

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

**A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600**

A STUDY IN RELIABILITY AND VALIDATION: WHAT IS THE NATURE OF
THE UNIQUE AND JOINT CONTRIBUTIONS OF THERAPIST ADHERENCE
AND THERAPEUTIC ALLIANCE TO TREATMENT OUTCOME?

by

Jennifer R. Patton

A dissertation submitted to the Graduate Faculty in
Psychology in partial fulfillment of the requirements for
the degree of Doctor of Philosophy, The City University of
New York

1997

UMI Number: 9807981

**Copyright 1997 by
Patton, Jennifer Ramsey**

All rights reserved.

**UMI Microform 9807981
Copyright 1997, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

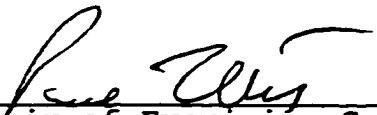
Copyright 1997

JENNIFER RAMSEY PATTON

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/15/97
Date


Chair of Examining Committee

9/22/97
Date


Executive Officer

Paul L. Wachtel, Ph.D.

Anderson J. Franklin, Ph.D.

J. Christopher Muran, Ph.D.

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

A STUDY IN RELIABILITY AND VALIDATION: WHAT IS THE NATURE OF THE UNIQUE AND JOINT CONTRIBUTIONS OF THERAPIST ADHERENCE AND THERAPEUTIC ALLIANCE TO TREATMENT OUTCOME?

by

Jennifer R. Patton

Adviser: Professor Paul L. Wachtel

In the last fifteen years in the area of comparative psychotherapy outcome research, the necessity for documentation of therapist fidelity to treatment protocol has been widely recognized. To evaluate the differential impact of treatments, it is necessary to demonstrate that clinicians have delivered treatments as intended and that treatments are reliably discriminable. A preliminary version of the Beth Israel Adherence Scale (Santangelo, 1995) was developed to evaluate therapist adherence to behaviors specified by protocol in three brief treatments: psychodynamic psychotherapy, cognitive-behavioral therapy, and interpersonal-experiential therapy. Patients were participants in The Beth Israel Brief Psychotherapy Research Project which has been conducting an NIMH-funded pilot project on the efficacy of psychotherapy treatments and their differential capacity to respond to ruptures in the therapeutic alliance. Three raters used a shortened and

refined version of the Beth Israel Adherence Scale to rate therapist adherence in 72 fifteen minute audiotaped segments excerpted from 24 cases, eight from each of the three therapies. Results demonstrated that therapists engaged in more behaviors reflective of their own orientations as opposed to those reflective of other treatment perspectives. Preliminary evidence indicated that the scale demonstrates psychometric properties to an acceptable degree: satisfactory to strong levels of interrater reliability and discriminant validity were obtained. A secondary and related goal of this study was to evaluate the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome to further an understanding of the nature of the contributions of both specific and non-specific factors. Results relating to this question did not yield significance, but were suggestive of the need for further research.

Acknowledgments

I gratefully acknowledge the support I received from my dissertation committee as well as from my friends and family. Dr. Paul Wachtel, the chair of my committee, provided encouragement and guidance throughout. Dr. Chris Muran, the committee member involved in the on-going research for this study, was responsive and attentive every step of the way with regard to both large and small details. His willingness to consult at nearly a moment's notice is much appreciated.

My dear friends, Debora Munczek and Nina Sanders, commiserated at many points throughout the process, offering kindness and concern, as well as good advice. My friends at Wendell Street listened to my worries and helped to make the the weight of this project more tolerable. Last, but never least, my beloved parents, Frank and Charlotte, and my beloved sisters, Martha and Charlotte, offered their reassuring presence, their hope for my future, and their belief in me. I thank you all.

Table of Contents

Copyright	ii
Approval	iii
Abstract	iv
Acknowledgments	vi
Table of Contents	vii
List of Tables	ix
CHAPTER	
1. Introduction	1
Review of the Literature	2
Brief Outline of Perspectives	2
Definition of Terms	8
History of Fidelity and Treatment Manuals	9
Advantages/Disadvantages of Treatment Manuals	12
Measures Used to Assess Treatment Fidelity	15
Difficulties with Treatment Manipulation Checks	17
Recommendations on Assessment of Fidelity	18
Relationship between Fidelity and Outcome	21
Relationship between Alliance and Outcome	22
Contribution of Adherence/Alliance to Outcome	23
Beth Israel Project: Context for Present Study	25
Theoretical Background of Adherence Scale	30
Statement of the Problem	38
Hypotheses	39
2. Method	40
Audiotapes	40
Therapists	41
Patients	42
Raters	43
Training of Raters	44
Development of the Scale	45
Criteria for Rating of Items	45
Rationale for Use of Rating Criteria	50
Measurement of Therapeutic Alliance	53
Measurement of Treatment Outcome	53
Data Analysis	54
3. Results	56
Hypothesis I	56
Hypothesis II	56
Hypothesis III	61
Multiple Regression Analyses	62
Correlational Analyses	71

4. Discussion	80
Hypotheses I and II	80
Psychometric Properties of the Scale	80
Limitations	84
Rating Process	84
Non-Independence of Segments	86
Differences in Data Sets	89
Hypothesis III	91
Joint Contribution of Adherence and Alliance	91
Unique Contributions of Adherence and Alliance	94
CBT and Positive Outcome/BAP and Poor Outcome	95
Alliance and Outcome/Alliance and Adherence	99
Conclusion	101
5. Appendices	102
A. Beth Israel Adherence Scale	102
B. Beth Israel Adherence Scale Rating Form	110
C. Scoring Directions for Rating	112
6. References	114

List of Tables

Table

1. ICC Means by Overall Scale, Subscale, & Therapy Stage	57
2. ANOVAs Comparing Subscale Scores Across Tx Type	60
3. MRA of Adherence & Alliance on Outcome: All Cases	66
4. MRA of Adherence & Alliance on Outcome: Early Cases	67
5. MRA of Adherence & Alliance on Outcome: Middle Cases	68
6. MRA of Adherence & Alliance on Outcome: Late Cases	69
7. f^2 , Effect Sizes, Power, & N for 80% Power for MRAs	70
8. 0-Order Correlations: Adherence, Alliance, & Outcome	75
9. First-Order Correlations: Adherence & Outcome	79

INTRODUCTION

Over the last 40 years, psychotherapy research has placed increasing emphasis on the evaluation of treatment outcome. Beginning in the 1950s, concern with therapist fidelity to treatment protocol as it related to the study of outcome became more pronounced in response to the field's desire to answer its critics (Garfield & Bergin, 1994). With the advent of the use of treatment manuals in the 1960s and 70s, interest in and commitment to demonstrating treatment fidelity grew (Luborsky & DeRubeis, 1984). Now, in the last fifteen years in the area of psychotherapy outcome research, the necessity for documentation of therapist fidelity to treatment protocol in comparative psychotherapy research has been widely recognized (Waltz, Addis, Koerner, & Jacobson, 1993).

There are now purported to be over 400 different psychotherapy treatments (Karasu, 1986). For comparative outcome research to be meaningful, attempts to evaluate the differential impact of specific treatments must result in those treatments being reliably discriminable and clinicians conforming to the principles of those treatments. In order to identify both the effects of specific treatments and the effects which transcend specific treatments, it is necessary to operationally define these treatments and to ensure that they are delivered as planned. Yeaton & Sechrest (1981) note that "any valid comparison of effectiveness of treatments simply cannot be made without considerable

attention to their strength and integrity" (p. 166). This is the principle motivating factor in treatment fidelity research at the present time.

The Beth Israel Brief Psychotherapy Research Project at Beth Israel Medical Center in New York City is conducting an NIMH-funded pilot project on the efficacy of three psychotherapy treatments and their differential ability to respond to problems in the therapeutic alliance. To evaluate the differential impact of these treatments, it is necessary to demonstrate that clinicians have delivered the treatments as intended and that the treatments are reliably discriminable. The primary focus of this dissertation is the evaluation of the reliability and validity of a shortened and refined version of the Beth Israel Adherence Scale now in use (Santangelo, 1995). A secondary and related goal of this dissertation is to evaluate the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome.

REVIEW OF THE LITERATURE

A Brief Outline of Perspectives in the Field of Psychotherapy Outcome Research

There are differing perspectives in the area of comparative psychotherapy outcome research regarding which factors contribute to efficacy and how these factors may be best understood and evaluated. It is not within the scope

of this study to discuss each of these perspectives. Some of the more widely held perspectives are outlined below, however, so that this dissertation and the Beth Israel study may be understood within the broader context of the current thinking in the field.

Considerable contradiction abounds in the field of psychotherapy (Arkowitz, 1992). One contradiction is reflected in the current trend towards psychotherapy integration which coexists with the continued pursuit of comparative psychotherapy research. It is evident now that the current trend over the past 15-20 years in the practice of psychotherapy is towards integration and/or eclecticism (Arkowitz, 1992; Wachtel, 1977, 1987). Most therapists currently describe themselves as working from an eclectic or integrative point of view although there remain some who are loyal to a single approach (Arkowitz, 1992). At the same time, Lambert & Bergin (1994) note that few studies have been done which assess the effectiveness of these integrative or eclectic treatments, making this an important area for future research investigation. Given the documented trend towards integration and eclecticism, one may question the relevance of comparative psychotherapy research. Indeed, it is reasonable to question the purpose of pursuing outcome studies which test the efficacy of different therapeutic treatments when there is considerable evidence that no specific treatment is more effective than

any other (e.g. Elkin et al., 1989; Luborsky, Singer & Luborsky, 1975; Smith & Glass, 1977; Stiles, Shapiro, & Elliot, 1986).

In support of this perspective, Lambert (1989) has noted that specific technical factors appeared to be responsible for less than 15% of the variance in outcome in a selected set of psychotherapy studies. Stiles, Shapiro, & Elliot (1986) state, "The paradoxical findings of outcome equivalence and content non-equivalence present a serious dilemma, because they seem to imply that no matter what a therapist does, the end result is the same" (p. 167). They note that researchers are motivated to resolve the outcome-equivalence/content-non-equivalence paradox because the implied advice (as stated by Rachman & Wilson, 1980, in Stiles, Shapiro, & Elliot, 1986) would be the seemingly untenable: "Regardless of the nature of your problem, seek any form of psychotherapy" (p. 167). Beutler (1991) notes that to fall prey to this line of thinking is to accept the "dodo bird verdict" (Luborsky, Singer, & Luborsky, 1975) which may be seen as an "indictment of comparative psychotherapy research" (p. 226). This "inertia laden belief" (that all therapies are the same)..."must be countered by any who dare venture the opinion that different psychotherapies exert different effects among different patients" (p. 226). From Beutler's perspective, it is unwise to give up the search for differential treatment

effects since there are a "large number of patient, therapist, and treatment variables that may mediate the effects of treatments" (p. 226). In general, in order to explain the no-difference result in comparative psychotherapy outcome studies, researchers have gathered empirical support for other explanations, which include "patient variability, variability in the quality of the patient/therapist alliance, or variability in the therapist's skillfulness" (Rounsaville, O'Malley, Foley, & Weissman, 1988, p. 682).

There is also evidence that the phenomenon of "common factors" or "non-specifics" is responsible for a large portion of psychotherapy's success (Howard & Orlinsky, 1986). This is in contrast to the hypothesis that specific technical factors may influence treatment efficacy. "Common factors" may be defined as those aspects of the treatment (e.g. empathy, warmth, genuineness, encouragement) which are often associated with the therapeutic alliance. They are noted to transcend distinct treatments and are at least partly responsible for treatment efficacy (Arkowitz, 1992; Frank & Frank, 1991; Karasu, 1986). One of the foremost writers in this area has been Jerome Frank who has suggested that all therapies share some curative features in common and that aspects shared by distinct treatments outweigh their differences (Frank & Frank, 1991).

In support of the study of specific factors,

however, there is some evidence that focused treatments in the cognitive-behavioral domain are more effective in the treatment of specific anxiety disorders than are the exploratory and open-ended therapies; however, research findings in this area remain somewhat equivocal (Lambert & Bergin, 1994). Further, Lambert & Bergin (1994) note that effect sizes for psychotherapy proper (specific factors and common factors combined) vs. no-treatment appear consistently larger than effect sizes for "placebo" (only common factors) treatment vs. no-treatment (Lambert & Bergin, 1994). Along these same lines, Jones, Cumming, & Horowitz (1988) present data indicating that specific factors can predict treatment outcome and "suggest that patient change is far more complex than the nonspecific hypothesis... implies" (p. 48). The above research would fit with Arkowitz's (1992) point of view that an exclusive focus on common factors in relation to treatment efficacy is misguided. In the same vein, Stiles, Shapiro, & Elliot (1986) note that the common factors explanation may be seen as only one possible resolution to the "dodo bird" paradox.

To further elaborate on the question of specifics vs. non-specifics, Butler & Strupp (1986) assert that this simplistic dichotomy (several decades old at this point) requires reevaluation. They urge that psychotherapy research abandon the simplistic notion that "specific" and "non-specific" factors may be siphoned out from "the

identity of psychotherapy with its interpersonal context" (p. 38) since psychotherapy is neither a medical treatment administered to a passive patient nor a myth (Frank & Frank, 1991) which cannot be substantiated. Further, Henry, Schacht, and Strupp (1990) note that the interpersonal impact of the therapeutic relationship on patient change is of central importance. Henry et al. (1990) note that when the impact of the therapeutic relationship is understood from this interpersonal perspective, there is no longer a need for the dichotomy between specific and non-specific factors.

In summary, there is evidence that both specific and non-specific factors play a role in psychotherapy outcome, and that the apparent equivalency of treatments should continue to be challenged. An approach to the study of outcome which considers common factors (reflected in the therapeutic alliance and its interpersonal context) and their interaction with specific techniques has gained support. This approach reflects a move towards a more complex and integrative research stance which emphasizes the study of both common factors and specific techniques, and coincides with the trend in clinical practice towards integration and/or eclecticism. An integrative approach to the study of psychotherapy outcome would appear to be the partial context for the Beth Israel study on problems in the therapeutic alliance. Indeed, in its evaluation of

differential treatment efficacy, the Beth Israel study examines both specific techniques (particularly those reflected in interpersonal-experiential therapy) and their interaction with common factors (which are those reflected in the therapeutic alliance).

In order for research to address the question of comparative outcome and differential treatment efficacy, it is essential that treatment fidelity be upheld. To this end, the primary goal of this dissertation is the evaluation of an adherence measure designed to contribute to the assessment of treatment fidelity. A discussion of the definition of terms associated with treatment fidelity is found directly below. A secondary question addressing the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome, will be discussed at a later point in this paper.

Definition of Terms

A critical prerequisite of comparative psychotherapy outcome research is to uphold the fidelity of the treatment (Kazdin, 1994). "Treatment integrity" (one aspect of treatment fidelity) may be defined as the degree to which a "treatment condition is implemented as intended" (Moncher & Prinz, 1991, p. 247; Yeaton & Sechrest, 1981). A second facet of treatment fidelity described as "treatment differentiation" refers to whether the delivered treatments

may be identified as different from each other in the intended manner (Moncher & Prinz, 1991). The terms "adherence" and "competence", understood to be "two separate phenomena" (Hill, O'Grady, & Elkin, 1992, p. 74), are seen as integral components of treatment integrity and may be subsumed under this comprehensive term. "Adherence" refers to the degree to which therapists used techniques appropriate to their treatment approach, while "competence" refers to the degree of skill with which these techniques are applied (Hill, O'Grady, & Elkin, 1992; Kazdin, 1994). Waltz et al. (1993) expand the above definition of competence by including the degree to which the therapist grasped the key aspects of the therapeutic context and responded appropriately to these identified "contextual variables" (p. 620). Finally, the use of treatment manuals (as a principle means to ensure treatment fidelity) is an integral component in the determination of outcome differences between psychotherapies (Lambert & Ogles, 1988; Luborsky & DeRubeis, 1984). A treatment manual is a written guide which outlines therapeutic procedures, including techniques, themes, and actions, to be followed by the therapist (Kazdin, 1994).

History of Fidelity and Treatment Manuals

Psychotherapy research continues to focus on the efficacy of specific treatments, more recently with regard

to how they interact with patient characteristics or disorders. With regard to treatment efficacy, Moncher & Prinz (1991) state that, "Verification of fidelity is needed to ensure that fair, powerful, and valid comparisons of replicable treatments can be made" (p. 247). Concern regarding treatment fidelity increased following the controversy surrounding the question "Does psychotherapy work?", generated by Eysenck in the 1950s when he asserted that psychotherapy was not effective (Eysenck, 1952). Eysenck's provocative 1952 review of psychotherapy studies generated an outpouring of outcome research designed to demonstrate the efficacy of psychotherapy (Moncher & Prinz, 1991). These efforts had their desired outcome as documented by a number of studies including the Smith & Glass (1977) meta-analysis demonstrating psychotherapy's effectiveness. Since that time, concern regarding treatment fidelity has grown substantially, with the last fifteen years evidencing particularly focused efforts. Nonetheless, in their review of 359 treatment outcome studies occurring during the years 1980-1988, Moncher & Prinz (1991) noted that 55% of studies basically disregarded the issue of treatment fidelity. In addition, only one of the eight studies in the most recent period of their review (1986-88) combined the recommended use of treatment manuals, supervision of therapists, and assessment of adherence to protocol (Moncher & Prinz, 1991).

Although a smattering of treatment manuals emerged in the 1960s (e.g. Wolpe, 1969), their formal entrance occurred somewhat later in the mid-to-late 1970s (Beck, Rush, Shaw, & Emery, 1979; Klerman & Neu, 1976; Klerman, Weissman, Rounsaville, & Chevron, 1984; Luborsky, 1976, 1984; Strupp & Binder, 1984). Luborsky & DeRubeis (1984) have described the proliferation of treatment manuals (during the late 70s and 80s) as a "small revolution" in the conduct of psychotherapy research (p. 5). Proponents of the development and use of treatment manuals included third-party insurance payers, Carter administration congressional committees, and many researchers in both psychology and psychiatry. The thinking at this time was that psychotherapy should be evaluated with the same rigorous approach used to evaluate the effectiveness of pharmacotherapy. In particular, researchers were concerned about exercising control over the "calibration of treatment amount and quality" and being able to assess treatment efficacy for a variety of different therapies (Luborsky & DeRubeis, 1984, p. 7).

Of course, "non-manual-based" psychotherapies have been studied for the past four decades (Luborsky, Woody, McLellan, O'Brien, & Rosenzweig, 1982; Luborsky & DeRubeis, 1984, p. 8). These studies have produced results which indicate that non-manual-based psychotherapies can be discriminated (Luborsky et al., 1982; Luborsky & DeRubeis,

1984). Still, there are prohibitive risks associated with the use of non-manual-based psychotherapies for the purposes of research, and these have led to the relatively recent appearance of treatment manuals accompanied by adherence/competence measures. Indeed, Luborsky & DeRubeis (1984) noted that as an indication of the sentiment in the early 1980s, the Cognitive Therapy and Research journal instructed its contributors on the requirement of the use of manuals in published outcome studies. Further, since the start of the NIMH Collaborative Study of Depression (Elkin, et al., 1989), the use of treatment manuals for the purposes of conducting sound psychotherapy efficacy research has become standard (Beutler, Machado, & Allstetter Neufeldt, 1994). Indeed, Lambert & Ogles in their (1988) review of psychotherapy treatment manuals named 10 different existing manuals outlining a variety of psychotherapies from the cognitive, behavioral, and psychodynamic persuasions. Since the Lambert & Ogles (1988) review, there has been a significant increase in the number of manuals in use. Psychotherapy manuals outlining humanistic approaches have contributed to the continued proliferation (Lambert & Bergin, 1994).

Advantages and Disadvantages of Treatment Manuals

Lambert & Bergin (1994) outline advantages to using treatment manuals in comparative outcome studies in

psychotherapy research. These include, but are not limited to, the enhancement of internal validity by "purifying" the technical features of the treatment; improved training and supervision of therapists; the facilitation of the development of adherence and competence measures; and the facilitation of the replication of treatments across time and location (Lambert & Bergin, 1994). It should be noted that the use of a treatment manual without the monitoring of treatment fidelity makes it difficult to assess whether the treatment has been implemented as intended. Manipulation checks provide the proof that the treatment as described in the manual has been delivered as planned (Waltz et al., 1993).

The importance of upholding treatment fidelity with the use of treatment manuals has been documented by a number of researchers (e.g. Yeaton & Sechrest, 1981; Jacobson, 1991; Crits-Christoph et al., 1991; Luborsky, McLellan, Woody, O'Brien, & Auerbach, 1985). Yeaton & Sechrest (1981) cite several studies (Kassebaum, Ward, and Wilner, 1971; Quay, 1977) in which treatment integrity is in question. They note that the results from studies in which treatment integrity has not been monitored are difficult to interpret, often leaving the reader with two possible, but different conclusions: either treatment was simply not effective, or if it could have been effective, it was administered in a manner unlikely to be effective. An analysis which

disregards treatment integrity may be seen as nearly worthless since it becomes virtually impossible to discern which factors were responsible for a given set of results (Yeaton & Sechrest, 1981).

Another advantage to the combined use of treatment manuals and manipulation checks may be found in the well-known example of the Snyder, Wills, & Grady-Fletcher (1991) follow-up study comparing two different marital therapies. The authors indicate that the Behavioral Marital Therapy (BMT) couples experienced a considerably higher rate of divorce in the years subsequent to treatment than did the Insight-Oriented Marital Therapy (IOMT) couples. A dialogue between Jacobson (1991) and the above authors ensued in which Jacobson asserted that the two treatments were not correctly represented in the manuals. Specifically, Jacobson (1991) noted that the BMT manual correctly reflected BMT as it existed in 1980, but did not include more up-to-date innovations. In addition, the IOMT manual was described as including therapist interventions which are reflective of BMT interventions as it is currently practiced. The "truth" in this controversy may be less relevant than the fact that such a dialogue could only have occurred because the carefully-controlled conditions of the Snyder, Wills, & Grady-Fletcher (1991) follow-up study included the use of treatment manuals which could be studied in retrospect (Lambert & Bergin, 1994).

A questionable aspect of the emphasis on the administration of manualized therapies has been articulated by Garfield (1992). He raises the question of whether research findings are pertinent to the "everyday" integrative and/or eclectic psychotherapies which are practiced in the typical clinical setting. Most therapists outside research settings do not consistently refer to therapy manuals in order to protect the purity of their treatment. Cohen, Sargent, & Sechrest (1986) note that "therapists learn overwhelmingly from experience with clients and only rarely consult therapy research" (as cited in Frank & Frank, 1991, p. 18). This raises the question of how to make research findings on manualized therapies relevant to everyday practice.

Measures Used to Assess Treatment Fidelity

Several studies (e.g. Hill, O'Grady, & Elkin, 1992) have demonstrated that adherence scales designed to assess treatment integrity and differentiation have achieved adequate reliability. Although the use of adherence and competence measures remains less frequent than is desirable, Waltz et al. (1993) note that a range of instruments has been developed. The simplest monitoring of fidelity which focuses only on therapist adherence involves the use of a presence/absence check in conjunction with a list of prescribed and proscribed treatment techniques. A more

complex instrument would rate the frequency or extensiveness of the intervention rather than simply its presence or absence. An example of the use of such an instrument can be found in the Luborsky, Woody, McLellan, O'Brien, & Rosenzweig (1982) study of three theoretically distinct therapies. Judges for the study used a 5-point continuous scale to rate the frequency of the intervention.

The well-known NIMH Treatment of Depression Collaborative Study (Elkin et al., 1989) generated a scale entitled the Collaborative Study Psychotherapy Rating Scale (CSPRS, Hollon et al., 1984; Hill, O'Grady, & Elkin, 1992) designed primarily to test adherence, although some items appear to address aspects of competence as well (Waltz et al., 1993). A more recent attempt at monitoring treatment fidelity combines an adapted version of the CSPRS with therapist session checklists for the treatment of cocaine abusers (Carroll, Rounsaville, & Nich, 1995). Another recent effort to measure adherence is reflected in the work of Butler, Henry, & Strupp (1995) whose Vanderbilt Therapeutic Strategies Scale (VTSS) evaluates therapist adherence to the tenets of Time-Limited Dynamic Psychotherapy (TLDP). Interestingly, these researchers assessed therapist competence using the traditional route of having supervisors rate the quality of therapists' performance.

There are, then, separate measures of competence

such as the Therapist Strategy Rating Form (TSRF) used in the Depression Collaborative Study designed to assess the quality of therapist interventions in interpersonal therapy (IPT, Klerman et al., 1984) and the Cognitive Therapy Scale (CTS) (Dobson, Shaw, & Vallis, 1985) which was designed to assess the skillfulness of therapists in cognitive therapy. Other measures which address issues of competence are reflected in the work of Silberschatz et al., 1986; Suh, O'Malley, & Strupp, 1986; and Svartberg, 1989.

Finally, a very recent effort to assess treatment fidelity focuses on the development of the Penn Adherence/Competence Scale for Supportive-Expressive Dynamic Psychotherapy (PACS-SE, Barber & Crits-Christoph, 1996). The 45 items in the scale are rated on both adherence and competence criteria. The authors note the value of the development of the PACS-SE as it is one of the few scales developed which facilitates the measurement of both adherence and competence. The PACS-SE scale may come closest to applying the recommendations suggested by Waltz et al. (1993) in their comprehensive description of guidelines for the development of treatment fidelity measures.

Current Difficulties with Treatment Manipulation Checks

Waltz et al. (1993) noted in their discussion of the current assessment of fidelity that "no widely accepted

methodology exists for this type of inquiry", which has led to an investment of efforts in a "reinvention of the wheel" (p. 620). They note the limitations of the methods currently in use. A key limitation associated with adherence assessment is the lack of the assessment of competence. When adherence checks are implemented without an assessment of the quality of the intervention, we still "do not know whether a fair test of the treatment has occurred" (p. 620). In addition, the inclination to equate adherence to protocol with successful treatment manipulation has resulted in the frequent use of adherence checks as a principle means of assuring treatment fidelity (Waltz et al., 1993). Ratings of therapist competence (Barber & Crits-Christoph, 1996; Shaw & Dobson, 1988) are noted to be an improvement over exclusive adherence checks, as competence implies adherence, while adherence may not necessarily presuppose competence (Waltz et al., 1993). Finally, criticisms of measures which assess competence consider the failure to incorporate an evaluation of the therapeutic context (Waltz et al., 1993). In general, it appears challenging to assess competence competently.

Current Recommendations Regarding the Assessment of Treatment Fidelity

In their review of the assessment of treatment fidelity, Waltz et al. (1993) suggest recommendations

regarding the development and use of adherence and competence measures. A general guideline is for researchers to tailor their selection of a specific manipulation check to the research problem in question. Specific recommendations include the following four: (1) When all that is needed is a simple test of whether two treatments are different, a "present or absent" checklist may suffice. Waltz et al. (1993) suggest that this method is economical and that interrater reliability is noted to be generally higher than can be observed from a continuous rating scale. Further, they observe that a present/absent adherence check also guards against the tendency to equate increased frequency of intervention with greater adherence, which may well not be the case. (2) Adherence measures should include four types of items: Interventions which are unique and essential to a treatment protocol; interventions which are necessary to that protocol but may be found in other treatments as well; interventions which are acceptable to that specific protocol, but not essential or unique to it; and interventions which are prohibited by that protocol. (3) To evaluate whether a treatment has been administered as intended, the assessment of both therapist adherence and competence must occur, and (4) Waltz et al. (1993) urge the incorporation of the therapeutic context in any evaluation of competence. They refer to Kiesler's (1973) discussion of the "uniformity myth" which points to the tendency to view

therapist interventions as "uniformly" competent across stage of therapy regardless of the therapeutic context. Considerations including client difficulty (e.g. hostile/angry client); the stage of the therapy; and the nature of the client's problems are all areas to reflect upon when making an assessment of the therapeutic context. They state that, "it seems clear that competency within a particular treatment modality is always linked to context" (p. 623).

(Additional recommendations regarding the nature of competence measures are suggested by Waltz et al. (1993) but will not be presented here since the current project's measure of treatment fidelity is primarily one of adherence.)

* * * * *

The above section concludes the discussion on treatment fidelity as it relates directly to the primary focus of this dissertation, which is the test of a measure designed to assess therapist adherence to treatment protocol. The discussion to follow provides the context for addressing a second and related question of this dissertation: What is the nature of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome? Areas to be discussed include treatment fidelity as it relates to outcome,

therapeutic alliance as it relates to outcome, and a possible relationship between therapist adherence and therapeutic alliance, and treatment outcome.

Relationship between Treatment Fidelity and Outcome

A preliminary review of the literature addressing the relationship between treatment fidelity and treatment outcome indicates that research thus far has been limited.

In a study of the efficacy of two distinct manual-based psychotherapies and a drug counseling treatment, Luborsky, McLellan, Woody, O'Brien, & Auerbach (1985) found that the purity of the two manual-based treatments "provided significant correlations with outcomes" (p.602). Luborsky et al. (1985) note that one possible inference from this finding may be that the purer the technique, the more effective the therapy. The researchers caution, however, that another interpretation might be that when the therapeutic alliance is established at a specific level (as it was in these two treatments), the therapist is able to adhere more closely to treatment protocol.

In their review of findings from manual-based psychotherapy training programs in short-term interpersonal psychotherapy (IPT) for depression, Rounsaville, O'Malley, Foley, & Weissman (1988) indicate that the evidence suggests that therapist adherence/competence is related to increased therapist efficacy (Rounsaville, O'Malley, Foley, &

Weissman, 1988). Although therapist efficacy is not necessarily synonymous with treatment outcome, there is an established link (O'Malley et al., 1988) making the relation between adherence/competence and therapist efficacy relevant to increased patient improvement or positive outcome.

In their use of the Collaborative Study Psychotherapy Rating Scale (CSPRS) for the purpose of measuring therapist adherence to three forms of treatment, Hill, O'Grady, & Elkin (1992) note that an important issue in the study of treatment fidelity is "whether adherence makes a difference" (p. 78). They also identify the need for future studies to address the interesting comparative treatment question of whether purity vs. eclecticism results in greater positive outcome.

Relationship between Therapeutic Alliance and Treatment Outcome

One of the most frequently cited works on the relationship between the therapeutic alliance and treatment outcome is the meta-analysis of 24 studies performed by Horvath & Symonds (1991). Horvath & Symonds (1991) note some general agreement on the critical features of the therapeutic alliance: (1) that it emphasizes the collaboration of the client and therapist and (2) that it includes both the client's and the therapist's ability to establish a contract which fits the "breadth and depth of

the therapy" (p. 139). The results of their meta-analysis suggest that the therapeutic alliance is a fairly robust factor connecting therapeutic process to treatment outcome (Horvath & Symonds, 1991).

The Nature of the Joint Contribution of Therapist Adherence and Therapeutic Alliance to Treatment Outcome: An Alternative Path to the Assessment of Therapist Competence

As was discussed above, there have been a number of recommendations in the psychotherapy research literature in recent years which urge the assessment of both adherence and competence in the evaluation of treatment fidelity (e.g. Barber & Crits-Christoph, 1996; Hill, O'Grady & Elkin, 1992; Rounsaville, O'Malley, Foley, & Weissman, 1988; Waltz et al., 1993). Adherence has been perceived as the simpler construct to measure in that it requires a "more objective" assessment of the presence, absence, or frequency of a therapist behavior. As Butler, Henry, & Strupp (1995) note, "At first glance, measuring adherence appears to be a simple task of counting occasions when the therapist follows prescriptions set forth in the manual" (p. 63). In contrast, the assessment of competence, which lies in subjective terrain, is more complex and requires that aspects of the therapeutic context be considered (Waltz et al, 1993). It is also agreed that the measurement of competence, which requires clinically experienced raters, is more costly than the assessment of adherence, which does not

require clinically experienced raters (Waltz et al., 1993). For some research programs this is an important consideration and may preclude the type of competence assessment suggested above.

Since the assessment of competence is thought to be a comparatively more complex and costly task than the assessment of adherence, it would seem reasonable to investigate avenues other than the use of an explicit competence measure to provide at least an indirect assessment of competence. There is, for example, some indication that treatment fidelity is predictive of outcome (Luborsky et al., 1985; O'Malley et al., 1988) and considerable evidence that therapeutic alliance is predictive of outcome (Horvath & Symonds, 1991). It may be warranted to attempt to evaluate competence indirectly by employing these or other well-defined constructs which have been demonstrated to influence treatment outcome (Muran, personal communication, August 1995). In short, alternative methods by which to evaluate competence may be possible. If a less costly and less subjective assessment of competence is attainable, it might shed light on additional means designed to uphold treatment fidelity. It may also further an understanding of the nature of therapist competence and its relationship to treatment efficacy. This is the basis for the present study's rationale for investigating the relationship between therapist adherence and therapeutic

alliance, and treatment outcome.

The NIMH-funded Beth Israel Brief Psychotherapy Research Study: Context for the Present Study

The following paragraphs provide some background and context for the present study on treatment fidelity and the development of an alternative measure of competence. The dissertation research was conducted under the auspices of the Beth Israel Brief Psychotherapy Research Project in New York City whose NIMH-funded pilot project on problems in the therapeutic relationship is currently on-going. The following discussion draws on the proposal submitted to NIMH by investigators Safran, Muran, and Winston (1992) for the purpose of studying the resolution of problems in the therapeutic alliance.

Safran, Muran, & Winston (1992) note that although there is evidence to show that psychodynamic and cognitive-behavioral treatments produce equivalent outcomes (Smith & Glass, 1977), the focus on equivalency obscures the fact that approximately 30% of patients do not demonstrate improvement in response to either treatment. The investigators note two other key findings: (1) that there is evidence that the quality of the therapeutic relationship is to date the best predictor of any psychotherapeutic outcome (Horvath & Symonds, 1991), and (2) that the non-improving 30% can be characterized by repeated negative

interactions between therapist and patient (Henry, Schacht & Strupp, 1986; 1990).

The Beth Israel Psychotherapy Research Project NIMH study seeks to accomplish the following: (1) To identify potential treatment failures by assessing the quality of the therapeutic relationship and (2) to evaluate the effectiveness of an interpersonal-experiential psychotherapy which specifically addresses problems in the therapeutic alliance.

Three modalities of psychotherapy for the treatment of personality disorders are the focus of the Beth Israel study (and also the focus of the present study's evaluation of an adherence measure and an alternative measure of competence.) The three therapeutic approaches are: Brief Adaptive Psychotherapy (BAP), (a dynamically oriented psychotherapy which links early relational patterns to current functioning); Cognitive-Behavioral Therapy (CBT); and Interpersonal-Experiential Therapy (IET), (the treatment noted above which emphasizes the quality of the relationship between therapist and patient in the Here-and-Now.)

Patients selected for the Beth Israel study received either cognitive-behavioral or psychodynamic short-term treatments through the Beth Israel Psychotherapy Research Project. Patients targeted as potential treatment failures early in their therapy were given the option of being randomly re-assigned to either interpersonal-experiential

therapy (IET) or one of the treatments they were not currently receiving (either CBT or BAP). Potential treatment failures were identified following the 10th session by consulting both patient and therapist versions of the Brief Psychotherapy Research Project Post-Session Questionnaire, which provides information on problems or tension in the therapeutic relationship. The PSQ included a version of the Working Alliance Inventory (WAI: Horvath & Greenberg, 1989; WAI-12: Tracey & Kokotovic, 1989) and the Session Evaluation Questionnaire (SEQ: Stiles, 1980).

Safran, Muran, and Winston (1992) state that "the primary goal of the (Beth Israel) study is to evaluate the hypothesis that patients reassigned to interpersonal-experiential therapy will show greater improvement than those reassigned to the other treatments" (p. 2). Additional goals of the Beth Israel study have included the development of a treatment manual for IET and the assessment of the reliability and validity of a refined version of the current Beth Israel Adherence Scale (Santangelo 1995; Turner & Muran, 1991; Winston et al., 1987). As has been noted, the latter is the primary focus of this dissertation. A final goal of the Beth Israel study is to extend research on an interpersonal-transactional model for understanding the processes involved in the resolution of ruptures in the therapeutic alliance. (A "rupture" may be defined as "an impairment or fluctuation in the quality of the alliance

between the therapist and client") (Safran, Crocker, McMain, & Murray, 1990, p. 154).

Evidence in psychotherapy research points to the importance of the therapeutic alliance as it relates to positive outcome as well as the value of modifying problematic therapist-patient interactions to support the alliance and guard against potential treatment failures (Safran, Muran, & Winston, 1992). The researchers note the value of considering contextual variables in the pursuit of treatment differences (Beutler, 1991; Kiesler, 1973). Making the specific intervention relevant to a non-static, contextual patient variable (i.e. not age or disorder) increases the possibility of identifying treatment differences (Safran, Muran, & Winston, 1992, p. 5). That is to say, administering IET's specific set of techniques to patients prone to problems in the alliance is more likely to lead to meaningful differences in treatment efficacy.

In a description of their rupture resolution model, Safran & Muran (1995) note that impasses in the therapeutic alliance "are viewed as inevitable aspects of the therapeutic process" (p. 85). In contrast to other well-known experiential approaches (e.g. Watson & Greenberg, 1995), the perspective outlined by Safran & Muran (1995) makes the handling of the impasse in the therapeutic relationship a "central focus for therapeutic exploration" (p. 85). A core interpersonal focus characterizes the

foundation of the approach. A primary perspective of the model is that difficulties in the therapeutic alliance reflect contributions made by both client and therapist. Safran & Muran (1995) note that the opportunity to "work through" a therapeutic impasse presents the client with the chance to develop a modified interpersonal schema in which the self is reflected as capable of relatedness.

Recent and current research at the Beth Israel Psychotherapy Research Project on the rupture resolution process identifies two elements critical to its perspective: (1) the principle of metacommunication and (2) a stage-process model of rupture resolution in the alliance. To metacommunicate is to communicate about the current interaction between client and therapist and to make this the primary focus of the therapeutic exploration in that moment. The stage-process model offers a "schematic representation of the different therapeutic stages that commonly emerge when alliance ruptures are resolved" (p. 86). This dual focus is central to the Beth Israel study on the resolution of ruptures in the therapeutic alliance (Safran & Muran, 1995).

The present study fits into the larger context of the Beth Israel study in terms of its goal of insuring that patients received one of the three forms of treatment outlined above (BAP, CBT, or IET). In order to demonstrate that patients received the intended treatment, it was

necessary to evaluate therapists for their adherence to treatment protocol. A preliminary version of a three-modality adherence scale for the Beth Israel project was demonstrated to have adequate levels of interrater reliability, internal consistency, and discriminant validity (Santangelo, 1995). The primary goal of this dissertation within the context of the larger project was to determine the interrater reliability and discriminant validity of a shortened and refined version of the current Beth Israel Adherence Scale. Due to constraints related to the measurement of competence, it was not possible to include a competence measure at this time. As has been noted, however, it may be possible to make some preliminary remarks about alternative methods of assessing therapist competence. This would be accomplished by examining the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome.

Theoretical Background of the Beth Israel Adherence Scale

A discussion of the theoretical basis for each of the therapy subscales in the Beth Israel Study is the focus of this section. The Beth Israel Adherence Scale (see Appendix A) is a 44 item measure which is reflective of each of the three modalities, providing 12 items to reflect each of the treatments, for a total of 36 specific therapist interventions. There are also 8 additional items (bringing

the total to 44 items) which are reflective of aspects which may be considered to cut across distinct theoretical orientations (i.e. "common factors", Frank & Frank, 1991). The "specific" items are randomly mixed throughout the scale and the "common factors" items are distributed evenly throughout. An explanation of the perspectives and procedures guiding the development and refinement of the current adherence scale is provided in subsequent sections.

The Interpersonal-Experiential Therapy (IET) subscale is composed of 12 items and is based on the work of Safran & Segal (1990), Greenberg & Goldman (1988), and Santangelo (1995). The 12 items reflective of Interpersonal-Experiential Therapy were: (1) explores the "how" or mechanism of a client's defense, (2) directs client's attention in non-confrontational manner to specific client behaviors, subtle non-verbal communications or paralinguistics, to increase client's awareness, (3) facilitates individuation and/or self-assertion, (4) directs or redirects the focus to the "here-and-now" (either with regard to the client's experience or with regard to the relationship between the client and therapist), (5) intervenes with skillful tentativeness, (6) tracks client's experience in a moment-to-moment fashion, (7) engages in empathic conjecture, (8) asks exploratory questions which probe for the feeling/experience underlying the client's utterance, (9) respects client as arbiter of experience,

(10) deepens client's awareness/experience through in or out-of session awareness exercise, (11) deepens client's experience through evocative reflection, and (12) metacommunicates by conveying own feelings to help client become aware of his/her role in the interaction or to probe for client's internal experience.

Interpersonal-Experiential Therapy (IET) as it is defined in the Beth Israel Project is informed by a number of traditions (Safran & Muran, 1995). In part, IET is comprised of features which are reflective of the interpersonal tradition (Kiesler, 1986; Levenson, 1991; Strupp & Binder, 1984; and Sullivan, 1953.) Experiential traditions (e.g. Greenberg, Rice, and Elliot, 1994; Perls, 1969; and Rogers, 1951) also contribute to the spirit of this approach. Finally, the IET model has been informed by the relational approach as discussed by Greenberg & Mitchell (1984) which has its origins in psychoanalytic theory, and by the social-constructivist tradition, as described by Hoffman, 1991. Safran & Muran (1995) note that defining aspects of this model include an emphasis on a (1) "two-person psychology" (p. 29) which focuses on the value of therapist/patient joint exploration of their contributions to the relationship, (2) the belief that patients are arbiters of their own experience, (3) the therapist's use of self-disclosure and metacommunication to enhance collaborative exploration, and (4) emotional immediacy

achieved by using phenomenological ("here and now") therapist interventions to explore the "particularities of the patient-therapist relationship" (p. 29), without the therapist's placing premature emphasis on the genetic past. A key principle in this therapy includes the therapist's use of metacommunication (as noted above) and an emphasis on mindfulness in the therapeutic relationship. In contrast to other traditional treatments (e.g. psychodynamic), the client is considered the "expert" on his or her own experience, and the therapist tentatively explores the interpersonal interactions with a focus on the client's immediate emotional experiencing (Safran & Segal, 1990).

The Brief Adaptive Psychotherapy (BAP) subscale is composed of 12 items which are based on a short-term dynamic psychotherapy for the treatment of personality disorders, developed at Beth Israel Medical Center by Pollack, Flegenheimer, Kaufman, & Sadow (1990). The 12 items reflective of Brief Adaptive Psychotherapy were: (1) Interprets other aspects of client's behavior or experience (not captured by more specific interpretation items), (2) reflects the content of client's statement, (3) frames symptoms in a relationship context, (4) links resistance (to the therapeutic process) to the maladaptive pattern, (5) confronts client, suggesting that he/she is saying, feeling, or thinking something different than what the client claims, (6) interprets/explores maladaptive patterns by linking

dynamics with parental/significant figures in the past to others in the present (not including therapist), (7) interprets and/or explores client's resistance or defenses, (8) explores and elucidates the unconscious aspects of major maladaptive patterns, thoughts, and behaviors, (9) frames symptoms as coping attempts (10) defines/identifies/specifies the maladaptive pattern, (11) interprets/explores maladaptive patterns by linking dynamics with others (past and present) to current dynamics with the therapist, and (12) interprets/explores maladaptive patterns by linking components of a conflict.

The BAP subscale has its theoretical origins in traditional psychoanalytic principles (e.g. Greenson, 1967). In addition, BAP employs a number of the traditional short-term psychotherapy techniques associated with Mann (1973) and Malan (1976) among others (as cited in Winston et al., 1991). It is an ego psychological approach (Freud, A., 1966; Hartmann, 1964) which differs from classical psychoanalysis in its short-term focus and in its use of specific techniques (as cited in Pollack et al., 1990). For example, where classical psychoanalysis employs free association, the BAP therapist "maintains an active focus" (p. 2). Where psychoanalysis emphasizes the development of the transference neurosis, the BAP therapist handles transference issues as they arise. It is a generally active and confrontational brief treatment which is "based on a

psychoanalytic understanding of character, character analysis, and of conflict and defense" (p. 2). The authors define character as reflecting adaptive or maladaptive patterns of beliefs and behavior.

Throughout the treatment, the BAP therapist identifies the "expectations, distortions, and behaviors" exemplified by the major maladaptive pattern. The transference is first explored in the context of the patient-therapist relationship, and then connected to other individuals in the patient's life. The authors note that it is this "continual working through of the pattern in the transference and in past and present relationships that leads to the cognitive and affective understanding that makes change possible" (Pollack et al., 1990, abstract).

The Cognitive-Behavioral Therapy (CBT) subscale is composed of 12 items which are based on a short-term, cognitive-behavioral treatment for personality disorders as described by Turner & Muran (1991). All of the items in this subscale were derived from the work of Beck and his colleagues (e.g. Beck et al., 1979) and from the Collaborative Study Psychotherapy Rating Scale (CSPRS, Hollon et al., 1984) used in the NIMH Treatment of Depression Collaborative Study (Elkin et al., 1989). The 12 items reflective of Cognitive-Behavioral Therapy were: (1) Assigns and reviews homework, (2) encourages client to distance him/herself from his/her thoughts, viewing them as

beliefs rather than facts, (3) probes for client's underlying beliefs or personal meaning behind client's thoughts, (4) explores the advantages and disadvantages of dysfunctional attitudes, (5) helps client identify cognitive distortions or errors that are present in his/her thinking, (6) facilitates client's consideration of alternative explanations for events, (7) asks client to report specific thoughts, (8) engages in socratic questioning aimed at guiding client's reasoning process, (9) engages in didactic persuasion, (10) helps client examine currently available evidence or information to test the validity as well as realistic consequences of the client's beliefs, (11) therapist and client practice rational responses to client's negative thoughts and beliefs, and (12) therapist works with client to plan or practice alternative overt behaviors for the client to use both inside and outside of therapy.

The CBT subscale has its theoretical origins in an integration of cognitive theory (e.g. Beck, 1976; Muran, 1991) and the conceptualization of the "self-schema" (e.g. Beck, 1967; Markus, 1990, as cited in Turner & Muran, 1991). Turner & Muran (1991) draw upon Guidano & Liotti (1983); Mahoney (1991); Muran (1991); and Safran & Greenberg (1986), when they discuss the roles of cognition and emotion in cognitive-behavioral therapy. They write, "...cognition is a manifestation of emotion concomitant with affective and motoric expressions. In other words, the experience of

emotion is a wholistic integration of cognitive, affective, and motoric components that are schematically structured" (Turner & Muran, 1991, p. 7).

Turner & Muran (1991) note that maladaptive self-schemas which are manifested cognitively may become representations of emotional disturbance. Maladaptive self-schemas are linked to information processing distortions. Beck (1976) refers to the products of these cognitive distortions as "automatic thoughts". Automatic thoughts are viewed as developing out of rigid belief systems or dysfunctional attitudes, which reflect emotionally laden knowledge about the self contained in the self-schema. Turner & Muran (1991) note that, "Since self-schemas are posited to be emotional in structure, it is integral that assessment be conducted in the context of the relevant emotion" (p. 12). Therapeutic change becomes possible within an interpersonal context in which there is an emphasis on the exploration and challenge of the emotion-laden, dysfunctional attitudes contained in the patient's self-schema (Turner & Muran, 1991).

The additional eight items which complete the 44 item Beth Israel Adherence Scale are examples of "non-specific" or "common factors" items. All eight items were derived directly from the CSPRS (Hollon et al., 1984) and were the following: (1) therapist's communication style (dull...interesting), (2) therapist conveys competence, (3)

therapist's involvement in the session, (4) therapist's warmth, (5) rapport between therapist and client, (6) therapist's receptive silence, (7) therapist's supportive encouragement, and (8) therapist works collaboratively with client to set and follow agenda.

The "non-specifics" described above have their origins in the research on common factors. Hollon et al. (1984) refer to these particular items as reflective of "facilitative conditions" and "explicit directiveness" (p. 7). They note the value of including these constructs (e.g. empathy, warmth, supportive encouragement, agenda setting) which have been "traditionally believed to be important in describing psychotherapies" (p. 7). Further, Hollon et al. (1984) note that their effort "to ensure that we are not simply picking up variation in non-specific factors represents one index of construct validity for the specific scales" (p. 7). Although there are conflicting perspectives on the usefulness of the specific/non-specific factors dichotomy, the inclusion of these "non-specific" items was believed to be warranted given the continuing research emphasis in this area.

STATEMENT OF THE PROBLEM

A reliable and valid, refined version of the Turner & Muran (1991), Pollack et al. (1990), and Santangelo (1995) Beth Israel Adherence Scale did not exist at the onset of

this study. Following an evaluation of the preliminary version of the Beth Israel Adherence Scale (Santangelo, 1995), specific items from each of the therapeutic modalities were identified as both less reliable and less representative of each of the treatments. These were removed from the scale. In addition, the number of items in each of the subscales was equalized, which was not the case in the preliminary version. In accordance with research recommendations (Waltz et al., 1993), eight "non-specific" items were introduced. It was hoped that the revised version of the Beth Israel Adherence Scale would bring the necessary improvements to the current scale and provide a reliable and valid instrument to be employed for the study of the comparative efficacy of interpersonal-experiential, cognitive-behavioral, and brief-adaptive therapy.

A secondary problem addressed by this study was the question of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome. It was hoped that research on this question would further an understanding of alternative methods by which to assess therapist competence.

HYPOTHESES

The proposed research hypotheses were as follows: (1) that the revised version of the Beth Israel Adherence Scale would be established as reliable between raters (2) that scale criteria would validly discriminate between the

three treatments and (3) that the joint contribution of therapist adherence and therapeutic alliance would be positively predictive of treatment outcome.

METHOD

Audiotapes

In past years, therapies in the Beth Israel Psychotherapy Research Project were 40 sessions in length, but since 1992 have been reduced to 30 sessions in keeping with current research trends. Research indicates that there is no significant difference between the two treatment lengths with regard to therapeutic change. Studies have shown that if change is to occur, it may generally be observed by the 26th session (Howard, Kopta, Krause, & Orlinsky, 1986).

Both treatments of 30 sessions and 40 sessions were sampled for this project since this made available a wider pool of therapies and therapists from which to select. Seventeen 30 session treatments and seven 40 session treatments comprised the sample of 24 treatments. The total of 24 treatments reflected an equal division of three, with 8 cases in each of the three treatment conditions (i.e. 8 BAP, 8 CBT, and 8 IET cases.) Three sessions from each of the BAP, CBT, and IET treatments were randomly selected from across the three stages of therapy (early, middle, and late), resulting in one session from each of the three

stages of therapy, for each of the 24 dyads. This equal balancing resulted in 72 sessions in total.

Fifteen minutes of each session was identified as the quantity of material to be assessed. Each of the seventy-two 15 minute videotaped segments was randomly assigned to one of three session conditions i.e, either the first 15 minutes (early), the middle 15 minutes (middle), or the last 15 minutes (late) of that session. Videotaped segments were transposed to audiotape to reduce the possibility of rater bias which could occur as a result of the rater's visually identifying the therapist. In sum, the complete data set was composed of seventy-two 15 minute audiotaped segments with 24 segments from each of the three modalities having been randomly selected from early, middle, and late stages of treatment.

In the 30 session therapies, the early stage was composed of sessions 3-11, middle, 12-19, and late, 20-28. In the 40 session treatments, the early stage was composed of sessions 3-14, middle, 15-26, and late, 27-38.

Therapists

Therapists for this project were members of the Department of Psychiatry at the Beth Israel Medical Center during the administration of their treatments. Therapists included psychiatrists, psychologists and social workers. They were either individually supervised by an expert in

their treatment modality or were themselves an expert in that particular treatment. Therapists in supervision regularly presented and discussed their on-going cases in a weekly clinical training seminar. The use of videotaped therapy sessions for the purposes of teaching and training was an integral component of the supervisory process. Treatment manuals for each of three treatment conditions were used as guides throughout the supervision and training process. All of the therapists in this project were selected for participation based on data gathered from adherence checks and feedback from supervisors and from the Beth Israel Psychotherapy Research Project director (Safran, Muran, & Winston, 1992). Therapists known to not be adequately adhering to treatment protocol were not included in the study. No training cases were used. Four therapists from each modality were selected to participate in the present study, with each therapist having one, two, or three cases, making 8 the total number of cases in each modality (as noted above).

Patients

Twenty-four patients from the Beth Israel Research Project participated in this study. 17 or 75% were women. Patients were between the ages of 18 and 65 and could identify at least one close personal relationship. Any person describing evidence of a significant medical

diagnosis (Axis III), the use of psychotropic medications, a history of suicidal or violent behavior, or on-going substance abuse, was excluded from the NIMH study, and thus from the present study.

The Structured Clinical Interview for the DSM-III-R (SCID: Spitzer, Williams, & Gibbon, 1987; Diagnostic and Statistical Manual, 3rd ed., revised) served as the screening and diagnostic tool for the formulation of patient diagnoses. 13 of the 24 (or approximately 1/2) had Axis I diagnoses of a depressive disorder, primarily dysthymia. 8 of the 24 (or 1/3) had Axis I diagnoses of anxiety disorders, primarily generalized anxiety disorder or social phobia. Two patients had a diagnosis of interpersonal problems on Axis I and one patient had no diagnosis on Axis I. All patients for this sample, except one, had a diagnosis of either a Cluster C personality disorder or personality disorder not otherwise specified with Cluster C features. One patient had a diagnosis of a Cluster A personality disorder.

Raters

Three raters were trained in all three treatment modalities and all rated the same 72 treatment segments. Two of the raters were advanced doctoral students in clinical psychology and one of the raters was at the Master's level in psychology. Two of the raters were women

in their early 30s and one of the raters was a man in his late 20s. The two female raters had an on-going affiliation with the Beth Israel Psychotherapy Research Project as research assistants, and the male rater had a past, brief affiliation with the Project.

Training of Raters

Two of the raters were trained by this author over a period of approximately 8 months with regular two hour meetings held on an average of every 2-3 weeks. The third rater, who entered the study subsequently, was trained by one of the above raters, who had assumed the responsibility from this author of maintaining adherence checks for the Beth Israel Project. The trainer who replaced this author was trained by this author, and subsequently briefly guided by this author in her new training position.

Both training groups were conducted along parallel lines. Each rater was guided by an adherence manual for the three modalities. The adherence manual contained explications and illustrations of each of its respective items. The viewing of the Psychotherapy Research Project's treatment videotapes was an integral component of the training. A training session consisted of the viewing of a previously selected videotape, which was juxtaposed with on-going discussion about the rating criteria and the nature of the items. The use of the adherence manual contributed to

the effort to achieve reliability among raters since the raters worked with the same manual.

Practice assignments were given from session to session, with a review and discussion of ratings accompanied by the viewing of the relevant videotape. Instruction on the definitions of all of the items was provided by both of the trainers. In this way, clarification about the parameters of specific items and consensus on the matching of items with interventions was achieved.

During the data collection period, which was scheduled to occur over a period of approximately four months, four monthly meetings were held to prevent rater drift. Specific anchors for each of the items were reviewed and clarified. The reviewing of selected audiotaped segments for the purposes of clarifying aspects of an item were interspersed throughout the meetings. The author summarized in written form the clarifications which had been discussed at each of the meetings. These were then distributed to each of the raters following each meeting so that raters could review the clarifications on an on-going basis as the data collection continued.

Development of the Scale:

Criteria for Rating of Items

Instructions to raters were to consider any therapist utterance in each of the 72 segments on two

dimensions: frequency and clarity. Frequency was considered an important dimension to assess since it reflects one method for establishing the presence of a particular intervention (Waltz et al., 1993). Since one of the defining aspects of treatment fidelity and treatment differentiation is whether interventions unique to a treatment modality are reflected in the administration of that modality, the measurement of frequency would seem an important dimension to assess. The assessment of clarity was introduced since it allowed for the evaluation of an aspect of the quality of an intervention without moving too far into the subjective realm of competence. Frequency was defined as the number of times an intervention occurred, while clarity was defined as the ease with which an intervention could be understood and recognized as a particular item. An intervention which was rated high on frequency occurred a number of times in a 15-minute segment, while an intervention rated high on clarity was a well-formed, easily recognized intervention occurring in that 15-minute segment. A single number on the 6-point, Likert-type adherence scale reflected a collapsed frequency and clarity rating (see p. 43).

In addition to rating interventions based on frequency and clarity, Raters were instructed to consider a third factor: the item's categorization in terms of the following three subcategories: (1) "global" items, which

represent a defining aspect of the treatment and tend to occur frequently and pervasively, (2) "moderately-occurring" items, which can occur not-at-all to frequently, and (3) infrequent items, which may occur infrequently, but are substantive in nature. An example of a global item would be the IET item, "Tracks client's experience in a moment-to-moment fashion." Examples of moderately-occurring items would be the CBT item, "Helps client identify cognitive distortions and errors present in his/her thinking" and the BAP item, "Links resistance (to the therapeutic process) to the maladaptive pattern". Examples of "infrequent, but substantive items" would be the CBT item, "Assigns and reviews homework" and the IET item, "Deepens client's experience through in or out-of session awareness exercise." Since global items generally occurred frequently and pervasively, raters were instructed to "weight" them at least one point lighter than moderately occurring items or infrequently occurring, substantive items. Infrequent, but substantive items, were "weighted" approximately 2-3 points heavier, since they might occur infrequently, if at all, in any one segment. Both the global and infrequently-occurring items were clearly identified as such, with the remaining items falling in the broader and mid-range category of "moderately-occurring" items. The "Directions for Scoring" (see appendix C), which accompanied the scale containing the individual descriptions of the items, outlined the rating

process in some detail.

The rating process proceeded in the following step-by-step way: The rater would listen to a segment, and note the presence of a specific item following each therapist utterance (see below). The Likert-type scale primarily assessed degree of occurrence: 1=not at all, meaning that the item did not occur; 2=slightly occurring, meaning that the item was observed to occur at least once with some degree of clarity - (An exception for this guideline would be the scoring of an infrequently occurring item, which if observed once, would be scored approximately 2-3 points higher); 3=somewhat, meaning that the item occurred at least 2-3 times with reasonable clarity; 4=moderately, meaning that the item occurred fairly frequently and with moderately good clarity; 5=considerably, meaning that the item occurred with high frequency and good clarity; and 6=extensively, meaning that the item occurred with very high frequency and excellent clarity.

A notation of a check (✓) representing the presence of an item, combined with either a minus sign (✓-) for less than average clarity, no marker for average clarity, or a plus sign (✓+) for more than average clarity, was made in the margin for each item. Throughout the rating process, raters were instructed to review the scale itself (which reflected condensed versions of each of the items, see Appendix A); the more extensive individual descriptions of

the items contained in the rater's manual; and the scoring directions.

Upon completion of a segment, the rater would tally the number of checks for each item, factoring into a final rating score the clarity "markers" for each item, and the "weight" of that item. Thus, a score of four checks (✓✓✓✓) for a global item might be scored a "3", while a score of three checks for a global item, all of which had minus signs, might be scored a "2". A score of one check with a plus sign (✓+) for an infrequently-occurring, but substantive item might be scored a "4", since these items are naturally "heavier" items. A score of 4 checks, two with plus signs (✓+✓+), for a moderately-occurring item, might be scored either a "4" or a "5" depending on the judgment of the individual rater. The scoring directions acknowledged the inherent subjective and ambiguous nature of the rating process, and urged raters to be guided by the above directions without feeling compelled to rigidly adhere when their intuition directed them to do otherwise.

In sum, raters were instructed to consider the items from the perspective of the three subcategories described above (global, moderately occurring, and infrequently occurring) as they rated each item on frequency and clarity. Anchors for each item, which considered all of the above factors, were discussed in collaborative fashion with the raters during both training and rater drift meetings.

Rationale for Use of Rating Criteria

In their review of the assessment of treatment fidelity, Waltz et al. (1993) recommended that researchers tailor their selection of a specific manipulation check to the research problem in question. They noted that a "present or absent" check list is a simple, economical test of therapist adherence which produces greater interrater reliability than can be observed from a Likert-type scale which measures degree of occurrence. They noted that the use of a present vs. absent manipulation check also guards against equating increased frequency of intervention with greater adherence, which may well not be the case.

The use of a Likert-type scale which measures a more differentiated degree of occurrence from 1 to 6 (the approach used for this study, see Appendix B), is an example of a more useful manipulation check. This is in contrast to the "present vs. absent" approach described above. A number of previous studies have selected the more differentiated method for the assessment of adherence (e.g. CSPRS: Hollon et al., 1984). Although it is less economical and leads to slightly lower interrater reliability than the present/absent check described above, it was the preferred choice for this study because its use provides a wider range of information regarding specific items as well as the option for refinement of the scale in the future. Since the adherence check is employed at least partly for training

purposes, it is particularly useful to have detailed information about a therapist's use of specific interventions to discuss with supervising clinicians. A supervisor may be given important details about the degree of occurrence and the quality of a therapist's intervention; a present/absent manipulation check would effectively rule out the above options.

It was useful to include a measure of "clarity" along with the measure of frequency, since adherence assessment is partly performed for the purpose of training therapists and improving their ability to deliver the treatment as intended. The assessment of clarity allowed raters to identify "clear" renditions of the items, without having to assess the degree of skillfulness (competence) reflected in the intervention, which requires that the overall therapeutic context be considered (Waltz et al., 1993). Again, it was believed that the introduction of a measure of "clarity" provided additional useful information, which would have been lost if "frequency" was the only criterion assessed.

Finally, this study followed the recommendation of Waltz et al. (1993) with regard to the importance of including four types of items in an adherence scale. They urged that an adherence measure include items which are unique and essential to a treatment protocol; interventions which are necessary to that protocol but may be found in

other treatments as well; interventions which are acceptable to that protocol, but not essential or unique to it; and interventions which are prohibited by that protocol. For example, the assignment of homework in which the client practices a specific assertive behavior outside the session would be considered unique to the CBT modality. It would also be considered "proscribed" or "off-task" by the BAP modality. The items unique to one of the three modalities are proscribed by the others, and thus meet the criteria for two of the recommendations outlined above (unique and proscribed items must be present). The remaining two recommendations are most closely met by the introduction of the eight "non-specific" items some of which are both essential (but not unique) and acceptable, but not necessary.

* * * * *

The above section concludes the discussion of methods used to address the primary goal of this study, which was the evaluation of the reliability and validity of the Beth Israel Adherence Scale. The sections below discuss the methods and measurement tools used to address a secondary question of this study, that of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome.

Measurement of the Therapeutic Alliance

The quality of the therapeutic alliance was assessed using a 12-item version of the Working Alliance Inventory (WAI: Horvath & Greenberg, 1989; WAI-12: Tracey & Kokotovic, 1989) a self-report, paper-and-pencil instrument designed to assess some of the "common" factors which influence the degree of success in therapy. The 12-item version of the WAI is embedded in the Beth Israel Post-session questionnaire (PSQ) which both patients and therapists completed following each session.

Horvath & Greenberg (1989) reviewed the results from three studies and noted that the data suggest that the WAI has demonstrated adequate reliability. They also noted that the data revealed preliminary support for the instrument's validity (Safran, Muran, & Winston, 1992).

The problem of missing WAI data was handled by selecting the PSQ directly above or below the missing PSQ (in numerical order), keeping the selection within the designated treatment stage (i.e early, middle, or late). Thus, if the PSQ for session 17 was missing, either PSQ session number 16 or 18 was selected in its place.

Measurement of Treatment Outcome

Treatment outcome was evaluated using data gathered from the Symptom Checklist - 90 (SCL-90: Derogatis, 1977; SCL-90-R: Derogatis, 1983) and from the Inventory of

Interpersonal Problems - 64 (IIP: Horowitz et al., 1988; IIP-64: Alden, Wiggins, & Pincus, 1990), both of which were administered at the onset of treatment and at treatment termination. The SCL-90-R is a self-report, paper-and-pencil inventory designed to identify and measure overall psychiatric symptoms. It is comprised of 90 items presented in Likert-type format and designed to assess degree of severity. Adequate psychometric properties have been demonstrated.

The IIP-64 is a self-report, paper-and-pencil instrument which measures the degree of interpersonal difficulty and social adjustment. It is a 64-item questionnaire set in Likert-type format, which reflects degree of distress. Adequate psychometric properties have been demonstrated (Safran, Muran, & Winston, 1992).

Due to missing outcome data, four patients were dropped from this portion of the study, bringing the sample size to 20 from the original 24. (Since outcome data was not relevant to the primary purpose of the study - the reliability and validation of the scale - the sample size remained at 24 for analyses pertinent to those questions.)

Data Analysis

Reliability between raters was assessed for significance using the Intraclass Correlation Coefficient (ICC: Shrout & Fleiss, 1979). The discriminant validity of

the adherence scale was determined by conducting a series of ANOVAs which compared the mean adherence subscale scores (dependent variable) from session samples of each of the three treatment conditions (independent variable). The hypothesis was that the adherence means for individual therapists' interventions would be reflective of the treatment which that therapist was administering. For example, in ratings of BAP treatment sessions, the adherence mean for the BAP subscale would be higher than the adherence means for the other two subscales.

A series of multiple regression analyses were conducted to evaluate the nature of the joint contribution of therapist adherence and therapeutic alliance to treatment outcome, with adherence and alliance as the independent variables and outcome as the dependent variable.

A series of zero-order and first-order correlational analyses were also performed to assess the unique and joint contributions of adherence and alliance to outcome. For the first order correlational analyses, specifically, the alliance measure was used as a covariate with the measure of adherence as the independent variable and the measure of overall treatment outcome as the dependent variable.

RESULTS

Hypothesis I:

Interrater reliability. Based on the 72 audiotaped 15-minute segments rated by three judges, analyses yielded adequate interrater reliability for a shortened and refined version of the Beth Israel Adherence Scale. These results confirm hypothesis I: that the Beth Israel Adherence Scale would be established as reliable between raters.

To assess interrater reliability, intraclass correlation coefficients (ICC(2,k), Shrout & Fleiss, 1979) were calculated, where "rater" was considered a random effect and "k" reflected the number of raters. For this study, ICC(2,3) reflected an estimate of how three raters would perform. The ICC coefficients for the overall scale ranged from .71 to .95, with a mean of .87 and a standard deviation of .05. This mean of .87 falls above Kraemer's (1981) standard of an ICC cut-off of .70. ICC means for the three subscales (BAP, CBT, and IET), averaged across stage of therapy, were also well above the Kraemer .70 level. ICC means observed by stage of therapy (early, middle, late), averaged across treatment type, were also above the .70 level (see Table 1). All ICC means were in the predicted direction.

Hypothesis II:

Discriminant validity. To test hypothesis II, which posited that the Beth Israel Adherence Scale could

Table 1. Interrater Reliability for the Beth Israel Adherence Scale: ICC Means by Overall Scale, Subscale, and Stage of Therapy

Variable (ICC)	Mean	Standard Deviation	Minimum	Maximum	N
Overall Scale	.87	.05	.71	.95	72
BAP	.85	.06	.71	.93	24
CBT	.87	.05	.76	.95	24
IET	.89	.03	.80	.94	24
Early	.87	.06	.71	.95	28
Middle	.87	.04	.75	.93	22
Late	.87	.04	.78	.91	22

Note. ICC: Intraclass correlation. BAP: brief adaptive therapy subscale; CBT: cognitive-behavioral therapy subscale; IET: interpersonal-experiential therapy subscale. ICC mean for overall scale was averaged across stage of therapy and subscale. ICC subscale means were averaged across stage of therapy. ICC means by stage of therapy were averaged across subscale.

distinguish between the three therapies, three one-way analyses of variance were conducted which compared adherence subscale scores from each of the three treatment conditions. These analyses confirmed that individual adherence means for each of the three subscales were greatest for the treatment that subscale was designed to reflect. Thus, adherence scores for each treatment-specific subscale distinguished between the three modalities.

Specifically, a significant difference between groups on the BAP subscale was found ($F(2,71) = 67.25, p < .001$) (see Table 2). Least Significant Difference post hoc tests revealed that the BAP therapists were rated higher on the BAP subscale ($M = 1.84, sd = .27$) than were IET therapists ($M = 1.30, sd = .14$) and CBT therapists ($M = 1.20, sd = .19$). The IET and CBT therapists did not differ significantly from each other.

Comparable findings were observed for the CBT subscale, for which a significant difference between groups was also noted ($F(2,71) = 43.38, p < .001$) (see Table 2). Least Significant Difference post hoc tests revealed that the CBT therapists were rated higher on the CBT subscale ($M = 1.61, sd = .33$) than were BAP therapists ($M = 1.17, sd = .11$) and IET therapists ($M = 1.11, sd = .07$). The BAP and IET groups did not differ significantly from each other.

As above, a significant difference between groups on the IET subscale was observed ($F(2,71) = 163.13, p < .001$)

(see Table 2). Least Significant Difference post hoc tests indicated that the IET therapists were rated higher on the IET subscale ($M = 2.22$, $sd = .30$) than were BAP therapists ($M = 1.28$, $sd = .16$) and CBT therapists ($M = 1.26$, $sd = .14$). The BAP and CBT groups did not differ significantly from each other.

Finally, a fourth one-way analysis of variance was conducted to assess whether there was a difference between adherence scores on the non-specific subscale items for the three treatments. As predicted, the analysis revealed that there were no significant differences between groups ($F(2,71) = 2.26$, $p = .11$) (see Table 2). Means were as follows: BAP therapists, $M = 3.51$; CBT therapists, $M = 3.52$; and IET therapists, $M = 3.67$. Thus, the non-specific items were rated as equally present in each of the three modalities.

Table 2. ANOVA Table Comparing BAP, CBT, IET, and CMFX Subscale Scores Across Treatment Type

<u>Source</u>	<u>D.F.</u>	<u>Sum of Squares</u>	<u>Mean Squares</u>	<u>F Ratio</u>	<u>F Prob.</u>
BAP					
Between groups	2	5.68	2.84	67.25	.00
Within groups	69	2.91	.04		
Total	71	8.59			
CBT					
Between groups	2	3.61	1.80	43.38	.00
Within groups	69	2.87	.04		
Total	71	6.48			
IET					
Between groups	2	14.57	7.28	163.13	.00
Within groups	69	3.08	.04		
Total	71	17.65			
CMFX					
Between groups	2	.40	.20	2.26	.11
Within groups	69	6.11	.09		
Total	71	6.51			

Note. CMFX: common factors subscale.

Hypothesis III:

To test hypothesis III, which posited that the unique and joint contributions of therapist adherence and therapeutic alliance would be positively predictive of treatment outcome, two sets of analyses were performed. The first series were multiple regression analyses. These were conducted to assess the joint contribution of therapist adherence and therapeutic alliance to treatment outcome, with an emphasis on the predictive capacity of the independent variables. The second series of analyses were correlational analyses, including both zero-order and first-order correlations, to assess the unique contributions of adherence and alliance to outcome.

For both the multiple regression and correlational analyses, outcome was assessed by taking the average of the two outcome measures for each case, the GSI (overall mean) on the SCL-90R (Derogatis, 1983) and the overall mean on the IIP-64 (Alden, Wiggins, & Pincus, 1990) (IIP-127: Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988) reflected in the Reliable Change (RC) score (Jacobson & Truax, 1991). A statistically significant reduction in patient symptomatology was identified by the RC index. An RC score for each patient was derived by taking the difference in intake and termination scores on both the SCL-90R and IIP-64, and then dividing by the standard error of the difference. An $RC \leq -1.96$ indicated statistically

significant patient improvement at the $< .05$ level of chance. Multiple regression analyses will be discussed first.

Evaluating the nature of the joint contribution of adherence and alliance to treatment outcome: multiple regression analyses

In order to assess the joint contribution of therapist adherence and therapeutic alliance to treatment outcome, a series of four multiple regression analyses was performed. None of these analyses yielded significant results. For the first multiple regression analysis, all segments by stage of therapy (early, middle, and late) for all cases were included. The dependent variable was the averaged RC score described above. The independent variables reflected overall adherence scores averaged across stage of therapy and alliance total scores averaged across stage of therapy. Average adherence was defined as the average mean score (across raters) on the relevant subscale for each segment. Thus, for example, in order to represent adherence to BAP within this overall score, only adherence scores for BAP segments were used. This applied to CBT and IET, as well. These "treatment-specific" adherence scores were then averaged across both treatment modality and stage of therapy to produce an overall adherence score. This score represented degree of overall adherence to any treatment modality, not treatment-specific adherence.

Adherence scores on non-relevant treatment scales were not used in this analysis since the focus was adherence to a specific treatment. (See correlational analyses for results relating to treatment specific adherence, i.e. adherence to BAP, CBT, or IET).

Average alliance was defined as the average of the WAI total scores across early, middle, and late segments. The results of the first analysis are presented in Table 3. Table 3 lists Multiple R, R square, and adjusted R square, the one-tailed correlations between variables (see zero-order correlation results in subsequent section for two-tailed tests), the unstandardized regression coefficients (B), and the standardized regression coefficients (Beta). R for regression was not significantly different from zero ($F(2,17) = .55$, signif $F = .59$) indicating that the IVs, therapist adherence and therapeutic alliance, could not be said to contribute significantly to prediction of treatment outcome in this analysis. Calculation of effect size, power for this analysis with $N = 20$, and the N required for 80% power, indicated that a larger N might yield significant results for the same analysis (See Table 7).

The three remaining multiple regression analyses examined "non-averaged" adherence to treatment modality by treatment thirds (i.e. within early, middle, and late segments). As before, adherence scores on non-relevant treatment scales were not entered in this analysis. As

noted above, none of the multiple regression analyses by stage of therapy yielded significant results. In all three analyses, the DV was the averaged RC score described above. The IVs were overall adherence and alliance by stage of therapy. Overall adherence for early stages was defined as the average mean score (across raters) on the relevant treatment-specific subscale for only early segments. Overall adherence for middle stages and overall adherence for late stages followed the same pattern. Alliance was defined as the WAI total score by stage of therapy. The results of these analyses are presented in Tables 4-6 and follow the same structure as Table 3 for the first multiple regression analysis. Tables 4, 5, and 6 list the one-tailed correlations between variables (see zero-order, two-tailed correlation results in subsequent section), Multiple R, R square, and adjusted R square, the unstandardized regression coefficients (B), and the standardized regression coefficients (Beta). Multiple R for regression was not significantly different from zero for any of the analyses (Early: $F(2,17) = .77$, signif $F = .48$; Middle: $F(2,16) = 1.85$, signif $F = .19$; Late: $F(2,16) = 2.31$, signif $F = .13$). As before, results indicate that the IVs, therapist adherence and therapeutic alliance, when considered by stage of therapy, cannot be said to contribute significantly to prediction of treatment outcome in these analyses. Effect size, power for the present analyses (with N (early) = 20; N

(middle) = 19; and N (late) = 19), and the Ns required for 80% power for each of these analyses, suggest that larger Ns may yield an increased likelihood of significant results. Multiple regression analyses employing variables associated with middle and late sessions, in particular, yielded medium-large effect sizes, suggesting that larger Ns may yield a strong likelihood of finding significant results for these analyses (See Table 7).

Table 3. Standard Multiple Regression of Therapist Adherence and Therapeutic Alliance Variables on Treatment Outcome: All Cases

Variables	Outcome (DV)	Adherence	Alliance	B	Beta
Adherence	.10			1.343	0.166
Alliance	-.19	.28		-0.563	-0.234
Means	-1.27	1.90	5.33		
Standard deviations	2.66	.33	1.11		
					$R^2 = .06$
					$R^2 = -.05$
					$R^2 = .25$

Note. Adherence was averaged across treatment-specific subscale and across stage of therapy. Alliance was averaged across stage of therapy.

Table 4. Standard Multiple Regression of Therapist Adherence and Therapeutic Alliance Variables on Outcome for Early Segments

Variables	Outcome (DV)	Adherence	Alliance	B	Beta
Adherence	.22			-1.623	0.262
Alliance	-.13	.21		-0.576	-0.185
Means	-1.27	2.02	5.29		
Standard deviations	2.67	.43	.86		
					$R^2 = .08$
					$R^2 = -.03$
					$R = .29$

Note. Adherence variable was averaged across treatment-specific subscale. Alliance variable reflected early segment WAI total scores.

Table 5. Standard Multiple Regression of Therapist Adherence and Therapeutic Alliance Variables on Outcome for Middle Segments

Variables	Outcome (DV)	Adherence	Alliance	B	Beta
Adherence	.39			1.867	0.345
Alliance	-.28	.25		0.275	0.191
Means	-1.59	1.86	5.35		
Standard deviations	2.30	.42	1.59		
					$R^2 = .19$
					$R^2 = -.09$
					$R^2 = .43$

Note. Adherence variable was averaged across treatment-specific subscale. Alliance variable reflected middle segment WAI total scores.

Table 6. Standard Multiple Regression of Therapist Adherence and Therapeutic Alliance Variables on Outcome for Late Segments

Variables	Outcome (DV)	Adherence	Alliance	B	Beta
Adherence	-.29			-1.614	-0.225
Alliance	-.42	.18		-0.801	-0.378
Means	-1.25	1.88	5.60		
Standard deviations	2.74	.38	1.29		
					$R^2 = .22$
					$R^2 = .13$
					$R^2 = .47$

Note. Adherence variable was averaged across treatment-specific subscale. Alliance variable reflected late segment WAI total scores.

Table 7. Multiple Regression Analyses: f^2 , Effect Size, Power, and N Required for 80% Power

Variable	f Square	Effect Size	Power	N Required for 80% Power
Across stages				
Outcome Adherence Alliance	.06	sm - med	12%	166
Early				
Outcome Adherence Alliance	.09	sm - med	<20%	196
Middle				
Outcome Adherence Alliance	.23	med - lg	39%	50
Late				
Outcome Adherence Alliance	.28	med - lg	46%	40

Note. Analyses reflect adherence as overall, non-treatment-specific variable. Adherence is averaged across subscale and across stage of therapy in first analysis; averaged across subscale and non-averaged (by stage of therapy) in analyses 2-4. Alliance reflects WAI total scores averaged across stage of therapy in first analysis and by stage of therapy in analyses 2-4. Outcome measured by RC index in all cases.

An assessment of the unique contributions of therapist adherence and therapeutic alliance to treatment outcome: correlational analyses

Four zero-order and eight first-order correlations were performed, 12 correlational analyses in total. The zero-order correlations were conducted to examine the relationships between all variables (therapist adherence as measured by subscale; therapeutic alliance; and treatment outcome). First-order correlational analyses were performed to examine the relationship between therapist adherence, controlling for the variance accounted for by therapeutic alliance, and treatment outcome. Although the total number of correlational analyses was high (12), it was considered too conservative to employ a Bonferroni type correction since the variables were all correlated to a greater or lesser degree, making the data non-orthogonal.

The first series of correlational analyses, which were zero-order correlations, examined the relationships between therapist adherence to treatment as measured by the four specific subscales; treatment outcome, as measured by the average Reliable Change (RC) score (described above), and therapeutic alliance, as measured by the WAI total scores as derived from patient responses to the WAI-12. As did the multiple regression analyses, the zero-order correlations assessed the association between variables both for all cases averaging across stage of therapy and for cases by stage of therapy (i.e. by early, middle, and late

segments).

For the first zero-order correlation, which examined the above relationships between all variables in all cases ($N = 20$), each of the variables (except outcome, which remained the same score across any single treatment) was averaged across early, middle, and late stages of therapy. Adherence variables reflecting average mean scores (across raters) for therapist adherence to treatment specific subscales were averaged across stage of therapy to produce a single treatment-specific adherence score for each of the treatment relevant cases. For example, the variable, average BAP adherence, was defined as the average mean score across raters for therapist adherence to the BAP subscale for each BAP case, averaged across stage of therapy. The variables average CBT adherence, average IET adherence, and average common factors adherence, were defined accordingly. The variable, average alliance, was defined as the average of the WAI total scores for each of the treatment-third segments. The outcome variable was defined (as above) as the average of the two outcome scores, the overall mean on the SCL-90R and the overall mean on the IIP-64 for each case, to produce a single outcome variable.

Thus, six variables in total were included in the first zero-order correlation. This analysis did not yield significant results. The correlation between average adherence to BAP and poor outcome, however, did reflect a

moderate-strong association although it was not significant ($r(18) = .40, p < .10$) (See Table 8). This relationship is in the opposite direction from the hypothesis III prediction that therapist adherence would be positively correlated with treatment outcome.

The remaining three zero-order correlation analyses were performed using only those variables associated with each treatment third. Each adherence variable was defined in the same manner as described above, except that variables were not averaged across stage of therapy. Rather, each adherence variable was reflective of the treatment specific subscale for only those treatment-relevant cases, for either early, middle, or late sessions. For example, the early BAP adherence variable was defined as the average mean score across raters for therapist adherence to the BAP subscale for each BAP case, for only early segments. The remaining adherence variables (CBT, IET, and CMFX) were defined in the same manner. The alliance variable was defined as the WAI total score for each third of treatment. The outcome variable remained the same as defined in previous analyses.

For early-stage segments, no zero-order correlations yielded significant results. The correlation between alliance and adherence to CBT reflected a strong association and effect size ($r(18) = -.41, p < .10$) (See Table 8) suggesting that adherence to CBT may be associated with poor alliance.

For middle-stage segments, the correlation between adherence to CBT and positive outcome also reflected a medium-to-large effect size (and approached significance) ($r(17) = -.40, p < .10$). This association, although not significant, is in the predicted direction. The correlation between adherence to BAP and increased symptomatology (poorer outcome) also revealed a strong association, again with a medium-to-large effect size ($r(17) = .44, p = .06$), reflecting the opposite of what was predicted (See Table 8).

For late-stage segments, the correlation between adherence to CBT and decreased symptomatology (better outcome) was significant with a sizable association ($r(17) = -.48, p < .05$). This relationship was in the predicted direction. Medium-to-large effect sizes were reflected in the sizable associations between adherence to common factors and decreased symptomatology (better outcome) ($r(17) = -.41, p < .10$), as well as between alliance and positive outcome ($r(17) = -.42, p < .10$). Although not significant, both associations were in the predicted direction. (See Table 8).

Table 8. Relationship of Treatment-Specific Adherence and Therapeutic Alliance to Treatment Outcome: Zero-Order Correlations

	Averaged Across Stage of Therapy	Early	Middle	Late
Variable	Outcome			
BAP	.40	.39	.44	.27
CBT	-.36	-.16	-.40	-.48*
IET	.12	.12	.31	.04
CMFX	-.21	.14	-.12	-.41
WAI	-.19	-.13	.28	-.42

Note. * $p < .05$. Adherence and alliance variables averaged across stage of therapy for column 1; variables non-averaged for columns 2-4. BAP = adherence variable for brief-adaptive psychotherapy subscale; CBT = adherence variable for cognitive-behavioral therapy subscale; IET = adherence variable for interpersonal-experiential therapy subscale; CMFX = adherence variable for common factors subscale; WAI = Working alliance inventory total score. $N = 20$ for columns 1 and 2; $N = 19$ for columns 3 and 4. (-) value indicates decreased symptomatology equivalent to positive outcome. All tests are two-tailed.

Eight first-order correlations were performed between therapist adherence, controlling for the variance accounted for by therapeutic alliance, and treatment outcome.

The first four partial correlations defined adherence as the average mean score (across raters) on the relevant subscale for each segment. Thus, for example, in order to represent adherence to BAP within this overall score, only adherence scores for BAP segments were used. This applied to CBT and IET, as well. These "treatment-specific" adherence scores were then averaged across treatment modality to produce an overall adherence score. This variable represented overall adherence combining all treatments. This set of four partial correlations examined the adherence variable averaged across stage of therapy and by stage of therapy, as well. Treatment outcome was again defined as the average of the two outcome scores, the overall mean on the SCL-90R and the overall mean on the IIP-64 for each case, to produce a single outcome variable. Alliance, the partialled-out variable in these analyses, was defined in the same way for both sets of four partial correlations: first as average alliance, which reflected the average of the WAI total scores across stage of therapy and, second, as a single WAI total score for each of the early, middle, and late segments.

These first-order correlations yielded non-significant results, suggesting that overall adherence, controlling for alliance, is not predictive of outcome (See Table 9). Results of these analyses are contrary to the Hypothesis III prediction.

The second set of partial correlations examined treatment-specific adherence (as measured by the treatment-specific subscale), both averaged across stage of therapy, and by stage of therapy. Thus, for example, the average BAP variable reflected treatment-specific adherence averaged across early, middle, and late segments. The early BAP variable reflected treatment-specific adherence only for early segments. The outcome variable and the alliance variables were defined as described above.

The first set of this series of four partial correlations examined all segments, averaged across the three stages of therapy. The predictor variable was averaged treatment-specific adherence, with averaged alliance partialled out, and outcome as the criterion variable. First-order correlations between adherence to BAP and poor outcome, and adherence to CBT and positive outcome, when controlling for alliance, yielded significant results (BAP, poor outcome: $(r(17) = .47, p < .05)$; CBT, positive outcome: $(r(17) = -.48, p < .05)$ (See Table 9).

The remaining three first-order correlations examined the relationships between non-averaged, treatment-

specific adherence and outcome, controlling for alliance (as defined by the WAI total scores for each treatment third). These analyses were by stage of therapy.

For early-stage segments, the first-order correlation between adherence to BAP and poor outcome, although non-significant, had a medium-to-large effect size ($r(17) = .43, p < .10$) (See Table 9).

For middle-stage segments, the first-order correlation between adherence to CBT and outcome, although non-significant, also had a medium-to-large effect size ($r(16) = -.42, p < .10$) (See Table 9).

For late-stage segments, the first-order correlation between adherence to CBT and outcome, partialling out alliance, yielded significant results, revealing a sizable association between the two variables ($r(16) = -.51, p < .05$). Also for late-stage segments, the first-order correlation between adherence to BAP and poor outcome (although not significant) revealed a medium-to-large effect size ($r(16) = .45, p < .10$) (See Table 9).

Table 9. Relationship of Overall and Treatment-Specific Adherence to Treatment Outcome, Controlling for Therapeutic Alliance: First-Order Correlations

	Averaged Across Stage of Therapy	Early	Middle	Late
Predictor	Outcome			
Overall Adherence	.16	.26	.35	-.24
BAP	.47*	.43	.37	.45
CBT	-.48*	-.23	-.42	-.51*
IET	.13	.12	.30	-.01
CMFX	-.17	.16	-.18	-.30

Note. * $p < .05$. $N = 20$ for columns 1 and 2; $N = 19$ for columns 3 and 4. Overall adherence variable is averaged across modality and across stage of therapy in first analysis, and across modality and by stage of therapy (non-averaged) in analyses 2-5. Alliance total scores are averaged across stage of therapy for first column, and non-averaged for columns 2-4. (-) value indicates decreased symptomatology equivalent to positive outcome. All tests are two-tailed.

DISCUSSION

The primary goal of this dissertation was to establish interrater reliability and discriminant validity for a shortened and refined version of the Beth Israel Adherence Scale. These aims were reflected in hypotheses I and II, which will be discussed in tandem. A secondary goal of this project was to evaluate the nature of the unique and joint contributions of therapist adherence and therapeutic alliance to treatment outcome, which was reflected in hypothesis III. A summary and interpretation of the results and a discussion of implications are presented for the three hypotheses. Limitations of the study and directions for future research are also discussed.

Psychometric properties of the Beth Israel Adherence Scale

Hypotheses I and II relate to the psychometric properties of the Beth Israel Adherence Scale, specifically those reflected by interrater reliability and the scale's capacity to discriminate between therapist behaviors associated with distinct treatments. Results offer preliminary evidence that the scale demonstrates psychometric properties to an acceptable degree.

Intraclass correlation coefficients indicated that agreement between raters was satisfactory to strong, both for the overall scale and the treatment-specific subscales

as well as across stage of therapy. Kraemer (1981) suggests that a range of .60 - .80 be considered "satisfactory" reliability; for both the overall scale and the treatment-specific subscales, all ICC scores were above the .70 level. By Kraemer's criteria, scores above the .80 level are considered to reflect very strong reliability; ICC means for overall scale, individual subscales, and by stage of therapy were above the .80 level.

Discriminant validity was assessed through a comparison of subscale scores from the three treatment conditions, demonstrating that the scale could distinguish between the three modalities. Results were in the predicted direction.

Interrater reliability levels from this study compare favorably with those cited by Butler, Henry, and Strupp (1995) in their study of the measurement of adherence in Time-limited Dynamic Psychotherapy (e.g. interrater reliability for the VTSS Specific Strategies subscale, in particular, ranged from .71 to .91). Subscale interrater reliability levels reported in the present study also compare favorably with by-item interrater reliability levels achieved for the previous version of the scale, particularly when items from the previous scale with poor reliability were removed. Finally, interrater reliability from the present study compares favorably with that reported by Hollon et al. (1984) for the CSPRS, an adherence measure

employed in the training and pilot phase of the NIMH Treatment for Depression Collaborative Study, from which a number of the items for the present scale were obtained.

The consistently high level of interrater reliability demonstrated for the present version of the scale is likely partly a function of the previously conducted item analysis. Specifically, reliability for the previous version of the scale was assessed by calculating by-item intraclass correlation coefficients. The broad range (.07 to .89) for individual items indicated that specific items were depressing the overall reliability of the previous version of the scale (Santangelo, 1995). In an item analysis, some of these items were removed. In addition, items which appeared to duplicate other items too closely or which were too difficult to operationalize were also removed. Thus, the remaining items were thought to well represent the unique treatments with which they were associated.

The high levels of interrater reliability for this scale may also reflect the emphasis placed on initial rater training and on the prevention of rater drift. A total of approximately 45-50 hours of training (including those hours devoted to the prevention of rater drift) were administered to all raters. Rater drift meetings focused primarily on the review and discussion of specific items about which raters differed. Consensus for each of these "problematic"

items was achieved, and this information was documented in written form by this author following each meeting and distributed to each rater, with the instruction to incorporate this "updated" information into the established rating process. Feedback from raters confirmed that this method of consistent and tangible reminders was useful in helping raters to recall and then incorporate changes in a comprehensive and organized manner.

Given the acceptable levels of psychometric properties established for the scale thus far, this measure represents another avenue for research into the investigation of therapist adherence. Of particular importance is the fact that the Beth Israel Adherence Scale includes a measure to assess therapist adherence to interpersonal-experiential therapy. Since interpersonal-experiential therapy in its current form is a relatively "new" treatment, the continued development of the scale offers the opportunity for further investigation into the particular "workings" of this modality. Of course, future development of the scale has particular relevance to the Beth Israel Research Project, which seeks to understand the differential impact of three treatment modalities. The development of the scale has considerably broader implications, however, as it contributes to the upholding of treatment fidelity in psychotherapy research, in general.

Limitations

Rating process

A factor which may have influenced interrater reliability and discriminant validity relates to the fact that raters were not completely "blind" to the rating process. As previously discussed, efforts were made to prevent raters from associating therapists with their respective modalities (e.g. videotaped segments were transposed to audiotape to minimize the possibility of raters identifying therapists). Precautions were particularly important since two of the raters were familiar with two of the therapies and several of the therapists. These raters were research assistants in the Beth Israel Psychotherapy Research Project, one having received training in BAP, the other in IET. Both reported that they thought they could recognize therapists by voice, although this was not consistent. Thus, there is some possibility that raters were inclined to "identify" interventions by modality associated with therapist, at least to some degree. At the same time, the third rater, who was involved in the Beth Israel Psychotherapy Project for just a few months, noted that he could only infrequently identify the modality associated with a specific therapist, by voice. Since all ICCs were consistently high (not just those associated with the less "naive" raters), and the contribution of the third

rater was likely to have been least biased in this respect, the degree of rater bias resulting from voice recognition probably did not exercise undue influence on overall scores. Still, this remains an area of concern.

In their report on adherence to treatment protocol as assessed by the Collaborative Study Psychotherapy Rating Scale, Hollon et al. (1984) noted that they hesitated "to use the term 'blind' to describe the raters, since the clinical management sessions were typically about half as long as the two psychosocial intervention sessions and so explicitly discussed medications that it is very unlikely that any rater failed to discern their nature" (p. 10). Clearly, the act of "blinding" raters to all aspects of the rating process is one which remains an important challenge in psychotherapy research.

Although the "naivete" of the raters was a concern, it is important to note that rater comments in rater drift meetings reflected raters' tendencies to consider interventions on an individual, non-modality-driven basis, rather than attempts to fit an intervention with a specific therapist. Raters also reported that the random distribution of the treatment-specific items throughout the overall scale helped to minimize the tendency to "lump" groups of items from one modality together, as raters had to proceed through the entire scale for each segment. Raters reported that they were less inclined to "stick" to a select

group of items from one modality when listening to a specific therapist. This factor may have helped to counter any tendency to identify therapies solely by therapist voice.

Some avenues for future research in this area are necessarily guided by the limitations associated with the study. For example, it would be useful to attempt to replicate these findings with raters who are more "naive" to the rating process. The risk of rater bias is an important one (Hill, O'Grady, & Price, 1988) and future research might do well to consider the implementation of rater bias checks in further refinement of the scale. Replication of the study with this criterion in mind would contribute to the psychometric value of the scale, in particular, and to the upholding of treatment integrity in psychotherapy research, in general.

Non-independence of segments

Another factor which may have influenced interrater reliability is the relative non-independence of therapy segments. An independent segment may be identified as a unique therapist/patient contact. That is, no therapist and no patient would be duplicated in any therapist/patient dyads. In this study, five of the 12 therapists were each represented with three patients, and thus observed by raters in nine segment units (rather than three). There is some

likelihood, therefore, that raters may have become somewhat more familiar with these therapists and better able to identify their modality through the increased exposure. Of course, the assessment of temporal consistency in this study would preclude complete segment independence, in any event, since nearly all therapist/patient dyads were observed at three junctures (i.e. in early, middle, and late stages of therapy).

A limitation of the study of the earlier version of the scale is discussed by Santangelo (1995), who cites segment non-independence as a possible confounding variable in the interpretation of the scale's psychometric properties. A recommendation of the previous study was to enhance segment independence for a study of the revised version of the scale (Santangelo, 1995). For example, in the earlier study, Santangelo (1995) notes that raters evaluated adherence for three therapists who represented three modalities and therefore could not "be considered separate variables" (p. 44). Santangelo (1995) observed that it was unclear whether raters were discriminating between therapists or modalities since there were three of each. Further, she noted that Luborsky et al. (1982) encountered a similar problem, "since the same therapists and often the same patients appeared in more than one session (Luborsky et al., 1982, p. 58). Luborsky et al. (1982) addressed this methodological difficulty by taking a

conservative approach and analyzing only "the first tape from each therapist that each judge heard" (p. 59), which guaranteed segment independence. Thus, these researchers felt justified in treating variables as "largely independent events despite the multiple representation of therapists" (p. 59).

Although the present study did not employ as conservative an approach as that described by Luborsky et al., the increase in the number of therapists in this study does represent a step toward enhanced segment independence. Specifically, modalities were represented by four therapists, rather than one, which prevents raters from associating any one modality with a single therapist. Thus, increasing the number of therapists in this study represents an improvement in segment independence relative to levels described in the previous study. In addition, it is important to note that having each therapist represented with more than one patient facilitates the testing of therapist effects. Crits-Christoph and Mintz (1991) note that "the use of only one patient per therapist produces complete experimental confounding that precludes testing for therapist effects at all" (p. 25). Future research could examine ways to further enhance segment independence, and might also consider the related issue of the testing of therapist effects.

Differences in data sets

A final limitation of the study in terms of the assessment of the scale's psychometric qualities could be reflected by the fact that the previous and revised versions of the scale were assessed on different sets of data. For example, the previous study used written transcripts of five-minute segments and the present study used audiotapes of 15 minute segments. For the previous study, CBT transcripts were obtained from the Minnesota Cognitive Therapy Transcript Library, while for the present study, audiotapes of CBT cases were from the Beth Israel Project. For the previous study, IET transcripts were taken from a psychotherapy study on interpersonal-experiential therapy conducted at Toronto's Clarke Institute, while IET audiotapes again came from the Beth Israel Project. In contrast, BAP transcripts and BAP audiotapes both came from the Beth Israel Project (reflecting the same approach across therapist intervention and comparable patient disorder type.)

While there are some differences in the origins of the data, it should be noted that there is also considerable overlap. Specifically, the Minnesota Cognitive Therapy Study was directed by Steve Hollon, who also directed the piloting of the CSPRS for the NIMH Collaborative Depression Study (1984; 1988), from which all of the Beth Israel CBT items are derived. In addition, Jeremy Safran, who directed

the Clarke Institute study, from which the IET transcripts were obtained, also directed supervision of IET cases at Beth Israel. It would therefore seem reasonable to assume that therapist interventions for CBT and IET transcripts and audiotapes are quite similar in content, although the medium for assessment remains a difference (i.e. transcript vs. audiotape). Finally, there is considerable consistency between the BAP transcripts and the BAP audiotapes since both were taken from the Beth Israel study.

Although there is overlap between the two data sets, enhanced consistency would be useful. Future research should focus on replication of the study in this respect in order to confirm the levels of interrater reliability and discriminant validity cited here. In addition, replication of the present study with a more similar data set would reduce the need for a test of internal consistency, which was not assessed for the revised version of the scale. Since the previously existing scale demonstrated high levels of internal consistency, it was believed that an exclusive focus on the remaining psychometric properties of the present scale would be justified. In addition, the evaluation of internal consistency is more often pursued as part of the assessment of overall psychometric adequacy for newly developed or newly tested scales. Still, this may reflect a limitation in terms of making assumptions about the degree of internal consistency for the present scale,

although methodological risk is probably fairly remote.

Hypothesis III

The joint contribution of therapist adherence and therapeutic alliance to treatment outcome

Hypothesis III addressed the question of the nature of the relationship of adherence and alliance to treatment outcome. Framed another way, this question attempts to enhance the understanding of the contributions of both specific (i.e. adherence/technique) and non-specific (i.e. alliance/relationship) factors. The specific/non-specific factors paradigm has informed the study of psychotherapy for several decades (Butler & Strupp, 1986). Some researchers believe that "the fundamental inappropriateness" of the specific/non-specific approach to the study of psychotherapy (which is derived from a medical or "clinical trials" model) has pointed research in inherently problematic directions (Butler & Strupp, 1986). From this perspective, there are limitations to the current approach which attempts to assess both the unique and joint contributions of specific and non-specific factors which appear highly interrelated. However, comparative outcome studies which examine both "specific" and "non-specific" factors are likely to continue to have a place in psychotherapy research as long as "alternative treatments exist for a particular clinical problem" (Kazdin, 1986, p. 96).

Hypothesis III was tested with two series of analyses (multiple regression analyses and correlational analyses). The results and implications of the multiple regression analyses will be discussed first. Since few of the analyses yielded significant results, a discussion of limitations will be integrated directly into a summary of the findings.

The multiple regression analyses were intended to assess the contributions of adherence and alliance, which taken together, could enhance an understanding of the nature of therapist competence. Although this rationale rests upon a solid conceptual foundation, results of the multiple regression analyses did not yield significance. At the same time, the magnitude of the effect (i.e. effect sizes) for two of the analyses were strong, which suggests the possibility of finding significant results with a larger sample size. It is interesting that effect sizes which involved specifically middle and late segments were sizable. This indicates that results were not temporally consistent, at least for this sample size. It is unclear why there is a difference between the effect sizes for all cases and early segments, on the one hand, and middle and late segments, on the other. This inconsistency across time is also revealed in the results of the correlational analyses.

Research on alternative methods of assessing competence remains an important area for investigation,

particularly since the "costs" associated with the assessment of competence can be considerable. Alternative methods to evaluate therapist competence could contribute to the greater facility with which this subjective construct is measured. The issue of therapist competence has generated concern in psychotherapy research, both from the perspective of the importance of upholding treatment integrity, and from the perspective of the impact of therapist competence on treatment outcome. There appears to be considerable support in the literature for conceptualizing the construct of treatment integrity as one which integrates both adherence and competence. Directions for future research concerning investigation of the nature of the contributions of therapist adherence and therapeutic alliance to treatment outcome must be initially guided by the need for a larger sample size. Considering the promising effect sizes, there remains clear incentive to continue to investigate the nature of therapist competence from this alternative perspective.

Related to this issue, Barber & Crits-Christoph (1996) note that there is the question of "whether judges can maintain the dimensions of competence and adherence separately" as they rate a series of interventions. For example, they question whether raters might "rate an intervention as more frequent because it was carried out skillfully" (p.91). Although there is no definitive

response to this question, the question itself points to a possible confound in the form of rater bias and may inadvertently support further investigation into alternative methods of competence assessment.

The unique contributions of adherence and alliance to outcome

An investigation of the associations between treatment-specific adherence, adherence to common factors, alliance, and outcome yielded mixed results. This was also true for analyses investigating the relations between adherence and outcome, while extracting the variance accounted for by alliance. While a substantial number of individual analyses yielded results at the non-significant level, several themes did emerge among those results which achieved significance and those which were non-significant, but had large effect sizes. One of these was in the predicted direction (e.g. the association between adherence to CBT and positive outcome), while another pattern reflected the opposite of what was predicted (e.g. the association between adherence to BAP and poor outcome.) In addition, significant results were not consistent over time, but strong associations were consistent over time.

At the simple correlation level, adherence to CBT, in late stages of treatment, was strongly associated with decreased symptomatology or positive outcome. In addition,

the magnitude of the effect for this same relationship across stage of therapy and in middle segments was medium to strong, which again suggests that a limitation of the present study is its small sample size. With a larger N, these analyses might have yielded significant results. Interestingly, when the variance accounted for by alliance was extracted, the association between adherence to CBT and outcome became more pronounced and more consistent over time. This suggests a strong relationship between these two variables, which is further bolstered by having extracted the variance accounted for by alliance, which is a construct solidly associated with positive outcome.

None of the other treatments were associated with positive outcome, which represented a departure from what was hypothesized. However, it seems a mistake to reject the null hypothesis of no association since there is indication that a larger sample size might yield significant results.

Associations between adherence to CBT and positive outcome and adherence to BAP and poor outcome

The sizable and yet inconsistent association between adherence to CBT and improved outcome requires an explanation. It is possible that there is an interaction effect between patient disorder and treatment type, although the data of this study do not provide any specific support for this idea. Future research might be designed with the

intention of investigating this possibility more closely.

In the present study, nearly all patients (approximately 20 out of 22) had a DSM-III-R Axis I diagnosis of either a depressive disorder or an anxiety disorder (in addition to an Axis II diagnosis of a personality disorder.) The anxiety disorders diagnosed (including generalized anxiety disorder and social phobia) are ones for which the use of CBT has considerable support in the literature (Hollon & Beck, 1994). Only two patients with anxiety disorders of this type, however, actually received CBT. It would seem unlikely, then, that the random assignment of patients to a specific treatment would favor an explanation of an interaction effect between patient disorder and treatment type in this respect, since so few anxiety-disordered patients received CBT.

Another possible explanation could lie with the fact that the remaining patients who received CBT all had diagnoses of some form of depressive disorder. It may be useful to recall that the CBT practiced at Beth Israel has its roots in Aaron Beck's Cognitive Therapy, the development of which was significantly informed by the study of "the role of cognition in depression" (Hollon & Beck, 1994, p. 430). Aaron Beck's Cognitive Therapy has been widely researched (particularly in relation to the treatment of depression) and has been shown to be "at least equal or superior to alternative psychosocial... approaches" (p.

428). Still, other therapies have been shown to be effective in the treatment of this disorder, as well. For example, interpersonal psychotherapy (IPT: Klerman, Weissman, Rounsaville, & Chevron, 1984) was developed to treat depression and employed (alongside cognitive behavior therapy) in the NIMH Treatment of Depression study. It, too, has demonstrated efficacy. Therefore, it would seem premature to conclude that the association between adherence to CBT and positive outcome represents an interaction effect between patient disorder (depression) and treatment type.

A general, and in some respects, more defensible explanation for the association between adherence to CBT and positive outcome may be related to the supportive and non-interpretive emphasis associated with this approach. Quite simply, it is possible that the supportive nature of this treatment was related to its effect on outcome. This hypothesis may also point to an explanation for the relationship between adherence to BAP and poor outcome.

In comparison to CBT, BAP is interpretive and confrontational in nature. Its authors note that it utilizes transference and defense interpretations, emphasizing "a continual working through of the (maladaptive) pattern..." (Pollack et al., 1990, abstract). Some authors have written on the purported risks associated with the use of interpretation (Henry et al., 1994; Strupp, 1989; Wachtel, 1993; and Wile, 1984). For example, studies

have linked increased use of transference interpretations with poorer outcome (Henry et al., 1994). Piper, Azim, Joyce, and McCallum (1991) observed that a higher frequency of transference interpretations may create in the patient the experience of feeling criticized. Wachtel (1993) has distinguished between "accusatory and facilitative" therapist comments and observes that interpretations, "not infrequently ...contain an implicit rebuke" (p. 70). Finally, Strupp (1989) has identified the "damaging consequences of communications that are experienced by patients as pejorative" (p. 717). Clearly, there exists both empirical and theoretical support for the belief that interpretation in general and transference interpretation in particular may be associated with poor outcome.

Two earlier studies conducted by the Beth Israel Psychotherapy Research Project examined the link between BAP and treatment outcome (e.g. Winston, Pollack, Laikin, Samstag, McCullough, & Muran, 1994). Interestingly, these studies yielded results in the opposite direction from those of the present study, supporting a link between BAP and positive outcome. The fact that the present study's results indicate strong associations in the opposite direction warrants attention from at least two perspectives. The first is that the examination of discrepancies between findings furthers an understanding of psychotherapy research, in general. To address the issue of discrepancies

in this case, one might consider whether therapists in this study employed a greater degree of interpretation than those in previous studies. From a speculative perspective, the degree of interpretation might account for the contrasting results. It would be useful for future research to consider this question.

Secondly, the unexpected (although inconsistent) finding that adherence to BAP is associated with poor outcome warrants further investigation from an obvious ethical perspective: patients have the right to receive a treatment which, at the very least, does no harm. It would be premature to discuss the "risks" associated with this treatment since the findings are mixed and must be considered preliminary, given the small sample size as well as results from previous studies. Future research is warranted to consider the present study's results in relation to previous findings as well as to evaluate further the possibility that risks may be associated with this treatment.

Associations between therapeutic alliance and treatment outcome and therapeutic alliance and therapist adherence

Another aspect of this study's investigation was related to the relations between therapeutic alliance and outcome, and therapeutic alliance and adherence. Few associations between these variables were significant, although this finding may be somewhat tempered by results

which reflected strong effect sizes. The most puzzling of these results was the finding that therapeutic alliance was not consistently associated with positive outcome. This is particularly puzzling as it is inconsistent with one of the most robust and frequently cited findings in the literature. Interestingly, although findings were not significant, the strength of the association between alliance and positive outcome in late stages was considerable ($-.41$), which again points to the need for increased sample size. Future research would do well to consider this puzzling discrepancy.

A somewhat less puzzling finding was reflected by the non-significant associations between adherence to common factors, in particular, and alliance. One might expect an association between these two variables, since it would make sense that there might be overlap in the measurement of these two constructs. Consistent with this finding were the findings of Barber and Crits-Christoph (1996) in their study of the development of a therapist adherence/competence rating scale. In explanation, they noted "the often reported low correspondence between different methods (specifically, self-report vs. judge) of assessing a similar construct (p. 92).

Conclusion

The principle aims of this study have been realized in the continued development of the psychometric properties of the Beth Israel Adherence Scale. These promising findings contribute to a body of research which seeks a fuller understanding of the assessment of treatment fidelity, an essential component of comparative outcome research.

The findings related to an investigation of the relationship of therapist adherence and therapeutic alliance to treatment outcome raise important conceptual and methodological questions which warrant further investigation. The measurement and assessment of psychotherapeutic variables is inherently complex and ambiguous since the constructs to be measured are difficult to operationalize and highly subjective. For this reason, the development of alternative methods of evaluating therapist competence, in particular, remains a important area for investigation, as explicit competence measures may continue to encounter the obstacles described above.

APPENDIX A

BETH ISRAEL ADHERENCE SCALE

1. **Assigns and reviews homework.** (The therapist reviews with the client the previous assignment from the week before. The therapist discusses with the client the assignment for the coming week.) (Rate on freq/clarity.)

2. **Interprets other aspects of client's behavior or experience. (Not captured in other items - General interpretation).** e.g. "It sounds like you have trouble figuring out who you are and what you want out of your life, separate from what your parents want." (Rate on freq/clarity.)

3. **Explores the HOW, or mechanism of a client's defense, not the WHY.** (Therapist focuses on the feelings underlying the client's defense and NOT the reasons for them. The goal is not to establish causal links but to help the client identify and experience the feelings which elicit certain defenses. e.g. "Are you aware of controlling your feelings in any way?", "What are you avoiding?", "Are you aware of stopping your feelings right now?" "How do you stop your feelings?"). (Rate on freq/clarity.)

4. **Reflects the content of client's statement.** (Therapist attempts to understand the meaning of the content of what the client has said and reflects this back to the client. It is often a summary or precis of what the client has just said rather than a reflection of feeling. Therapist conveys that client's meaning has been understood. (Rate on freq/clarity).)

5. **Therapist's communication style.** How interesting is the therapist's style of communication? (Consider (1) the vividness of her/his language; (2) the originality of her/his ideas; (3) the liveliness of her/his manner of speaking.) Rating: "1" = dull, uninteresting; "3" = less interesting than average; "6" = very interesting. (Evans et al., 1984)

6. **Directs client's attention in non-confrontational manner to specific client behaviors, subtle non-verbal communications or paralinguistics, to increase client's awareness.** (This can be an observation of facial expression, body movement or posture, or voice inflection, etc. Therapist does this in a supportive and nonjudgmental manner. e.g. "I'm aware of a particular tone in your voice. "When you say this, you have a very angry expression on your

face.") (Rate on freq/clarity).

7. **Encourages client to distance him/herself from his/her thoughts, viewing them as beliefs rather than facts.** (Did the therapist urge or challenge the client to consider his/her thoughts as beliefs which may or may not be true. Therapist urges the client to consider his/her thoughts as testable hypotheses rather than as proven facts. This item can be coded if the therapist makes direct statements to this effect OR if the therapist less directly encourages this, as well. e.g. "What's that belief about?", "What is that thought?" - NOT, "What do you think?" or "What do you believe?" BUT, more "this or that thought", "this or that belief", "Do you see how thinking of it in this way allows you to see it as a hypothesis rather than a carved-in-stone fact?"). (Rate on freq/clarity).

8. **Facilitates individuation and/or self-assertion.** (Therapist encourages the client to either ask for what s/he wants or to express his/her feelings directly to therapist. e.g. "Do you have a sense of what you want from me right now?" "I wonder if you could tell me how disappointed you are in me now?" (Rate on freq/clarity).

9. **Frames symptoms in a relationship context.** (Therapist shows client that particular symptoms are associated with aspects/events in client's relationships. Symptoms are believed to be a result of previously dysfunctional relationships. For example, forgetting is a "symptom" of memory dysfunction; anxiety and depression are also examples of symptoms. e.g. Therapist notices that every time a client's attractiveness is mentioned, she feels very sad. Father would show little interest when client would get recognition for an achievement or attribute, etc. Therapist says, "You felt depressed in response to your father's losing interest in you. And now you feel sad with me because you perceive that I, too, have lost interest in you.") (Rate on freq/clarity).

10. **Therapist conveys competence.** Does the therapist convey that she/he understands the client's problems and is able to help the client? Rating: "1" = "not at all"; "3" = some; "6" = very much. (Evans et al., 1984).

11. **Probes for client's underlying beliefs or personal meaning behind client's thoughts.** (e.g. "What does that mean to you?" "What does that thought mean to you?" "If you think that he doesn't want to talk with you, what does that mean to you", "It sounds like you believe that in order to feel good about yourself, you must be liked by everybody".) (Rate on freq/clarity).

12. **Links resistance (to the therapeutic process) to the maladaptive pattern.** (e.g. "You're tuning out here with me just like you tune out elsewhere when things get tough" - links behavior with the therapist to behavior in other situations/with other people. "You're shutting down with me just like you do with your family when you get angry".) (Rate on freq/clarity).

13. **Explores the advantages and disadvantages of dysfunctional attitudes.** (e.g. "What's the advantage to believing that?" "How useful is the belief that you will never get ahead?" "Is there a disadvantage to that thinking style?"). (Rate on freq/clarity).

14. **Directs or redirects the focus to the "here and now" (either with regard to the client's experience or with regard to the relationship between the client and therapist.)** (e.g. "What's happening for you right now?", "What would satisfy you with me right now?" "What's your fear of exploring those feelings with me right now?" (Rate on freq/clarity).

15. **Therapist Involvement.** How involved is the therapist? Rating: "1" = very detached; "6" = very involved. (Evans et al., 1984).

16. **Confronts client, suggesting that he/she is saying, feeling, or thinking something different than what the client claims.** (e.g. "You say you're not angry and yet your expression looks angry", "You say you're not anxious, but you're talking very quickly and you've been twisting your hands back and forth." Includes all client behaviors, not just non-verbal. (Rate on freq/clarity).

17. **Helps client identify cognitive distortions, errors that were present in his/her thinking.** (Magnifying or minimizing; catastrophizing; personalizing; generalizing. e.g. "Do you see how this all-or-none thinking actually decreases the number of options you see?", "It sounds like you believe that the only possible result of your effort is going to be complete failure. Is there a more accurate way to think about this problem? Do you see how you are singling out the "worst-possible-case" scenario?"). (Rate on freq/clarity).

18. **Intervenes with skillful tentativeness.** (Refers to quality of therapist attitude of exploration and subjectivity. Therapist uses words like "perhaps", "it seems", "possibly"). (Rate on freq/clarity).

19. **Facilitates client's consideration of alternative explanations for events.** (Did the therapist help the client

consider alternative explanations for events besides the client's initial explanation for those events? e.g. "Is there any other explanation for this event than the one you've come up with so far?", "What would be another way to explain why Bill reacted in that way?" "What about considering another perspective on this situation and seeing if it fits", "Are there other factors which could have played a role in your not getting the position?"). (Rate on freq/clarity).

20. **Interprets/Explores maladaptive patterns by linking dynamics with parental/significant figures in the past to others in the present (not including therapist. (i.e. carrying past parental relationship dynamics into the present in a way that is not productive).** e.g. "One of the things we've learned from looking at your relationship with your mother is that you tried to do the accommodating thing in order to get her approval. It seems that you do a similar thing with Bob, never crossing him, so that he won't be angry with you.") (Rate on freq/clarity).

21. **Interprets and/or explores client's resistance or defenses.** (An interpretation provides a new understanding or offers a label of an inner state; it presumes knowledge by the speaker of the client's experience and places it in the speaker's frame of reference. e.g. "You try to avoid situations which make you feel confused." "When you feel anxious, you tend to withdraw.") (Rate on freq/clarity).

22. **Therapist Warmth.** Did the therapist convey warmth? (Rating: "1" = not at all or very little; "6" = very much. (Evans et al., 1984).

23. **Tracks client's experience in a moment-to-moment fashion.** (The act of following client's perceptions, thoughts, feelings as they emerge in the moment. Therapist does not make reference to client processing that is not currently being experienced.) (Rate on freq/clarity).

24. **Asks client to report specific thoughts.** (Asks client to report specific thoughts as verbatim as possible. In order to code this item, the therapist must have attempted to elicit the client's thoughts as specifically and as verbatim as possible. e.g. "What specific thoughts do you have about that?" "What are your thoughts about that feeling?" "Let's see what the specific thought is that you have about this?", "Let's get to the thought that you're having about this feeling.") (Rate on freq/clarity).

25. **Engages in empathic conjecture: Hypothesizing, exploring the nature of the client's experience and then "checking in"**

after making the conjecture (often, but not always, interrogative). (The conjecture is about inner experience not about psychogenetic causes or patterns in behavior or experience. Therapist takes a "guessing" or "hypothesizing" stance with client and asks client to "check" therapist's hunch with client's experience. e.g. "and so this is when I guess the hopelessness sets in... Is that true for you?" "Powerful, right? It's like the only power you have, right?" (Rate on freq/clarity.)

26. **Explores and elucidates the unconscious aspects of major maladaptive patterns, thoughts, and behaviors.** (e.g. "What's that need you have to feel frustrated?" "Why do you think you do that?" "What's that about when you act that way?" "Why do you think you're so frightened of competition?" "When you feel sacred, you act hostile. Why do you think that is?" Here, the therapist is probing for the unconscious aspects of the client's behavior/feelings.) (Rate on freq/clarity.)

27. **Engages in socratic questioning aimed at guiding client's reasoning process.** (e.g. "And what do you think would happen if you did that?", "Where is that kind of thinking going to take you?", "How likely is that to happen?", "Where's the evidence for that?". This is guided questioning. This often involves "disputing" or "challenging" the client's beliefs or ideas.) (Rate on freq/clarity.)

28. **Rapport.** How much rapport was there between therapist and client (i.e. how well did the therapist and the client get along?) Rating: "1" = total absence of rapport; "6" = excellent rapport. (Evans et al., 1984).

29. **Asks exploratory questions which probe for the feeling/experience underlying the client's utterance including feelings about the feeling/experience or utterance itself - feeling ashamed about feeling this way, etc.** (Therapist makes inquiries into what the client is experiencing or has experienced. "What does that feel like?", "What was it like for you when he went away?", "What was that like for you?", "What's your feeling about being so anxious?") (Rate on freq/clarity.)

30. **Frames symptoms as coping attempts.** (The therapist recognizes and points out that particular symptoms can be understood as faulty and costly attempts at a problem-solution. e.g. "You really want someone to soothe you but nobody is there so you eat as a way of feeling better." (Rate on freq/clarity.)

31. **Engages in didactic persuasion.** (The stance is teaching, guiding, persuading. It is a goal-directed stance that is meant to, through examining evidence, convince the client that his/her way of thinking is maladaptive or erroneous. e.g. "This plan we were talking about allowed you to test out the predictions you had. Do you see how you were able to disprove those predictions and thus get more accurate information?") (Rate on freq/clarity.)

32. **Defines/Identifies/Specifies the maladaptive pattern.** (e.g. "You have a tendency when you're feeling scared to pull back. We've seen how this happens in your close friendships and in your relationships with people at work." "When you get angry with people you care about, you have a tendency to react impulsively. This happens in relationships with people you're close to. We need to understand what this pattern is about.") (Rate on freq/clarity.)

33. **Receptive Silence.** (Did the therapist allow silence to continue or use minimal encouragements such as "uh-huh," "mm-hmm," "okay") as a means of encouraging the client to talk?) Rating: "1" = not at all; "6" = very much. (Evans et al., 1984.)

34. **Helps client examine currently available evidence or information to test the validity as well as realistic consequences of the client's beliefs.** (Does the therapist help the client use evidence from (1) client's past experience and (2) his/her knowledge of the way the world works, to test his/her beliefs for validity. This same idea can also be applied when examining with the client the realistic consequences of an event, etc. e.g. "Let's look at what actually happened and see if your belief still fits", "Let's see what the evidence is for this belief. How do you know your friends are tired of you?"). (Rate on freq/clarity.)

35. **Respects client as arbiter of experience.** (Therapist maintains a humble, subjective, exploratory stance. Therapist is not the expert on the client's feelings; s/he is facilitating their unfolding). (Rate on freq/clarity.)

36. **Interprets/Explores maladaptive patterns by linking dynamics with others (past and present) to current dynamics with the therapist.** (Therapist tries to show the client that patterns that existed in relationships with significant others are similar to patterns in the relationship with the therapist. e.g. "So you presented yourself in a dependent way with John as a way to get taken care of, and now we see that the same thing is happening here with me."). (Rate on freq/clarity.)

37. **Deepens client's awareness/experience through in or out-of session awareness exercise.** (Often, when the client has expressed an emotion, the therapist will say, e.g. "Try saying that to me directly", "Try saying, "I'm angry at you" or "Over the week, be aware of when you get sad or close off and withdraw".) (Rate on freq/clarity.)

38. **Therapist and client practice rational responses to client's negative thoughts and beliefs.** (Rational responses represent more accurate or reasonable ways of thinking about an event or issue than the client's original thought or belief. e.g. "Let's try to generate some thoughts that may be more reasonable than concluding that you are a failure. "I'll come up with the negative thoughts and you try to counter them with more reasonable thoughts. What would you say if I said that I can't make a decent meal?"). (Rate on freq/clarity.)

39. **Supportive Encouragement.** (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? Rating: "1" = not at all; "6" = extremely. (Evans et al., 1984.)

40. **Deepens client's experience through evocative reflection.** (Therapist takes the client's either implicitly or explicitly expressed feelings and empathizes with these feelings to amplify/elaborate the client's felt experience of them. e.g. "So, you're feeling a bit shut down and angry", "So you're feeling like no one really understands how hard it is for you".) (Rate on freq/clarity.)

41. **Works with client to plan or practice alternative overt behaviors for the client to use both inside and outside of therapy.** (Overt behaviors refer to "observable" behaviors rather than "covert" or cognitive behaviors. The therapist may help the client to develop a plan for getting a new job. The therapist might discuss and role-play with the client how to interact with someone outside of therapy.) (Rate on freq/clarity.)

42. **Interprets/explores maladaptive patterns by linking components of a conflict.** (Therapist provides a construction that links different components of an internal conflict. For example, drives or wishes can be linked with anxiety which can be linked with defensive processes which can be linked with affect. e.g. "You felt anxious and that made you pick a fight with your wife" "You want to leave but you are afraid to so you stay.") (Rate on freq/clarity.)

43. **Metacommunicates by conveying own feelings to help client become aware of his/her role in the interaction or to**

probe for client's internal experience (general metacommunication item). (Includes acknowledging own role in the interaction. e.g. "I think I've been acting hostile towards you", "I feel shut out right now", "I'm feeling put down right now", "I feel like I'm playing a game of chess. Does that make any sense to you?") (Rate on freq/clarity.)

44. Set and follow agenda. (Did the therapist work collaboratively with the client to formulate and follow a specific agenda for the session? Rating: "1" = "not at all"; "6" = "thoroughly"). (Evans et al., 1984).

Adapted from:

Brief Psychotherapy Research Project Rater's Manual (1996). Unpublished manuscript. (Available from Beth Israel Medical Center, New York, NY 10003.)

Evans, M. D., Piasecki, J. M., Kriss, M. R., & Hollon, S. D. (1984). Raters' Manual for the Collaborative Study Psychotherapy Rating Scale-Form 6. Minneapolis: University of Minnesota and the St. Paul-Ramsey Medical Center. (Available from the U.S. Department of Commerce, National Technical Information Service, Springfield, VA 22161.)

Greenberg, L. S. & Goldman, R. L. (1991). Can experiential therapy be successfully discriminated from cognitive and dynamic therapy? The testing of an adherence measure. Unpublished manuscript.

Luborsky, L. (1984). Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment. New York: Basic Books.

Safran, J. D. & Segal, Z. V. (1990). Interpersonal process in cognitive therapy. New York: Basic Books.

Santangelo, C. M. (1995). The development of the Beth Israel Adherence Scale. Unpublished doctoral dissertation, Derner Institute of Advanced Psychological Studies, Adelphi University, New York.

APPENDIX B

BETH ISRAEL ADHERENCE SCALE: RATING FORM

Rater Name: _____ Tape #: _____ Segment #: _____
 Rating Date: _____

Please rate each of the following 44 items on a 1 - 6 frequency/clarity scale based on both the general scoring directions and the directions for each individual item. Note that although the overall rating criteria for most of the items is the same (a frequency/clarity rating ranging from "not-at-all" to "extensively"), there is some deviation from this for several of the items. For example, item 5, (Th.'s communication style) is evaluated on its degree of interest ("dull" to "very interesting"), not on its frequency. The coder should refer to the accompanying description of the items to review the rating criteria for each item.

		not at all	-	extensively
1.	Homework.....1	2	3	4 5 6
2.	General interpretation.....1	2	3	4 5 6
3.	Explores "how" of defense...1	2	3	4 5 6
4.	Reflects content.....1	2	3	4 5 6
5.	Th.'s communication style...1	2	3	4 5 6
6.	Non-verbal.....1	2	3	4 5 6
7.	Distance.....1	2	3	4 5 6
8.	Individuation.....1	2	3	4 5 6
9.	Frames symptoms.....1	2	3	4 5 6
10.	Th. conveys competence.....1	2	3	4 5 6
11.	Probes meaning.....1	2	3	4 5 6
12.	Links resist./mal. patt.....1	2	3	4 5 6
13.	Advan./Disadv.....1	2	3	4 5 6
14.	Here and Now.....1	2	3	4 5 6
15.	Therapist involvement.....1	2	3	4 5 6
16.	Confronts.....1	2	3	4 5 6
17.	Cognitive distortion.....1	2	3	4 5 6
18.	Tentative.....1	2	3	4 5 6
19.	Alternative explanation.....1	2	3	4 5 6
20.	Links sig. past/present.....1	2	3	4 5 6
21.	Interprets def./resist.....1	2	3	4 5 6
22.	Therapist warmth.....1	2	3	4 5 6
23.	Tracks.....1	2	3	4 5 6
24.	Specific thoughts.....1	2	3	4 5 6
25.	Empathic conjecture.....1	2	3	4 5 6
26.	Unconscious aspects.....1	2	3	4 5 6
27.	Socratic questioning.....1	2	3	4 5 6
28.	Rapport.....1	2	3	4 5 6
29.	Probe feeling.....1	2	3	4 5 6
30.	Symptoms as coping.....1	2	3	4 5 6

31. Didactic.....1	2	3	4	5	6
32. Maladaptive pattern.....1	2	3	4	5	6
33. Th.'s receptive silence.....1	2	3	4	5	6
34. Examine evidence.....1	2	3	4	5	6
35. Arbiter.....1	2	3	4	5	6
36. Links sig. past/therapist...1	2	3	4	5	6
37. Awareness exercise.....1	2	3	4	5	6
38. Rational responses.....1	2	3	4	5	6
39. Th.'s supportive encour.....1	2	3	4	5	6
40. Evocative reflection.....1	2	3	4	5	6
41. Plan/Practice altern. behs..1	2	3	4	5	6
42. Links parts of conflict.....1	2	3	4	5	6
43. Metacommunication.....1	2	3	4	5	6
44. Set and follow agenda.....1	2	3	4	5	6

APPENDIX C

Scoring Directions for Adherence Ratings

General Directions:

On the rating sheets, use a check (✓) to reflect frequency (1 to 6 on the likert-type scale) and average clarity for an item. Use a check minus (✓-) to reflect frequency and less than average clarity for an item. Use a check plus (✓+) to reflect frequency and more than average clarity for an item. Base your numerical score (1 - 6) for each item on the averaging together of the number of checks you have for each item. Frequency may be defined as the number of times an intervention, identified as an item, occurs. Clarity may be defined as an intervention's being clearly recognizable and easy to identify as that item.

Specific Directions:

- (1) If an item occurs NOT AT ALL, it receives a rating of "1".
- (2) If an item occurs very frequently and with clarity it receives a rating of "6".
- (3) If an item occurs once and it occurs with "decreased clarity", it receives a rating of "2". More generally, its rating score falls between "not at all" and "somewhat". An example of an exception to this guideline would be item #1, "homework", which if it occurs at all should receive a higher score, because it is an infrequently occurring intervention.
- (4) Item #19 "tracks", an example in contrast, would receive the lowest rating above "not at all" ("2"), if it occurs once because it is, by definition, a "global" item and is expected to occur frequently and pervasively if it is applied as defined. Again, if the therapist "tracks" at all, these interventions should be evaluated for both FREQUENCY and CLARITY.
- (5) The same would apply for item #26, "didactic", which is also a "global" item and thus would be expected to occur with regularity as a defining feature of Cognitive-Behavioral therapy. Therefore, if one "didactic" intervention is made, it would be scored or "weighted" as described above for "tracks".

(6) In general, the items which CAN occur with high frequency in almost any session are the "global" items ("#6 distance, #15 tentative, #19 tracks, #26 didactic, #29 arbiter"). These items should be weighted slightly differently; they need to be weighted at least 1 point LIGHTER in comparison with the other items in the scale, since, by definition, they occur more frequently. It should be noted that the "global" items have a tendency to define the flavor of a modality, and thus may be evident by the TONE they set rather than their quantifiable presence.

(7) As a guide, keep in mind the concept of "WEIGHT". The weight of an item (which is measured by its FREQUENCY AND CLARITY) determines the RATING. Raters need ask, is this item a "global" item; is it an "infrequently-occurring" item (e.g. #1, homework), or does it fall within the broader range of "can-occur-not-at-all-to-frequently-occurring", like most of the items in the scale.

In sum:

(8) In general, the items can be divided into three WEIGHT groups: (1) Global items (listed above) which, by definition, are "hallmark" items which tend to occur frequently and/or pervasively. (2) Middle-range items which can occur "not-at-all to frequently". (3) "Infrequent, but substantive" items ("homework", "awareness exercise", and "metacommunication") which may not occur at all, but when do occur, occur in prominent form.

(9) Finally, consider that the act of coding psychotherapy interventions is inherently a complex and subjective process, based largely on intuition. The above guidelines are guidelines, and are not meant to be applied rigidly. Items should be coded as described above as closely as possible, but the coding guidelines themselves should be prevented from making the coding process cumbersome and rigid.

REFERENCES

- Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. Journal of Personality Assessment, 55, 521-536.
- Arkowitz, H. (1992). Integrative theories of therapy. In D. Freedheim (Ed.), History of psychotherapy: A century of change. (pp. 261-297). Washington, D.C.: American Psychological Association.
- Barber, J. P. and Crits-Cristoph, P. (1996). Development of a therapist adherence/competence rating scale for supportive-expressive dynamic psychotherapy: A preliminary report. Psychotherapy Research, 6(2), 81-94.
- Bartko, J. J. (1976). On various intraclass correlations reliability coefficients. Psychological Bulletin, 83, 762-765.
- Bartko, J. J. & Carpenter, W. T. (1976). On the methods and theory of reliability. The Journal of Nervous and Mental Disease, 163, 307-317.
- Beck, A. T. & Emery, G. (1985). Anxiety disorders and phobias: A cognitive perspective. New York: Basic Books.
- Beck, A. T. & Freeman, A. (1990). Cognitive therapy for personality disorders. New York: Guilford Press.
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). Cognitive therapy of depression: A treatment manual. New York: Guilford Press.
- Bergin, A. E. & Garfield, S. L. (1994). Handbook of psychotherapy and behavior change (Eds., 4th Ed.). New York: Wiley.
- Beutler, L. E. (1989). Differential treatment selection: The role of diagnosis in psychotherapy. Psychotherapy, 26, 271-281.
- Beutler, L. E. (1991). Have all won and must all have prizes? Revising Luborsky et al.'s verdict. Journal of Consulting and Clinical Psychology, 59, 226-232.
- Beutler, L. E., Machado, P. P., & Neufeldt, S. A. (1994). Therapist variables. In A. Bergin and S. Garfield (Eds.) Handbook of psychotherapy and behavior change (Fourth Ed.) pp. 229-269. New York: John Wiley & Sons, Inc.

Bootzin, R. R. & Ruggill, J. S. (1988). Training issues in behavior therapy. Journal of Consulting and Clinical Psychology, 56, 703-709.

Butler, S. F., Henry, W. P., & Strupp, H. H. (1995). Measuring adherence in TLDP. Psychotherapy, 32(4), Winter, 629-638.

Butler, S. F. & Strupp, H. H. (1986). Specific and nonspecific factors in psychotherapy: A problematic paradigm for psychotherapy research. Psychotherapy, 23, 30-40.

Butler, S. F., & Strupp, H. H. (1993). The effects of training experienced dynamic therapists to use a psychotherapy manual. In N. Miller, L. Luborsky, J. Barber, & J. Docherty (Eds.), Psychodynamic Treatment Research (pp. 191-210). New York: Basic Books.

Carroll, K., Nich, C., & Rounsaville, B. J. (1995). Use of observer and therapist ratings to monitor delivery of coping skills treatment for cocaine abusers: Utility of therapist session checklist. Unpublished manuscript. Yale University, New Haven.

Cicchetti, D. V. (1976). Assessing inter-rater reliability for rating scales: Resolving some basic issues. British Journal of Psychiatry, 129, 452-456.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155-159.

Crits-Cristoph, P., Baranackie, K., Kurcias, J. S., Beck, A. T., Carroll, K., Perry, K., Luborsky, L., McLellan, A. T., Woody, G. E., Thompson, L., Gallagher, D., & Zitrin, C. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. Psychotherapy Research, 1, 81-91.

Crits-Cristoph, P. & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. Journal of Consulting and Clinical Psychology, 59, 20-26.

Derogatis, L. R. (1983). SCL-90: Administration, scoring, and procedures manual for the revised version. Baltimore: Clinical Psychometric Research.

Derogatis L. R. & Meliseratos, N. (1983). The brief symptom inventory: An introductory report. Psychological Medicine, 13, 595-605.

DeRubeis, R., Hollon, S., Evans, M., & Bemis, K., (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.

Dobson, K. S., Shaw, B. F., & Vallis, T. M. (1985). Reliability of a measure of the quality of cognitive therapy. British Journal of Clinical Psychology, 24, 295-300.

Elkin, I., Shea, T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., Glass, D. R., Pilkonis, P. A., Leber, W. R., Docherty, J. P., Fiester, S. J., & Parloff, M. B. (1989). NIMH Treatment of Depression Collaborative Research Program: General effectiveness of treatment. Archives of General Psychiatry, 46, 971-982.

Evans, M. E., Piasecki, J. M., Kriss, M. R., & Hollon, S. D. (1984). Raters' Manual for the Collaborative Study Psychotherapy Rating Scale-Form 6. Minneapolis: University of Minnesota and the St. Paul-Ramsey Medical Center. (Available from the U.S. Department of Commerce, National Technical Information Service, Springfield, VA 22161.)

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation. Journal of Consulting Psychology, 16, 319-324.

Fiedler, F. E. (1950). The concept of an ideal therapeutic relationship. Journal of Consulting Psychology, 14, 239-245.

Frank, J. D. (1979). The present status of outcome studies. Journal of Consulting and Clinical Psychology, 47, 310-316.

Frank, J. D. & Frank, J. F. (1991). Persuasion and Healing. (Third edition). Baltimore, Md.: Johns Hopkins University Press.

Garfield, S. L. (1992). Major issues in psychotherapy research. In D. Freedheim (Ed.), History of psychotherapy: A century of change (pp. 335-359). American Psychological Association: Washington, D.C.

Garfield, S. L. & Bergin, A. E. (1994). Introduction and historical review. In A. Bergin and S. Garfield (Eds.) Handbook of psychotherapy and behavior change (Fourth Ed.) pp. 3-18. New York: John Wiley & Sons, Inc.

- Goldfried, M. R. (1991). Research issues in psychotherapy integration. Journal of Psychotherapy Integration, 1, 5-25.
- Greenberg, J. R. & Mitchell, S. A. (1983). Object relations in psychoanalytic theory. Cambridge: Harvard University Press.
- Greenberg, L. S. & Goldman, R. L. (1988). Training in experiential therapy. Journal of Consulting and Clinical Psychology, 56, 696-702.
- Greenberg, L. S. & Safran, J. D. (1989). Emotion in psychotherapy. American Psychologist, 44(1), 19-29.
- Greenson, R. R. (1967). The technique and practice of psychoanalysis, vol. 1. New York: International Universities Press.
- Hartley, D. E. (1985). Research on the therapeutic alliance in Psychiatry. Psychiatry Update: American Psychiatric Association Annual Review, 4, 532-549.
- Henry, W. P., Schacht, T. E., & Strupp, H. H. (1986). Structural analysis of social behavior: Application to a study of interpersonal process in differential psychotherapeutic outcome. Journal of Consulting and Clinical Psychology, 54, 27-31.
- Henry, W. P., Schacht, T. E., & Strupp, H. H. (1990). Patient and therapist introject, interpersonal process, and differential psychotherapy outcome. Journal of Consulting and Clinical Psychology, 58, 768-774.
- Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993a). Effects of training in TLDP: Changes in therapist behavior. Journal of Consulting and Clinical Psychology, 61(3), 434-440.
- Henry, W. P., Schacht, T. E., Strupp, H. H., Butler, S. F., & Binder, J. L. (1993b). Effects of training in TLDP: Mediators of therapists' responses to training. Journal of Consulting and Clinical Psychology, 61(3), 441-447.
- Henry, W. P., Strupp, H. H., Schacht, T. E., & Gaston, L. (1994). Psychodynamic approaches. In A. Bergin and S. Garfield (Eds.) Handbook of psychotherapy and behavior change (Fourth Ed.) pp. 467-508. New York: John Wiley & Sons, Inc.
- Hill, C. E., O'Grady, K. E., & Elkin, I. E. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive behavioral therapy,

interpersonal therapy, and clinical management. Journal of Consulting and Clinical Psychology, 60(1), 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, 3, 346-350.

Hollon, S. D. & Beck, A. T. (1994). Cognitive and cognitive-behavioral therapies. In A. Bergin and S. Garfield (Eds.) Handbook of psychotherapy and behavior change (Fourth Ed.) pp. 428-466. New York: Wiley.

Hollon, S. D., Evans, M. D., Elkin, I. E., & Lowery, H. A. (May 1984). System for Rating therapies for depression. Paper presented at the Annual meeting of the American Psychiatric Association, Los Angeles, CA.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. (1988). Inventory of interpersonal problems: psychometric properties and clinical applications. Journal of Consulting and Clinical Psychology, 56 (6), 885-892.

Horvath, A. O. & Goheen, M. D. (1990). Factors mediating the success of defiance and compliance-based interventions. Journal of Counseling Psychology, 37, 363-371.

Horvath, A. O. & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology, 36, 223-233.

Horvath, A. O. & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Counseling Psychology, 38(2), 139-149.

Howard, K., Kopta, M., Krause, M., & Orlinsky, D. (1986). The dose-effect relationship in psychotherapy. American Psychologist, 41, 149-164.

Howard, K. I. & Orlinsky, D. E. (1986). Process and outcome. In S. Garfield & A. Bergin (Eds.), Handbook of psychotherapy and behavior change (3rd ed., pp. 311-381). New York: Wiley.

Iberg, J. R. (1991). Applying statistical control theory to bring together clinical supervision and psychotherapy research. Journal of Consulting and Clinical Psychology, 59, 575-586.

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

Jacobson, N. S. & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 59(1), 12-19.

Jones, E. E., Cumming, J. D., & Horowitz, M. J. (1988). Another look at the nonspecific hypothesis of therapeutic effectiveness. Journal of Consulting and Clinical Psychology, 56(1), 48-55.

Kachigan, S. K. (1982). Multivariate statistical analysis: a conceptual introduction. New York: Radius Press.

Karasu, T. B. (1986). Specificity versus nonspecificity. American Journal of Psychiatry, 143, 687-695.

Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. Journal of Consulting and Clinical Psychology, 54(1), 95-105.

Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In A. Bergin & S. Garfield (Eds.), Handbook of psychotherapy and behavior change, fourth edition (pp. 19-71). New York: John Wiley & Sons, Inc.

Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. Journal of Consulting and Clinical Psychology, 58, 729-740.

Kiesler, D. J. (1973). The process of psychotherapy. Chicago: Aldine.

Kiesler, D. J. (1988). Therapeutic metacommunication: Therapist impact disclosure as feedback in psychotherapy. Palo Alto, CA: Consulting Psychologists Press.

Kiesler, D. J. & Watkins, L. M. (1989). Interpersonal complementarity and the therapeutic alliance: A study of relationship in psychotherapy. Psychotherapy, 26(2), 183-194.

Klerman, G. L. & Neu, C. (1976). A manual for interpersonal treatment of depression. Unpublished manuscript, Yale University.

Klerman, G. L., Weissman, M. M., Rounsaville, B. J., & Chevron, E. S. (1984). Interpersonal psychotherapy of depression. New York: Basic Books.

Kraemer, H. C. (1981). Coping strategies in psychiatric clinical research. Journal of Consulting and Clinical Psychology, 49, 309-319.

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

Lambert, M. J. & Bergin, A. E. (1992). Achievements and limitations of psychotherapy research. In D. Freedheim (Ed.), History of psychotherapy: A century of change (pp. 360-390). Washington, D.C.: American Psychological Association.

Lambert, M. J. & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. Bergin and S. Garfield (Eds.) Handbook of psychotherapy and behavior change (Fourth Ed.) pp. 143-189. New York: John Wiley & Sons, Inc.

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

Liberman, R. L. & Eckman, T. (1981). Behavior therapy vs. insight-oriented therapy for repeated suicide attempters. Archives of General Psychiatry, 38, 1126-1130.

Luborsky, L. (1984). Principles of psychoanalytic psychotherapy: A manual for supportive-expressive treatment. New York: Basic Books.

Luborsky, L. & Barber, J. P. (1993). Benefits of adherence to psychotherapy manuals, and where to get them. In N. Miller, L. Luborsky, J. Barber, & J. Docherty (Eds.), Psychodynamic treatment Research (pp. 211-226). New York: Basic Books.

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., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "Everybody has won and all must have prizes?" Archives of General Psychiatry, 32, 995-1008.

Luborsky, L., Crits-Christoph, P., McLellan, A. T., Woody, G., Piper, W., Liberman, B., Imber, S., & Pilkonis, P. (1986). Do therapists vary much in their success? Findings from four outcome studies. American Journal of Orthopsychiatry, 56, 501-512.

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., McLellan, A., O'Brien, C., & 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. (1990). Psychotherapy process research: Progress, dilemmas, and future directions. Journal of Consulting and Clinical Psychology, 58, 265-272.

McGraw, K. O. & Wong, S. P. (1992). A common language effect size statistic. Psychological Bulletin, 111(2), 361-365.

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

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.

O'Malley, S. S., Suh, C. S. & Strupp, H.H. (1983). The Vanderbilt Psychotherapy Process Scale: A report on the scale development and a process-outcome study. Journal of Consulting and Clinical Psychology, 51, 581-586.

Piper, W. E., Azim, F. A., Joyce, S. A., & McCallum, M. (1991). Transference interpretations, therapeutic alliance and outcome in short-term individual psychotherapy. Archives of General Psychiatry, 48, 946-953.

Pollack, J., Flegenheimer, W., Kaufman, J., & Sadow, J. (1990). Brief adaptive psychotherapy for personality disorders: A treatment manual. Unpublished manuscript, Beth Israel Medical Center, New York.

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., O'Malley, S., Foley, S. & Weissman, M. M. (1988). Role of manual-guided training in the conduct and efficacy of interpersonal psychotherapy for depression.

Journal of Consulting and Clinical Psychology, 56, 681-688.

Safran, J. D., Crocker, P., McMain, S., & Murray, P. (1990). Therapeutic alliance rupture as a therapy event for empirical investigation. Psychotherapy, 27(2), 154-165.

Safran, J. D., Muran, J. C., & Winston, A. (1992). Resolving problems in the therapeutic alliance. Grant proposal submitted to the ADAMHA Small Grant Program of NIMH.

Safran, J. D. & Muran, J. C. (1995). Resolving therapeutic alliance ruptures: Diversity and integration. In J. Safran & J.C. Muran (Guest Eds.), In Session: Psychotherapy in Practice, 1(1), 81-92.

Safran, J. D. & Segal, Z. V. (1990). Interpersonal process in cognitive therapy. New York: Basic Books, Inc.

Santangelo, C. M. (1995). The development of the Beth Israel Adherence Scale. Unpublished doctoral dissertation, Derner Institute of Advanced Psychological Studies, Adelphi University, New York.

Shaw, B. F., & Dobson, K. S. (1988). Competency judgments in the training and evaluation of psychotherapists. Journal of Consulting and Clinical Psychology, 56, 689-695.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-429.

Silberschatz, G. & Curtis, J. T. (1993). Measuring the therapist's impact on the patient's progress. Journal of Consulting and Clinical Psychology, 61(3), 403-411.

Silberschatz, G., Curtis, J. T., & Nathans, S. (1989). Using the patient's plan to assess progress in psychotherapy. Psychotherapy, 26(1), 40-46.

Silberschatz, G., Fretter, P. B., & Curtis, J. T. (1986). How do interpretations influence the process of psychotherapy? Journal of Consulting and Clinical Psychology, 54(5), 646-652.

Smith, M. L. & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32, 752-760.

Snyder, D. K., & Wills, R. M., & Grady-Fletcher, A. (1991). Long-term effectiveness of behavioral versus insight-oriented marital therapy: A four-year follow-up study.

Journal of Consulting and Clinical Psychology, 59, 138-141.

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

Stiles, W. B. (1980). Measurement of the impact of psychotherapy sessions. Journal of Consulting and Clinical Psychology, 48, 176-185.

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

Strupp, H. H. (1986). Psychotherapy: Research, practice, and public policy (how to avoid dead ends). American Psychologist, 41, 120-130.

Strupp, H. H. (1989). Psychotherapy: Can the practitioner learn from the researcher? American Psychologist, 44(4), 717-724.

Strupp, H. H. (1993). Special Feature: The Vanderbilt Psychotherapy Studies: Synopsis. Journal of Consulting and Clinical Psychology, 61(3), 431-433.

Strupp, H. H. & Binder, J. L. (1984). Psychotherapy in a new key: A guide to time-limited psychotherapy. New York: Basic Books.

Strupp, H. H. & Butler, S. F., & Rosser, C. L. (1988). Training in psychodynamic therapy. Journal of Consulting and Clinical Psychology, 56(5), 689-695.

Strupp, H. H. & Howard, K. I. (1992). A brief history of psychotherapy research. In D. Freedheim (Ed.), History of psychotherapy: A century of change pp. 309-334. Washington D.C.: American Psychological Association.

Suh, C. S., O'Malley, S. S., & Strupp, H. H. (1986). The Vanderbilt process measures: The Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS). In L.S. Greenberg & W.M. Pinsof (Eds.), The psychotherapeutic process: A research handbook (pp. 285-323). New York: Guilford Press.

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

Svartberg, M. & Stiles, T. C. (1991). Comparative effects of short-term psychodynamic psychotherapy: A meta-analysis. Journal of Consulting and Clinical Psychology, 59, 704-714.

Svartberg, M. & Stiles, T. C. (1992). Predicting patient change from therapist competence and patient-therapist complementarity in short-term anxiety-provoking psychotherapy. Journal of Consulting and Clinical Psychology, 60, 304-307.

Tabachnick, B. G. & Fidell, L. S. (1996). Using multivariate statistics, 3rd edition. New York: Harper Collins College Publishers.

Tracey, T. & Kokotivic, A. (1989). Factor structure of the Working Alliance Inventory. Psychological Assessment, 1, 207-210.

Turner, A., & Muran, J. C. (1991). Cognitive-behavioral therapy for personality disorders: A treatment manual. Unpublished manuscript, Beth Israel Medical Center, New York.

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.

Wachtel, P. L. (1977). Psychoanalysis and behavior therapy: Toward an integration. New York: Basic Books.

Wachtel, P. L. (1987). Action and insight. New York: Guilford Press.

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(4), 620-630.

Wile, D. B. (1984). Kohut, Kernberg, and accusatory interpretations. Psychotherapy, 21, 353-364.

Winston, A., Flegenheimer, W., Pollack, J., Laikin, M. Kestenbaum, R., & McCullough, L. (1987). A brief psychotherapy fidelity scale: Reliability, validity, and relation to outcome. Paper presented at the meeting of the Society for Psychotherapy Research. Ulm, FRG. June.

Winston, A., Pollack, J., Laikin, M., Samstag, L., McCullough, L., & Muran, J. C. (1994). Short-term psychotherapy of personality disorders. American Journal of Psychiatry, 151, 190-194.

Winston, A., Pollack, J., McCullough, L., Flegenheimer, W., Kestenbaum, R., Trujillo, M. (1991). Brief psychotherapy of personality disorders. Journal of Nervous and Mental Disease, 179(4), 188-193.

Wolpe, J. (1969). The practice of behavior therapy. New York: Pergamon.

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