Important comments, gestures or experiences are not ignored, but responded to with creativity, using natural impulses or feelings. Relevant accidental or incidental events are included into the therapeutic process. Rigid, repetitive, restrictive therapist patterns are indicative of a low score on spontaneity.

Spontaneity in the client's behavior is actively encouraged. The therapist may draw attention to an offhanded comment, or a certain gesture which might easily have gone unnoticed. This comment or gesture may be explored as reflecting a basic assumption upon which the clients base much of their behavior relating to the problem, or as a creative, spontaneous solution to the problem.

EXAMPLES

The following are examples of spontaneous therapist behavior. Spontaneity is not expected to occur constantly, but is expected to occur regularly. It is a quality of the therapist that reflects a creative way of being in relationship with clients.

> One client gestures with both hands towards his partner when describing his frustration concerning their bitter arguments. The therapist asks the client to exaggerate the gesture and talk a bit about what it means, then explores the partners reaction to the gesture as well.

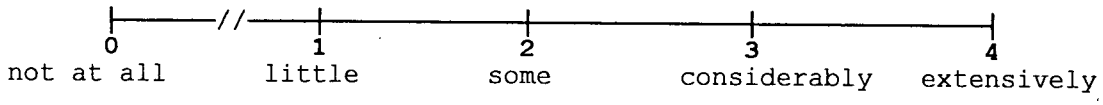
> The client is talking about her difficulty letting go of the expectations of her now dead mother, represented by a set of knitting needles. She says "I think I haven't really buried her yet." The therapist pursues that offhand comment, and suggests they perform a symbolic burial of her mother's expectations together. When the client agrees, the client and therapist arrange to meet at a suitable burial site, chosen by the client, to perform a short burial service.

Important Distinctions for Item #6

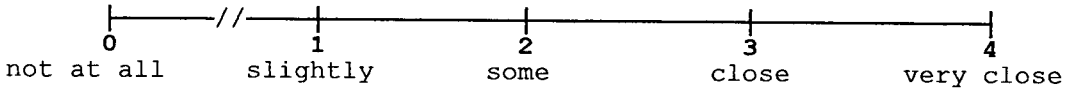
(see Item #37, *Therapist Involvement.*)

7. SEMI-STRUCTURED SESSION FORMAT:

7a. Extensiveness: To what extent did the therapist follow a semi-structured format?



7b. Emulation: How well did the therapist emulate following a semi-structured format?



This item measures the extent to which the therapist uses a particular semi-structured session format. The structure for the session need not rigidly follow the format described below, or the whole sequence may not be included in the tape segments examined. The rater should attempt to judge whether the session seems to be semi-structured, using these basic guidelines, as opposed to unstructured and free-flowing. The general structure for sessions follows the rhythm outlined below:

a. **Greeting and social chat**. Therapists should warmly welcome clients back to each meeting and feedback charts should be displayed on the wall prior to the client's entrance. Time should be allowed for some social chat and to orient the client back to the work space.

b. **Review of significant events**. Therapists initiate session by inquiring about how the past week has gone and by inviting clients to share with them anything they have written in the Weekly Situation Diary (WSD). Therapists are expected to respond empathically and nonjudgmentally to the information which clients present at this time and should feel content that they understand the significant aspects of the weeks to be charted in the meeting before proceeding.

c. **Charting from WSD's**. Therapists and clients should come to some agreement regarding how they will chart the WSD's. Some clients prefer to chart the two weeks on each graph before moving on, while others like to do the charting on all graphs for the first week being reviewed before repeating the process for the second week. Both of these approaches have their respective merits and clients and therapists should collaboratively decide which procedure to adopt. However, it is suggested that the weeks be charted separately and viewed as units unto themselves. This is particularly relevant when a "good" week has preceded a "bad" week. Clients are often drawn to minimize or disregard the important feedback potential contained in the "good" week, opting to focus rather on what went wrong in the "bad" week. In these situations it is critical that therapists affirm and highlight the significance of the "good" week and fully develop the feedback messages that are available when things are improving prior to moving on to the troubles of the next week.
As quantities are charted on each graph, they should be contextualized by previous scores and the way each new score fits into the patterns which evolve over time should be articulated. Clients should be asked to explore the specific meaning of scores and encouraged to be descriptive in their reports (e.g., "Frank, last week you rated your family satisfaction level very low and this week it's gone way up... Can you describe what's happened to bring about this change in your level of satisfaction and are they specific examples of this?").

d. **Feedback generation and chart integration**. Each chart should be given the opportunity to speak directly to clients and charts should also be compared with the associations between them investigated. Therapists should direct client's awareness to aspects of charts or relationships between charts that are curious or important to the recovery process. This procedure represents an extensive review of the past two weeks. Therapists must be aware of feelings or experiences triggered in the review (e.g., "you seem disappointed in yourself at having to report that nothing has changed in your marriage this week. Seeing it in black and white can be pretty hard sometimes..."). In this way, therapists should attend to the clients' ongoing emotional state, validating and naming reactions as they emerge in the session.

e. **Meaning derivation and pattern interpretation**. Therapists should join with clients in sitting back from the charts once the new WSD information has been recorded to consider the unfolding patterns and to exhaust the feedback potential contained in the charts. The entire set of charts should be viewed separately and also in unison so that the various feedback messages can be received. Therapists should ask clients to articulate central messages received from the charts (e.g., "If you look at all the charts, and if they could speak to you in one voice, what would that voice, which is your life pattern, say to you today?"). At this time, clients may be asked to generate alternative meanings or interpretations available in the charts, (e.g., "You've told me that the charts tell you that because you have a dissatisfying marriage that you drink a lot. I wonder if the charts may be saying something different. What else might they be saying?" Therapists should avoid making chart interpretations or stressing explanations that clients do not generate themselves.

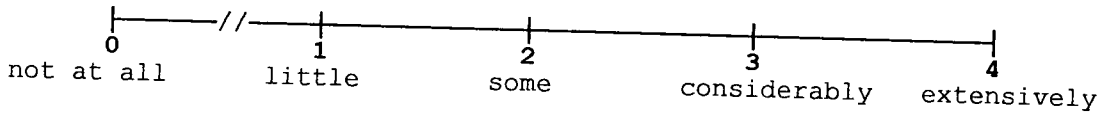
f. **Summarization and closure**. Sessions should conclude with clients and/or therapists summarizing the feedback given and received in the session. In addition, therapists should include some process comments about how clients have been affected by sessions and reiterate any significant comments the clients have made regarding intentions for the upcoming weeks. Finally, sessions should be formally brought to a conclusion with therapists thanking clients for their hard work and establishing dates for the next therapy meetings.

Important Distinctions for Item #7
with Item #13 *Task Oriented*

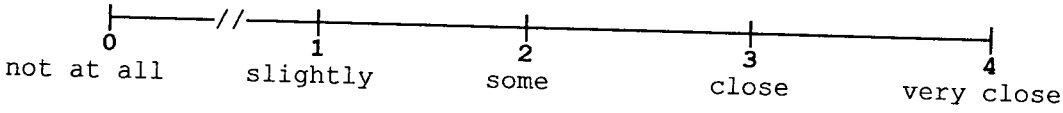
Semi-Structured Session Format considers the session format as a whole. The format described includes specific tasks which would also be rated on Item #13. However, it is possible to have a task orientation, without a semi-structured format, or a semi-structured format with a minimal task orientation. For Item #13, pay attention only to the focus on tasks, not the session format.

8. TRACKING LEARNING:

8a. Extensiveness: To what extent did the therapist keep track of, acknowledge, or highlight important events, learnings or achievements?



8b. Emulation: How well did the therapist emulate keeping track of, acknowledging, or highlighting important events, learnings or achievements?



In order for this item to have a rating greater than "0", the therapist must have pointed out an accomplishment or acknowledged a special event for the client.

The therapist may keep track very well, receiving a high emulation score and yet only receive a low extensiveness score if tracking is done infrequently.

EXAMPLE

The following are examples of a therapist tracking learning:

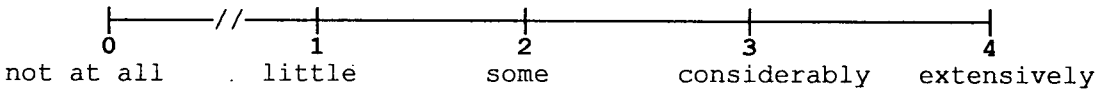
T: "So, you and your ex-husband have been able to have several discussions without the bickering you described earlier. Congratulations."

T: "You were able to find your own way to ask for what you wanted with your new friends. Last week you were afraid you'd just go along with the crowd. I think you've learned you are an important part of the group too."

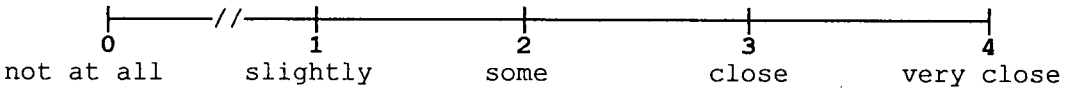
T: "You've expressed your feelings very well indeed. I'm feeling touched by your experience. You've been paying attention to what's going on inside you for the last few weeks, and it seems to help you express your feelings."

9. SYSTEMIC/RELATIONAL RATIONALE:

9a. Extensiveness: To what extent did the therapist use a systemic or relational rationale in working with the client?



9b. Emulation: How close to the model was the therapist in using a systemic or relational rationale in working with the client?



This item is intended to measure the use of a systemic or relational rationale in the therapeutic process. Interpretations may be made of dynamics in relationships, explorations of interpersonal expectations made, and similarities between disputed relationships to other relationships considered. Relationships and the systemic quality of relationships are considered central. Reality is viewed as contextual, and the context of behaviors within the relationship/system is the basis for interventions.

The rater should look for evidence of therapeutic work that deals directly or indirectly with relationships in any level of the system, or the experience of the client as part of a system. This work will tend to emphasize the whole experience of the client, and not be limited to a cognitive understanding of the various levels of the client system. Interventions based on a systemic/relational rationale may increase conscious awareness of various levels of the client's ecosystem, but also may simply be framed with those assumptions implicit.

EXAMPLES

The following are examples of the therapist using a systemic/relational rationale in working with the internal world of the client.

> The client mentions judging herself harshly, and the therapist suggests the client put the judge in a chair. They explore the judge as a part of the client, then the therapist asks if the judge reminds her of anyone earlier in her life. The client says her mother was also quite critical of her.

> The therapist is working with a couple, who have come because the husband has been uninterested in sex for several months. The therapist encourages the clients to enact metaphors that describe their relationship around this issue. The implicit assumption that the couple's relationship is central to the problem is not directly stated.

> The client's relationship with alcohol is explored using a symbol to represent the alcohol. The problem is described by the therapist as a troubled relationship, rather than a characteristic of the client. Later, the impact of the client's relationship with alcohol on his other relationships is considered. This may include his relationship with himself, his partner, family, friends, career, etc.

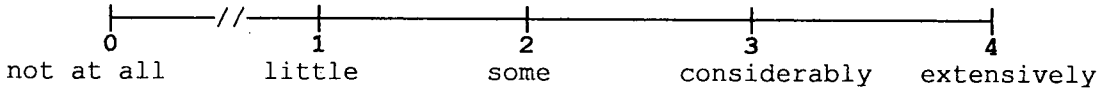
A couple begins to get excited about some changes they want to make about their time together. The therapist suggests using restraint in making the changes in the relationship, to go slow. The therapist points out the balance they have established up to this point, and how radical changes might have effects they hadn't anticipated.

Important Distinctions for Item #9
with Item #44, *Ecosystemic Focus*.

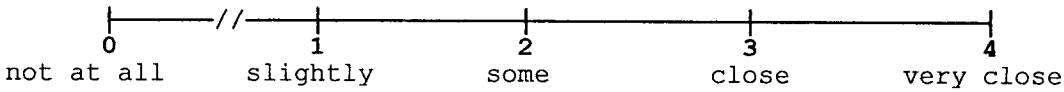
Ecosystemic Focus is concerned with increasing an awareness of the various levels of the client's ecosystem. Parts of self, immediate family and extended family, as well as friends and community are focused on. With Item #9, *Systemic/Relational Rationale*, therapeutic interventions are based on the importance of systemic and relational issues.

10. TIMING:

10a. Extensiveness: To what extent was the therapist timing non-coercive and sensitive to client readiness to engage in the therapeutic process?



10b. Emulation: How well did the therapist emulate the description of non-coercive and sensitive timing of the therapeutic process?



This item measures the therapist's sensitivity to the client's readiness for a therapeutic intervention. A high rating on this item will occur when the client seems ready for the interventions chosen throughout the session. The therapist may clarify, provide encouragement, or describe possible benefits of a process in order to help a client prepare for an activity, but the therapist should not coerce the client. This item also refers to the ability of the therapist to respond to the changing needs of the client while engaged in a process. Timing then, refers to both beginning and ending a process, in a way that is sensitive to the client. This also refers to the therapist's ability to go as deeply as the client is ready for, and not prematurely finishing a process.

The rater is cautioned that this item is not intended to judge the intensity of the intervention, but simply the therapist's sensitivity to the client's readiness. One therapist may at times sensitively employ an intervention that deeply intensifies the client's experience, where another therapist never allows the intensity to go beyond mild or moderate levels. The rater is cautioned not to consider either level of intensity better for this item, but to focus on the timing involved.

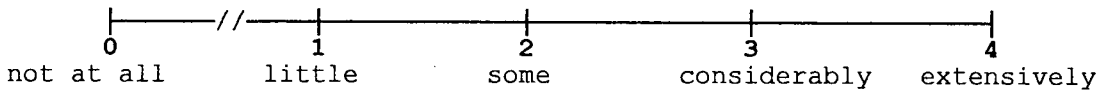
Important distinctions for Item #10

with #21, *Appropriate Intensification of Experience*

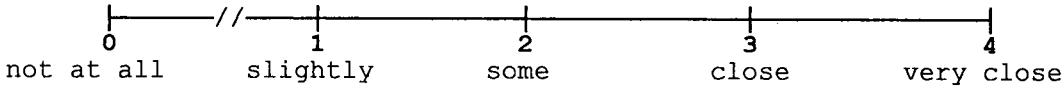
and with #27, *Low to Moderate Intensity*

11. WARMTH AND CARING:

11a. Extensiveness: How much did the therapist convey warmth and caring?



11b. Emulation: How well did the therapist emulate warmth and caring according to the description?



Warmth, which has been equated with unconditional positive regard, has been defined by Rogers and Truax (1967) as "the therapist communicating to his client a deep and genuine caring for him as a person with human potentialities, a caring uncontaminated by evaluations of his thoughts, feelings, or behaviors." This communication need not be explicit but the therapist's caring should be made evident by her/his behavior. Raters who have developed their own operational definition of *Warmth and Caring* are encouraged to use it only if it is consistent with how *Warmth and Caring* is defined above.

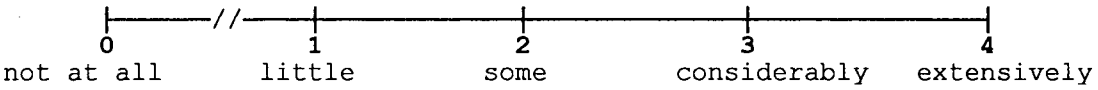
The rater must be careful not to assume that the therapist conveys warmth merely because she/he is a therapist. The rater must also remember that she/he is not rating how warm and caring the therapist is in general, but rather how much warmth and caring the therapist conveyed in the session being rated.

This item is one in which it is particularly important for the rater to avoid giving ratings of greater than "0" as default values. The rater must be able to justify from the therapist behavior any rating on this item which is greater than "0".

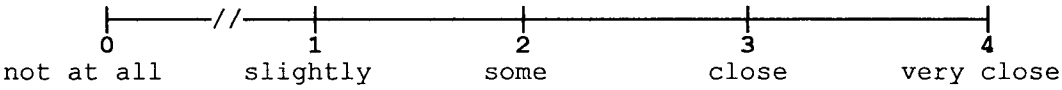
Important distinctions for Item #11
with Item #31 *Rapport*
The rater should rate Item #31 and Item #11 independently. It is possible for the therapist to be warm and caring and yet not get along with the client. Conversely, it is possible for the therapist to not demonstrate warmth or caring for the client and yet develop strong rapport.

12. CHANGES DESIRED IN RELATIONSHIPS:

12a. Extensiveness: To what extent did the therapist explore with the client those changes the client would like to see in an important relationship (or in role expectations within a relationship)?



12b. Emulation: How well did the therapist emulate a good exploration with the client of desired changes the client would like to see in an important relationship (or in role expectations within a relationship)?



This item is intended to measure attempts by the therapist to explore with the client what changes the client would like in a relationship.

EXAMPLE

The following example should receive a rating greater than "0" on this item because the therapist explores with the client desired changes in her relationship:

T: *"It sounds to me like you are not especially satisfied with the way things are between you and your sister."*

C: *"I'm tired of her treating me like her little sister. I'm an adult and I think I deserve to be treated like one."*

T: *"How does she treat you like a `little sister'?"*

C: *"She's too protective. She's so worried that I will make a mistake and be hurt by it that she's constantly butting in where she shouldn't be."*

T: *"What would you like to see her do differently than she does now?"*

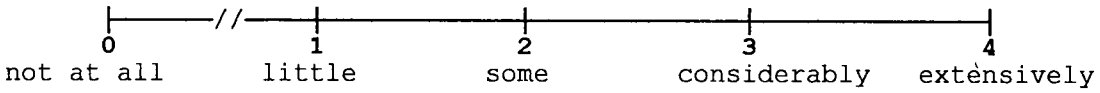
C: *"I'd like her to let me choose for myself who to get involved with and who not to get involved with. She doesn't have to like the people I associate with but she can keep her concerns to herself and let me make mistakes if I will."*

Important Distinctions for Item #12
with Item #45, *Discussing Ways of Changing Relationships.*

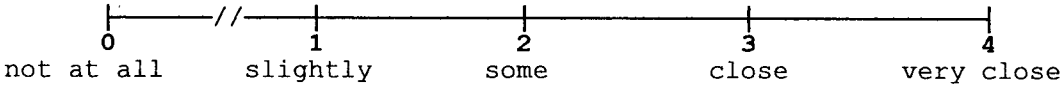
Discussions of how to make changes are rated in Item #45, whereas discussions of what changes are desired are rated in Item #12. If in the process of considering options for interpersonal change, the therapist helped the client decide or clarify what changes the client would like to see occur, then both Item #12 and Item #45 should receive a rating greater than "0".

13. TASK ORIENTED:

13a. Extensiveness: To what extent was the therapist task oriented, giving explicit guidance?



13b. Emulation: How close to the ideal in terms of the description was the therapist at being task oriented, and giving explicit guidance?



This item is intended to measure the focus on task oriented activities by the therapist. The therapist gives explicit guidance in performing these tasks. A high rating on this item would indicate a consistent emphasis by the therapist on completion of tasks. Tasks are highly structured, specific, concrete, and functional activities.

For example, therapists may guide and direct client's attention to filling in and reviewing feedback charts in a structured way, or may encourage the completion of their weekly situation diary (WSD).

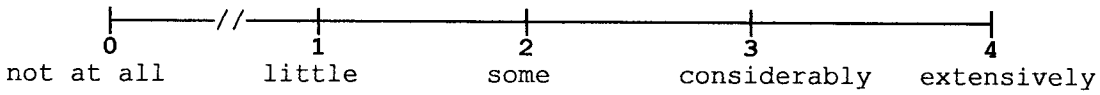
Extended discussion of past or present experiences, insights or goals, or any experiential, expressive or ceremonial activities, even if structured, are not to be considered tasks.

Important Distinctions for Item #13
with Item #7 *Semi-Structured Session Format*

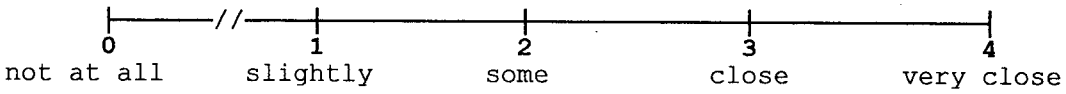
Semi-Structured Session Format considers the session format as a whole. The format described includes specific tasks which would also be rated on Item #13. However, it is possible to have a task orientation, without a semi-structured format, or a semi-structured format with a minimal task orientation. For Item #13, pay attention only to the focus on tasks, not the session format.

14. PRIMARY EMPHASIS ON RELATIONSHIP BETWEEN ALCOHOL & CLIENT SYSTEM:

14a. Extensiveness: How much did the therapist primarily help the client explore the relationship between alcohol and the client system?



14b. Emulation: How well did the therapist emulate primarily helping the client explore the relationship between alcohol and the client system?



This item is intended to measure the therapist's emphasis on the relationship between alcohol and the client system in the session. Ths client system can include parts of self, family, friends, or community. The therapist helps the client focus on how alcohol relates to these various levels of the client system. A very high rating on this item would indicate a very strong focus on this relationship (between alcohol and the system), and minimal exploration of other interrelationships not directly related to alcohol.

EXAMPLE

The following is an example of an exploration that would receive a rating of "0" if typical of the session as a whole.

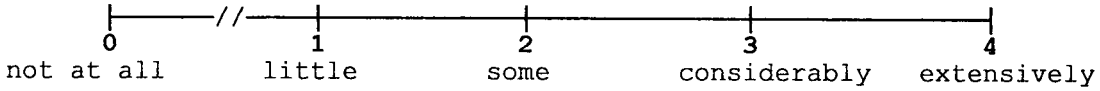
> The therapist notices that the client has mentioned the impact of his mother on his relationship several times. The therapist spends about twenty minutes discussing the client's feelings, goals and possible alternatives regarding dealing with his mother. No mention of alcohol is made during this time.

The following would receive a rating greater than "0", if typical of the session as a whole.

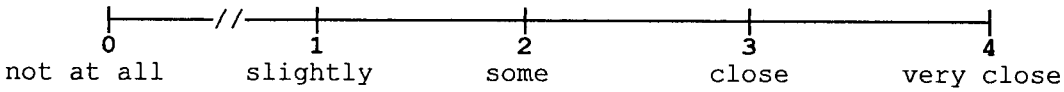
> The therapist notices that the client has mentioned the impact of his mother on his relationship several times. The therapist spends some time exploring how the client feels about his mother, then also looks for information on how alcohol influences the situation. The client notes that his mother suggests his wife makes him drink. He feels defensive and often ends up drinking after talking with his mother, which creates some problems at home.

15. RECOVERY DIRECTED BY CLIENT:

15a. Extensiveness: To what extent was the client's recovery directed by the client, leaving the therapist less involved?



15b. Emulation: How well did the therapist emulate allowing the client's recovery to be directed by the client, leaving the therapist less involved?



This item measures the therapist's ability to allow the client's recovery be directed by the client, leaving the therapist less involved. To receive a high rating on this item, the focus of the sessions should remain on the client, with the therapist's involvement kept to a warm, empathic listener. The therapist is also responsible for maintaining the session structure and promoting insight, but essentially allows the client's recovery to progress at the pace set by the client. The client is treated as the one responsible for the progress and direction of therapy, and the therapist is more in an assisting role.

The therapist should not initiate new areas of discussion, intensify the client's experience, confront or challenge the client to grow in any particular ways. The therapist is not likely to discuss the personal growth she/he has gained through the sessions. Self-disclosure or immediacy may occur, but the overall style of the session is client-focused, with minimal personal involvement by the therapist.

The rater is cautioned not to consider lack of involvement undesirable on the part of the therapist, and the therapist may be deliberately attempting to remain detached. Thus, the rater should not let her/his ratings of the therapist's involvement be influenced by whether the rater considers therapist involvement to be good or bad.

EXAMPLES

The following is an example where the client's recovery was directed by the client, with the therapist relatively uninvolved.

C: "I'm beginning to see that my relationship with drinking is very important. I want to quit at times, but I really feel like I need to drink other times."

T: "Sometimes you feel comforted by drinking."

C: "Yeah. This week in particular I realised that I wanted to drink whenever Barbara and I were arguing, like the bottle would somehow make me feel better. I know that sounds stupid, but..."

T: "You are beginning to see a connection between how you feel in your relationship and your drinking patterns, and you sound like you feel embarrassed to realize it."

C: "I feel like an idiot. Yet I still want to drink so bad at times."

T: "This realization has really made you question your behavior. And the comfort you've found in the bottle is obviously important to you. I wonder what your next step is going to be?"

C: "I've already cut down on my drinking. And the other day I went for a walk down by the beach, and that seemed to calm me down for a while."

T: "So you've been trying out some new ways to comfort yourself, while paying attention to what do feel and do."

The following example illustrates the therapist more involved and more actively directing the pace of recovery, and would receive a very low rating or "0"

C: "I'm beginning to see that my relationship with drinking is very important. I want to quit at times, but I really feel like I need to drink other times."

T: "Sometimes you feel comforted by drinking."

C: "Yeah. This week in particular I realised that I wanted to drink whenever Barbara and I were arguing, like the bottle would somehow make me feel better. I know that sounds stupid, but..."

T: "It sounds like you might have some concern about me judging you for getting comfort through drinking. Would you like some honest feedback?"

C: "I think so."

T: "Actually I'm feeling suprised and delighted you made that connection. It wasn't clear to me yet, though it makes sense. I've often looked for comfort in other ways when I'm feeling stressed in my relationship. Some of those ways, like lack of sleep, or obsessive TV watching, or affairs really caused more problems than they solved. I'm impressed that you could be so clear about the connection so quickly."

C: "Well at least I haven't been sleeping around."

T: "I felt ashamed of my affairs when I did that. Now I realize I need comfort from time to time and try to find less destructive ways to give that to myself. I haven't used affairs for several years now, but its taken some work."

C: "Well, I guess I feel ashamed of my drinking, but I don't know if I can stop, either."

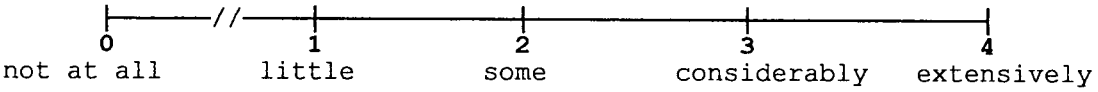
T: "Would you like to take a few minutes to talk about some other ways you can support and comfort yourself that you might feel better about?"

Important Distinctions for Item #15
with Item #29, *Therapist Relationship with Client*
and with Item #37, *Therapist Involvement*.

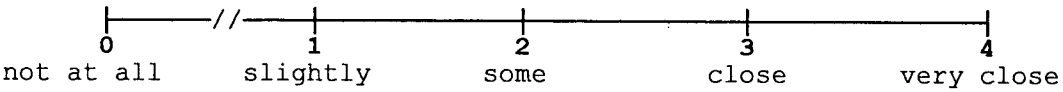
The relationship between the therapist and the client may be central to the therapy and the therapist may be less involved on a personal level, leaving more direction to the client. Item #15 does not limit the importance of the relationship, although it may suggest a very different style of relationship than Item #37.

16. GENERATING RELATIONAL NOVELTY:

16a. Extensiveness: How often did the therapist help to generate relational novelty in the session?



16b. Emulation: How well did the therapist emulate helping to generate relational novelty in the session?



This item measures the therapist's collaborative generation of relational novelty. The overarching focus is on perturbing change in client interaction sequences. Since clients in distress are characterized by highly predictable, rigid patterns of relationship with a narrow range of possible deviation from the "way it always is", they frequently feel confined and trapped within unsatisfactory relationships in which they experience little, if any, direct authority or influence. The therapist strives to provide opportunities for clients to experience "relational novelty" and directly influence these relationships rather than repeatedly engage in seemingly invariant sequences of behavior that conclude with the same unsatisfactory cyclical result.

Particular segments of interaction are focused on and given symbolic significance. The interactional sequences may be between the therapist or between client subsytem members (or parts of self, objects, etc). Framed as important contributing features of clients' relationships, specific aspects of the sequence are slowed in therapeutic time, magnified and intensified in client awareness. The predominant purpose of therapist activities in generating relational novelty is to challenge entrenched sequences of interaction, thereby triggering the emergence of new patterns of client interaction. In so doing, therapists assist clients in altering their sense of constriction, through an immediate experience of expansion and potentiality, thereby empowering them with experiences of influence.

Reframing, positive connotation, reciprocal metaphors, sculpting, roleplaying, etc are all possible activities designed to provide the clients with intense novel experiences with symbolic relational implications. These new experiences challenge clients with new directly experienced information about self and the relationships they are involved with.

Other possible issues that relate to generating relational novelty include developing new relationships, considering options in relationships, or exploring substantive relational themes. Substantive relational themes refer to the underlying assumptions about self in relationship that an individual may enact in many different relationships over time.

EXAMPLE

The following is an example of the therapist generating relational novelty.

T: "So, from what you've described, it sounds like alcohol has been a pretty close friend the last few years."

C: "That's interesting, I never thought of it that way. I certainly find the bottle to be good company, in a way. But, it's been getting in the way of my relationship with my wife."

T: "The bottle's been getting in the way?"

C: "Well, it's more like my desire for the bottle that gets in the way. A lot of times I'd rather spend time at the pub than hang out at home."

T: "Sounds like you are feeling closer to the alcohol than to your wife."

C: "Sometimes, yeah."

T: "Could we use this bottle to be alcohol? Could you show me how you relate to alcohol and your wife? We've already used this chair to represent your wife."

C: "Well, the alcohol is right here in my hand. It protects me from her sometimes. I guess it gets in the way too. Huh, that's right. It gets in the way. But I'm afraid to put it away."

T: "You are feeling anxious about changing your relationship with alcohol. Somehow it feels better to hang on than let go."

C: "When I put the bottle to the side, I feel really exposed. I'm just waiting for her to jump on me."

T: "So moving the bottle makes you very aware of your feelings of not being safe. And it looks like you feel pretty uncomfortable. How does your relationship with the bottle help you out?"

C: "I can count on drinking to blot her out. I can't think clearly enough to worry about her. It's like a sheild."

T: "I wonder if it seems like your wife is a warrior attacking you, that you need a sheild."

C: "I know she's not. She just wants to be closer. But it sure feels like that."

Important Distinctions for Item #16

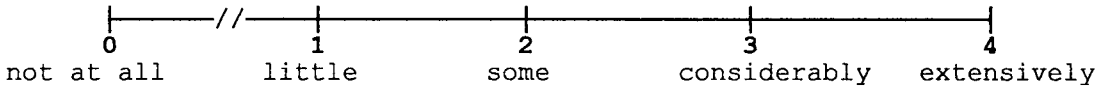
with Item #12, *Changes Desired in Relationships*

and with Item #45, *Discussing Ways of Changing Relationships*

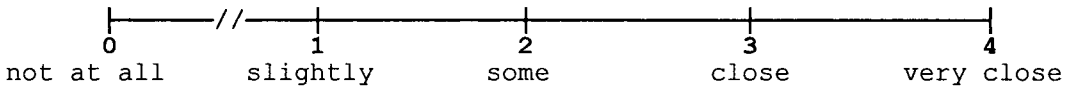
Generating Relational Novelty is focused on experiences in the here and now that involve new interactions. Item #12 and Item #45 are focused on desired changes or possible methods of creating change, and are not here and now experiences of change.

17. ASSESSING GENERAL FUNCTIONING:

17a. Extensiveness: To what extent did the therapist survey the client's general level of functioning in major life spheres (e.g., work, intimate relationships, family life, social life, etc.)?



17b. Emulation: How well did the therapist emulate surveying the client's general level of functioning in major life spheres (e.g., work, intimate relationships, family life, social life, etc.)?



This item is intended to measure the thoroughness with which the therapist assesses the client's functioning in each of the major aspects of the client's life (e.g., intimate relationships, family matters, friendships, other social relationships, vocational pursuits, etc.).

This assessment need not have been done all at one time in the session. It could have been accomplished gradually over the course of the entire session. The main criterion is the extent to which the therapist concluded the session with knowledge of how the client is functioning in each of the major aspects of the client's life. Over the course of a session, it is likely that at least one of these areas of the client's life will have been discussed regardless of whether the therapist attempted to survey the client's major life spheres. In order to rate this item greater than a "0", however, the therapist must have gathered enough information to have gained at least a superficial understanding of the client's functioning in her/his major life spheres.

EXAMPLES

The following example would receive a low rating on this item:

> The therapist spent some time talking with the client about a situation that occurred at work concerning how the client was getting along with his boss. Later, the therapist asked the client how he was getting along with his spouse and his children but did not pursue the client's responses to either of the latter questions.

The following example should receive a medium to high rating on this item:

> The therapist did all of the things in the above example and also asked about additional areas of the client's life such as: the client's pursuit of his previous hobbies, interactions with co-workers, time spent with friends, family, activity in social groups, etc. Although the therapist need not ask about all of these areas, the criterion for assigning a higher rating on this item is that the therapist did considerable assessment of the client's functioning in her/his major life spheres.

The following example should receive a high rating on this item if:

> In the second example above the therapist asked enough follow-up questions or gleaned enough information from her/his initial questions to obtain a thorough assessment of the client's functioning in all or most of the major spheres of the client's life. The therapist is not expected to explore the client's interpersonal relationship to score highly on this item; rather the therapist should conduct a thorough survey of the client's functioning. Although the rater will be unlikely to know what are the major spheres of the client's life and thus will not know whether they were all covered, the rater should still give this item a high rating if (1) the therapist thoroughly assessed the client's functioning in several major life spheres and (2) it is not obvious to the rater that other major life spheres were not explored.

Important Distinctions for Item #17

with Item #44, *Ecosystemic Focus*

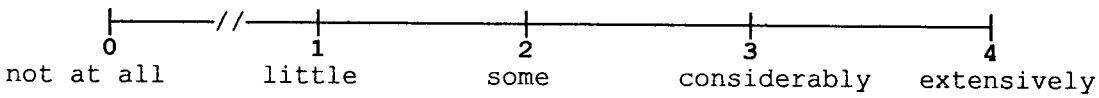
Ecosytemic Focus emphasizes the exploration of patterns of relationship in different levels of the client's life. *Assessing General Functioning* is more of a survey of the client's level of functioning. Item #17 could receive a high rating through gathering a thorough description of functioning level, while Item #44 requires an exploration of the systemic interactions between the client and various levels of their relationships (with self, others, family, friends, colleagues, symbols, etc.)

and with #25, *Client Capability and Resources*

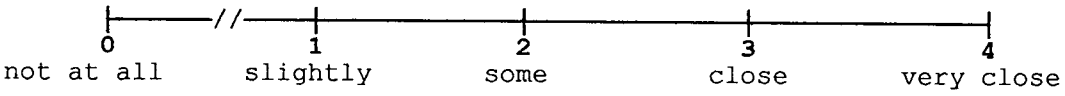
To rate highly on #25 the therapist focuses on the client's resources and capabilities, and may or may not also encourage independence from the therapist. For Item #32 the therapist may comment on the client's abilities or resources, but the intention is to encourage the client to make decisions independently from the therapist. It is quite possible that both items could rate highly in a session. The rater is asked to distinguish those elements that specifically emphasize resources, from those elements that specifically encourage independence, or to note those interactions that seem to do both.

18. WHOLISTIC APPRECIATION:

18a. Extensiveness: How extensively did the therapist use a wholistic appreciation of the client, including the body, thoughts, and emotions as a part of the whole experience?



18b. Emulation: How well did the therapist emulate using a wholistic appreciation of the client's experience, including the body, thoughts, and emotions as a part of the whole experience?



This item measure the therapist's use of a wholistic perspective to explore the client's experience. The whole experience is fedback and shared with the client, including body tension, gestures, voice tone, or other physical responses, as well as considering thoughts and feelings, memories, expectations, or visions of the future.

The client is approached and responded to as a whole. A rating higher than "0" would indicate the therapist had appreciated several aspects of the client's experience, and was not focused primarily on one aspect. Body reactions, thoughts, and emotions are seen as equally important means of communication and validated. Interventions may also act on the whole person, through movement, discussion of cognitions and acceptance of emotions. Memories or visions of the future are responded to as if they were a part of the present time experience of the client.

EXAMPLE

The following example illustrates the therapist working with a wholistic appreciation.

C: "He's so unreliable. It doesn't bother me any more, but I just can't count on him."

T: "You feel deeply disappointed that you can't count on him, and as you say that, I recognize how dissatisfied you are with that. I hear a tightness in your voice when you talk about him right now, and I notice your shoulders are tense too."

C: "I'm tired of him getting to me. I don't want to care any more."

T: "You are feeling disappointed with his behavior, and angry with yourself for trusting him."

C: "Yeah, I'm a sap, and he's a jerk."

T: "I also hear you talking about how you want to feel less caught up in reacting to him. You have a goal to stay calm and relaxed when you see him."

C: "Right! I've been through all the anger and frustration so many times before. I just want to let go of needing him to be different. I don't think he's going to change."

T: "Right now, you know what you want, and you're feeling impatient because you are still upset with him."

C: "I can't stop thinking of all the times I counted on him, and he screwed up."

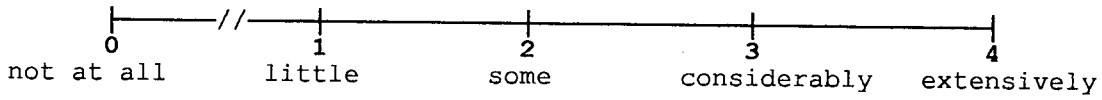
T: "When I see the tears in your eyes, I guess you are feeling the hurt and disappointment of those memories as if they were happening right now."

C: "It hurts."

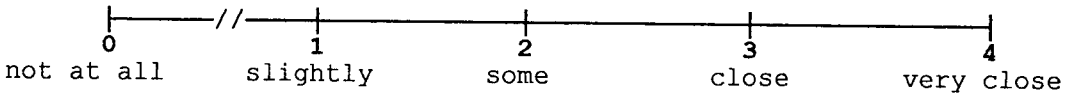
T: "The pain feels like a deep wound."

19. CUMULATIVE LEARNING FORMAT:

19a. Extensiveness: To what extent did the therapist use a cumulative learning format with the client?



19b. Emulation: How well did the therapist emulate using a cumulative learning format with the client?



This item measures the therapist's use of a cumulative learning format with the client. Cumulative learning may be based on information from charts or client's stories, and refers to learning that builds over time with the accumulation of new information. Understanding is expected to increase as patterns become apparent in a cumulative learning format. Cumulative learning is rational, information-based and gradual. Cumulative learning may include information about feelings, body reactions, relationships at any systemic level, dreams, goals, memories, etc (as long as there is not a here and now focus).

Cumulative learning is not insight-oriented, experiential, metaphorical or intuitive. Cumulative learning also does not emphasize expressive, symbolic or developmental interventions.

EXAMPLE

The following is an example of an exchange based on a cumulative learning format.

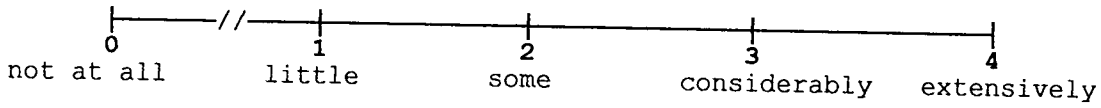
T: "I notice that over the last few weeks, your alcohol consumption has decreased, and there has been some improvement in your relationship with your partner. What do you see in your chart in terms of the effect not drinking has on your relationship?"

C: "I notice that I'm more interested in her when I'm not drinking so heavily. But, I also can see, that when I cut it out completely three weeks ago, that I had some real conflict with her. I'm thinking that finding a balance might be the best thing for me right now."

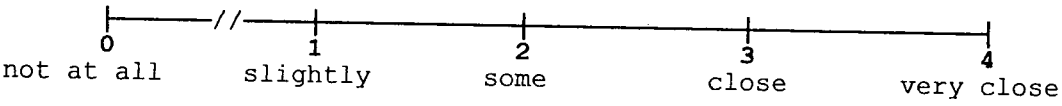
T: "The impact of drinking and not drinking stands out when you can chart the effects on your relationship."

20. APPROPRIATE EXPRESSION OF THERAPIST'S FEELINGS:

20a. Extensiveness: To what extent did the therapist express her/his feelings related to therapy with the client?



20b. Emulation: How well was the therapist able to emulate expressing her/his feelings related to therapy with the client?



This item measures the therapist's expression of personal feelings in the session. The rater should look for both self-disclosure and relationship immediacy in rating this item. Expression of feelings should be related to therapy, and sensitive to the client's needs. This item indicates a level of involvement, and a real relationship between the client and therapist. The rater should look for specific examples of the therapist expressing feelings related to therapy.

Therapist disclosures that seem unrelated to therapy, or are overly extensive, would receive a lower rating.

EXAMPLES

T: "I feel excited to hear about your promotion. I really enjoy your success."

T: "As you've been talking, I've had some reactions to what you were saying. However, I feel nervous about sharing my feelings. I'm afraid you'll get angry if I'm not very careful. I remember you said a couple of weeks ago, that you intimidated your wife sometimes, to keep her quiet. I'm wondering if you have any sense of wanting me to keep quiet."

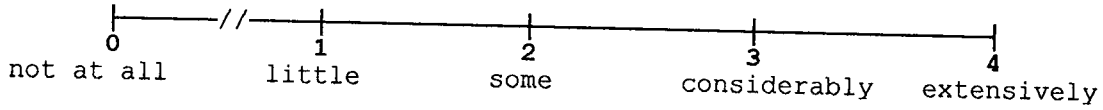
C: "When I look at the chart, it seems clear to me that the times when I'm drinking less, are the times when I get along with the kids better. I never used to care that much about them, but paying attention to them has made me appreciate them more. I really love them."

T: "I'm deeply touched by your caring. I'm also warmed by the progress we've seen in some of your other relationships too, with your partner and at work. It feels like a very special journey we are sharing."

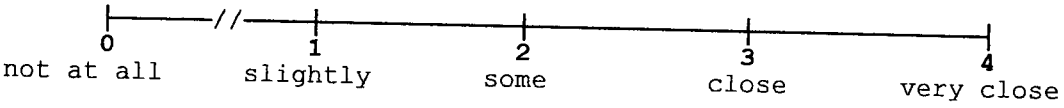
T: "I'm feeling annoyed that you are late again. I'd like to explore what's going on to make you late so consistently. Perhaps there's something I'm doing to support you being late. Is that something you'd be willing to talk about?"

21. APPROPRIATE INTENSIFICATION OF EXPERIENCE:

21a. Extensiveness: To what extent did the therapist actively attempt to intensify the client's experience, to a level acceptable to the client?



21b. Emulation: How well did the therapist emulate intensifying the client's experience, to a level acceptable to the client?



This item refers to the principle that therapeutic change involves a deepening, enhancing and broadening of the client's experience. The focus is on increasing or intensifying the client's awareness of current feelings, perceptions and physical states. The therapeutic experience is designed to let it happen more, more fully and more deeply, with greater depth and breadth and saturation.

It is important to distinguish between enhancing and deepening feelings which often accompany therapeutic experiencing, and deepening the essence or substance of the experience. The therapist focuses on the substance of the experience rather than on the emotional response which accompanies the experience. It may be more important for a client to intensely recognize and feel the power imbalance in a relationship, rather than to simply overtly express anger or cry, for example.

There are numerous methods, both verbal and action oriented, that may be used to enhance and intensify the client's experience. Intensification tends to involve experiential activities that intensify the client's experience on multiple levels. The language of the therapist will tend to be metaphorical, concrete and pictorial in detailing the specifics of the situation. An interaction between significant others, blocked parts of self, loved objects, or other self introjects is likely to be evoked. Talking "about" the situation is not enough. The situational context is created vividly enough that the other is so alive and real that there is a meaningful response, a genuine being here in the interaction. A therapeutic trance-like state may be fostered. Intensity may also be increased by clear and precise representations and descriptions of the scene, feelings, thoughts, etc. Rather than leaving a situation, intensity may be increased by staying with it, in order to understand and fully appreciate it. The therapist may amplify the bodily aspects of the client's experiences, through repetition, exaggeration, or personalizing body reactions (for example, "I am shaking", not "My hand is shaking"). Intensification may also be approached through dramatizing relationships in the past, present or future. Simply identifying the emotions involved through advanced empathy, may also intensify the experience. Finally, metaphors, symbols or symbolic actions may also be ways of intensifying experience.

Timing is an important element of intensification of experience. A high rating on this item would occur if intensification is developed at a level that seems to feel acceptable to the client (though not necessarily comfortable for them). A high rating would not occur if the client seems ready for more intensification, or seems overwhelmed with the intensity.

Important Distinctions for Item #21

with Item #30, *Emphasis on Awareness*

Item #30 focuses on increasing awareness, and #21 focuses on intensification. It is possible to increase awareness in a way that does not intensify the experience. Intensification would be

expected to increase awareness, although not necessarily at the present moment. For example, a burial ceremony may not immediately create deeper awareness, while it may deepen the experience of letting go. It may be only on reflection that conscious awareness may increase.

with Item #10, *Timing*

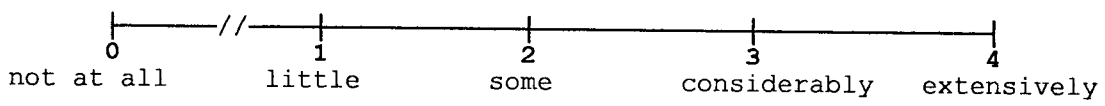
Timing should be used well when intensifying the client's experience. However, when rating timing, intensification is only one possible point that timing is considered. The session as a whole should be considered when rating timing. Likewise, intensification is more than good timing. For each example of intensification, timing is one element of the final rating.

with Item #4, *Advanced Empathy*

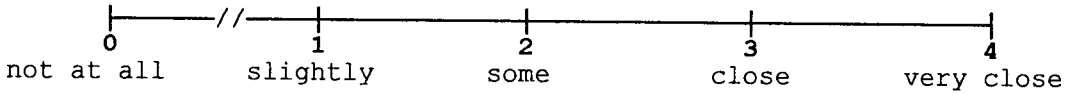
Advanced Empathy is one way to intensify the client's experience, but should not be the only intervention used to receive a high rating on this item.

22. COGNITIVE ORIENTATION:

22a. Extensiveness: To what extent did the therapist use a cognitive orientation, promoting insight for the client?



22b. Emulation: How well did the therapist emulate using a cognitive orientation, promoting insight for the client?



This item examines the use of a cognitive orientation by the therapist. The rater should look for an emphasis on cognitions of the client by the therapist, even though the therapist may also respond wholistically and empathically. The therapist works to promote insight for the client. Understanding or recognition of patterns is developed, rather than therapeutically experiencing them.

EXAMPLE

The following example is one where the therapist uses a cognitive orientation to promote insight in the client.

T: "I notice that there is less of a gap between your level of satisfaction with your family and your ideal charts the last few weeks. What do you think the charts are telling you?"

C: "As we've been getting along, I've changed some of my expectations of my family, too. I was expecting too much and that was causing some of our conflicts."

T: "Does that mean you don't get what you want from your family?"

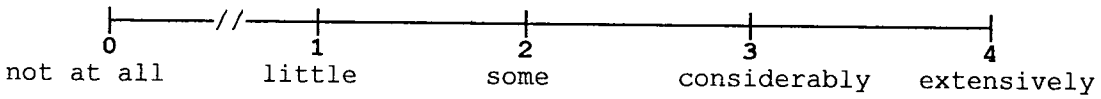
C: "No, I've been getting more of what I want, because I don't scare the kids by yelling at them so much. And I don't get my wife angry with me by complaining as soon as I notice something isn't done my way."

Important Distinctions for Item #22
with Item #30, *Emphasis on Awareness*

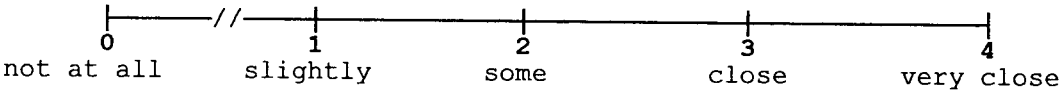
One aspect of awareness for #30 is cognitive, but it also includes awareness of body sensations, gestures, emotions, memories or expectations experienced in the present moment. #22 is not necessarily focused on the present moment, and does only pursues cognitive aspects of the client's experience.

23. RECENT PAST ORIENTATION:

23a. Extensiveness: How often did the therapist use a recent past orientation?



23b. Emulation: How close to the ideal in terms of the description was the therapist at using a recent past orientation?



This item measures the therapist's use of a recent past orientation. The therapist monitors past behavior, and encourages the client to learn from experiences in the past couple of weeks. The therapist may focus on behaviors, emotions, or thoughts that occurred in the recent past. The events of the recent past are described as useful information and is not explored as a present time experience.

The rater is looking for the therapist's overall orientation through the session. A high rating on this item would indicate a pervasive emphasis on the recent past.

EXAMPLES

The following example is one where the therapist uses a recent past orientation.

C: "I'm still angry about what happened at her mother's house last week."

T: "Can you tell me a bit about what happened?"

C: "I had two drinks while I was there, and already her mother was looking down her nose at me."

T: "So you were aware of her disapproval, and I get the impression you had been feeling pretty good about limiting your drinking that day."

C: "That's right. I'll tell you, I just about blew it then. When she looked at me like that, I wanted to drink a whole bottle, just to piss her off."

T: "What stopped you?"

C: "I just thought about some of the things we've talked about here, and I remembered how often I've let other people's attitudes push me into drinking."

T: "You decided not to let her get to you?"

C: "Exactly."

C: "When my kid talked back to me after coming home late, I was seeing red."

T: "You felt furious."

C: "That's right, and that was just the beginning that day. Later on the vacuum cleaner broke down and I can't afford to replace it, and when I did the dishes, I broke my favorite bowl."

T: "Sounds like a pretty tough day. I wonder if you felt overwhelmed with all that going on?"

The following is an example of the therapist not using a recent past orientation. If this type of response were typical of the therapist through the session, a rating of "0" would be appropriate.

C: "I'm still angry about what happened at her mother's house last week."

T: "You're really angry. Even your shoulders are tense."

C: "Yes, whenever I remember her mother looking down her nose at me, I get mad. I only had two drinks the whole evening."

T: "It sounds like you are also feeling hurt by her reaction."

C: "I do. I wish she'd get off my back. It just makes this harder."

T: "You're feeling like she's right on top of you. If this pillow were her mother, where would it be?"

C: "On my back. But only when I have a drink."

T: "Hm, so when we move the bottle closer..."

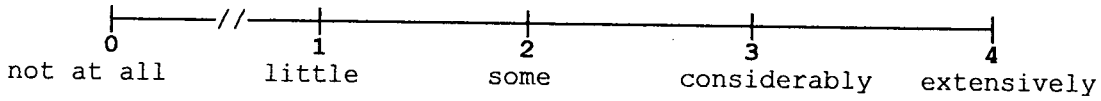
C: "The pillow jumps on me."

Important Distinctions for Item #23
with Item #39, *Here and Now Experiential Focus*

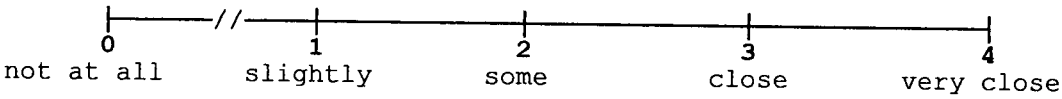
Recent past orientation emphasizes memories of past events, and thoughts or feelings about them. Here and Now Experiential Focus emphasizes the present time experiencing of thoughts, feelings or behavior. It is unlikely that a single session would receive a very high rating on both items. However, it is quite possible to receive a rating greater than "0" on both.

24. CLIENTS DEFINE THEIR OWN EXPERIENCE:

24a. Extensiveness: To what extent did the therapist encourage the clients to define their own experience?



24b. Emulation: How well did the therapist emulate encouraging the client(s) to define their own experience?



This item is intended to measure the way the therapist deals with the meaning a client attributes to their experience. The therapist encourages the client to define their own experience, and uses the client's interpretation rather than imposing their own interpretations. The client's definition is accepted as valid, although the therapist may challenge the meaning a client attributes to an experience, acting as a kind of custodian of therapeutic meaning. Meaning is seen as constructed by the individual, and changeable as new perspectives develop.

When working with couples, this may be seen as supporting the acceptance of alternate perspectives as valid for each individual. Each person's interpretation is considered important and expected to be unique.

EXAMPLE

The following example would rate greater than "0" if it occurred consistently in the session.

C: "When he left me there to go sleep with his new girlfriend, I felt devastated."

T: "You feel disappointed and rejected."

C: "I feel like I'm being abandoned again, just like before."

T: "Your boyfriend didn't call for you and you felt let down and this really hurt."

C: "Yes, I feel lost, empty."

T: "So when your boyfriend violates your relationship, you feel lost and empty and rejected."

C: "It means I'm not good enough."

T: "I can hear that you're really doubting yourself. I'm not sure I agree with the interpretation though. I can see how much you feel like that when he's gone, but I think it might mean more about him than you. Do you have any sense of that?"

C: "Well, I'm sure tempted to just call him a selfish bastard."

T: "Hm, what might that mean about him leaving you?"

C: "It would mean he's too immature to realize what he's losing by leaving me."

T: "That's certainly a different view of things. How do those two views fit together?"

C: "I guess it's not all my fault, and it's not all his fault. We both need to grow up."

The following example would receive a very low rating, because the therapist is imposing a definition on the client.

C: "My girlfriend has been getting very demanding lately. I keep taking time away from her, because I don't want to have to deal with her. She's pissed off with her job right now, and she takes out her anger on me, even though I know she doesn't want to."

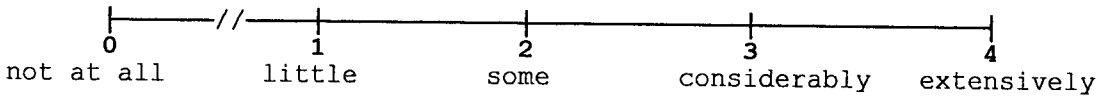
T: "It sounds to me like you are still struggling with some feelings about your mother. Your girlfriend reminds you of her."

Important Distinction for Item #24
with Item #4, *Advanced Empathy*

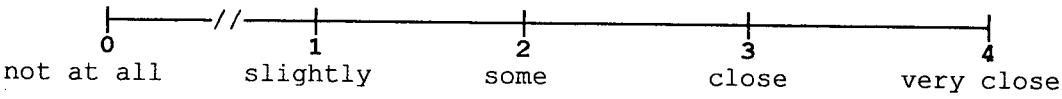
It is possible to confuse *Advanced Empathy* and a focus on underlying feelings and issues with the therapist imposing a definition on the client. The rater should look for the therapist's sensitivity to accepting the client's interpretation rather than forcing the therapist's perspective, even when the therapist is accurately intuitive and empathic. Offering another viewpoint is quite different from imposing one.

25. CLIENT CAPABILITY AND RESOURCES:

25a. Extensiveness: To what extent did the therapist stress the client's own resources and capacity for change?



25b. Emulation: How close to the description of stressing the client's own resources and capacity for change was the therapist?



The therapist should be considered to have stressed the client's own resources and capacity for change if she/he:

1) explicitly focused on the client's abilities to handle the situation, or
2) encouraged independence from the therapist, or
3) helped the client acknowledge their personal resources, or review the various ways they support themselves, or
4) pointed out or reviewed the capacity for change the client has.

In rating this item the rater must consider what transpired over the entire session, and be able to justify from the therapists behavior a rating of greater than "0".

EXAMPLES

T: "Given that you've developed some good skills over the last several weeks, I'm confident that you can make a good decision for yourself. I'll be anxious to hear what you decide."

T: "I'm really glad to hear about the way you handled that situation. That's a nice example of how you didn't need my help to come up with a way of dealing with a difficult situation. You came up with a plan on your own, carried it out, and resolved the problem."

C: "I'm thinking about quitting my job. I just can't handle the pressure."

T: "Quitting your job would be a pretty big move on your part. How can you go about deciding whether that is the best thing for you to do?"

C: "I'm always being asked to make important decisions without enough time to weigh it all out. I get so nervous I dread going to work."

T: "I wonder what it would take to do that job well?"

C: "Actually, I've done pretty well, but I just feel so nervous."

T: "So you seem to be capable of handling the job, but your anxiety is your main concern? Have you ever been in a spot like that before?"

C: "My last job was pretty fast paced as well, but I really enjoyed it."

T: "How did you handle your nervousness in that situation?"

C: "It was tough at times, but I guess I could really count on my friends to talk to when things got hairy. That's the biggest difference now."

T: "I know this is a new city for you, but do you have any friends you could talk to?"

C: "Actually, I do. I just haven't done it. I guess I just didn't really think of it."

T: "You know, I get a strong impression of how much you are able to adapt to situations when you have a little support. Is that right?"

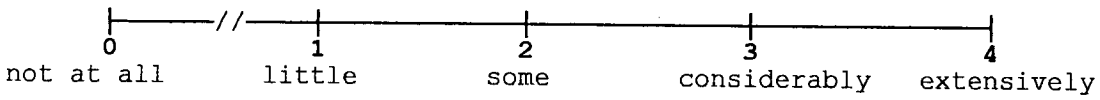
C: "Yes, I guess that's right."

Important Distinctions for Item #25
with #32, *Encourages Independence*

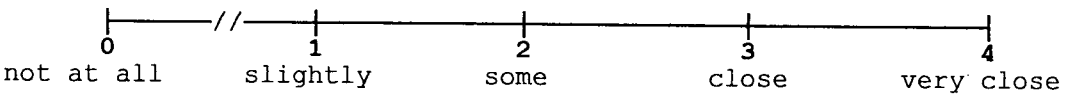
To rate highly on #25 the therapist focuses on the client's resources and capabilities, and may or may not also encourage independence from the therapist. For Item #32 the therapist may comment on the client's abilities or resources, but the intention is to encourage the client to make decisions independently from the therapist. It is quite possible that both items could rate highly in a session. The rater is asked to distinguish those elements that specifically emphasize resources, from those elements that specifically encourage independence, or to note those interactions that seem to do both.

26. COLLABORATION OR MUTUALITY:

26a. Extensiveness: To what extent did the therapist actively attempt to engage the client in working together to explore therapeutic issues?



26b. Emulation: Compared to the descriptions, how well did the therapist actively attempt to engage the client in working together to explore therapeutic issues?



How much of a shared venture was the session? To be rated higher than a "0" on this item, the therapist must have done more than simply ask questions. She/he must have attempted to engage the client in a mutual process of exploration of the issues. There should have been a mutual exchange of information and effort between the therapist and the client. The role of the therapist is to befriend, honor and defend the client as they explore themselves.

This item is one in which it is particularly important for the rater to avoid rating greater than "0" unless she/he heard the therapist attempt to involve the client in a collaborative process of making decisions, exploring issues, etc. Collaboration can be a way of the client and therapist working together in an ongoing co-operative manner, which may not always be made explicit in words. Ratings greater than "0" would indicate some explicit effort on the therapist's part, as well as unspoken collaboration.

EXAMPLE

The following is an example of mutual exchange of information and working together which, if it occurred throughout the session, would warrant a high rating on this item.

C: *"I've had it with my husband."*

T: *"You feel discouraged and upset about your husband."*

C: *"He's so inconsiderate of me, he sits around and lets me make meals, do the cleaning and the laundry and yet I work just like he does!"*

T: *"You don't feel appreciated. In fact, you feel pretty annoyed with him."*

C: *"I've been considering leaving him."*

T: *"Things have become so bad in the relationship that you feel like leaving."*

C: *"Well, I don't really want to leave him, I just wish he would help out more at home."*

T: *"You are really bothered about his not sharing the household chores?"*

C: *"Its a perfect example of him not really participating."*

T: *"This issue seems pretty important to you. I'd like to try something that might help me get a good picture of what the problem feels like for you and might give us some ideas of what a solution might look like. Would you like to spend some time on this?"*

C: *"Sure, I'd like to try and feel better about this."*

T: *"Could you think of a metaphor or a picture that would describe the relationship at home the way things are now?"*

C: *"The picture that comes to mind is that I'm trying to set sail, and he's keeping one foot on the dock."*

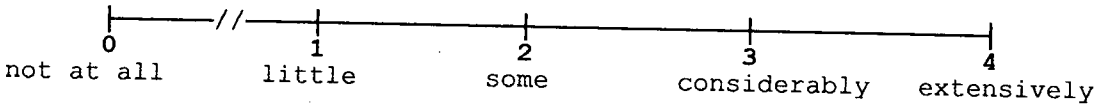
NOTE: Do not rate this item greater than "2" if the therapist attempted to involve the client only in setting an agenda but not at other times during the session.

Important Distinctions for Item #26
with Item #32 *Encourages Independence*

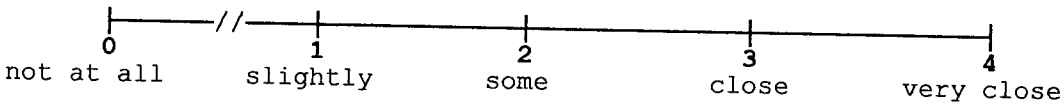
The purpose of Item #26 is to measure how much the therapist worked with the client (i. e., shared in problem solving), whereas the purpose of #32 is to measure how much the therapist encouraged the client to solve problems with little or no help from the therapist.

27. LOW TO MODERATE INTENSITY:

27a. Extensiveness: To what extent did the therapist maintain low to moderate intensity in the session?



27b. Emulation: How well did the therapist emulate maintaining low to moderate intensity in the session?



This item measures the extent to which the therapist was able to maintain a low to moderate intensity level in the session. The focus in this item is on creating a learning environment with little or moderate anxiety. The rater should look for a general intensity level as well as possibly specific examples of the therapist limiting or reducing intensity levels.

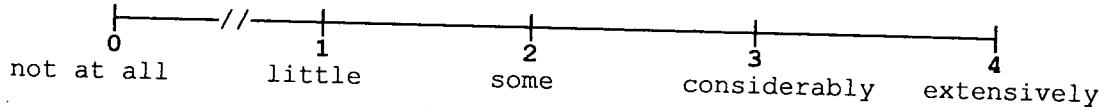
Important Distinctions for Item #27

with Item #21, *Intensification of Experience*

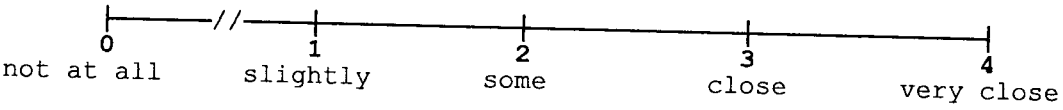
Item #21 measures the intensification of experience within the level acceptable for the client. This may be low or moderate through a specific session, however the emphasis is on the therapists attempts to help intensify the experience. In Item #27, the therapist may attempt to limit or lessen the intensity of the immediate experience in the session, or may simply maintain a low to moderate intensity level with little emphasis on intensification. Perhaps a useful criteria in rating these two items is the impact of the therapist on the intensity. If the therapist's interventions increased the intensity consistently through the session, they would receive a high rating on Item #21. If however, the therapist's interventions did not intensify (maintaining a low to moderate level of intensity) or lessened the intensity level, they would receive a high rating on Item #27.

28. RELATING INTERPERSONAL CHANGE TO THERAPY:

28a. Extensiveness: To what extent did the therapist relate changes in the client's interpersonal relationships (or role expectations within relationships) to the emphasis in therapy on understanding and changing the client's interpersonal functioning (or role expectations within relationships)?



28b. Emulation: How well did the therapist emulate relating changes in the client's interpersonal relationships (or role expectations within relationships) to the emphasis in therapy on understanding and changing the client's interpersonal functioning (or role expectations within relationships)?



In order for this item to be rated greater than "0", the therapist must have made a connection between the interpersonal focus in therapy and changes that have occurred in the client's relationships or in role expectations within the client's relationships.

EXAMPLES

The following is an example of the therapist clearly having related interpersonal change to therapy. This example would warrant a medium to high rating depending on how this discussion continued.

T: "We've spent a lot of time in therapy on improving your communication with your wife and on ways of spending more time with your children. Those relationships seem to be vastly improved since you first came here. It seems to me that doing the things we've discussed has led to this improvement. Is that how you see it?"

The following example should be rated greater than "0" on this item because the therapist helped the client to relate what was discussed in the therapy to her resolution of a specific interpersonal problem.

T: "Last week we talked about a problem you were having with your brother. It seems to me that you came to some decisions about what you wanted to say to him. Did you bring those things up?"

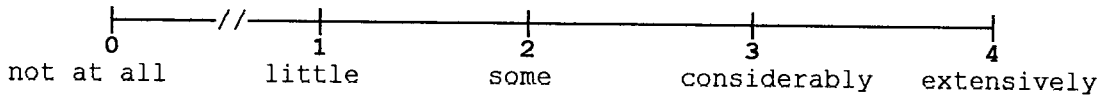
C: "Yes, I did. I think the discussion you and I had helped me to sort things out so when I did talk to him I could tell him all of the things that were bothering me."

T: "So it seems like our discussion had a positive impact on your relationship with your brother."

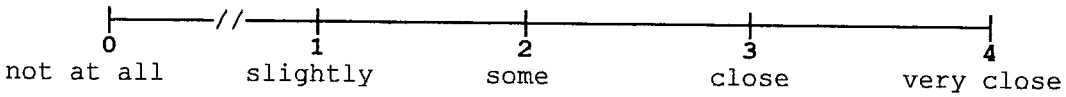
Important Distinctions for Item #28
(See Item #9, *Systemic/Relational Rationale* and
Item #44, *Ecosystemic Focus* and
Item #14 *Relationship Between Alcohol and Client System Explored.*)

29. FOCUS ON THE THERAPIST AND CLIENT RELATIONSHIP:

29a. Extensiveness: How often was the relationship between the therapist and the client a central part of therapy?



29b. Emulation: How well developed as a central part of therapy was the relationship between the therapist and the client?



This item refers to the quality of the relationship between the therapist and client as it relates to therapy. The rater should consider the level of *involvement* between the therapist and client. As well, the therapeutic relationship may serve as a *model* for understanding other relationships, or for trying out new behaviors in a relationship.

Involvement: The rater should consider how carefully the therapist paid attention to what the client was saying and how responsive the therapist was to questions or comments by the client. Involvement may also be reflected in the therapist's very real and personal reactions to the client.

This item is not a measure of frequency of therapist verbal response, nor a measure of therapist style.

Therapeutic relationship as a model: The rater should also consider the therapist's efforts to use the relationship which they have with the client as:

1) a vehicle for helping the client to understand how she/he comes across or reacts to others (via focusing on the client's behavior towards or feelings about the therapist); or
2) a means of modeling behaviors which will be helpful to the client in forming or changing relationships outside of therapy. The therapist may use role playing as a means of accomplishing these goals, if the therapist and client reverse roles (not if either roleplays a third person).

If the therapist discusses some aspect of the way the clients relate or come across to the therapist without making any reference to other relationships, the rater should assign a low rating to this item. If the therapist (or the client at the therapist's urging) makes some connection between an aspect of the therapeutic relationship which is being discussed and an aspect which has been or will be present in another relationship, this item should receive a rating greater than "0". If the therapist is highly involved, but does not use the relationship as model, this item should not receive a "4".

EXAMPLE

The following example should receive a rating of greater than "0" on this item because it illustrates the use of the therapeutic relationship by the therapist as a vehicle for helping the client to understand how she comes across to others:

T: "I've noticed that when we discuss what other options you would have for dealing with a problem in addition to the one you initially came up with, your voice changes, it takes on the kind of `edge' that my voice takes on when I'm irritated. Do you notice that?"

C: "I guess I didn't know that it showed, but it bugs me when you start telling me what I should do."

T: "I'm sorry if it sounds to you like I'm trying to tell you what to do. That's not my intention. My intention is to help you see that you have other options besides the first one you came up with. The reason I think it is important to do that is so you consider several options, before you choose which one you want to exercise in dealing with a problem."

C: "I see. I guess I'm kind of sensitive about other people telling me what to do."

T: "Does that sensitivity emerge with other people as well as with me?"

C: "I don't know if others are aware of it but it certainly is there."

T: "Does it seem to you like others are trying to tell you what to do when you have discussions with them about a problem you are faced with?"

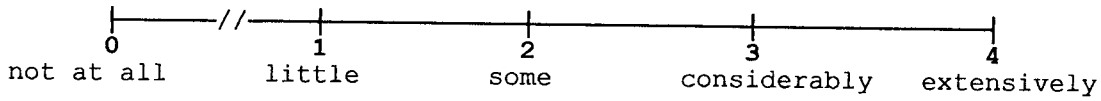
C: "I know they aren't, but it often seems to me at the time like they are."

T: "You mentioned to me that you were not aware of the `edge' in your voice which I noticed and which suggested to me that you were irritated. Do you think that some `edge' appears in your voice when you are discussing issues like these with others?"

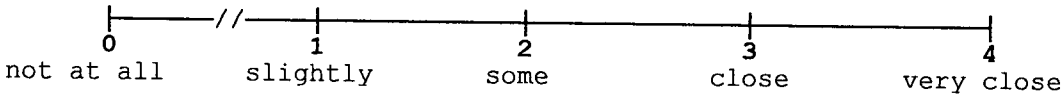
C: "Yes, I suppose it probably does, and others may be picking up on it like you did."

30. EMPHASIS ON AWARENESS:

30a. Extensiveness: To what extent did the therapist actively assist the client in becoming aware of their own present experience?



30b. Emulation: How close was the therapist to the model of assisting the client in becoming aware of their own present experience?



The rater should look for an ongoing focus by the therapist on the client's here and now experience in the session as a whole. Therapeutic processes focus on increasing awareness for the client. This item should be rated on the combination of the following two components.

Experience of the client. Feelings, thoughts, behavior, expectations and/or memories are all potential parts of the client's experience in the moment. The rater should see a consistent focus by the therapist on stimulating the client's awareness of their experience on several of these levels, throughout the session. The therapist will emphasize awareness of experiences in the session, rather than abstractly discussing the past.

Here and now. The emphasis is on awareness in the present moment, rather than the near or distant past. Memories of the past or visions of the future are explored primarily in terms of how they impact the here and now.

EXAMPLES

The following examples demonstrate specific instances of the therapist emphasizing awareness of the client's present experience.

T: "I really respect the way you are looking after yourself by staying on guard against falling into the depression today. It sounds like you are paying careful attention to the real possibilities of danger you are feeling right now. In fact, when you say the 'pit of depression' is by the door, I'm wondering if you are feeling a bit nervous about finishing the session with me today."

C: "I get queasy whenever I think about how he touched me when I was six. I don't think I'll ever get over it."

T: "Even now, you feel sick when you are remembering. I notice you are holding your stomach. Is that where you feel the queasiness?"

C: "I feel she doesn't respect me when she comes home late. She's only 13."

T: "I understand you want her home on time, but I'm not quite clear how you are feeling when you think of her coming home late."

C: "I'm not going to put up with much more of her attitude!"

T: "When you say that so forcefully, it sounds like you are feeling angry."

C: "Ya, I'm furious. I could never get away with this kind of crap when I was a kid. My dad would beat me within an inch of my life."

T: "You know, on the one hand, I'm hearing your anger about your daughter being late, but on the other hand I'm hearing what I think is a fear that your anger might get out of hand with her. How does that fit for you right now?"

C: "I still hate my old man for beating me, so I don't want to hit her. But I don't want to lose her, and sometimes getting mad seems like all I can do."

T: "You seem to feel sad, scared and helpless right now. You're just not sure what to do. There don't seem to be any easy answers."

Important Distinctions for Item #30

with Item #22, *Cognitive Orientation*

One aspect of awareness for #30 is cognitive, but it also includes awareness of body sensations, gestures, emotions, memories or expectations experienced in the present moment. #22 is not necessarily focused on the present moment, and only pursues cognitive aspects of the client's experience.

with Item #39 *Here and Now Experiential Focus*

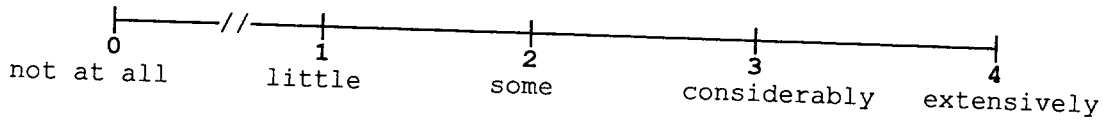
Item #39 *Here and Now Experiential Focus*, emphasizes the therapist's evocation of here and now experiences for the client, where Item #30, *Emphasis on Awareness*, focuses on the therapist assisting the client's awareness of their own present experience.

with Item #41, *Primary Empathy*
and Item #4, *Advanced Empathy*

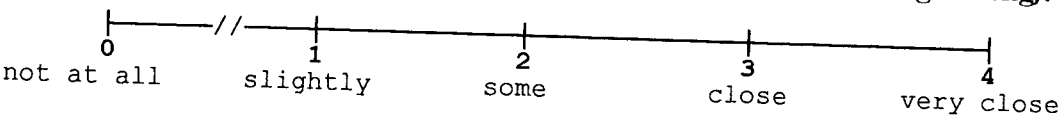
Both #41 and #4 focus on the therapist's communicating their understanding of the experiences and feelings of the client. This may increase the client's own awareness of their own experience. Item #30 focuses on the therapist assisting the client in becoming aware of their experience. In other words, a high score on #30 indicates the increased awareness of the client, but not necessarily the empathy of the therapist. A high score on #41 and/or #4 indicates the empathy of the therapist, but not necessarily the increased self-awareness of the client. These items will often be related, so care must be taken to distinguish the unique aspects of each item..

31. THERAPEUTIC RAPPORT:

31a. Extensiveness: How much rapport was there between therapist and client (i.e., how well did the therapist and client get along during the session)?



31b. Emulation: How close to the ideal in terms of the description was the rapport between client and therapist (i.e., how well did the therapist and client get along)?



This item is intended to measure the extent to which the relationship between the therapist and client is marked by harmony and accord (i.e., how well the therapist and client got along in the session). Raters who have developed their own operational definition of *Rapport* are encouraged to use it only if it is consistent with how *Rapport* is defined above.

Among the items in this scale, this item is clearly the most dependent on client behavior as well as therapist behavior. Although the rater should assign a low rating to this item if she/he believes that the therapist made efforts to get along with the client without success, this item should not be given a high rating unless rapport clearly existed between the therapist and client.

EXAMPLES

The following example indicates a strong rapport between client and therapist.

C: "I wanted to talk about how Ann and I get along, you know, sexually."

T: "I really appreciate that you feel safe enough to say that, when I can see how difficult it is for you."

C: "I do feel awkward and uncomfortable bringing it up. I'm not sure how you feel about talking about it."

T: "I think it takes courage for you to face your discomfort. I want to let you know that I am very much willing to discuss this with you. I also hope we can both be sensitive to those moments that feel uncomfortable and that we can take our time."

C: "Thanks. That sounds good. I'm not really sure where to begin, but I'm feeling more comfortable."

T: "I can see your body relaxing, and you just took a deep breath. Looks like you're ready to dive in."

C: "Dive in, that's just what it feels like. Well, here goes..."

The following is an example of poor rapport between therapist and client.

C: "I wanted to talk about how Ann and I get along, you know, sexually."

T: "Just before we get into that, I wanted to check and see if you were still planning to come next week."

C: "Yes, I'll be in town after all."

T: "So you wanted to talk about your sexual relationship?"

C: "It's been on my mind."

T: "Sex has been a kind of obsession for you?"

C: "Well, no not an obsession, I've just been thinking a lot about it."

T: "So what do you want to talk about?"

C: "It's probably not that important."

T: "What is really important for you?"

C: "I don't know."

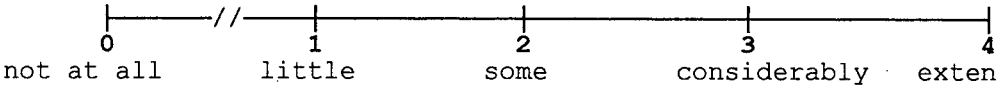
T: "You're feeling uncertain, and unsure of what you want?"

Important distinctions for Item # 31

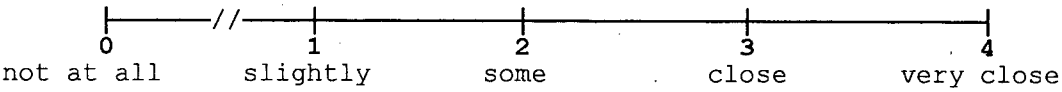
See Item #11 *Warmth and Caring*

32. ENCOURAGES INDEPENDENCE:

32a. Extensiveness: How much did the therapist encourage the client's independence from the therapist in dealing with her/his problems?



32b. Emulation: How close to the ideal in terms of the description was the therapist at encouraging the client's independence from the therapist in dealing with her/his problems?



The therapist should be considered to have encouraged the client's independence (from the therapist) if she/he:

(1) explicitly stated that it is the therapist's goal for the client to do things on her/his own, or
(2) Instead of supplying solutions to the client, subtly urged the client to come up with solutions to her/his problems.

In rating this item the rater must consider what transpired over the entire session, and be able to justify from the therapist's behavior a rating of greater than "1"

EXAMPLES

The following are examples of explicit encouragement of independence:

T: *"As we work together I'd like you to become less and less reliant on my input such that eventually you really don't need my help on dealing with problems you come up against."*

T: *"Given that you've developed some good skills over the last several weeks, I'm confident that you can make a good decision for yourself. I'll be anxious to hear what you decide."*

T: *"I'm really glad to hear about the way you handled that situation. That's a nice example of how you didn't need my help to come up with a way of dealing with a difficult situation. You came up with a plan on your own, carried it out, and resolved the problem."*

The following is an example of more subtle encouragement of independence:

C: *"I'm thinking about quitting my job. I just can't manage people."*

T: *"Quitting your job would be a pretty big move on your part. How can you go about deciding whether that is the best thing for you to do?"*

C: *"The most important thing about my job is to manage the people who work under me. But I just had another of my key staff members leave the company to take another job and in general, the morale around the office is not great."*

T: *"Do you think that these occurrences are due to how you are managing your staff?"*

C: *"Well, yes. I'm the manager after all, my job is to take the responsibility."*

T: *"Could there be any other reasons, beside your managing ability, that you lost another key staff member and the office morale is low?"*

C: (No response).

T: *"If another manager in your company were to come to you and tell you that she had just lost another key staff member and that the office morale was low, what would you attribute that to?"*

C: *"It could be that she's not a good manager, but if she was in an area like mine, it could be that other companies are raiding her staff and luring away the good ones."*

T: *"How would that impact on the morale of the others who were left?"*

C: *"If she couldn't replace the people she lost, that would increase the burden on those who were left, leaving them feeling under pressure."*

T: *"Do you think any of those things could be going on in your situation?"*

C: *"Yes, I guess they could be."*

T: *"As a manager, what can you do to deal with that situation?"*

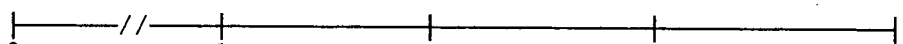
T: *"I'd like you to give some thought to what the best thing would be for you to do in this situation, quit or take some other action. Then I'd like to hear what you come up with."*

Important Distinctions for Item #32

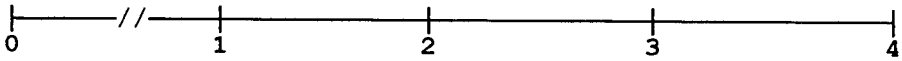
See Item #26, *Collaboration or Mutuality.*

33. RESPECTFUL:

33a. Extensiveness: How often was the therapist respectful of the client and the therapeutic process?



33b. Emulation: How well did the therapist emulate expressing respect for the client and the therapeutic process?



This item is intended to measure the quality of respect the therapist brings to the session. Respect may be expressed in several ways.

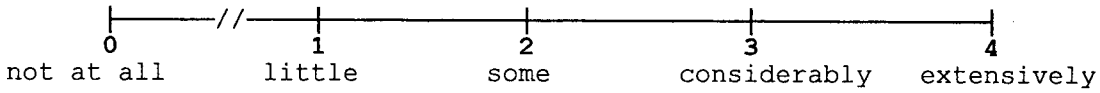
1) Treating the client as an important and worthy individual
2) Listening in a nonjudgmental way
3) Treating the therapeutic process as a special event worthy of care and commitment.
4) Treating symbols, chairs, charts, etc., that have been used as a part of therapy with care, as if they were valuable.

Important Distinctions for Item #33
with Item #5, *Accepting and Validating Client's Experience*

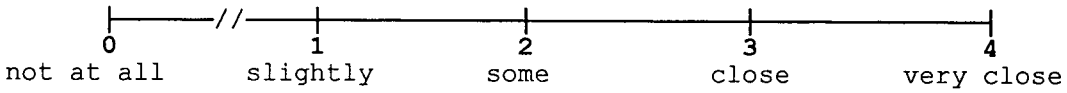
Accepting and validating the client's experience is one way of expressing respect for the client, but on its own would not indicate a rating greater than "0". It is important when rating Item #33, *Respect*, to consider the session as a whole. The rater should see evidence of respect (more than just acceptance) for both the client and the process.

34. THERAPIST ENSURES CLIENT COMPLETES WSD:

34a. Extensiveness: How much did the therapist ensure that the client complete their Weekly Situation Diary (WSD)?



34b. Emulation: How well did the therapist emulate ensuring that the client complete their WSD?

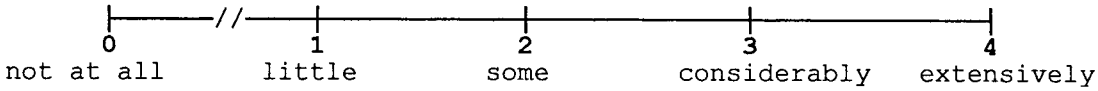


This item measures the effort the therapist made in the session to ensure the client completed their WSD. This effort may be seen in various forms. The therapist may acknowledge successful completions, encourage, remind or schedule the client's completions, enlist their cooperation, help them catch up, or use other ways to ensure they complete their WSD's.

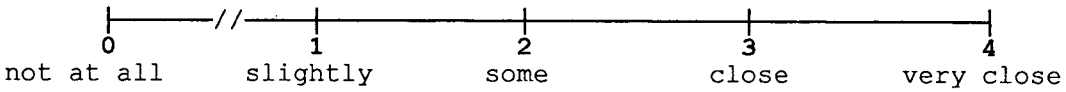
This item need not take a lot of time in the session, and would receive a rating higher than "0" if the therapist uses any of the efforts described above. A high rating would indicate the therapist had thoroughly ensured the client would complete their WSD's and would not indicate the length of time spent on doing so.

35. THERAPIST GATHERS QUANTITATIVE INFORMATION FROM CLIENT:

35a. Extensiveness: To what extent did the therapist gather quantitative information from the client?



35b. Emulation: How well did the therapist emulate gathering quantitative information from the client?



This item measures the therapist's use of quantitative information provided by the client. The use of charts, scales, amounts, or ratios are all considered quantitative information. How many, how often, or how much would be common questions appropriate for this item. The therapist relies primarily on this information throughout the session in developing discussion. The important criterion for this item is the amount of quantitative information used in the session. This includes information from the Weekly Situation Diaries (WSD). A high rating on this item indicates a strong focus on gathering quantitative information throughout the session, where the therapist minimalizes the gathering of abstract, metaphorical or qualitative information.

EXAMPLE

The following example would receive a rating greater than "0", if similar information was gathered by the therapist.

> The therapist notes that according to the client's Weekly Situation Diary for the past two weeks, the client drank three times: the first time he/she drank heavily, consuming 26 ounces of gin; the second time less heavily, 4 ounces of gin after dinner; and the third time, more heavily again, about 14 ounces of gin. This information forms the basis of a short discussion related to the client's desired consumption level.

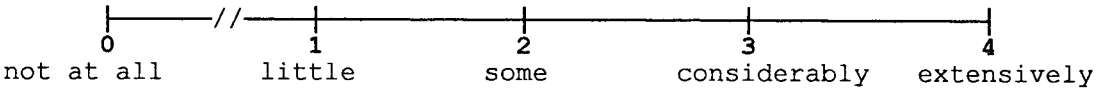
Important Distinctions for Item #35

with Item #38, *Accents Feedback From Charts*

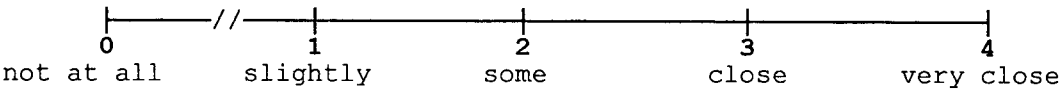
Item #38 focuses on the feedback of information using the charts, while Item #35 focuses on the type of information gathered from the client.

36. SYMBOLIC ORIENTATION:

36a. Extensiveness: How often did the therapist use a symbolic orientation in dealing with client problems or goals?



36b. Emulation: How well did the therapist emulate the use of a symbolic orientation in dealing with client problems or goals?



This item taps the use of symbols in therapy as a means of externalizing an internal problem, goal, desired state, or relationship. Relationships which can be expressed symbolically include relationships with other people, parts of the self (like one's anger, critical judge, or inner child) or an object (like alcohol). The rater should look for the therapist asking for, suggesting or supporting the development of symbols by the client, including through dreamwork.

Typically, the disturbed part of the relationship for a client is embodied in and given concrete form (i.e., an empty chair, a photograph, a letter, a bottle, etc.), when symbolically externalizing . The symbol may be derived from an expressive transaction (art work, etc), or a metaphor emerging from the therapeutic discourse, or suggested by the therapist. Once given form, the externalized other provides an opportunity for client and therapist to directly encounter the specific qualities of the relationship between it and the experiencer. The breadth of possible externalization is essentially unlimited and can range from abstract concepts (e.g.,. peace, society, time frames, parts of self) to real objects (e.g.,. liquor bottle, food needles, money).

Symbols may also include metaphors, imagery, gestures, labels, rituals, or body postures, etc. which represent an issue the client is dealing with. Symbols need not be represented concretely.

Symbols may be used in several ways:

1) for describing the relationship,
2) engaging directly in the relationship dialogue,
3) heightening the experience of rigidity of relationship,
4) exploring possible changes
5) directly experiencing relational novelty with the other object, or
6) deciding what to do with the externalized symbol.

It is important for the therapist to treat externalized objects as significant symbols, to be handled with the real integrity and respect that one affords to any other person. The same quality of respect also applies to responding to metaphors or other symbols.

EXAMPLES

The following examples illustrate the therapist working with a symbolic orientation.

T: "You've talked about alcohol as a kind of friend that gives you solace, companionship and comfort. Could we use this wine bottle to represent alcohol?"

C: "Okay, I guess. What do I do with it?"

T: "If you were going to show me where alcohol is in your life right now, using this bottle, where would you put it in this room?"

C: "I'd hold onto it. Keep it close."

T: "So right now as you hold the alcohol in your arms, how does that feel?"

C: "I like the alcohol being so close. But you know, I feel like it's hard to be close to my wife when I've got alcohol right here. It's almost like having a lover."

T: "So this is a very important relationship you have with alcohol, and sometimes you feel your relationship with alcohol affects your relationship with your wife."

C: "Yes, but I wouldn't want to let it go. I feel like sometimes its the only thing that helps me cope."

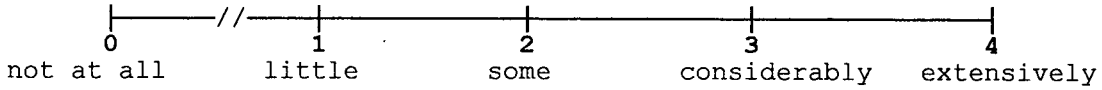
T: "It feels like the right thing to do for now, to keep the alcohol in your arms."

C: "I brought in a symbol of my desired state, that you asked for last week. It was hard at first, but suddenly I knew just the right thing. I made this picture with felt pens and magazine ads. I hope its okay."

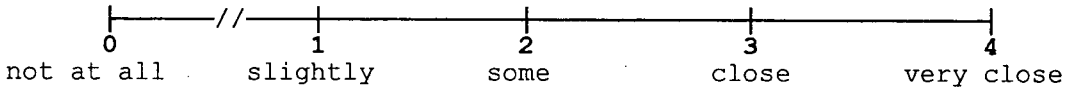
T: "I can see this means a great deal to you. I feel privileged that you would share it with me. I'm very interested in what this picture expresses for you. Could you tell me how you came to create it?"

37. THERAPIST INVOLVEMENT:

37a. Extensiveness: To what extent was the therapist involved in the session, with a lively, vivid, original and interesting communication style?



37b. Emulation: How well did the therapist emulate showing their involvement through their interest and liveliness, and their vivid, original, interesting communication style?



This item is intended to measure how involved the therapist appears to be in the session. The rater should consider how carefully the therapist paid attention to what the client was saying, and how vivid, lively or original the therapist's communication style is. The therapist may indicate involvement by mentioning isomorphic or parallel processes that she/he experiences that are similar to client experiences.

Ratings on this item should not be based on frequency of therapist verbal response. The therapist may be paying attention to the client, but not frequently interjecting her/his comments. In such cases this item should receive a high rating, despite the fact that the therapist was not verbally active.

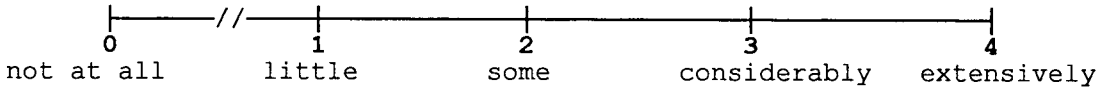
The therapist's communication style should show vividness, liveliness, and originality. Vividness refers to how colorful and descriptive the therapist's language is. Liveliness refers to an engaging manner of speaking on the part of the therapist. Originality can be conveyed in the therapist's choice of examples, use of analogies, or design of engaging homework assignments, among many other things.

Involvement may also be expressed by the isomorphism of therapist and client issues. The therapist may, for example, mention a fear of sadness that is similar to a client's fear of grief, and point out that the therapist is also involved in learning how to face that fear. Transference or countertranference issues that emerge due to the therapist's involvement will be dealt with directly by the therapist.

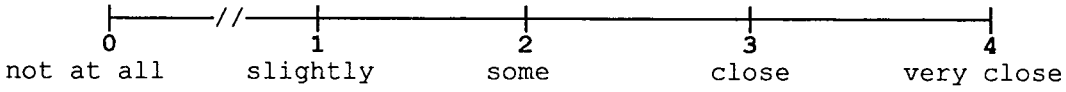
The rater is cautioned not to consider lack of involvement undesirable on the part of the therapist, as the therapist may be deliberately attempting to remain detached. Thus, the rater should not let her/his ratings of the therapist's involvement be influenced by whether the rater considers therapist involvement to be good or bad.

38. ACCENTS FEEDBACK FROM CHARTS:

38a. Extensiveness: How much did the therapist accent feedback from the charts, following the guidelines for the feedback process?



38b. Emulation: How well did the therapist emulate following the guidelines for the feedback process when using the charts?



This item measures the use of a structured feedback process based on charts developed from the weekly situation diaries. Feedback guidelines are outlined below. The therapist acts as a custodian of the feedback process and assumes responsibility for the conduct of the session, directing the client's attention to aspects of the feedback charts that are particularly relevant to the client's ongoing recovery process.

Guidelines for Giving Feedback

1. **Readiness of the Giver.**
   Give the feedback only when there are clear indicators the receiver is ready to listen to it. If not ready, the receiver will be apt not to hear it or to misinterpret it.

2. **Descriptive, Not Interpretive.**
   Giving feedback should be like acting as a "candid camera". It is a clear report of the facts, rather than the giver's ideas about why things happened or what was meant by them. It is up to the receiver to consider the whys or the meanings or to invite the feedback giver to do this with the therapist.

3. **Recent Happenings.**
   The closer the feedback is given to the time the event took place the better. When feedback is given immediately, the receiver is most apt to be clear on exactly what is meant. The feelings associated with the event still exist so that this, too, can be part of understanding what the feedback means.

4. **Appropriate Times.**
   Feedback should be given when there is a good chance it can be used helpfully. It may not be helpful if the receiver feels there is currently other work that demands more attention. Critical feedback presented in front of other pressing issues may be seen as damaging rather than helpful.

5. **New Things.**
   There is a tendency in giving feedback to say only the obvious. The therapist should consider whether what they are reacting to really may be new information for the receiver. Many times, the thing which may be helpful new information is not simply a report of what the therapist saw the receiver doing, but rather the way it caused them to feel or the situation they felt it put them in.

6. **Changeable Things.**
   Feedback should be about things which can be changed if the receiver chooses to do so.

7. **Not Demand a Change.**
   The concept of feedback should not be confused with requesting a person to change. The receiver can consider whether he/she wishes to attempt a change on the basis of new information. The therapist may wish to include the changes they would like to see, but is not likely to be helpful to say in effect, "I told you what's wrong with you, now change!"

8. **Not an Overload.**
   When learning how to give feedback, we sometimes tend to overdo it. It's as though we were telling the receiver, "I just happen to have a list of reactions here and if you'll settle back for a few hours I'll read them off to you." The receiver may prefer time to consider the feedback.

9. **Given to be Helpful.**
   The therapist should consider their own reasons for giving their reactions. Are they trying to be helpful to the receiver? Or, are they unloading some of their own feelings, or using the occasion to try to get the receiver to do something the would be helpful for them? For example, if the therapist feels angry at the client and wishes to express it, they may appropriately say so, if they include a description of the behaviors that caused the anger.

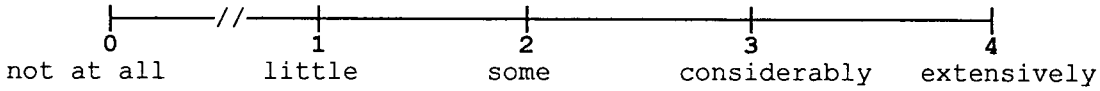
10. **Giver Shares Something.**
    Giving feedback can become "one-up-manship". The receiver goes away feeling as though he/she is "not as good" as the giver because it was his/her potential for improvement that was focused on. The giver may see themselves as having given a lecture from the lofty pinnacle of an imaginary state of perfection. The exchange often can be kept in better balance by the giver including some of his/her own feelings and concerns.

11. **Is Specific, not General.**
    Use quotes and give examples of what you are referring to.

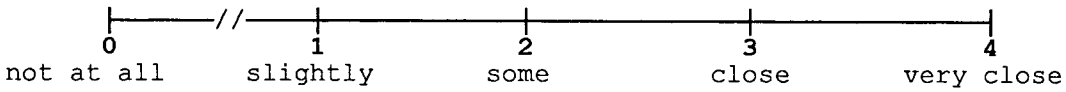
Not all therapists use charts. Therapists not using charts would automatically receive a rating of "0" on this item.

39. HERE AND NOW EXPERIENTIAL FOCUS:

39a. Extensiveness: To what extent did the therapist use a here and now experiential focus throughout the session?



39b. Emulation: How well did the therapist emulate maintaining a here and now experiential focus throughout the session?



The rater should look for a focus by the therapist on creating an ongoing here and now experience for the client throughout the session as a whole. Therapeutic processes focus on change occurring through experiences in the present. The experience for the client must be appropriately contextualized and prepared for by the therapist. This item should be rated on the combination of the following two components.

Experience of the client. Feelings, thoughts, behavior, expectations and/or memories are all potential parts of the client's experience in the moment. The rater should see a consistent focus by the therapist on stimulating the client's awareness on several of these levels during the session. The therapist will emphasize creating experiences in the session, rather than abstractly discussing the past.

Here and now. The emphasis is on evocation of an experience in the present moment, rather than simply describing the near or distant past. Memories of the past or visions of the future may be explored through enactment, rolepaying, working with symbols, or other techniques that create an experience in the here and now.

EXAMPLES

The following examples indicate specific times when the therapist created a here and now experience for the client.

C: "My dad just wasn't there for me to talk to when I was growing up. He was always off on trips. When he did come home, he was too tired to talk."

T: "I notice when you talk about your childhood with your dad that your shoulders seem to tighten up. Could you exaggerate that tension in your shoulders and tell me again about your dad's absence?"

T: "You've been telling me what you think will happen if you begin looking for a new job, and it sounds like you're feeling afraid right now, just thinking about it. Could you draw a picture of how that fear feels?"

C: "I'm feeling afraid of my depression coming back. It's like a pit, waiting to swallow me up."

T: "You experience your depression as lurking like a trap, waiting to catch you?"

C: "Yes, I always have to be on guard."

T: "So you are feeling on guard even here, at this moment?"

C: "Yes."

T: "If we were to put that pit of depression somewhere in the room, where would it be right now?"

C: "It's waiting by the door, ready to get me when I leave."

T: "You sound terrified."

C: "I am. Its a big, black pit, just waiting to get me."

T: "How big is it when we bring it into this room?"

C: "It's not as big as it used to be, I could sneak past it if I was careful today."

T: "So you are beginning to think you might be able to avoid it today, if you stay on guard?"

C: "Yes, at least for now."

T: "Is there anything you'd like to do with that pit now?"

C: "I'd like to make it smaller."

T: "Why don't you show me how you'd make it smaller? Just go ahead and pretend you can change it. How would you do it?"

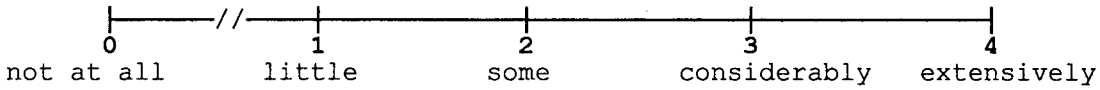
Important Distinctions for Item #39
with Item #30 *Emphasis on Awareness*

Item #39 *Here and Now Experiential Focus*, emphasizes the therapist's evocation of here and now experiences for the client, where Item #30, *Emphasis on Awareness*, focuses on the therapist assisting the client's awareness of their own present experience.

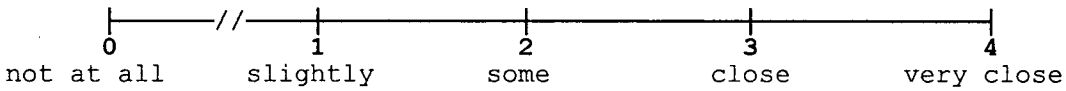
(see also Item #16, *Generating Relational Novelty*)

40. SETTING AND FOLLOWING AGENDA:

40a. Extensiveness: Did the therapist work collaboratively with the client to formulate and follow a specific agenda for the session?



40b. Emulation: How well did the therapist emulate working collaboratively with the client to formulate and follow a specific agenda for the session?



There are three aspects to consider when rating this item:

1) Did the therapist work with the client to set a specific agenda for the session?
2) Did the therapist work with the client to follow the agenda during the session?
3) Did the therapist attempt to involve the client in a collaborative effort?

To the extent that all of the above are present, the item will be rated highly. Note that the rater is not being asked how collaborative this effort was but rather how thoroughly the agenda was set and followed as a result of a collaborative effort on the part of the therapist.

How to rate the three aspects of this item:

Setting Agenda. A specific agenda involves the therapist setting aside a significant amount of time at the outset of the session to compile agenda items contributed by the therapist OR the client. These agenda items should have been prioritized and ideally a specific amount of time should have been allotted for each item (although the latter need not have been present in order for this item to be rated highly).

EXAMPLES

This item would receive a low rating if the therapist and client set and followed an agenda in a collaborative fashion, but the agenda was vague. For example:

T: "It seems like it would be important for us to spend some time talking about your family. What do you think about doing some of that today?"

The following are examples of therapist behaviors that constitute a complete agenda and thus would be likely to result in a high rating on this item:

1) Generating items to be discussed.

2) Choosing which of the items generated will be discussed.

3) Determining the order in which items are to be discussed.

4) Allotting the amount of time to be spent discussing each item.

All of these aspects need not be present for this item to receive a high rating if the rater believes that the aspects which were present were done thoroughly.

Following Agenda. In order to determine if the agenda was completely followed, the rater should consider:

1) If the items agreed on in the agenda were discussed in the session.

2) If the priority and the time allotted for each item (if time was allotted for each item) matched that specified in the agenda.

3) If major deviations from the original agenda occurred, those deviations were clearly agreed upon by the therapist and client.

If the therapist worked collaboratively with the client to formulate a very specific agenda but did not work to see that it was followed, this item should be rated lower than if the negotiated agenda had been followed.

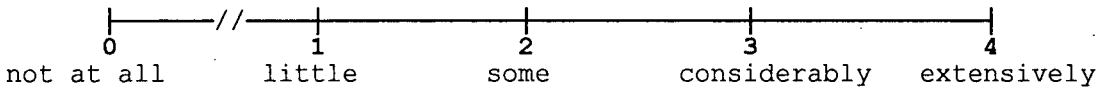
Collaboration. Collaboration involves the client contributing items to the agenda, participating in prioritizing them and helping to decide how much time should be allotted to each (if time was allotted for each item). If deviations from the agenda occurred, they should have been negotiated by the client and the therapist.

NOTE: If the therapist set and followed a very specific agenda, but did not collaborate with the client, do not rate this item higher than "3".

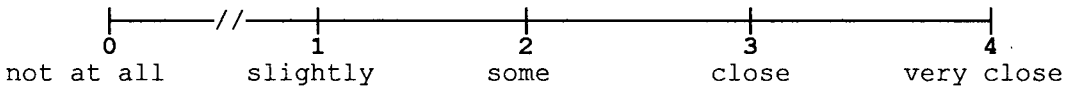
NOTE: Session excerpts may not include the setting of agenda, if this was done in the first ten minutes of the session. The rater is asked to consider the evidence in the session excerpt only in coming up with a final rating. If there is clear indication that a collaborative agenda was set earlier in the session, (and the agenda was being followed) then this item would receive a rating greater than "0".

41. PRIMARY EMPATHY:

41a. Extensiveness: How extensively was the therapist empathic toward the client (i.e. did she/he thoroughly convey an intimate understanding of and sensitivity to the client's experiences and feeling)?



41b. Emulation: How well did the therapist emulate the description of being empathic toward the client (i.e. did she/he convey an intimate understanding of and sensitivity to the client's experiences and feeling very well)?



Accurate empathy has been described by Rogers and Trux (1967) as the "ability of the therapist accurately and sensitively to understand experiences and feelings and their meaning to the client during the moment-to-moment encounter of psychotherapy." This item is intended to measure the extent to which the therapist conveyed to the client that she/he had an intimate understanding of the client's experiences and feelings and their meaning to the client.

In many cases the therapist will not yet have an intimate understanding of the client's experience. This item should still be rated highly if the therapist actively attempted to understand the client's experiences and feelings.

To rate this item, the rater will need to rely to some degree on nonverbal therapist cues including tone of voice, expressiveness, etc. However, it is not correct to assume that the therapist is empathic merely because she/he is a therapist. This item is one on which it is particularly important for the rater to avoid giving ratings of greater than "0" as default values. The rater must be able to justify from the therapist's behavior any rating on this item which is greater then "1". In other words, don't assume that the therapist is empathic, at least a little.

EXAMPLES

This item should receive a rating of "0" if throughout the session the therapist:

a) ignored or seemed disinterested in the client's experiences and feelings.
b) was unable to and did not attempt to understand the client's experiences and feelings.
c) devalued or dismissed the client's experiences or feelings or the meaning that the client attached to them.

This item should receive a high rating if throughout the session the therapist:

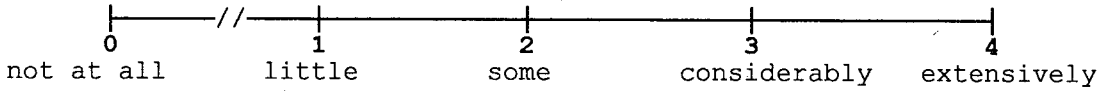
a) attempted to recognize and validate the client's experience with her/his thoughts or feelings. (The therapist may ultimately disagree with the client's beliefs or interpretations of events yet still demonstrate empathy depending on how she/he voices her/his disagreement to the client).
b) asked questions of the client in order to understand the client's experiences and feelings or their meaning to the client (e. g., T: "You look very hurt right mow. Tell me what that feels like... I know it's very painful.").

Important distinctions for Item #41

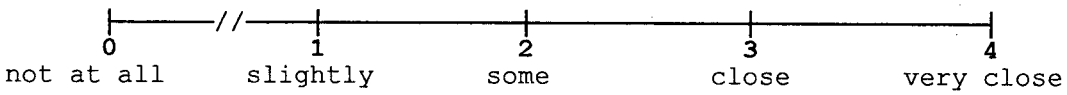
Compare with item #4 *Advanced Empathy.*

42. CLIENTS DERIVE OWN MEANING FROM CHARTS:

42a. Extensiveness: To what extent did the therapist encourage the client to derive their own meaning from the charts?



42b. Emulation: How well did the therapist emulate encouraging the client to derive their own meaning from the charts?



This item is intended to measure the therapist encouraging or supporting the client in deriving their own meaning from the charts. The therapist's interpretation of the charts is minimalized. The rater should look for specific examples of therapist behavior that promotes the client in deriving their own meaning from the charts.

This item is particularly focused on the use of charts, so therapists not using charts would receive a rating of "0".

EXAMPLE

The following example shows the therapist encouraging the client to derive their own meaning from the charts.

T: "So, we've charted your extended family relationships, your relationship with your family and your partner, as well as your drinking patterns for the last two weeks. What stands out for you in looking at the charts?"

C: "When I had that low period while my parents were visting, my drinking seemed to increase a bit, and all my family scores also went down too."

T: "What does the chart tell you?"

C: "Having my parents around is more trouble than I realized. It affects everything. Even my relationship with myself suffered."

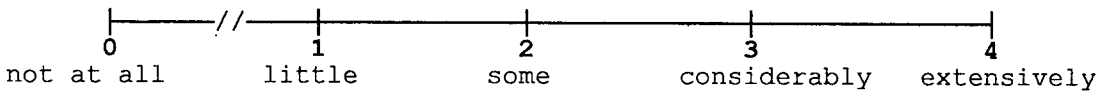
Important Distinctions for Item #42

with Item #24, *Clients Define Their Own Experience*

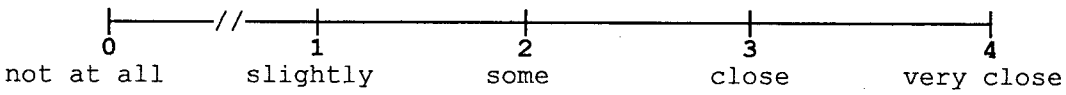
Item #42 refers to the meaning a client derives from their charts, while Item #24 refers to the meaning a client derives from their experiences in the past or present. The charts may refer to the client's experiences in the past, but to receive a high rating on this item, the client is focusing on deriving meaning from the charts themselves. If the client is also focusing on deriving meaning from the experiences more directly, they may also receive a rating higher than "0" on Item #24.

43. EXPRESSIVE/CEREMONIAL ACTIVITIES:

43a. Extensiveness: To what extent did the therapist actively attempt to engage the client in expressive or ceremonial activities?



43b. Emulation: How well did the therapist emulate actively attempting to engage the client in expressive or ceremonial activities?



This item measures the use of expressive or ceremonial activities by the therapist. A rating greater than "0" is indicated when either type of transaction is seen.

Expressive Activities. Expressive activities involve making public experiences that have been restricted to private domains. This may induce the exploration, discovery, naming and owning of experiences that may not have had expression within a client's own awareness as well as other members of the client subsystem. Therapists must be sensitive to the potentially profound personal and group significance of moving to the social domain matters previously unacknowledged or confined to private domains.

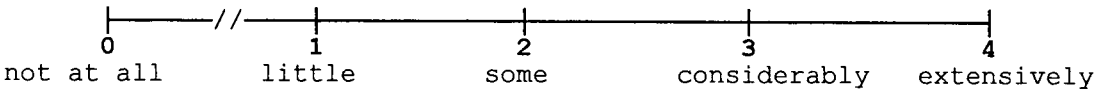
Expressive activities frequently include the use of mediums other than verbal exchanges as part of their procedural consideration. For example, therapists may ask the client(s) to do some art work (drawing, painting, singing, dance/movement) focused on a specific relationship felt important to bring out in the light of the social domain. Alternatively, activities might center upon the use of other mediums such as metaphors, story telling, mind mapping, and sculpts to provide an avenue of symbolic expression.

A key process feature of this transaction class is that they only include unfolding, creative, and expressive experiences. They do not include the working through or transformation of the expressed symbols and are used solely for the projecting out to the social domain of private matters.

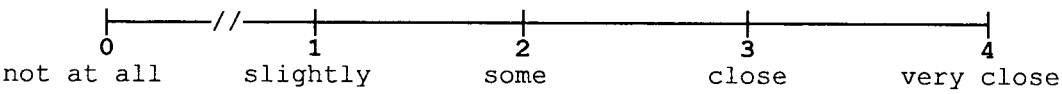
Ceremonial Activities. Ceremonial activities share an orientation towards therapeutic sequences imbued with symbolic significance. The therapist is engaged in a formal process of recognition and acknowledgement of change. Highlighting transformation in either the client(s) or the therapeutic system, ceremonial activities represent experiential markers punctuating the old from the new. Because these activities are meant to be memorable, they are enacted with a degree of reverence by system members with careful attention being paid to procedural details. Generally, the therapist acts as an "emcee" of a kind during the activity, proposing a ceremony in order to demark the present and future from the past. The therapist generates, with the client(s), an occasion designed to function as a rite of passage. The therapist's intention is to symbolize and legitimize the resolution of originally troubled relationships. Since the client must participate wholely in the symbolic activity, therapists may include clients in the planning and preparation of the ceremonial event. In this way, therapists may ensure that the activity fits coherently with the client subsystem. When this is not done, the therapist structures the occasion, drawing on metaphoric and symbolic material evident in past sessions. Ceremonial activities are best implemented sparingly in order to maximize the intensity of their significance. They include such techniques as burials, celebrations, cremations, penance, wake, confession, or rituals.

44. ECOSYSTEMIC FOCUS:

44a. Extensiveness: How often did the therapist use an ecosystemic focus (i. e., explore the client's patterns of relationship with various levels of the system, intrapersonal, marital, extended family, friends and community, etc.)?



44b. Emulation: How close to the model of using an ecosystemic focus was the therapist (i. e., explore the client's patterns of relationship with various levels of the system, intrapersonal, marital, extended family, friends and community, etc.)?



The purpose of this item is to measure the extent to which the therapist used an ecosystemic focus. The focus could be intrapersonal, or interpersonal, and include the immediate or extended family, friends or community. The rater should look for specific examples of times when the therapist explored the client's pattern of relationship with various levels of the system.

The exploration could be very literal, as in charting effects on family interactions, or involve more metaphorical explorations, as in drawing pictures to represent parts of self.

It is important to consider the range of systems considered within a session, as well as the extent to which each system is explored.

EXAMPLES

The following are examples in which the therapist addressed different levels of the system.

T: "You've been describing how you felt not drinking for a week. I'm curious about how your partner has reacted. Could you tell me a bit about that?"

T: "It seems like considering your friends is important in making this decision about changing jobs. Do you have a sense of the different opinions they might have?"

C: "One time I just want to leave him, and the next minute I'm desperately clinging to him. I don't know what to do. I know he wants me to stay, but..."

T: "It sounds like you are hearing two voices inside yourself, maybe a bit like the two of you in the relationship. Perhaps we could take a few minutes and have a kind of discussion between those parts of yourself. That inner dialogue might shed some light on the relationship between you two."

T: "You are feeling angry when your boss criticizes you in front of your colleagues. It seems like that is a familiar feeling for you. Have you ever felt like that when you were younger?"

C: "Well, yes. My dad would make fun of me when I was learning to play soccer. I felt furious and humiliated then, too. It's a lot like the way I feel now."

T: "You feel hurt by the lack of support."

C: "I just want to be accepted."

T: "You felt like you couldn't get your dad's approval, and now it seems like you can't get your boss's approval?"

C: "It seems to be the theme of my life right now."

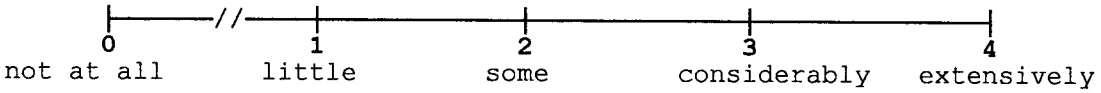
T: "Are there other relationships in your life that feel like this to you?"

Important Distinctions for Item #44
with Item #29, *Focus on the Therapist and Client Relationship*

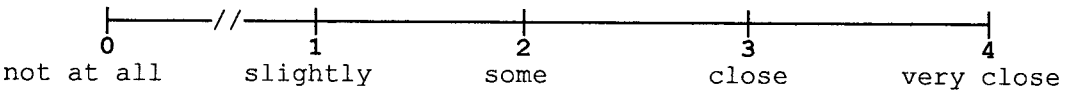
When the therapist uses the relationship with the client to explore other relationships, they are using an ecosystemic focus, however the therapist relationship could be central with little evidence of an ecosystemic focus. It is important to consider the entire session when rating Item #44, *Ecosystemic Focus*. The rater should look for evidence of more than one level of the system being considered.

45. DISCUSSING WAYS OF CHANGING RELATIONSHIPS:

45a. Extensiveness: How much did the therapist help the client to consider ways in which the client can bring about a desired change in an important relationship (or role expectations within a relationship)?



45b. Emulation: How well did the therapist emulate helping the client to consider ways in which the client can bring about a desired change in an important relationship (or role expectations within a relationship)?



This item measures attempts by the therapist to discuss with the client options she/he has to bring about changes in a relationship. In rating this item, the rater should not consider activities or enactments that create relational novelty in the session (See Item #16 *Generating Relational Novelty*); the rater should consider discussions of possible changes.

EXAMPLE

T: *"So, you are concerned your wife is having an affair."*

C: *"I'm pretty sure she is; she has been going out a lot without me."*

T: *"What do you plan to do about it?"*

C: *"I don't know what to do about it. I've been depressed for so long now that I guess I can't blame her if she is having an affair. I guess I'll just have to hope that it burns itself out."*

T: *"Okay, so one thing you could do is wait it out and see what happens. Is there anything else you could do?"*

C: *"I suppose I could confront her and tell her that she better stop or get out...I'm afraid to do that though because she probably will leave."*

T: *"Is there a way you can talk about it with her without giving her an ultimatum?"*

C: *"Yes, I suppose so...but I don't know that doing that would change anything."*

T: *"What is it about that option that might not lead to change?"*

C: *"Well, suppose I do talk to her and she says, 'I don't love you anymore.' She might decide then that she might as well leave."*

T: *"It sounds like that might be one reason you'd have for not talking to her about it?"*

C: *"Yeah...I'm afraid of what she'll do. We don't have much of a relationship anymore."*

T: *"Have you thought about other things you could do to improve your relationship? For instance, are there things that you used to enjoy doing together that you don't do anymore?"*

C: *"I was thinking that we don't go to any movies or concerns anymore and that's one thing we used to enjoy. We also used to enjoy going for walks in the countryside."*

T: *"Are those things that you would be able to arrange to do?"*

C: *"I think so, if my wife is willing to spend time with me."*

T: *"Are these things that she'd enjoy?"*

C: *"Yes, I think so...she used to talk about wanting to go skiing. I've never been interested because I haven't had enough energy. I bet she'd enjoy doing that and I might even enjoy it once I got out there."*

T: *"So you have some ideas of things you can do with your wife that you think might be fun for both of you. Do you think that doing those things will have a positive impact on your relationship?"*

C: *"I'm not sure, but it's probably worth a try."*

Important Distinctions for Item #45

See Item #16 *Creating Relational Novelty*

APPENDIX C

List of TARS Items and Subscale Items

TARS ITEMS

1) Supportive Encouragement
2) Concrete Language
3) Developmental Focus
4) Advanced Empathy
5) Accepting And Validating Client's Experience
6) Spontaneity
7) Semi-Structured Session Format
8) Tracking Learning
9) Systemic/Relational Rationale
10) Timing
11) Warmth And Caring
12) Changes Desired In Relationships
13) Task Oriented
14) Primary Emphasis On Relationship Between Alcohol And Client System
15) Recovery Directed By Client
16) Generating Relational Novelty
17) Assessing General Functioning
18) Wholistic Appreciation
19) Cumulative Learning Format
20) Appropriate Expression Of Therapist's Feelings
21) Appropriate Intensification Of Experience
22) Cognitive Orientation
23) Recent Past Orientation
24) Clients Define Their Own Experience
25) Client Capability And Resources
26) Collaboration Or Mutuality
27) Low To Moderate Intensity
28) Relating Interpersonal Change To Therapy
29) Focus On The Therapist And Client Relationship
30) Emphasis On Awareness
31) Therapeutic Rapport
32) Encourages Independence
33) Respectful
34) Therapist Ensures Client Completes WSD
35) Therapist Gathers Quantitative Information From Client
36) Symbolic Orientation
37) Therapist Involvement
38) Accents Feedback From Charts
39) Here And Now Experiential Focus
40) Setting And Following Agenda
41) Primary Empathy
42) Clients Derive Own Meaning From Charts
43) Expressive/Ceremonial Activities
44) Ecosystemic Focus
45) Discussing Ways Of Changing Relationships

ExST items

3) Developmental Focus

4) Advanced Empathy

6) Spontaneity

9) Systemic/Relational Rationale

16) Generating Relational Novelty

18) Wholistic Appreciation

21) Appropriate Intensification Of Experience

24) Clients Define Their Own Experience

26) Collaboration Or Mutuality

28) Relating Interpersonal Change To Therapy

30) Emphasis On Awareness

36) Symbolic Orientation

37) Therapist Involvement

39) Here And Now Experiential Focus

43) Expressive/Ceremonial Activities

SFT Items

2) Concrete Language

7) Semi Structured Session Format

13) Task Oriented

14) Primary Emphasis On Relationship Between Alcohol And Client System

15) Recovery Directed By Client

17) Assessing General Functioning

19) Cumulative Learning Format

22) Cognitive Orientation

23) Recent Past Orientation

27) Low To Moderate Intensity

32) Encourages Independence

34) Therapist Ensures Client Completes Wsd

35) Therapist Gathers Quantitative Information From Client

38) Accents Feedback From Charts

42) Clients Derive Own Meaning From Charts

Overlap Items

1) Supportive Encouragement

5) Accepting And Validating Client's Experience

8) Tracking Learning

10) Timing

11) Warmth And Caring

12) Changes Desired In Relationships

20) Appropriate Expression Of Therapist's Feelings

25) Client Capability And Resources

29) Focus On The Therapist And Client Relationship

31) Therapeutic Rapport

33) Respectful

40) Setting And Following Agenda

41) Primary Empathy

44) Ecosystemic Focus

45) Discussing Ways Of Changing Relationships

Appendix D

Tables of Mean and Standard Deviation Scores of TARS Items

Table 29

Mean & Standard Deviation Scores of ExST Items for Entire Population

	Extensiveness						Emulation					
	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=120)		Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
3	2.4	1.2	1.0	1.1	2.5	0.6	2.7	1.3	1.5	1.7	2.7	0.8
4	1.2	1.0	1.7	1.0	1.9	1.1	2.1	1.7	2.9	1.5	2.2	1.2
6	2.9	1.2	2.4	1.4	2.9	0.9	3.0	1.2	2.7	1.5	3.0	1.0
9	2.5	1.3	2.9	1.4	2.8	1.1	2.8	1.4	3.0	1.4	2.9	1.1
16	1.5	1.5	1.4	1.3	2.1	1.3	1.8	1.7	2.2	1.8	2.5	1.6
18	2.3	1.0	2.4	1.2	2.0	1.0	2.4	1.1	2.8	1.4	2.2	1.2
21	2.5	1.3	2.5	1.6	2.8	1.0	2.7	1.3	2.7	1.6	3.0	1.2
24	2.7	1.1	3.4	0.8	3.2	0.6	2.8	1.1	3.5	0.8	3.3	0.8
26	2.6	1.1	2.9	1.4	3.0	0.8	2.7	1.2	3.0	1.4	3.1	0.9
28	1.1	1.1	0.2	0.5	1.4	0.7	1.2	1.3	0.4	1.1	1.7	1.0
30	2.3	1.4	2.8	1.2	2.6	1.3	2.5	1.5	3.1	1.2	2.7	1.5
36	1.9	1.5	1.3	1.4	1.9	1.3	2.1	1.6	2.0	1.9	2.3	1.6
37	2.9	1.0	3.5	0.8	3.4	0.8	3.0	1.0	3.4	0.8	3.5	0.8
39	2.0	1.5	2.6	1.5	2.2	1.3	2.2	1.6	3.0	1.5	2.5	1.5
43	1.3	1.3	0.7	1.0	1.7	1.2	1.7	1.6	1.3	1.7	2.0	1.5

Table 30

Mean & Standard Deviation Scores of SFT Items for Entire population

	Extensiveness						Emulation					
	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=120)		Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
2	2.0	1.0	1.7	1.4	2.3	1.1	2.0	1.0	2.4	1.5	2.3	1.1
7	1.0	1.6	1.0	1.6	1.3	1.7	1.0	1.6	1.0	1.6	1.3	1.7
13	0.9	1.5	0.7	1.3	1.2	1.2	1.0	1.5	1.0	1.6	1.2	1.2
14	1.6	1.2	1.4	1.3	2.2	1.0	1.7	1.3	1.9	1.5	2.4	1.1
15	1.8	1.3	1.4	1.5	2.1	0.9	1.7	1.3	1.4	1.5	2.2	0.8
17	2.1	1.3	1.4	1.4	1.9	0.8	2.1	1.3	1.6	1.6	2.0	0.9
19	1.0	1.5	1.0	1.5	1.3	1.2	1.1	1.6	1.2	1.7	1.4	1.4
22	1.8	1.1	1.3	1.7	2.0	1.1	1.8	1.2	1.4	1.7	2.1	1.1
23	2.1	1.4	1.4	1.6	2.2	0.9	2.0	1.4	1.5	1.7	2.3	1.0
27	2.3	1.2	1.7	1.6	2.2	1.0	2.3	1.2	1.9	1.6	2.2	1.1
32	1.0	1.0	1.1	1.2	2.1	0.9	1.1	1.1	1.3	1.4	2.1	0.9
34	0.3	0.7	0.9	1.5	0.3	0.6	0.3	0.8	1.1	1.7	0.3	0.7
35	1.1	1.6	0.9	1.5	1.2	1.4	1.1	1.6	1.1	1.8	1.3	1.6
38	1.0	1.6	0.8	1.4	0.9	1.4	1.0	1.6	0.9	1.5	0.9	1.5
42	0.9	1.4	0.9	1.6	0.8	1.3	0.9	1.5	1.0	1.7	0.8	1.3

Table 31

Mean & Standard Deviation Scores of Common Items for Entire Population

	Extensiveness						Emulation					
	Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=120)		Rater 1 (n=120)		Rater 2 (n=120)		Rater 3 (n=30)	
Item#	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1	0.9	1.0	0.9	1.0	2.5	0.8	1.2	1.3	1.6	1.6	2.7	0.9
5	3.6	0.7	3.8	0.4	3.5	0.6	3.6	0.7	3.8	0.5	3.5	0.6
8	1.2	1.2	0.8	1.0	1.9	0.9	1.5	1.5	1.6	1.7	2.2	0.9
10	3.4	0.9	3.6	0.8	3.1	0.7	3.5	0.8	3.6	0.8	3.3	0.8
11	3.7	0.6	3.7	0.6	3.5	0.6	3.7	0.5	3.7	0.6	3.7	0.5
12	1.0	1.1	1.0	1.0	1.9	0.9	1.2	1.3	1.4	1.5	2.2	1.0
20	1.7	0.9	0.8	0.9	2.1	1.0	2.3	1.3	1.7	1.8	2.4	1.2
25	1.4	1.1	1.6	0.9	2.5	0.7	1.6	1.3	2.3	1.2	2.5	0.8
29	2.8	0.9	3.1	1.0	2.7	0.8	2.9	0.9	3.3	0.9	2.8	0.8
31	3.6	0.7	3.7	0.6	3.5	0.6	3.6	0.7	3.6	0.7	3.4	0.7
33	3.9	0.3	4.0	0.1	3.6	0.5	3.9	0.3	4.0	0.3	3.7	0.5
40	1.6	1.4	0.9	1.4	1.0	1.0	1.8	1.5	0.9	1.5	1.0	1.1
41	2.8	1.0	3.5	0.7	3.5	0.6	3.3	1.0	3.6	0.7	3.7	0.5
44	2.5	0.9	3.0	0.8	2.1	1.0	2.6	1.0	3.3	0.7	2.2	1.0
45	0.5	0.8	0.2	0.5	1.2	0.9	0.6	1.0	0.3	0.8	1.2	1.0