

Theses and Dissertations

2009

Concordance between therapists' self-reports and observers' ratings of adherence to marital therapy

Arlene A. Cruz

Follow this and additional works at: <https://digitalcommons.pepperdine.edu/etd>

Recommended Citation

Cruz, Arlene A., "Concordance between therapists' self-reports and observers' ratings of adherence to marital therapy" (2009). *Theses and Dissertations*. 37.
<https://digitalcommons.pepperdine.edu/etd/37>

This Dissertation is brought to you for free and open access by Pepperdine Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Pepperdine Digital Commons. For more information, please contact bailey.berry@pepperdine.edu.

Pepperdine University
Graduate School of Education and Psychology

CONCORDANCE BETWEEN THERAPISTS' SELF-REPORTS AND OBSERVERS'
RATINGS OF ADHERENCE TO MARITAL THERAPY

A clinical dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Psychology

by

Arlene A. Cruz

October, 2009

Kathleen Eldridge, Ph.D. – Dissertation Chairperson

This clinical dissertation, written by

Arlene A. Cruz

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirement for the degree of

DOCTOR OF PSYCHOLOGY

Kathleen Eldridge, Ph.D., Chairperson

Joy K. Asamen, Ph.D.

Clarence Hibbs, Ph.D.

Robert A. de Mayo, Ph.D., ABPP
Associate Dean

Margaret J. Weber, Ph.D.
Dean

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	v
ACKNOWLEDGMENTS.....	vi
VITA.....	viii
ABSTRACT.....	xiv
INTRODUCTION.....	1
Methodological Issues in Assessing Treatment Adherence.....	1
Characteristics of Treatment, Therapists, and Clients that Affect Treatment Integrity.....	3
Recommendations for Assessing Treatment Adherence.....	5
Observers vs. Therapist Self-Reports in Assessing Treatment Adherence.....	7
Research Questions.....	10
METHOD.....	13
Participants and Procedures.....	13
Design.....	16
Measures.....	17
RESULTS.....	19
Data Preparation for Analyses of SRT and BCTRM.....	19
Data Analysis with SRT and BCTRM.....	20
Data Preparation for Analysis of SRT and CTRS.....	21
Data Analysis with SRT and CTRS.....	21
Correlation Results.....	22
Correlation between Therapist Self-Reports and Graduate Student Ratings of Adherence.....	22
Correlation between Therapist Self-Reports and Undergraduate Student Ratings of Adherence.....	23
DISCUSSION.....	24
Importance of Findings.....	24
Methodological Limitations.....	24

Previous Studies that Replicate and Contrast with Present Findings.....	26
Implications for Psychotherapy Research.....	27
Implications for Clinical Supervision.....	28
REFERENCES.....	32
APPENDIX A: Literature Review Table.....	36
APPENDIX B: Request for Use of Data from the Original Study.....	58
APPENDIX C: Copy of Measure BCTRM.....	61
APPENDIX D: Copy of Measure CTRS.....	79
APPENDIX E: Copy of Measure SRT.....	85

LIST OF TABLES

	Page
Table 1. Correlations Between Therapist and Graduate Observer Ratings on TBCT Interventions	22
Table 2. Correlations Between Therapist and Graduate Observer Ratings on IBCT Interventions.....	23

ACKNOWLEDGEMENTS

I would like to acknowledge those who contributed to the completion of this project, and the ones who have been with me throughout my doctoral program. First and foremost, I am grateful to Dr. Kathleen Eldridge for her continued help throughout my dissertation. Thank you for offering your expertise, and nurturing support that inspired me to complete this work. I am fortunate to have you as my chairperson, professor, and mentor throughout this journey. I extend my sincere gratitude and appreciation to the committee members. Thank you to Dr. Joy Asamen for helping me develop my skills as a quantitative researcher; and to Dr. Clarence Hibbs for your practical knowledge and fresh perspectives in marital therapy. Both of you ensured that I got my work finished in decent time. I am also grateful to Regina Meister for her writing support, and for ensuring this work is APA compliant. Finally, thank you to Dr. Yuying Tsong for your help with the statistics, and data analysis.

This work would have been an arduous journey without the support of my colleagues. Thank you to my dissertation lab mates, Jessica Nelson and Diane Olarte, who sat with me for hours writing, practicing, studying, and eating. Special thanks to Atorina Benjamin, Natalie Feinblatt, Nina Grayson, and Shana Spangler, who have been my support system during my practicum and externship years. Thank you to the rest of my classmates. You have made me feel welcomed despite joining your class later on in the program—you are all superstars.

Thank you to my friends, who have grown to understand me these past years. Your enduring support provided the focus and concentration I needed throughout this doctoral program. Special thanks to Rowena Gandionco, Candice Urruttia, Rhea Decolongon, and

Dan Definney. Our friendship has been tested, and I am truly blessed to have you still with me.

Last but not least, I would like to acknowledge my family. I offer my special thanks to my dad, mom, Teejay, Anna, and Rae for always being there, and for celebrating this work just because it was important for me. You have been a constant source of support and encouragement ever since my younger years growing up in Manila, Philippines. I feel your love oceans away. *Mahal ko kayo.*

VITA

Arlene A. Cruz
 Email: arlene_06@yahoo.com

EDUCATION

- 09/2004 - Present Doctoral Student (expected completion date 9/2009) Psy.D.
 (A.P.A.-accredited program in clinical psychology)
Pepperdine University, Los Angeles, CA
- 04/2003 Masters of Arts in Psychology (M.A.)
Pepperdine University, Los Angeles, CA
- 03/1994 Bachelors of Arts in Economics (B.A.)
University of Ateneo de Manila, Quezon City, Philippines

CLINICAL TRAINING EXPERIENCE

- 09/2008 – Present Psych Assistant; Intern
Switzer Learning Center, Torrance, CA
 Duties: Facilitated individual therapy for students (5th to 12th
 grade levels) with emotional disturbance (i.e., mood disorders,
 substance-related disorders, pervasive developmental disorders)
 and specific learning disabilities; conduct intensive behavioral
 observation and assessment including several neurological,
 cognitive, behavioral assessment tools and error analysis of
 achievement testing to extract neuropsychological implications;
 facilitate initial evaluations and intakes; prepare and attend
 IEP meetings.
 Supervisor: Owen K. Fudim, Ph.D., ABPP, Clinical Director
- 04/2008 – 09/2008 Residential Counselor; Psych Tech
Malibu Horizon, Malibu, CA
 Duties: Facilitated group therapy in an inpatient alcohol/drug
 rehabilitation center; administer medication and check vital signs
 for safe detoxification; conduct initial evaluations and intakes;
 assist with admissions and discharge, orientation and transfer of a
 resident.
 Supervisors: Clinton Weyand, Psy.D., Clinical Director
 Akikur Mohammad, M.D.
- 08/2007- 8/2008 Student Therapist; Testing Clerkship
Los Angeles County Department of Mental Health

Specialized Foster Care Program (Former: ICAT-Pasadena, CA)

Duties: Provide psychodiagnostic assessment (testing) and intake evaluations of foster children; work with a multi-disciplinary team to make the best placement possible for children who are removed from their homes; receive year-long-seminar experiences on family therapy team training, testing, cultural competence, and other clinical issues.

Supervisor: Tulip Mitchell, Ph.D., Coordinator

09/2004 – 8/2008

Student Therapist

Pepperdine Community Counseling Center, Encino, CA

Duties: Conduct initial evaluations and intakes; formulate treatment plans; provide brief and long-term psychotherapy to adults, adolescents and children from diverse populations in an outpatient setting; attend and present clinical presentations in weekly supervision.

Supervisor: Anat Cohen, Ph.D., Clinic Director

8/2006 – 8/2007

Extern: Student Therapist

Children's Hospital Los Angeles, CA

Duties: Practicum at the Hematology-Oncology, Behavioral Sciences; facilitate smooth school transition for children and teens fighting cancer and blood diseases; worked with the entire healthcare team, patients, families, and schools to identify and manage complex medical and psychological factors that impede learning and social interactions; offered clinical services and direct counseling of children and families, individual therapy, as well as facilitated group therapy for grieving parents.

Supervisor: Betty Gonzalez-Morkos, Ph.D., S.T.A.R. Coordinator

03/2000 – 12/2006

Psychiatrist Assistant

Las Virgenes Behavioral Health & Medical Clinic, Agoura Hills

Duties: Assist in counseling patients suffering from substance addiction and abuse; help in administering MMPI-2 and BDI; provide medication management to patients diagnosed with depression, bipolar disorder, anxiety disorders, PTSD and drug dependence.

Supervisors: Akikur Reza Mohammad, M.D., Medical Director and Vicki Santos, Nurse

10/1998 – 05/1999

Assistant to Clinic Director

Mission Family Clinic, Panorama City, CA

Duties: Volunteer position to assist in counseling Tagalog-speaking veteran patients diagnosed with PTSD and children suffering from various life adjustments; acquired knowledge of medications and ability to read medical charts; participated in

rounds with the physician and assisted in facilitating evaluations of patients.

Supervisor: Dr. Erlinda Grey, M.D., Clinic Director

RESEARCH EXPERIENCE

- 01/2006-09/2006 Research Assistant
Behavioral Sciences
Children's Hospital Los Angeles, CA
 Duties: Investigated the effects of the built environment (hospital) on health outcome of 150 pediatric oncology patients (ages 5 to 18 years), their parents (n=150), and nursing staff (n=175). Collected data in Children's Hospital of Los Angeles' oncology department; administered questionnaires PedsQL Healing Environment Modules, PedsQL Present Functioning Modules, PedsQL Core Scales, PANAS or PANAS-C.
 Supervisors: Sandi Sherman, Ernie Katz, Ph.D., Principal Investigators
- 04/2003 – 04/2004 Research Assistant
Psychology Department
Compton Community College, CA
 Duties: Coauthored Gardening Activity as an Effective Measure in Improving Self-Efficacy and Self-Esteem (*The Community College Enterprise, 2004*). Collected and analyzed data on college students' gardening activity; results of this study showed collaborative work such as gardening helped elicit pro-social behaviors and increase self-efficacy among students of Compton Community College.
 Supervisor: August Hoffman, Ph.D., Principal Investigator, Academic Senate Vice-President ECC- Compton Center
- 06/2003 – 1/2004 Research Assistant
Graduate School of Education & Psychology
Pepperdine University, Encino, CA
 Duties: Closely examined published manuals on adolescent HIV prevention programs and helped identify common themes presented across literatures; studied five evidence-based adolescent HIV interventions, processes and target behaviors in order to identify underlying structural components across HIV interventions. Based on a grounded theory approach, nineteen common process were identified and categorized into structural features, group management strategies, competence building, and addressing developmental issues of adolescence. 10-15 hours per week.

Supervisors: Barbara Ingram, Ph.D. and Mary Jane Rotheram-Borus, Ph.D., Principal Investigators

- 04/2003 – 01/2004 Research Assistant
Southern California Twin Project
Psychology Department, University of Southern California
 Duties: Voluntarily participated in a longitudinal study on the hereditary factors of learning disabilities, aggression and conduct disorders in 600 identical and fraternal twins, ages 9-12 years old; scored and administered WASI, Woodcock-Johnson III, Porteus Maze, Trails A and B, Wisconsin Card Sort, CTOPP, and structured interviews; monitored twins' neuro-physiological responses (brain activities, perspiration, blinking, and heart rates) via electrodes on their scalp, fingers, eyelids and chest, respectively.
 24 hours per week.
 Supervisors: Laura Baker, Ph.D. Southern California Twin Project Director; Dora Lozano, Ph.D., Research Associate

TEACHING EXPERIENCE

- 06/1999 – Present Self-Employed: Private Tutor
 Duties: Administer one-on-one educational sessions in literacy and mathematics at students' homes in Los Angeles and San Fernando Valley areas; implement behavioral techniques for efficient learning and procedures to 5 students (ages 6- to 20-years-old) diagnosed with learning disabilities; prepare students for standardized tests (SAT, SSAT and ISEE) using innovative learning strategies.
- 04/2004 – 5/2005 Remediation Tutor
The Kelter Center, Los Angeles, CA
 Duties: Engaged 3rd-12th grade students in creative tutorial sessions with emphases on reading comprehension, written expression, reading fluency, spelling, test preparation and study skills; reviewed students' psychological reports and cognitive assessments (such as WISC-IV, Woodcock Johnson- III and CTOPP) to create a set of learning goals and objectives; designed lesson plans to operationalize each student's goals.
 Supervisor: Erin Smith, M. Ed., Mathematics Department Head; Sasha Borenstein, Ed. D., Director
- 04/2003 – 04/2004 College Mentor; Teacher Assistant
Compton Community College, Compton, CA
 Duties: Volunteered in a mentoring program designed to help students transfer to four-year universities; assisted students with

their coursework and advised on their academic and personal goals; helped foster a sense of community by working collaboratively with students on a garden project; assisted professor in teaching psychology classes.

Supervisor: August John Hoffman, Ph. D., Professor of Psychology

06/2003 – 01/2004 Teacher Assistant
California State University Northridge, Northridge, CA
 Duties: Voluntarily assisted professor in teaching psychology classes composed of 50 college students each; areas of content included introductory psychology, group dynamics, experimental research and developmental psychology; helped correct students' test results and term papers; conducted tutorial sessions composed of 10-20 students to review concepts prior to tests.
 Supervisor: August John Hoffman, Ph. D., Adjunct Faculty, California State University Northridge

08/1999 – 09/2002 Social Studies Teacher & Literacy Tutor
Westmark School; Learning With A Difference, Encino, CA
 Duties: Taught Social Studies classes to 12th graders; facilitated one-on-one tutorial sessions in math, economics, language arts and English to 3rd to 12th grade students diagnosed with ADD and learning disabilities; created individualized programs specifically tailored for each student's needs.
 Supervisors: Lori Zaragoza, Social Studies Department Head; Diane Seaman, Tutorial Director

WORK EXPERIENCE

06/1998 – 08/1999 Administrative Assistant
*Graduate School of Business and Management
 Pepperdine University, Malibu, CA*
 Duties: Assisted business graduate students in pursuing their academic and career goals; worked with groups of international students to improve their GMAT scores for MBA admissions; directed new student orientation to inform incoming students of school policies; assisted marketing department to recruit prospective business graduate students; acted as a liaison between faculty, staff and students.
 Supervisors: Marcie Braddock, MBA-Preparatory Program Administrator; Louise Pratt, MBA Program Director

06/1995 - 01/1998 Office Manager
RJ Financial, Los Angeles, CA
 Duties: Directly led a personnel of 10 employees; designed an efficient system to reduce company's collection accounts; interacted with customers on a daily basis; interviewed, selected and oriented newly hired employees to company's policies and procedures.
 Supervisor: Aileen Ablan, Personnel Supervisor

PUBLICATIONS

Hoffman, A. J., Cruz, A. & Thompson, D. (2004). Gardening activity as an effective measure in improving self-efficacy and self-esteem: Community college students learning effective living skills.
The Community College Enterprise.

PROFESSIONAL AFFILIATIONS

10/2008 – Present California Latino Psychological Association
 01/2004 – Present American Psychological Association
 01/2004 – Present American Psychological Association
 Division 40 - Clinical Neuropsychology
 08/2007 – Present American Association for Marriage and Family Therapy

ABSTRACT

In the context of psychotherapy research, investigators often assume that the assessment of adherence to treatment protocols should be done by trained observers, who are viewed as more neutral or objective than the therapists themselves. The aim of this present study is to check the concordance between therapist self-reports and observer ratings of adherence to marital treatment (IBCT and TBCT). The data for the current study were obtained from an archive of adherence data for 35 randomly selected therapy cases, collected by Andrew Christensen and colleagues (2004) in the context of a large clinical trial of marital therapy.

For both the TBCT and IBCT interventions, there was a consistent and high concordance between the therapist self-reports and observer ratings, suggesting that therapists accurately self-reported the same interventions seen by the observer raters. Results of this study challenge the widely accepted notion that observer ratings are superior to therapist self-evaluations. Present findings reveal that therapist self-reports on adherence to marital treatment can prove comparable to the revered “gold standard” observer reports, and can serve as a valuable supplement to other adherence ratings. Therapist self-reports are not only cost-efficient, but can also provide a unique perspective in understanding nuances of psychotherapy that often go unnoticed by distant observer raters. Ways to maximize the accuracy and reliability of therapist self-reports are discussed.

INTRODUCTION

Assessing the degree to which therapists adhere to manual-specified interventions is an ongoing challenge in treatment outcome studies. Knowing if a specific treatment intervention is delivered with integrity is important in determining what constitutes real change in treatment (Gresham, MacMillan, Beebe-Frankenberger, & Bocian 2000; Moncher & Prinz, 1991; Kazdin, 1986). As one of the three components of treatment integrity, adherence is defined as the degree to which the therapist utilizes specified procedures while foregoing other treatment techniques (Margison, Barkham, & Evans, 2000; Waltz, Addis, Koerner, & Jacobson, 1993). Knowing if a therapist adhered to the prescribed treatment intervention is an essential aspect of treatment integrity, without which the results obtained would likely be of questionable value in evaluating treatment effectiveness (Gresham et al., 2000; Waltz et al.).

Methodological Issues in Assessing Treatment Adherence

A number of methodological issues arise when conducting research on therapists' adherence to treatment. First, there is a lack of uniformity across studies in the methods used to measure adherence (Heaton, Hill & Edwards, 1995; Waltz, et al., 1993). For example, methods have included a checklist in which the rater indicates the occurrence-nonoccurrence of interventions that are prescribed and proscribed by the treatment protocol (Waltz et al.), ratings of frequency or extensiveness of interventions that are delivered, as well as core elements of the interventions that were present in the session (Luborsky, Woody, McLellan, O'Brien, & Rosenzweig, 1982), and ratings on both the adherence and the quality of the interventions delivered (O'Malley, Foley, Rounsaville & Watkins, 1988; Wills, Faitler, & Snyder, 1987). Second, the measures differ in

complexity, particularly in the level of training required of neutral observers and therapists, which poses a significant challenge in obtaining an accurate assessment of adherence (Dobson & Singer, 2005; Kazdin, 1986). Less sophisticated measures usually do not discriminate between different treatment modalities, and fail to differentiate critical dimensions of treatment interventions (Kazdin; Miller & Binder, 2002). Adherence rating scales that are carefully designed to measure unique treatment components can help discriminate between various treatment models, and serve as a necessary tool in honing in on what treatment component facilitated the change (Kazdin). Third, the source of data varies, which can affect the ratings collected (Waltz et al.). For example, ratings may be based on therapist self-report or examination of process notes, transcripts of therapy sessions, and audiotaped or videotaped sessions (Miller & Binder; Waltz et al.). According to Waltz et al., the videotaped session is the most effective source of data, as videotapes offer the most comprehensive information about a therapy session. Finally, another challenge in assessing treatment adherence is the different ways in which the unit of analysis is defined. Some studies, for example, use segments of sessions while others code entire sessions in their studies (Gresham, MacMillan, Beebe-Frankenberger & Bocian, 2000; O'Malley et al., 1988). The sampled period of time may not accurately represent all of the therapy sessions. Randomly selecting periods of observation is necessary to ensure that no differences occur across the available periods of assessment (Kazdin; Wilkinson, 2000).

Given the diverse approaches to assessing adherence, it would be interesting to note how these methods correlate. In a study by Heaton, Hill and Edwards (1995), two methods for measuring treatment adherence were examined. The molecular method

determines how often a therapeutic technique is implemented by analyzing transcripts sentence-by-sentence. The second method, which is called the molar or global method, estimates the frequency of use of techniques across an entire session or segment. The data for the two molar measures used were found to be positively related, but the results were different when correlated with the molecular data, indicating that different methods for measuring adherence do not always produce similar findings (Heaton et al.). The purpose of this study is to determine the concordance between two methods of measuring therapist adherence in delivering marital therapy, one utilizing therapist self-report and the other based on neutral observer ratings.

Characteristics of Treatment, Therapists, and Clients that Affect Treatment Integrity

In treatment outcome studies, characteristics of the treatment, therapists and clients influence the degree of treatment integrity (Perepletchikova & Kazdin, 2005). Treatment characteristics, such as the number of treatment agents, required resources and the duration of the treatment intervention, may potentially compromise treatment integrity. Complex treatments that require various treatment agents other than the therapist may be at risk for procedural degradation. The collaboration of numerous treatment agents, such as a client's spouse and family members, may pose as a source of variation in treatment integrity. Literature shows that the more treatment agents involved in the process, the higher the probability of failure to follow the specified protocol (Perepletchikova & Kazdin). Still another challenge in integrity is that certain treatment approaches require materials and resources that may not be readily available or cost effective. Treatments that require special resources such as expensive technical equipment and supplies tend to be delivered with less integrity depending on cost issues

and availability of these special resources (Gresham, et al., 2000; Perepletchikova & Kazdin). The duration of therapy also influences the degree of treatment integrity. Longer in-depth therapies may require greater variation in treatment protocol than brief symptom-focused therapies (Perepletchikova & Kazdin). For example, in a study of supportive expressive dynamic therapy, early symptomatic improvement predicted higher adherence to the specified treatment protocol than less symptomatic improvement (Barber, Crits-Christoph, & Luborsky, 1996). In other words, the more the patient showed immediate benefits from treatment, the easier it was for the therapist to adhere to the treatment manual (Barber et al.), suggesting that treatments resulting in quick symptom-relief are delivered with greater integrity than slower-acting treatments (Barber et al.).

The characteristics of clients and therapists must also be taken into account when studying treatment integrity (Waltz et al., 1993). For example, therapists are also more likely to follow through with set protocol if they see that the client is invested in the process (O'Malley, et al., 1988; Patterson & Chamberlain, 1994). In addition, the more a client is perceived to benefit from treatment, the easier it is for the therapist to adhere to the specified treatment protocol (Barber, et al., 1996; Perepletchikova & Kazdin, 2005). Client's characteristics such as resistance and defensiveness may cause therapists to be more emotionally distant and less likely to administer specified treatment protocols (Patterson & Chamberlain). Other client characteristics, such as client's anger, hostility and how severe or long-standing his or her problems are, may affect therapist adherence to a prescribed technique (O'Malley et al.; Perepletchikova & Kazdin).

Level of therapist training and experience may or may not impact adherence. For example, Fals-Stewart and Birchler (2002) studied the delivery of behavioral couple therapy (BCT) to alcoholic men and their partners. The authors studied the BCT delivery of paraprofessionals (bachelor's-level) and professionals (master's-level; Fals-Stewart & Birchler). Regardless of one's level of experience, findings suggested that both paraprofessionals and professionals adhered closely to the BCT manual (Fals-Stewart & Birchler). Others have found that therapists with more experience tend to integrate other techniques and are less likely to adhere to a specified treatment protocol (Margison, et al., 2000), and that training and supervision tend to solidify therapists' working styles and hinder new learning (Henggeler, Schoenwald, & Liao, 2002; Henry, Strupp, Butler, Schacht & Binder, 1993). However, in the body of literature reviewed by Miller and Binder (2002), manual-based training promotes adherence overall, and yields an increase in competence (Siqueland et al., 2000).

Perhaps more important than the therapist's level of experience is finding out under what conditions different therapists perform most competently (Christensen & Jacobson, 1994). For example, professionals, who tend to be older than the paraprofessionals, perform slightly better delivering briefer treatments to older patients, whereas paraprofessionals were slightly better when working in longer treatments with younger patients (Berman & Norton, 1985).

Recommendations for Assessing Treatment Adherence

Recommendations for improving the quality of treatment adherence assessment include controlling for confounding variables such as characteristics of therapists and clients (Perepletchikova & Kazdin, 2005). Quality supervision and monitoring adherence

via utility of videotaped sessions is recommended (Miller & Binder, 2002; Perepletchikova & Kazdin). Research has shown that supervision and monitoring treatment delivery helped reduce therapeutic drift and contributed to increasing adherence to the specified treatment protocol (Miller & Binder).

Waltz et al. (1993) suggest some features that should be considered in developing an effective adherence measure. First, the measure should include therapist behaviors that are specific to the treatment modality being studied. Second, the measure must include behaviors that are compatible with the modality that will be studied but are neither necessary nor unique to it. These would be interventions found in Treatment A if administered appropriately, but may also be present in Treatment B. Third, in order to detect reduced dosage or potency, adherence measures must include therapist behaviors such as chatting with the client at the beginning of the sessions. Lastly, in order to detect possible protocol violations, the measure should include behaviors that are proscribed by the treatment modality. If an adherence measure confined itself to essential and unique items only, the treatment distinctiveness may be overestimated.

Rater biases such as availability and anchoring heuristics can occur when using neutral raters or therapist self-reports, but may be more likely when using a global method of adherence assessment (Heaton, et al., 1995). An availability heuristic occurs when raters overestimate the occurrence of an intervention because they are basing their ratings on ease with which the intervention can be brought to mind as opposed to actual reality. This bias may lead raters to overestimate the occurrence of a particular intervention because it is mentally available. The anchoring heuristic happens when rater assessments are based on an initial bias or a preconceived notion that is used as the

criteria set when rating the therapy sessions (Heaton et al.; Tversky & Kahneman, 1974). The global method may be more sensitive to cognitive biases because the raters are required to infer and judge large chunks of data with no firm anchors for the scale points. Because the global method considers the entire therapy session as the unit of analysis, it is less precise compared to more operational methods (i.e., molecular method), which measure explicit content (i.e., grammatical sentence structure) that leaves less room for personal interpretation (Heaton et al.).

Observers versus Therapist Self-Reports in Assessing Treatment Adherence

Research has found that ratings of adherence vary by informant, as each informant has his or her own way of viewing psychotherapy (Huey, Henggeler, & Brondino, 2000; Mintz, Auerbach, Luborsky & Johnson, 1973; Waltz et al., 1993). Most studies have gathered data on treatment adherence using observer ratings rather than affiliated raters (such as therapist self-reports) as it is assumed that there is less concern with rater bias (Waltz et al.; Xenakis, Hoyt, & Marmar, 1983). This may be a reasonable assumption if the treatment manual is comprehensive, and if the observers are properly trained (Luborsky, et al., 1982; Waltz, et al.).

On the other hand, therapist self-reports offer a unique perspective that is not available to neutral observers. Observers tend to be less sensitive to nuances of treatment and interventions that are not explicitly described in the treatment manual (Carroll, Nich, & Rounsaville, 1998). Furthermore, observers often do not view the entire session but base their ratings on a small segment of the session, rely on videotaped sessions that may have poor sound quality, and are removed from the therapeutic process, and therefore may miss subtle interventions that are not evident in videotapes (Carroll et al.).

Studies of treatment outcome that rely on adherence ratings provided by therapists themselves are considered problematic by many researchers due to potential for bias (Jacobson & Addis, 1993; Moncher & Prinz, 1991; Waltz et al., 1993). For example, because some therapists may wish to portray themselves as adhering to the treatment manual more closely than they actually do, the use of therapist self-reports may result in inflated treatment integrity levels (Perepletchikova & Kazdin, 2005; Waltz et al.).

One study that used therapist self-reports of treatment adherence yielded conflicting results when correlated with observer ratings (Carroll et al., 1998). Therapists reported delivering 7 out of the 10 techniques they were asked to rate, which was not corroborated by the observer raters (Carroll et al.). These findings imply that different raters offer varied perspectives on adherence that may not be interchangeable (Carroll et al.; Mintz et al., 1973). Low agreement among different raters (i.e., therapist, external rater and patient) implies that each rater has his or her own perspective on the psychotherapy process (Mintz et al.). The problems in obtaining poor agreement regarding the treatment process may also be due to factors other than differences in training, role, personal distortions or motives (i.e., unreliability of an item in the measure; Mintz et al.). Similarly, Hill, O'Grady, and Price (1988) indicate that treatment adherence has varied dimensions, and is best understood when evaluated from multiple perspectives.

However, in a study by Singer (2002), therapists rated their own adherence to the Multidimensional Family Prevention treatment, as did observer raters. Intraclass correlation coefficients fell within the moderate to strong range, implying that observers and therapists shared a similar perspective of treatment adherence (Singer). These

findings imply that therapist self-report can be a reliable source of information regarding treatment adherence (Singer).

The Value of Therapist Self-Reports in the Assessment of Treatment Adherence

Although there is some disagreement as to the reliability of using therapist self-reports in assessing treatment adherence, there are clearly some benefits that should not be ignored. For example, Carroll et al. (1988) found that by asking therapists about the treatment they were delivering, they were able to identify therapeutic techniques that were more clinically effective as well as more frequently employed, which may help in refining the intervention offered. Furthermore, therapist self-reports may also provide information that can enhance the training of therapists as well as contribute to understanding the in-session dynamics between therapist and client (Carroll et al.).

Therapists' contributions are essential to understanding the subjective intentionality behind the therapeutic intervention that is not captured in video- and audio recordings (Xenakis et al., 1983). Therapists have unique access to many observational sources, such as nonverbal cues and personal reactions that may influence treatment delivery (Xenakis et al.). Moreover, therapist self-reports are an inexpensive alternative to the much more involved process of garnering objective ratings. Therapist self-reports are more cost efficient and easily attained, as therapists immediately fill out the measure after the therapy session (Carroll et al., 1998; Xenakis, et al.).

To address the issue of therapist bias, Carroll et al. (1988) suggests a protocol in which therapist self-ratings completed immediately after a session are associated with ratings made by the therapists of their videotaped session at a later point in time. Also suggested is the use of reinforcements to encourage therapists to adhere to treatment, and

provide more candid assessments of their treatment adherence. Finally, Carroll et al. suggest that therapist self-report ratings can be improved through extensive training of therapists, use of detailed treatment manuals, reviewing the accuracy of therapist self-reports, and evaluating sources of disagreement between therapists and neutral observers (Carroll et al.).

Research Questions

The present study sets out to explore couple therapists' self-perceived adherence to two marital treatments, Traditional Behavioral Couple Therapy (TBCT) and Integrative Behavioral Couple Therapy (IBCT) by correlating their ratings with those of neutral observers. Due to criticisms about the reliable use of therapist self-reports for measuring treatment adherence, these ratings tend to be underused. Furthermore, in a study done by Carroll et al. (1998) that correlated therapist self-reports to the ratings conducted by neutral observers, there appeared to be an overall poor agreement between the therapists and the raters regarding the specific interventions delivered. Yet there are many valuable reasons to utilize therapist self-ratings of adherence, as it offers a different perspective of the psychotherapy process (Carroll et al.; Xenakis et al., 1983).

This present study hopes to determine the reliability of therapist self-reports by correlating them with observer ratings of treatment adherence. It does not examine the relationship between therapist adherence and treatment outcome that most outcome studies explore (Hogue et al., 2008). It is important to determine if concordance exists between therapist self-reports and neutral observer ratings in order to further strengthen findings, and/or add to the unique facets of the therapeutic process. A high correlation would imply that therapist self-reports may be used as a reliable tool for measuring the

therapeutic process, and researchers may be able to choose the more cost-effective therapist self-reports instead of using neutral observer ratings. On the other hand, low concordance would imply that therapist self-reports offer a different perspective of the therapeutic process and should be used as a supplement to observer ratings, in order to effectively hone in on what produces change (Gresham, et al., 2000). Therefore, the current study might point to the necessity for multiple adherence measures, and prompt future research on the factors associated with varying levels of concordance among multiple measures.

In addition, the current study utilizes two distinct groups of external observers, naïve undergraduates and informed graduate students. Undergraduate students were chosen based on the rationale that they were less informed on marital therapy compared to the graduate students. It is assumed that having no prior knowledge of IBCT or TBCT would account for more neutrality as raters, and thus generate more unbiased findings. On the other hand, the informed graduate students can bring value-added expertise, and they can readily identify the treatment conditions while using a more detailed and thorough measure. These two groups of observers also differ in the coding system used. The graduate students used a molecular or more detailed system of coding, whereas the undergraduate students used a more generalized rating system. The two observer groups also differ in their method of review. The graduate raters watched videotaped sessions while the undergraduate raters listened to audiotaped therapy sessions. Using two distinct groups of observers allows for a more robust design to formulate assumptions and/or conclusions appropriate to each observer group.

The research questions that will be investigated in this study are:

1. What is the correlation between therapists' self-reports and graduate students' molecular ratings of adherence?
 - a. What is the concordance on Traditional Behavioral Couple Therapy interventions, including Behavior Exchange, Communication Training, and Problem Solving Training?
 - b. What is the concordance on Integrative Behavioral Couple Therapy interventions, including Empathic Joining, Unified Detachment, and Tolerance?
2. What is the correlation between therapists' self-reports and undergraduate students' global ratings of adherence?
 - a. What is the correlation between the therapists' ratings of change-focused interventions including Behavior Exchange, Communication Training, and Problem Solving Training, and the undergraduates' global ratings of Change Interventions?
 - b. What is the correlation between the therapists' ratings of acceptance-oriented interventions including Empathic Joining, Unified Detachment, and Tolerance, and the undergraduates' global ratings of Acceptance interventions?

METHOD

The current study's participants were obtained from an archive of data collected by Andrew Christensen and colleagues (2004). The methods and procedures that follow are pertinent to the present study. For further information on the original study, please refer to Christensen et al.

Participants and Procedures

Couples

A total of 134 heterosexual married couples were recruited in Los Angeles and Seattle for a clinical trial of marital therapy. The participants were solicited through radio, newspaper, television advertisements, as well as letters and brochures sent to clinics and practitioners, describing the study in detail. The study was conducted simultaneously at the University of California, Los Angeles (UCLA) and the University of Washington with about half of the participants recruited at each site. The study sample had a mean age of 41.62 years for wives ($SD=8.59$), and 43.49 years for husbands ($SD=8.74$). Wives and husbands had a mean number of 16.97 ($SD=3.23$) educational years (including kindergarten). Couples were married an average of 10 years ($SD=7.60$) with a median of 1 child ($SD=1.0$). The participants were predominantly Caucasians (husbands: 79.1%, wives: 76.1%). The other ethnic groups included African Americans (husbands: 6.7%, wives: 8.2%), Asians or Pacific Islander (husbands: 6.0%, wives: 4.5%), Latino or Latina (husbands: 5.2%, wives: 5.2%), and Native American or Alaskan Native (husbands: 0.7%).

The participants were screened in a three-phase process via telephone interview, mailed questionnaires, and a pre-treatment in-person assessment session to assess

whether they met inclusion and exclusion criteria. Couples were required to voluntarily seek out couple therapy, be legally married, and in severe and chronic marital distress, which was assessed by a score of at least one standard deviation below the population mean ($SD < 98$) on the Dyadic Adjustment Scale, and a *T score* of 59 or higher on the Global Distress Scale. Participants had to be between the ages of 18 and 65, have a minimum high school education, and be fluent in English. Participants who were currently in psychotherapy or met current criteria for substance dependence, moderate to severe domestic violence, or severe psychopathology were excluded from the study.

Couples that met all inclusion criteria and consented to participate were randomly assigned to one of the two treatment conditions, TBCT or IBCT. The TBCT group had a total of 68 couples, while the IBCT group had a total of 66 couples. Each couple received up to 26 free treatment sessions. The average number of sessions reached 22.9 ($SD = 5.35$) that occurred over an average period of 36 weeks. For the present study, 35 out of the 134 couples were randomly selected (stratified across time, therapists, and treatment conditions) for observational adherence coding.

Therapists

Four doctoral-level clinical psychologists in Los Angeles and three in Seattle were selected on the basis of their reputation and expertise in the field. The therapists had 7 to 15 years of experience post-licensure. They received extensive training and supervision to monitor adherence and competency in treatment delivery. Before treating cases, the therapists studied the treatment manuals, and they attended workshops led by Andrew Christensen and Neil Jacobson.

The therapists received supervision in both TBCT and IBCT from Christensen and Jacobson, two experts in TBCT and IBCT. They also received additional supervision from Peter Fehrenbach, a therapist in the pilot study of TBCT and IBCT (Jacobson, Christensen, Eldridge, Prince, & Cordova, 2000), and Don Baucom, a published expert on TBCT. All seven therapists submitted audio and videotaped sessions to their supervisors. The therapists received weekly feedback prior to their next session either through telephone or email.

External Raters

Graduate students. Three advanced graduate students were selected to rate the therapy sessions for adherence, using the Behavioral Couple Therapy Rating Manual (BCTRM) described in the preliminary pilot study of TBCT and IBCT (Jacobson et al., 2000). At least one early, one middle, and one late therapy session were randomly selected from each of the 35 selected cases. For training purposes, a few of the cases had additional sessions rated, ranging from four to eight, yielding a total of 115 sessions. The graduate students viewed the videotaped sessions, and made independent ratings at the end of each session.

Undergraduate students. Eleven undergraduate students who had no knowledge of IBCT or TBCT were also selected and extensively trained as adherence raters. Since the graduate-level raters were familiar with TBCT and IBCT, and were able to recognize the treatment conditions, the undergraduate raters who were unfamiliar with TBCT and IBCT were selected in order to prevent rater bias. The undergraduates were told that they were examining the correlates of different types of therapist activities. They had weekly training sessions with an advanced graduate student, who consistently checked the inter-

rater reliability throughout the coding process. Prior to allowing them to independently rate, the undergraduates were required to consistently reach acceptable levels of inter-rater reliability on sessions selected for training purposes.

The undergraduate observers rated the same 35 cases for adherence, using the Couple Therapist Rating Scale (CTRS; Jacobson et al., 2000). At least two early, two middle, and two late therapy sessions were coded by the 11 undergraduate observers, yielding a total of 208 sessions (101 TBCT sessions and 107 IBCT sessions; Christensen et al., 2004). The undergraduate raters listened to audiotapes of the sessions, and made independent overall ratings at the end of each session (Jacobson et al.).

Design

This is a correlational study done to establish the relationship between the therapists' self-reports (the Sessions Ratings by Therapist; SRT), and the observer ratings of adherence (the graduate students' Behavioral Couple Therapy Rating Manual; BCTRM, and the undergraduate students' Couple Therapist Rating Scale; CTRS). The predictive variable will be the observer ratings, while the outcome variable will be the therapist self-ratings. The concordance rate will demonstrate the strength of the relationship between these two variables. This study's primary goal is to simply examine the concordance between the raters, and does not suggest or assume in any way that high adherence leads to better treatment outcome. Furthermore, this study examines relationships between observer ratings and therapist self-reports, while associations between undergraduate and graduate observer ratings are beyond the scope of this dissertation.

Measures

Behavioral Couple Therapy Rating Manual (BCTRM)

The Behavioral Couple Therapy Rating Manual (BCTRM) is an adherence scale that was developed for the original study of TBCT and IBCT (Jacobson et al., 2000). The BCTRM (see Appendix C) consists of 28 items measured on a 5-point Likert scale ranging from 0 (*not at all*) to 4 (*extensively*). These 28 items include 8 TBCT-content items reflecting change-oriented behavioral marital therapy (i.e., Behavior Exchange, Communication Training), such as *Therapist taught or initiated practice of active listening or expressive communication skills*. Out of the 28 items, 9 items reflect IBCT's acceptance-based interventions (i.e., Empathic Joining, Unified Detachment). For example, an IBCT item describing unified detachment is: *Therapist explored reasons why partners might differ regarding preferences for intimacy, time alone, need for reassurance, ways of showing affection, etc.* Additional items proscribed in either intervention included items such as Cognitive Interventions and Genogram. Lastly, interventions compatible with either treatment modality are also included in the measure such as Ordinary Conversation and Reviewing Homework.

After watching a session, the graduate student observers independently rated the extent to which each of the 28 items occurred during the session. Two major summary scores of IBCT interventions and TBCT interventions were taken, and alpha reliabilities computed across coders were .93 and .97, respectively (Christensen et al., 2004).

Couple Therapist Rating Scale (CTRS)

This adherence rating manual was also used in the original study of TBCT and IBCT (Jacobson et al., 2000). The CTRS (see Appendix D) consists of four 9-point scales

ranging from 1 (*does not instigate change/does not promote acceptance*) to 9 (*instigates change very much/extensively promotes acceptance*). The four items assess the extent to which the therapists (a) set and followed an agenda, (b) engaged in change-oriented strategies, (c) engaged in acceptance-based strategies, and (d) assigned and checked homework (Christensen et al., 2004; Jacobson et al., 2000).

In this global rating system, the code for *instigating change* consisted of any of the change-oriented interventions defined in the earlier molecular system used by the graduate-level observers. Similarly, the code for *acceptance* consisted of any of the acceptance-oriented interventions in the earlier molecular system. Having no knowledge that they were rating two different forms of therapy, the undergraduate raters listened to audiotapes of the sessions, and made an independent overall rating at the end of each session (Jacobson et al., 2000). The four global ratings yielded the following alpha reliabilities: .75 for agenda, .93 for change, .92 for acceptance, and .83 for homework.

Sessions Ratings by Therapist (SRT)

This self-rating adherence measure (see Appendix E) includes a checklist of specific treatment interventions that the therapists used to indicate whether they did or did not use each intervention listed during the session. Three of the items represent TBCT interventions (Behavior Exchange, Communication Training, and Problem Solving) and three of the items represent IBCT interventions (Emphatic Joining, Unified Detachment, and Tolerance Intervention).

The seven therapists completed the SRT immediately following every session. Although therapists completed this measure after each session, only the data from the sessions that were also rated by the observers will be used.

RESULTS

Data Preparation for Analyses of SRT and BCTRM

The dichotomous yes/no variable in the *Session Ratings by Therapist (SRT)* measure was assigned a value of 1 to the *yes* category (when therapists checked the item to indicate they had delivered the intervention), and a value of 0 to the none event or *no* category (if the item was not checked by therapists, it indicates that they did not deliver the intervention).

Four of the therapist self-rating items have a one-to-one correspondence with items in the coding system utilized by the graduate student observers. The observer ratings from these items (Behavior Exchange, Communication Training, Problem Solving Training, and Unified Detachment) can fall anywhere within the range of zero to four. The lowest rating attained was a zero (0 = *not at all*), implying that the treatment procedure was not utilized. The highest rating was a four (4 = *extensively*), signifying that the treatment procedure was extensively used.

Two of the therapist self-report items (Empathic Joining and Tolerance Intervention) correspond to multiple items in the coding system utilized by graduate student observers. Specifically, the therapist self-report Empathic Joining item corresponds to three items in the coding system: Problems as Differences, Reasons for Partner Differences, and Soft Disclosures (item numbers 10, 11, and 15, respectively). These three items were summed. The lowest sum that was attained was a zero with the observers rating zero (0 = *not at all*) on all three scales. The highest sum attained was a total of twelve, with the observers rating 4 (4 = *extensively*) for all three of the Empathic Joining treatment procedures (4 x 3 = 12). Similarly, the therapist self-report Tolerance

Intervention item corresponds to five items on the coding system: Preparing for Slip-ups and Lapses, Positive Features of Negative Behavior, In-Session Rehearsal of Negative Behavior, Instructing Couple to Fake Negative Behavior at Home, and Self-Care (items 19, 20, 22, 23, and 24, respectively). These five items were summed. Zero was the lowest sum attained, indicating that the raters did not observe Tolerance Intervention (0 (*not at all*) x 5 items equal 0). The highest sum attained was 20 (= 4 x 5) with the raters observing the treatment procedure extensively (4 = *extensively*).

The items within and between the two measures (the therapists' SRT and the graduate student raters' BCTRM) were matched according to the theoretical concepts of IBCT and TBCT. There was no factor analysis or content analysis done when matching items within or across measures.

Data Analysis with SRT and BCTRM

Two measures of correlation were used in this present study. A point-biserial correlation was the initial method used to measure the degree of relationship between the dichotomous therapist self-ratings and the scale scores of the observer ratings. Second, item analyses were conducted to see what point on the scale scores clearly differentiated between the *yes* and *no* therapist ratings. Once this point of discrimination was determined, it was used to divide the scale scores into two sections. These two sections were recoded into a 0 for a *no*, and a 1 for a *yes*. Finally, a kappa correlation was calculated to see the extent to which the dichotomous graduate ratings are in concordance with the dichotomous therapist ratings.

Data Preparation for Analysis of SRT and CTRS

The CTRS as rated by the undergraduates has two global items: *Instigating Change* (item 2) and *Displaying or Promoting Acceptance and Understanding of Client Behavior and Feelings* (item 3). Both use a 9-point Likert scale (1 = *did not happen* and 9 = *happened extensively*). The CTRS' item *Instigating Change* corresponds to three items on the therapist self-reports (Behavior Exchange, Communication Training, and Problem Solving Training). If the therapist did not check any of these items (indicating that they did not deliver any of the Change interventions), this was coded as a 0. If the therapists checked any of these three items (indicating that Change interventions were delivered) this was coded as a 1.

Similarly, the CTRS item *Displaying or Promoting Acceptance and Understanding of Client Behavior and Feelings* corresponds to three items on the therapist self-reports (Empathic Joining, Unified Detachment, and Tolerance Intervention). As above, we coded 0 if none of the items were checked by the therapist. If any of these items were checked, this was coded as a 1.

Data Analysis with SRT and CTRS

Similar to the analyses described above, a point-biserial correlation was initially used to measure the degree of relationship between the dichotomous therapist self-ratings and the scale scores of the observer ratings. Second, an item analysis was then conducted to determine where on the undergraduate scale (1 to 9) clearly demarcated between the 1 and 0 ratings of the therapists. After determining this point of discrimination, the undergraduate ratings were divided into two sections. These two sections were then recoded into a 0 for a *no*, and a 1 for a *yes*. Finally, a kappa correlation was applied to

see the extent to which the dichotomous undergraduate ratings were in concordance with the dichotomous therapist ratings.

Correlation Results

Cohen (1988) and Hemphill, Simon, Burnaby & Canada (2003) gave the following guidelines for interpreting strength of correlation coefficients: .10 = weak; .30 = medium; and .50 = strong. For both the TBCT and IBCT interventions, there is a consistent and high concordance between the therapist and observer (both graduate and undergraduate student raters) ratings.

Correlation Between Therapist Self-Reports and Graduate Student Ratings of Adherence

There was a strong positive relationship between the therapist reports and the graduate ratings on TBCT interventions (see Table 1).

Table 1
Correlations Between Therapist Reports and Graduate Observer Ratings on TBCT Interventions (n=104)

TBCT Interventions	<i>r</i>	<i>p</i>	<i>kappa</i>	<i>p</i>
Behavior Exchange	.61	.00	.56	.00
Communication Training	.66	.00	.66	.00
Problem Solving	.79	.00	.79	.00

For the IBCT interventions, the correlations between the therapist and graduate ratings ranged from weak to strong positive relationships (see Table 2).

Table 2
Correlations Between Therapist Reports and Graduate Observer Ratings on IBCT Interventions (n=104)

IBCT Interventions	<i>r</i>	<i>p</i>	<i>kappa</i>	<i>p</i>
Empathic Joining	.49	.00	.50	.00
Unified Detachment	.55	0.00	.59	.00
Tolerance Intervention	.25	0.01	.21	.03

Correlation Between Therapist Self-Reports and Undergraduate Student Ratings of Adherence

For the change-focused interventions, there was a significant correlation between therapist self-reports and observer ratings ($r = .85$, $n = 187$, $p = .00$; $kappa = .40$, $p = .00$). For the acceptance-oriented interventions, a strong positive relationship also existed between the therapist and observer ratings ($r = .84$, $n = 187$, $p = .00$; $kappa = .40$, $p = .00$).

In sum, there was only one weak positive relationship (therapist and graduate observer ratings on IBCT's tolerance intervention), and the rest of the findings showed an overall strong positive relationship. In general, high correlations existed between the therapist and the observer ratings (for both graduate and undergraduate student raters), suggesting that the therapist reports were in concordance with the interventions observed by observer raters.

DISCUSSION

The results confirm that the therapist self-reports were in concordance with the observer ratings. Strong positive correlations existed between the therapist reports and observer ratings, suggesting that the therapist endorsed the same interventions seen by the observers. This finding is important as it suggests that therapist self-report measures can stand on their own, and are as reliable as the frequently used observer ratings.

Out of these significant findings, there was only one weak positive relationship. Relatively lower agreement in IBCT's Tolerance Intervention could be due to the possibility that this technique may seem to appear in both TBCT and IBCT. For example, a therapist may endorse Tolerance Intervention's self-care strategies (IBCT; designed to foster acceptance and increase tolerance of other's upsetting behavior) while to the more distant observer, the therapist may appear to be instructing specific skill-building exercises (TBCT; more change-oriented). These findings highlight the therapists' unique access to the subjective intentionality behind the intervention often undetected by observers, and not captured in video- and audio recordings sources.

The present study has some limitations that must be considered in interpreting the results. First, it is difficult to generalize the results because of the study's specific sample. In terms of diversity, for example, only 20% of the sample represented ethnic groups. Our sample may not accurately represent the population of ethnic groups in Southern California, wherein the largest ethnic group (Hispanics alone) represent 40.57% (Kirchner & Chen, 2008). Second, most of the participants were highly educated and distressed couples, which limited the generalizability of the findings to other couples. In addition, because the present study used data that was part of a large, ongoing clinical

trial, only a subset of randomly selected sessions were rated for adherence. Concordance between therapists and raters on these particular sessions may not represent concordance on all other sessions.

Furthermore, it is important to note that the present study's purpose was to add to the literature about research methodology. Our high concordance rates are specific to controlled clinical trials. For example, the therapists recruited for the study had 7 to 15 years of experience post-licensure, and were carefully trained and closely supervised. In addition, they were chosen for their prior professional relationship with the principal investigators, which meant they had strong research backgrounds, possibly a high regard for research, had been taught to develop thorough conceptualization of clients, and demonstrated investment in this study with their willingness to comply with all procedures (taping each session, intense supervision, completion of measures, etc.). Therefore, high concordance between our raters and therapists may not represent concordance between other kinds of pairings. Lastly, the present study does not examine the relationship between adherence and treatment outcome (McMurray, 2007; Hogue et al., 2008), but instead provides the preliminary results for further research on adherence.

Regarding the adherence measures, using three measures provides access to various perspectives of the same concept, allowing us to corroborate ratings between three sources of data (therapists, graduate and undergraduate raters). However, one limitation is that the adherence measures have different item rating scales; the therapists used a 2-point (Yes or No) scale, whereas the graduate and undergraduate observers used 5-point and 9-point scales respectively. Similarly, the graduate and undergraduate ratings may not directly complement each other due to different coding systems used; the

graduate raters used a molecular system while the undergraduate raters used a global or molar system.

Nonetheless, our consistent findings that therapist self-reports are in high concordance with observer ratings replicate the findings of the study conducted by Singer (2002). Correlations falling within the moderate to strong range suggest that observers and therapists shared a similar perspective of treatment adherence (Singer). Similarities between Singer's and the present study may have contributed to the paralleled results. First, both studies used therapists and observers with similar educational background, extensive training, and supervision. Another contributing factor could be that the measures used in both studies had items on specific therapist behaviors and skills (e.g., the therapist assigned homework, the therapist reinterpreted negative behavior). Perhaps higher concordances were found due to the fact that the therapists and observers were rating more objective or concrete aspects of therapy.

Conversely, the results of the present study contrast with the findings of Carroll, Nich, and Rounsaville (1998). Carroll et al. found that eighty percent of the therapist-observer correlations were in the poor to fair range while the present study found a moderate to high degree of correlation. Specifically, Carroll et al. found that the therapists tend to self-report implementation of interventions more frequently than did observers. The difference in strength of correlations may be due to a number of factors. First, the therapists in both studies received different levels of training and supervision. Therapists selected for the present study had between seven to fifteen years experience post-licensure. However, the therapists from Carroll et al. only averaged 4.4 years of postdoctoral experience, and had a shorter course of training (2-day didactic seminar).

More years of training afforded firsthand exposure to various treatment strategies. More importantly, the therapists from the present study received intense training and supervision from the pioneers and primary authors of both TBCT and IBCT treatment manuals. Supervision from experts likely provided a more thorough training that left little room for ambiguity, which helped increase raters' proficiency in recognizing various techniques. In addition, the intense supervision was maintained throughout the treatment course to prevent therapist drift. The therapists were required to send in audio- and/or videotapes of their sessions to the supervisors each week of the treatment duration. Yet another influential factor is the difference in length of treatment course. While the present study had a length of treatment of up to 26 weeks, Carroll et al. had only a 12-week course of treatment. Perhaps a longer treatment course provided prolonged supervision that exposed raters to different treatment interventions.

Present findings have significant implications for psychotherapy research. They emphasize the importance of therapist self-report as a unique and reliable tool in measuring the therapeutic process. The present findings that both groups of observer raters with different levels of training and sophistication (graduate students using a molecular coding system, and the less trained undergraduate students using a molar coding system) had high agreements with the therapist ratings add credibility to therapists as reliable reporters. High correlations with two groups of observers (with different coding system, levels of education, and methods of review) suggest that therapists can provide similar views as the often highly regarded observer raters. Consequently, future researchers may be able to choose the more cost-effective therapist self-reports over the much more time-consuming observer ratings.

In addition, high therapist-observer agreements in both treatments (TBCT and IBCT) further challenge the assumption that the observer raters are superior to therapist reports. It was no surprise that there was high concordance between the TBCT therapists and observers since TBCT is more structured, and the behavioral interventions are clearly demarcated. However, high IBCT therapist-observer agreements further strengthen the above findings since IBCT is more fluid and integrative. IBCT may appear ambiguous, and interventions may seem to flow back and forth into each other; thus, it is impressive that high IBCT therapist-observer correlations were found, suggesting that therapists can still be objective even in transcendental moments in IBCT.

Therapist self-ratings seem to be comparable to the gold standard observer ratings when the following factors are present: when there is thorough training, quality supervision, and when rating more concrete and specific aspects of therapy. Moreover, since psychotherapy is a complex endeavor, having multiple adherence measures can be beneficial, and therapist self-reports can be a valuable supplement to other adherence ratings.

The present study also has important implications particularly when it comes to supervision in the context of psychotherapy research. Quality supervision and intensive training may have fostered high concordance between therapist and observer raters. High agreement between therapist and observer ratings suggests that therapists can be relied upon to report what they do in therapy. The intense supervision that was maintained throughout the treatment course likely prevented therapist drift, and encouraged adherence to the specific treatments. This indicates that with quality supervision, therapists can accurately implement techniques, as well as be forthcoming in providing

cost-efficient and less time consuming self-evaluations. Note however, that the present study has a unique sample in that the therapists knew that they were being videotaped and closely monitored. In comparison to regular practitioners, the therapists used in the present study may have been more attentive to treatment delivery due to the rigorous supervision that is usually difficult to replicate in reality.

It is important to note that the present study's purpose was to add to the literature about research methodology, not to draw conclusions about other contexts in which therapists provide services and self-report to supervisors. However, even outside the context of psychotherapy research, the type of supervision offered appears to influence the extent to which trainees implement supervisor feedback in subsequent therapy sessions. It appears that the more variety in teaching methods, the more likelihood generalization takes place. In a study conducted by Milne, Pilkington, Gracie, & James (2003), the generalization of supervision themes to actual therapy was more evident when supervisors include "hard" instructional techniques (e.g. reviewing recordings of therapy, lectures, corrective feedback and specific direction). Specifically, the authors found that the transfer of supervision material to the therapy process was found to occur more for video- and/or audio-taped sessions (Milne & James, 2000; Milne et al., 2003).

While direct methods are traditionally used, indirect methods such as therapist self-evaluations are also shown to be invaluable, as these "soft" methods uniquely tap into supervisee's processes (Carter, Enyedy, Goodyear, Arcinue & Puri, 2009; Falender & Shafranske, 2007; James, Milne, Marie-Blackburn & Armstrong, 2006). Since supervision is a dynamic relationship, consideration of supervisee's (in our case, the therapists) perspective fosters a more authentic learning environment, and increases

competency-based supervision (Carter et al., 2009; Falender & Shafranske; James et al., 2006). Indirect methods such as self-assessment can provide knowledge and insight into a therapist's capability to perform, and entrustability (capacity to which a supervisee can be trusted to carry out a professional activity; Falender & Shafranske). Overall, competency-based supervision can be achieved if a wide range of supervision techniques is used (Falender & Shafranske; Norcross & Halgin, 1997).

Future research studies are necessary to determine the specific situations in which therapist and observer perspectives will be similar and/or different. If the present study's results were replicated, it would challenge the assumption that observers are the gold standard or superior raters. Replicating the findings under different conditions would lend credence to therapist ratings as a valuable supplement to other adherence ratings. Perhaps future studies with different parameters can be used, such as quality of supervision, different therapist rating methods and times (e.g., video or audio-review days after the therapy session), characteristics of therapists and raters (level of expertise or education), and using less structured treatment interventions with complex dynamics.

In sum, a number of biases regarding the use of therapist self-reports (Jacobson & Addis, 1993; Moncher & Prinz, 1991; Waltz et al., 1993) have prevented researchers from using this source of data. The present findings challenge the assumption that observer ratings are superior to therapist self-reports. The latter might be preferred if cost is a concern since observer ratings tend to be more expensive and time-consuming. More importantly, therapist self-rating is a unique tool that taps into the subjective intentionality distinct in more complex or dynamic therapies. Researchers would benefit from the unique contribution of therapist ratings, as this provides a valuable outlook on

psychotherapy. Therapists as raters are not only cost-efficient, but could help detect therapy nuances that may often go unnoticed by more distant or objective raters.

REFERENCES

- Barber, J. P., Crits-Christoph, P., & Luborsky, L. (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of Consulting and Clinical Psychology, 64*(3), 619-622.
- Berman, J. S. & Norton, N. L. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98*, 401-407.
- Carroll, K. M., Nich, C., & Rounsaville, B. J. (1998). Utility of therapist session checklists to monitor delivery of coping skills treatment for cocaine abusers. *Psychotherapy Research, 8*(3), 307-320.
- Carter, J. W., Enyedy, K. C., Goodyear, R. K., Arcinue, F., & Puri, N. N. (2009). Concept mapping of the events supervisees find helpful in group supervision. *Training and Education in Professional Psychology, 3*(1), 1-9.
- Christensen, A., Baucom, D. H., Atkins, D. C., Berns, S., Wheeler, J., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72*, 176-191.
- Christensen, A. & Jacobson, N. S. (1994). Who (or what) can psychotherapy do: The status and challenge of nonprofessional therapies. *Psychological Science, 5*(1), 8-14.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. In Erlbaum (Ed.) Hillsdale, NJ.
- Dobson, K., & Singer, A. (2005). Definitional and practical issues in the assessment of treatment integrity. *Clinical Psychology: Science and Practice, 12*(4), 384-387.
- Falender, C. A., & Shafranske, E. P. (2007). Competence in competency-based supervision practice: Construct and application. *Professional Psychology: Research and Practice, 38*(3), 232-240.
- Fals-Stewart, W., & Birchler, G. R. (2002). Behavioral couples therapy with alcoholic men and their intimate partners: The comparative effectiveness of bachelor's- and master's-level counselors. *Behavior Therapy, 33*, 123-147.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*(4), 198-205.
- Heaton, K. J., Hill, C. E., & Edwards, L. A. (1995). Comparing molecular and molar

- methods of judging therapist techniques. *Psychotherapy Research*, 5, 141-153.
- Hemphill, J. F., Simon, F., Burnaby, B. C., & Canada, C. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, 58(1), 78-79.
- Henggeler, S. W., Schoenwald, S. K., & Liao, J. G. (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child and Adolescent Psychology*, 31(2), 155-167.
- Henry, W. P., Strupp, H. H., Butler, S. F., Schacht, T. E., & Binder, J. L. (1993). Effects of training in time-limited dynamic psychotherapy: Mediators of therapists' response to training. *Journal of Consulting and Clinical Psychology*, 61, 441-447.
- 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.
- Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A. & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, 76(4), 544-555.
- Huey, S. J., Henggeler, S. W., & Brondino, M. J. (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of Consulting and Clinical Psychology*, 68(3), 451-467.
- Jacobson, N. S., & Addis, M. E. (1993). Research on couples and couple therapy: What do we know? Where are we going? *Journal of Consulting and Clinical Psychology*, 61(1), 85-93.
- Jacobson, N. S., Christensen, A., Eldridge, K., Prince, S. E., & Cordova, J. (2000). Integrative behavioral couple therapy: An acceptance-based, promising new treatment for couple discord. *Journal of Consulting and Clinical Psychology*, 68, 351-355.
- James, I. A., Milne, D., Marie-Blackburn, I., & Armstrong, P. (2006). Conducting successful supervision: Novel elements towards an integrative approach. *Behavioural and Cognitive Psychotherapy*, 35, 191-200.
- Kazdin, A. E. (1986). Comparative outcome studies of psychotherapy: Methodological issues and strategies. *Journal of Consulting and Clinical Psychology*, 54, 95-101.
- Kirchner, J. & Chen, A. (2008). *Census data for ethnic groups in Southern California*. Retrieved September 12, 2009, from Southern California Association of Governments Website: <http://www.scag.ca.gov/census/>

- Luborsky, L. (1992). Does psychotherapy research really offer good ideas for psychotherapists? *Journal of Psychotherapy Practice & Research*, 1(4), 310-312.
- Luborsky, L., Woody, G. E., McLellan, A. T., O'Brien, C. P., & Rosenzweig, J. (1982). Can independent judges recognize different psychotherapies? An experience with manual-guided therapies. *Journal of Consulting and Clinical Psychology*, 50(1), 49-62.
- Margison, F. R., Barkham, M., & Evans, C. (2000). Measurement and psychotherapy: Evidence-based practice and practice-based evidence. *British Journal of Psychiatry*, 177, 123-130.
- McMurray, S. K. (2007). Adherence to treatment and treatment outcome in marital therapy: Are therapist's interventions related to couple's success? *Fuller Theological Seminary, School of Psychology*. AAT 3262584.
- Miller, S. J., & Binder, J. L. (2002). The effects of manual-based training on treatment fidelity and outcome: A review of the literature on adult individual psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, 39, 184-198.
- Milne, D. L. & James, I. (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology*, 39, 111-129.
- Milne, D. L., Pilkington, J., Gracie, J. & James, I. (2003). Transferring skills from supervision to therapy: A qualitative and quantitative N=1 analysis. *Behavioural and Cognitive Psychotherapy*, 31, 193-202.
- Mintz, J., Auerbach, A., Luborsky, L., & Johnson, M. (1973). Patient's, therapist's and observers' views of psychotherapy: A "Rashomon" experience or a reasonable consensus? *British Journal of Medical Psychology*, 46(1), 83-89.
- Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247-266.
- Norcross, J. C. & Halgin, R. P. (1997). Integrative approaches to psychotherapy supervision. In C.E. Watkins, Jr. (Ed.), *Handbook of psychotherapy supervision*. Chichester: Wiley.
- O'Malley, S. S., Foley, S. H., Rounsaville, B. J., & Watkins, J. T. (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression. *Journal of Consulting and Clinical Psychology*, 56(4), 496-501.
- Patterson, G. R., & Chamberlain, P. A. (1994). Functional analysis of resistance during parent training therapy. *Clinical Psychology: Science and Practice*, 1(1), 53-70.

- Perepletchikova, F., & Kazdin, A. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology, 12*(4), 365-383.
- Singer, A. J. (2002). Therapist and observer ratings of therapist fidelity to a family-based prevention model. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 62*(7-B), 3389.
- Siqueland, L., Crits-Christoph, J. P., Barber, J. P., Butler, S. F., Thase, M., Najavitz, L. et al. (2000). The role of therapist characteristics in training effects in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence. *Journal of Psychotherapy Practice and Research, 9*, 123-130.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology, 61*, 620-630.
- Wills, R. M., Faltler, S. L., & Snyder, D. K. (1987). Distinctiveness of behavioral versus insight-oriented marital therapy: An empirical analysis [Special issue]. *Journal of Consulting and Clinical Psychology, 55*(5), 685-690.
- Wilkinson, J. (2000). Direct observation. In G.M. Breakwell, S. Hammond & C. Fife-Schaw (Eds.), *Research methods in psychology* (2nd ed.). Thousand Oaks, CA: Publisher.
- Xenakis, S. N., Hoyt, M. F., & Marmar, C. R. (1983). Reliability of self-reports by therapists using the Therapist Action Scale. *Psychotherapy: Theory, Research & Practice, 20*(3), 314-320.

APPENDIX A
Literature Review Table

Author, Year, Title	Publication Type	Objectives	Sample	Variables/ Instruments	Research Design	Results/ Statistics	Major Findings
I. Treatment Integrity and Adherence							
Moncher & Prinz (1991). Treatment fidelity in outcome studies	Journal article	To evaluate outcome studies and determine the extent to which treatment fidelity was considered.	359 outcome studies from 1980 to 1988	The outcome studies were evaluated according to (1) the training of treatment implementers, (2) the procedures used to facilitate fidelity, (3) verified treatment aspects, (4) methods used to assess fidelity, and (5) how fidelity assessment was used in the interpretation of the results.	Review study	55% of the outcome studies essentially overlooked treatment fidelity. The present authors found that only 1 out of 8 studies combined the use of treatment manuals, supervision of treatment agents, and did an adherence check to protocol. Specific recommendations were given to promote treatment fidelity.	Treatment adherence is best understood from multiple perspectives. 23% of the reviewed study used therapists or clients as raters of adherence. Only 32% of the studies included data from more than one treatment session to obtain their ratings. Therapists are likely to be biased in their perceptions of sessions. Clients can also be problematic, as they are likely to not have the training necessary to describe what happened in the session.
Waltz, Addis, Koerner, Jacobson, (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence	Journal article	Purpose: To review different methods for assessing adherence and competence and provide some guidelines for future use of adherence and competence scales.	N/A	Adherence is defined as the extent to which a therapist used interventions and approaches prescribed by the treatment manual, whereas the term competence refers to the level of skill shown by the therapist in treatment delivery. By skill, the authors mean the extent to which the therapists conducting the interventions took the relevant aspects of the therapeutic	Review study	N/A	Some assessments done on adherence may be problematic, as they used therapists for adherence raters. Challenges in adherence measures: (1) Degree of complexity requires various levels of expertise (raters and therapists). (2) measures used different sources of information: process notes, transcripts, audiotapes and videotapes. Choose the source that will retain the most information (videotape), (3) unit of analyses vary across studies. Segments vs. entire sessions. (4) Different methods used in measuring adherence: (a)

				context into account and responded appropriately to these contextual variables.			checklist technique (occurrence-nonoccurrence) (b) requiring inference about the quality of the intervention (which puts adherence and competence together).
Gresham, MacMillan, Beebe-Frankenberger & Bocian (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented?	Journal article	Purpose: To review three major learning disabilities journals in the last 5 years, and to outline technical issues behind treatment integrity.	N/A	N/A	Review study	Only 18.5% of the 65 studies mentioned in the 3 studies measured the integrity with which treatments were delivered.	Difficult to know what treatments are effective/ineffective since treatment integrity was not measured thoroughly. Issues that affect treatment integrity: (1) specification of treatment components (i.e., in behavioral terms; do we use global or molecular levels of independent variable specification?), (2) deviations from treatment protocols (i.e., therapist drift) and their relation to the amount of behavior change, and (3) measurement issues [i.e., direct (videotaped sessions) vs. indirect assessment (self-reports)]. Indirect assessments may be problematic but are useful especially if complemented with direct assessment tools. (4) Complexity of treatment (i.e., more complex treatments that need help from third parties tend to be delivered with less integrity).
II. Methodological Issues							
Heaton, Hill, & Edwards	Journal article	To compare the	23 cases of 6-	2 methods compared:	Experimental	The two molar measures were	HCVRCS may be prone to cognitive

<p>(1995). Comparing molecular and molar methods of judging therapist techniques</p>		<p>molecular (via sentence-by-sentence transcripts) and molar (global) methods of measuring therapist techniques.</p>	<p>therapy sessions. 23 therapist who rated themselves on adherence to techniques. 23 clients. 15 judges were female undergraduate students who were unaware of the study's hypotheses.</p>	<p>(1) molecular method-analyzes sentence-by-sentence transcripts; assessed by HCVRS which looks at the grammatical structure of therapist's verbal behavior. (2) The molar or global method-estimates the frequency of use by simply viewing or listening to a session; assessed by TPI-R and PQS.</p>		<p>positively related, but when compared with the molecular, they were found not related, which may be due to the different metrics used.</p>	<p>biases since it has no firm anchors, leading to judges overrating the frequency of events. Availability heuristic occurs when raters overestimate on the basis of the ease with which the instances of events can be brought to mind. This bias may lead raters to overrate the frequency of a particular class of events. The anchoring heuristic happens when raters estimate based on an initial value and then adjust them to yield a final answer.</p>
<p>O'Malley, Foley, Rounsaville, & Watkins (1988). Therapist competence and patient outcome in interpersonal psychotherapy of depression</p>	<p>Journal article</p>	<p>Purpose: to explore the relationship of therapist performance and outcome</p>	<p>15 therapists and 35 patients</p>	<p>THERAPIST S: self-report ratings of effectiveness after the 4th session. SUPERVISORS: used Therapist Strategy Rating Form and the Process Rating Form. OUTCOME MEASURES : Hamilton Rating Scale for Depression and the Social Adjustment Scale administered by independent clinical evaluators, Patient Attitudes and Expectations Form</p>	<p>Experimental</p>	<p>Therapist performance was associated with patient self-report change and apathy, but it was not significantly related to social adjustment.</p>	<p>Other factors in the relationship may affect outcome. For example, low correlation to social adjustment may be due to the short duration of the study (4-month duration may have been too brief to detect significant changes in social functioning).</p>
<p>Wills, Faitler, & Snyder (1987). Distinctiveness of behavioral versus insight-oriented</p>	<p>Journal article</p>	<p>Purpose: a comparative outcome study done on behavioral marital</p>	<p>4 therapists, 17 couples, and 24 sessions (12 BMT</p>	<p>The Therapist Intervention Coding System (TICS) included 8</p>	<p>Experimental</p>	<p>Therapists did not use any BMT-specific techniques in their IOMT sessions (and vice-versa). Interventions</p>	<p>Therapists were found to deliver each treatment modality in an uncontaminated manner, and authors found that</p>

marital therapy: An empirical analysis		therapy (BMT) and insight-oriented marital therapy (IOMT). Authors attempt to delineate the specific, nonspecific interventions appropriate to BMT and IOMT.	and 12 IOMT). The 1 st and 2 nd authors of the study coded the therapy sessions.	categories specific to BMT and IOMT (and other nonspecific interventions).		could be reliably coded into specific techniques appropriate for each treatment modality, as well as into other techniques not specific to either treatment modality.	the two techniques were inherently different in structure (BMT is more structured). The two approaches also differed in their relative proportion of specific vs. nonspecific techniques.
Dobson & Singer, (2005). Definitional and practical issues in the assessment of treatment integrity	Journal article	To comment on issues related to defining and assessing treatment integrity and its components. Its clinical applications are also discussed.	N/A	N/A	Discussion Article	N/A	An issue in assessing adherence is the variety of expertise/training required from the raters. A major concern for practitioners is how manual-based treatments can constrain their ability to use techniques flexibly. Practitioners concerned more with competence, and less with treatment differentiability in actual practical settings.
Kazdin, (1986). Comparative outcome studies of psychotherapy : Methodological issues and strategies	Journal article	Considers multiple design issues in comparative studies (i.e., conceptualization, implementation, and evaluation of alternative treatment, assessment of treatment processes).	N/A	N/A	N/A	Results of outcome studies are mixed depending on the research design and how it is conducted. Special issues to consider: conceptualization, assessment, and evaluation of the results. Other variables: time variable (brief vs. long-term therapy), focus of treatment benefits and costs in conclusions drawn.	Outcome assessment must be taken from multifaceted perspectives (therapist, raters, patient, significant others) and from different modalities of assessment (self-report, clinician ratings, direct observation). Certain measures may be biased and only reflect changes in a specific treatment. Things that can increase precision of the test include therapist training (to prevent therapist drift), supervision, and selected sessions must be assessed.
Miller & Binder (2002).	Journal article	To examine training	N/A	N/A	Review study	N/A	5 methodological issues that impact

The effects of manual-based training on treatment fidelity and outcome: A review of the literature on adult individual psychotherapy		issues that influence treatment outcome.					manual-based training: the lack of a uniform definition of manual-based treatment, trainer variables, therapist variables, the experimental design and the conceptualization or operational definitions of dependent variables. Effective adherence measures include: (1) therapist behaviors essential and unique to the treatment, (2) that are essential but not unique, (3) compatible with the specific modality, not prohibited, but neither necessary nor unique, (4) behaviors that are proscribed. Competence, on the other hand, involves a more flexible and creative application of techniques.
Wilkinson (2000). Direct Observation	Book chapter	Explains the different types of observation and the practical issues of each.	N/A	N/A	Discussion Article	N/A	Observations can be done via videos or audiotapes. Problem arises in the selected portion that the observers choose to observe. Observer bias occurs when the observer conform to preconceived notion that may very well affect their perception of the event, and reduce the reliability of the data. Author recommends doing "practice runs" or training sessions before conducting the actual study.
Pereplechikova & Kazdin, (2005). Treatment integrity and therapeutic	Journal article	The study aims: (1) to discuss characteristics of treatment,	N/A	N/A	Review study	N/A	Factors that deviate from treatment integrity outcome studies: Indirect methods (i.e., therapist

change: Issues and research recommendations		therapist and client associated with treatment integrity and outcome, (2) to emphasize the need for empirical examination on the relationship between treatment integrity and outcome, and (3) to recommend changes needed in evaluating integrity.					reports that may be prone to demand characteristic and/or a need for social approval), reliance on post treatment data, deviations from treatment protocol in order to better suit client's needs, and specific treatment characteristics (i.e., complexity, required resources/materials, time/duration, treatment agents, as well as clients' and therapists' acceptability of treatment. Therapist's characteristics: Level of experience, and therapist's motivation may affect treatment integrity. Client's Characteristics: client's difficulty, hostility, problem severity, duration and comorbidity influence integrity of treatment delivery.
Barber, Crits-Christoph, & Luborsky (1996). Effects of therapist adherence and competence on patient outcome in brief dynamic therapy	Journal article	Purpose: to test whether adherence and competence in the specific techniques of supportive-expressive (SE) dynamic was associated with patients' change in symptoms. The main components of SE therapy are expressive (interpretative) and supportive techniques.	29 patients diagnosed with major depression, treated by 4 therapists. 2 doctoral-level judges blind to patient outcome, and to therapist identity.	Measures: The Penn Adherence-Competence Scale for Supportive-Expressive therapy (PACS-SE). Three major theoretically derived subscales include: general therapeutic skills; supportive skills and expressive skills.	Experimental	Early symptomatic improvement predicted adherence to expressive techniques, indicating that the less symptomatic improvement by Session 3 (i.e., the more depressed on the BDI), the less adherence to expressive interventions was evident.	Change in patients' depressive symptoms was associated with competency rather than frequency of SE delivery. The more the patient shows immediate benefits from treatment, the easier it is for the therapist to adhere to the SE treatment manual.
Fals-Stewart & Birchler (2002).	Journal article	Purpose: to compare the bachelor's-	Alcoholic men and their	The Timeline Followback	Experimental	Equivalency test results show no significant	Bachelor's-level and master's-level counselors were

Behavioral couples therapy with alcoholic men and their intimate partners: The comparative effectiveness of bachelor's- and master's-level counselors		and master's-level counselors in their treatment delivery (behavioral couples therapy), and to compare their patient outcomes.	intimate partners (N=48). Four bachelor's-level and four master's-level counselors	Interview (TLFB) assesses the participants' substance and drug use. Dyadic Adjustment Scale. Client Satisfaction Questionnaire		differences between couples assigned to either condition ($p > .20$), and counselor type was not significant ($p > .05$), as couples' DAS scores were equivalent for couples treated by bachelor's-level and master's-level counselors.	found to have comparable treatment adherence. Bachelor's-level counselors, however, were found to have lower competence. Additionally, partners reported to have equivalent levels of satisfaction being treated by the bachelor's- and master's level counselors.
Margison, Barkham, & Evans (2000). Measurement and psychotherapy : Evidence-based practice and practice-based evidence	Journal article	Purpose: To review measurement relevant to psychotherapy.	N/A	N/A	Review study	N/A	Seven measurement strategies in psychotherapy were shown how these might be applied in clinical practice. Authors review the ff domains: (1) interventions, (2) case formulation, (3) treatment integrity, (4) performance--which includes adherence, competence and skillfulness, (5) treatment definitions, (6) therapeutic alliance, (7) routine outcome measurement.
Henry, Strupp, Butler, Schacht & Binder (1993). Effects of training in time-limited dynamic psychotherapy : Mediators of therapists' response	Journal article	Purpose: Authors examined therapist variables, patient variables, and training variables that may influence therapist responses to a TLDP (time-limited dynamic psychotherapy) training program.	2 trainers led 2 groups of therapists (n=16)	INTREX questionnaire was given to therapists. CDPS (Capacity for Dynamic Psychotherapy Scale) given to patients. Independent raters who were unfamiliar with the training status of the therapists were used to rate both instruments.	Experimental	Previous supervision hours and competence at hand (prior to the present training) significantly correlate with posttraining changes. More hours of previous supervision was associated with less TLDP adherence ($r = -.48, p < .06$). Previous supervision hours have a stronger negative effect (than years of experience) on technical changes.	Therapist with more previous training/supervision hours was more resistant to present training/change. Personality factors may weaken the link between manual adherence and therapeutic outcomes (positive therapeutic outcome may be due to other factors and NOT due to the treatment modality). Patient variables are also influential; the more "difficult" and more educated a patient is, the more the therapist adhered to TLDP.

<p>Siqueland, Crits-Christoph, Barber, Butler, Thase, Najavitz & Onken, (2000). The role of therapist characteristics in training effects in cognitive, supportive-expressive, and drug counseling therapies for cocaine dependence</p>	<p>Journal article</p>	<p>To examine the role of therapist characteristics such as demographics (age, gender, race, professional degree-- M.A., Ph.D.) and level of experience on their therapy training.</p>	<p>60 therapists, 200 patients (crack cocaine smokers).</p>	<p>3 treatment conditions: SE (supportive-expressive therapy), CT (cognitive therapy) and IDC (Individual Drug Counseling). The Adherence (frequency)/Competence (quality) Measures: Penn Adherence and Competence Scale for Supportive-Expressive Therapy (PACSE), Cognitive Therapy Scale (CTS), Adherence and Competence Scale for Addiction Counseling.</p>	<p>Experimental</p>	<p>None of the demographic variables were significantly correlated with training competence. But the therapist's level of experience prior to the present study was found to be strongly associated with change after training.</p>	<p>CT therapists with more pre-training supervision hours showed less change, which may be due to hardening of "their own style." Regarding the finding that more supervision predicts less change with training, the authors suggest considering personality variables such as therapist flexibility and willingness to try new things as part of their selection process.</p>
<p>Christensen and Jacobson (1994). Who (or what) can psychotherapy do: The status and challenge of nonprofessional therapies</p>	<p>Journal article</p>	<p>Purpose: To review research on nonprofessional psychological treatment</p>	<p>N/A</p>	<p>N/A</p>	<p>Review study</p>	<p>N/A</p>	<p>Reviewed research studies that showed no significant relationship between therapist's level of experience and therapy outcome; no overall difference in effectiveness between professional therapists, graduate-student therapists and paraprofessional therapists.</p>
<p>Berman & Norton 1985). Does professional training make a therapist more effective?</p>	<p>Journal article</p>	<p>Purpose: To review research studies on the effectiveness of professionals and paraprofessionals.</p>	<p>43 studies were reevaluated</p>	<p>N/A</p>	<p>Review study</p>	<p>No significant difference between the paraprofessionals and the professionals, and they improved comparably at the same rate. They found therapist effectiveness may be dependent on the length of</p>	<p>There was a slight difference between the paraprofessionals and the professionals. The professionals were more effective when working with older patients, and were better in brief treatments.</p>

						treatment.	
Patterson & Chamberlain (1994). Functional analysis of resistance during parent training therapy.	Journal article	Purpose: Review a decade of studies on parental resistance, and its effects in parent training therapy.	N/A	N/A	Review study	N/A	Findings show increase in resistance when therapists intervened. Increases in therapists' interventions were accompanied by increases in parental resistance. Studies also show that resistance changed therapist behavior, influenced therapist effectiveness, inducing more drift from set treatment protocol.
Henggeler, Schoenwald & Liao, (2002). Transporting efficacious treatments to field settings: The link between supervisory practices and therapist fidelity in MST programs	Journal article	Purpose: To validate measures for the following: therapist adherence, supervisory practices, and the association of both.	74 therapists organized into 16 teams. 12 MST Supervisors. 285 youths	Measures: SAM (Supervisor Adherence Measure) and the TAM (Therapist Adherence Measure). SAM was completed by the therapists, who rated their supervisors at 2-month intervals. TAM was completed by the caretakers. TAM data were nested; each therapist was evaluated by multiple families, and families rated therapist on multiple occasions.	Experimental	Supervision and therapist behavior can have an inverse relationship. High supervisory fidelity is likely more associated with low therapist MST adherence.	Findings show that the predictors of therapist fidelity to treatment protocols may be associated with characteristics of individual clinicians (i.e., types of professional training), and go beyond supervisory constructs.
Tversky & Kahneman, (1974). Judgment under uncertainty: Heuristics and biases	Journal article	Purpose: to describe heuristics commonly used when making judgments under uncertainty.	N/A	N/A	Discussion article	N/A	Three kinds of heuristics used: (1) representativeness - when people judge whether an event A belongs to process B, (2) availability of instances used when people assess the frequency of a class, (3)

							adjustment from an anchor-numerical/value prediction .
Huey, Henggeler, & Brondino, (2000). Mechanisms of change in multisystemic therapy: Reducing delinquent behavior through therapist adherence and improved family and peer functioning	Journal article	To study the association between MST adherence and outcome (e.g., improved functioning, reduced delinquent behavior).	2 Samples: (1) Diffusion Sample-155 violent juvenile offenders and their primary caregivers . 10 MST therapists. (2) CDA Sample-118 substance-abusing offenders. 3 therapists.	MST Adherence Measure was completed by the primary caregiver, youth and therapist. They evaluated the extent to which the therapist engaged in MST behaviors. 5 factors based on therapist ratings, 6 factors based on caregiver ratings, and 4 factors from youth ratings.	Experimental	Mostly insignificant correlations between the three informants (for the Diffusion sample average $r=.19$. CDA sample $r=.18$). Therapist adherence to MST contributes directly and indirectly to reductions in delinquent behaviors, but a closer inspection on individual factors based on youth reports reveals adherence may be counterintuitive.	Positive changes in family functioning predicted changes in delinquent behaviors. There was a lack of interrespondent agreement among participants. Factor analyses reveal that adherence varies depending on the source of information, which suggests that informants held own notions regarding how adherence should be construed. Therapist report on adherence should be interpreted with caution due to possible biases that could influence ratings of their own in-session behavior.
Mintz, Auerbach, Luborsky & Johnson (1973). Patient's, therapist's and observers' views of psychotherapy : A Rashomon experience or a reasonable consensus?	Journal article	Purpose: Attempts to view psychotherapy from 3 perspectives (patient, therapist and external observers). Would there be a reasonable agreement?	4 cases (12 sessions each). 2 neutral observers.	Measures: The Therapy Session Report (TSR)- filled out by patients, therapists and observers.	Experimental	Interjudge agreement of all ratings revealed that none of the view-pairs reached statistical significance (none of the pair agreed more than any other).	Regarding what is considered as effective, there was an overall poor agreement obtained. Judges agreed across views about as much as neutral observers agree with each other. Each group have their own unique perspective.
Jacobson & Addis (1993). Research on couples and couple therapy: What do we know? Where are we going?	Journal article	Purpose: to discuss the outcome and process research on couple therapy. Which treatments work, how do they work, and what factors predict outcome?	N/A	N/A	Review study	N/A	Findings from previous studies show success rates, and that every tested treatment seems to show a reliable change. All approaches seem to help couples, and no published study failed to outperform a control group. Authors discuss the assets and liabilities in doing specific methods

							in comparative clinical trials (generally asks which approach is better), and in doing intramodel comparisons (contrast treatments derived from the same theoretical model). It is often quite difficult to enforce blindness on those involved in the assessment procedure. Therapist bias likely plays a factor, and can affect the treatment process and outcome.
Xenakis, Hoyt, & Marmar, (1983). Reliability of self-reports by therapists using the Therapist Action Scale	Journal article	To assess whether therapists could reliably rate their own actions using the Therapist Action Scale.	RATERS: 8 therapists and 2 independent judges. PATIENTS: 3 men and 22 women.	The TAS (Therapist Action Scale) and PAS (Patient Action Scale) were used.	Experimental	When the therapist-independent judge ratings were correlated, only 4 out of the 26 exceeded the .6 coefficient level and only 11 of the 26 exceeded the .4 level. The independent judges' ratings showed a stronger and more consistent array of interrater reliability correlations. In addition, the independent judges also rated the therapists as engaging in a lower percentage of actions than did the therapists themselves for 22 of the 26 items. The general lack of consensus can stem from 3 possible sources: (1) raters; (2) instruments and; (3) the clinical phenomena.	There is a general lack of interrater reliability, suggesting caution must be practiced when interpreting therapist self-reports on therapeutic actions and/or process. However, authors of the present study evaluated dynamic treatment that was not manual-guided, and agreement might be higher for manual-based treatments. Compared to the independent judges, the therapists have access to more observational sources such as nonverbal cues, personal reactions and additional information about the case. Even if sessions were videotaped, videotaped sessions do not show the therapist's subjective awareness.
Carroll, Nich, & Rounsaville (1998). Utility of therapist session checklists to	Journal article	Purpose: to examine the concordance between therapist and observer reports of	121 cocaine abusers. 5 CBT therapists. 6 CM therapists.	Patients randomly assigned to one of the four treatment sessions (CBT with	Experimental	Out of 7 of the 10 items, there were more therapists reporting that they had delivered a technique when the raters indicated	Therapists tended to overestimate their use of specific interventions relative to the raters. Ratings

monitor delivery of coping skills treatment for cocaine abusers		therapist delivery of cocaine interventions	5 graduate student raters.	desipramine, CM with desipramine, CBT with placebo, and CM with placebo).		they did not. Results show an overall poor agreement between the raters and the therapists on the interventions delivered in sessions.	from different perspectives (therapists, observers, patients) are not interchangeable.
Luborsky, Woody, McLellan, O'Brien, & Rosenzweig (1982). Can independent judges recognize different psychotherapies? An experience with manual-guided therapies	Journal article	To examine the degree to which therapists adhere to SE, CB and DC techniques.	Patients were male veterans with narcotic addiction. Therapists have experience in the specific form of psychotherapy (4 SE, 3 CB and 6 DC). Sessions were selected randomly from any session after the third.	Judges rated the extent to which the 3 criteria were present in a session using a 5-point Likert scale (1=none; 5=very much). The same 5-point scale was used for assessing the global content of each segment studied: the degree to which the treatment fits the specification of the three therapies. A 3rd judge, using the Temple Content Categories that noted the objective counts of well-specified components of speech, did analysis of speech content.	Experimental	Study 1: 70% accuracy was found in discriminating three forms of therapy. Study 2: 80% accuracy due to revisions of the treatment rating form and after practice gained by the judges in the first study.	The manuals served as helpful guides in training the therapists, and as a reference in their therapeutic work. Thus, the manuals increased adherence to the respective treatment techniques. This is evidenced by the judges' successful ratings of the treatment approaches using the manual-derived criteria. The study concludes that independent judges can reliably recognize the manual-guided therapies.
Hill, O'Grady, & Price (1988). A method for investigating sources of rater bias	Journal article	Purpose: To outline a method in examining rater bias.	826 sessions of the Treatment of Depression Collaborative Research Program. Clients: 59 in CBT, 61 IPT (Interpersonal	5 major scales of the Collaborative Study Psychotherapy Rating Scale (CSPRS): CM, IPT, CBT, Explicit Directiveness (EC), and the Facilitative Conditions (FC) scale. A	Experimental	FC scale (extent of the therapist's supportive encouragement, warmth, etc.) was mostly highly related to rater-bias scale, indicating that these FC ratings (done by raters) were highly associated with how much they liked and felt similar to the therapist and	Overall, authors found only minimal observer bias in this study. When measures have high specificity and are easy to rate, the authors found that there is less concern with rater bias. With measures that are more abstract and subjective (like the FC items), however, the

			Psychotherapy), 119 in Clinical Management (CM). 28 therapists. 8 graduate students as raters.	rater-bias scale was also used to measure the attitudinal similarity between raters and ratees.		client.	authors advise using a measure such as the rater-bias scale to determine if bias was present.
Singer (2002). Therapist and observer ratings of therapist fidelity to a family-based prevention model	Journal article	Purpose: to compare therapist report with observer ratings.	110 sessions (CBT and MDFP)	Therapist Behavior Rating Scale-2nd version (TBRS-2) filled out by observers. Therapists used the Therapist Self-Report Checklist (TSRC) to rate their own adherence.	Experimental	Results show that therapists adhered to MDFP overall, but violated adherence to a certain extent because they spent too much time with the adolescents alone. When therapists rated their own adherence (using TSRC), intraclass Correlational coefficients found moderate to strong concordances, inferring that observers and therapists shared a similar perspective of adherence. When observer and therapist ratings were both compared with objective ratings, high concordance was found between therapist ratings and objective ratings, as well as high concordance between observer ratings and objective ratings.	High concordance reveals that therapist report can be a reliable source of information regarding treatment adherence.
III. Method Section							
Christensen, Atkins, Berns, Wheeler, Baucom, Simpson (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples.	Journal article	Purpose: examine efficacy of TBCT v. IBCT	134 chronically distressed married couples	Outcome measures: Marital Adjustment Test, Marital Satisfaction Inventory, Dyadic Adjustment Scale; Conflict Tactics Scale-Revised, Structured Clinical Interview for DSM-IV.	Experimental	Therapists were adherent and competent based on alpha reliabilities across coders. TBCT couples seem to improve earlier in the treatment process, but quickly reached plateau. IBCT couples, on the other hand, slowly improve throughout treatment, and 65% of them showed reliable	Changes in outcome suggest that IBCT and TBCT can be used with very severely distressed couples. Increased marital satisfaction in IBCT couples may be due to the emphasis on central themes. TBCT, on the other hand, focused more on problem behaviors.

						change.	
Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum)	Book chapter	Purpose: to provide a rationale for various statistical analyses.	N/A	N/A	N/A	3 Parameters of a statistical test's power: 1. the significance criterion 2. the effect size 3. degree H1 exists Criteria for significance: 1. What's the standard of proof the H1 exists? The rate of rejecting the true Ho? "If null hypothesis is rejected, the probability of the obtained sample result is no more than .05" which is a statistically significant result. Since .05 is small, and researcher is able to reject Ho, it means that he is able to reject Ho within the .05 significance level (means that there is only .5 chance of wrongly rejecting Ho). If the probability is greater than .05, he would have to accept Ho. 2. the exact reason why H1 exists. To assess the reliability of a statistic: use SE (standard of error) of the statistic.	Pp. 77-81: correlation coefficients of: .10 = small .30 = medium .50 = large in terms of magnitude of effect sizes. The ff. is used to gauge if it is reasonable in effect size: - large number of subjects (n), large number of meta-analytic studies for which these values are based, other studies with similar magnitudes of correlation coefficients
Hemphill, Simon, Burnaby & Canada (2003). Interpreting the magnitudes of correlation coefficients.	Journal article	Purpose: discussion of empirical guidelines for interpreting magnitude of correlation coefficients.	N/A	N/A	Discussion article (cited Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum)	J. Cohen (1988) gave an operating definition for interpreting correlation coefficients (see record 1987-98267-00).	Magnitude of correlation coefficients is considered a key index of effect size (the size of the observed relationship; important in practical situations).
Jacobson, Christensen, Eldridge, Prince, & Cordova (2000). Integrative behavioral	Journal article	Purpose: to provide data on IBCT treating marital distress	N=21 couples seeking therapy for marital distress	ADHERENCE MEASURES : Behavioral Couple Therapy Rating Manual	Experimental	A second manipulation check was used by using naïve raters (undergraduate students), who used a global system that helped	IBCT is found to be more effective than TBCT due to the acceptance factor, which is absent in TBCT.

couple therapy: An acceptance-based, promising new treatment for couple discord				(BCTRM). 3 graduate students were selected as raters for adherence, and were trained to use the BCTRM. Nine undergraduate students served as naïve raters and were trained to use a global system, using Couple Therapist Rating Scale (CTRS). OUTCOME MEASURES : Global Distress Scale, Marital Satisfaction Inventory, Dyadic Adjustment Scale.		ensure treatment adherence. Results from the two adherence ratings indicate that IBCT and TBCT are distinct approaches, as IBCT therapists demonstrated more acceptance while TBCT therapists instigated more change strategies. In addition, the therapists were able to keep the two approaches distinct, and they avoided acceptance interventions in TBCT sessions, and incorporated acceptance techniques in IBCT sessions. TBCT was competently given based on a rating scale and rated by an expert; pre- and post-test scores on GDS and DAS. 80% of IBCT couples improved or recovered.	
IV. Discussion							
Falender & Shafranske (2007). Competence in competency-based supervision practice: Construct and application	Journal article	To review/define competence as a construct; to define competency-based clinical supervision.	N/A	N/A	Review study	N/A	APA requires competence in supervision, which takes into account supervisee's skill set and entrustability (when she/he is trusted to carry out the task). Competence is a dynamic construct that requires dyad to commit to self-assessment that pushes own limits.
Carter, Enyedy, Goodyear, Arcinue & Puri (2009). Concept mapping of the events supervisees find helpful in group supervision	Journal article	Purpose: to identify what is helpful in group supervision.	Responses from 49 graduate students in psychology.	Cluster analysis and multidimensional scaling were used for the concept map.		Helpful events fell into 5 clusters that included supervisor impact, specific instruction, self-understanding, support and safety, and peer impact. They differed on two dimensions (perceiving supervisor vs. peer impact &	Other dimension in the vertical axis (Perceiving supervisor vs. peer impact and Acquiring objective vs. self-knowledge range from entire focus on supervisee (process own counter-transference, increase self-

						acquiring objective vs. self-knowledge). These clusters can be used as specific goals for supervision. First cluster (Supervisory Impact) shows that helpfulness is seen in both content (specific advice) and process (making supervisees comfortable). Cluster 5 (Peer Impact)- encourages collaborative problem-solving, encourages diverse viewpoints and techniques.	awareness) to the other; more objective and didactic (watching videos and lectures). Other helpful aspects: supervisor's open and validating style, competent and even-handed personality, didactic expertise, and use of in vivo techniques. Unhelpful aspects: disorganization, supervisor's domineering style, unprofessionalism, inexperience and conflicts with co-supervisors. Important in supervision (individual or group): the opportunity for self-exploration developed through the relationship(s).
Milne & James (2000). A systematic review of effective cognitive behavioural supervision	Journal article	Purpose: To review impact of supervision and its effectiveness .	N/A	N/A	Review study	Pyramid approach was beneficial. Supervisor's close supervision benefits supervisee. Systematic/Direct methods such as modeling competence, providing instructions benefit supervisee. Specifically, direct methods such as viewing videos of therapy, providing info, feedback and directions had more beneficial outcomes. Most common direct method was the corrective feedback.	However, generalizability is limited since population used is learning disabilities, wherein effective supervision might call for more direct methods. Reading training manuals worked for therapists but authors suggest that separate manuals be written for supervisors to develop competence. Developing a manual for supervisors, would also send out the message that competence does not develop automatically.
Milne, Pilkington, Gracie & James (2003). Transferring skills from supervision to therapy: A qualitative and quantitative N=1 analysis	Journal article	Purpose: to analyze CBT supervision, and its effectiveness .	N=1 (supervisor and the therapist-patient dyad)	10 videotaped supervision sessions and 10 therapy sessions were used for grounded theory analysis. 2 observers coded the 20	Experimental	14 supervisory themes were extracted using grounded theory.	Generalization or transfer of supervision themes were observed to happen after supervision took place (ex. If information gathering happened in supervision

				tapes.			sessions, therapist would also engage in the same behavior in subsequent therapy with patient). Mirroring of behaviors happened. Out of the 14 categories, the most common mirrored behavior or transfer is the “informing/gathering,” & “agenda setting and managing the session.”
James, Milne, Marie-Blackburn & Armstrong (2006). Conducting successful supervision: Novel elements towards an integrative approach	Journal article	Purpose: to summarize supervision theories from CBT perspective. Suggestions made on how to promote successful learning in supervision.	N/A	N/A	Discussion Article	N/A	Raises the question of how to promote learning in supervision. Authors propose the following sequentially: 1. Assess learning needs 2. Establish a baseline (with the use of scales) 3. Work at the right developmental level 4. Apply supervisory techniques (i.e., listening, observing, questioning) 5. Evaluation of progress (i.e., segments of tapes might be reviewed). 7 theories are put forth.
McMurray, S.K. (2007). Adherence to treatment and treatment outcome in marital therapy: Are therapist’s interventions related to couple’s success?	Journal article	To study the effects of therapist treatment adherence on outcome.	35 couples	TBCT and IBCT treatment were given.	Experimental	No relationship between TBCT and outcome. Strongest relationship was seen in (early and late) IBCT delivery and outcome.	Timing of treatment delivery is important; successful therapy is dependent on therapist’s awareness of when certain interventions should be delivered.
V. Miscellaneous							
Carroll, Nich, Siffry, Nuro, Frankforter, Ball, Fenton, & Rounsaville (2000). A general system for evaluating therapist	Journal Article	To describe the development of the Yale Adherence and Competence Scale (YACS).	N/A	N/A	Discussion Article	N/A	YACS is a general system for rating therapist adherence and competence in delivering behavioral treatments of substance abuse. 3

adherence and competence in psychotherapy in the research addictions							scales of YACS are assessment, general support and goals of treatment, which capture the general aspects of drug abuse treatment.
Hogue, Liddle, & Rowe (1996). Treatment adherence process research in family therapy: A rationale and some practical guidelines	Journal article	To present practical guidelines for conducting observational-based adherence research on family therapy models.	N/A	N/A	Review Study	N/A	Two important reasons for studying treatment adherence: they verify the level of treatment fidelity, and they provide a manipulation check of the independent variable. Things to consider in adherence research: (A) What will be coded: segment vs. whole session that will be coded (B) Who will code: non-participant raters vs. therapist and client who rate their own behaviors (C) Kind of coding system (simple occurrence vs. non-occurrence of an intervention, frequency counts, etc.).
Chevron, & Rounsaville, & Rothblum, (1983). Selecting psychotherapists to participate in psychotherapy outcome studies: Relationship between psychotherapist characteristics and assessment of clinical skills	Journal article	Purpose: describes the process involved in assessing psychotherapist skills, and also studies the relationship between judges' ratings of therapist skills and characteristics.	27 male and 7 female therapists; 3 evaluators	Videotaped recordings of therapist-patient sessions that are evaluated on two dimensions: (1) empathy, (2) potential as an IPT therapist. Raters were blind to therapists' professional discipline and level of experience.	Experimental	Gender comparisons and professional discipline did not reach statistical significance at the $p < .05$ level of confidence. On the other hand, therapist's age and level of clinical experience were highly correlated ($r = .94, p < .001$).	There is a positive relationship between ratings of therapist skill and age and level of experience. Older and more experienced therapists were judged as more empathic and rated to be more effective in short-term psychotherapy. Psychotherapists' gender and professional degree, on the other hand, failed to predict judges' ratings.
Jacobson, & Christensen (1996). Acceptance and change in couple therapy	Book chapter	Chapter title: from change to acceptance	N/A	N/A	N/A	Only two-thirds of the TBCT couples improved, and of those, one-third relapsed within two years after treatment. Some factors that influence outcome	Only half of the couples were helped with the traditional behavioral approach. Acceptance of incompatibilities is the missing link in

						include traditional values, commitment, age, emotional engagement, and convergent marriage goals. Initial pilot data shows significantly increased satisfaction in IBCT couples when compared with the TBCT couples.	TBCT.
McIntosh, Jordan, McKenzie, Luty, Carter, F., Carter, J. et al., (2005). Measuring therapist adherence in psychotherapy for anorexia nervosa: Scale adaptation, psychometric properties, and distinguishing psychotherapies	Journal Article	Compare therapist adherence to three psychotherapies for treating anorexia.	56 female participants randomly chosen. Therapists were 2 psychologists, and 1 psychiatrist experienced in CBT, IPT, and nonspecific supportive clinical management. Raters were 2 graduate students.	CSPRS (Collaborative Study Psychotherapy Rating Scale)	Experimental	No significant differences between the means of the 3 treatment modalities, indicating satisfactory therapist adherence.	Therapists were rated exhibiting significantly more behaviors according to their therapy, indicating very satisfactory adherence to therapy. The two raters found the three psychotherapies clearly distinguishable, even though they were unaware of therapy conditions.
Milne, Aylott, Fitzpatrick, & Ellis (2008). How does clinical supervision work? Using a "Best Evidence Synthesis" approach to construct a basic model of supervision	Journal article	Purpose: to conduct an empirical review to codify a basic model of clinical supervision.	N/A	N/A	Review Study	N/A	Successful supervision outcomes incorporated multiple instructional methods (i.e., lectures, corrective feedback, observing). Live or video-based observation of the supervisee occurred in 42% of the studies. More than one supervision method was used, with feedback and specific skills training as the most common methods.
Thompson, B. (2002). "Statistical" "practical," and "clinical": How many	Journal article	Purpose: to review and categorize 3 types of significance, to review	N/A	N/A	Review Study	N/A	Authors recommend that both "practical" and "clinical" indices are included in

kinds of significance do counselors need to consider? Journal of Counseling & Development, 80(1), 64-71		various indices of practical and clinical significance					reports. Framework for conceptualizing effect sizes is presented. Standardized differences vs. variance-accounted-for indices. It is necessary to present the SD. For example, Cohen's d represents the SD pooled from both groups using score scaling. .2 = small. .5 = medium. .8 = large.
Shavelson, R. (2 nd ed.; 1988). Statistical Reasoning for the Behavioral Sciences. Allyn & Bacon, Inc.	Book Chapter	Purpose: to give a conceptual and procedural knowledge of statistics.	N/A	N/A	N/A	In correlational studies, researcher does not manipulate IV to see its effect on DV rather she selects 2 variables to see how they covary. Subjects cannot be randomly assigned; nature or experience has already performed the treatment.	p. 159: Misleading correlation coefficient: 1. check on variances or SD to determine if restriction of range happened (small variance or SD suggests restriction of range). 2. Extreme groups are only used. 3. Combining groups or samples from 2 or more populations 4. outliers.
Korin, B.P. (1975). Statistical concepts for the social sciences. Winthrop Publishers, Inc.	Book Chapter	Purpose: to discuss different research problems and various statistical methods.	N/A	N/A	N/A	There are linear correlations. If bivariate data is spread out, slope of regression line is close to 0 or correlation is close to 0 (answers for SRT and BCTRM are spread out); doesn't mean no strong relationship between x & y, it just means no strong straight-line relationship. GOODMAN & KRUSKAL'S LAMBDA-association between nominal scales. CORRELATION RATIO (E squared); related to Pearson's r squared: association between interval and nominal. Type I errors.	P. 95: Bivariate situation-association between two characteristics. Scatter diagram-graph for bivariate data; when 1 or both scales are nominal or ordinal. P. 113: SPEARMAN'S RHO: Correlation/Association between ordinal scale variables. Based on the ranks assigned to the variables. PEARSON PRODUCT MOMENT CORRELATION- assoc. between interval scales; coefficient of determination (r^2) . P. 118: -1 < r < 1. When r is near +1 or -1: correlation is

							positively (negatively) high. There's no specific value that marks high from low. The labels of high, moderate or low depends on type of data and how result compares to other similar sets of data.
Miller, D.C. (1991; 5 th ed.) Handbook of research design and social measurement. Sage Publications, Inc.	Book Chapter	Purpose: To discuss procedures in basic, applied and evaluation research.	N/A	N/A	N/A	p. 245 Nominal and ordinal scales require nonparametric test. Interval and ration use parametric tests. Interval scales- has all the characteristics of ordinal scales (objects stand in some kind of relation to the categories) and the distance between any two numbers on the scale are of known size.	p. 244: table: Pt. Biserial r - for measuring the relationship between a truly dichotomous variable and a continuous variable. Pearson Product Moment r - for measuring relationships between 2 variables when both are continuous and the relationship is rectilinear. The coefficient of correlation is most reliable when based upon a large number of pairs of observation.

APPENDIX B
Request for Use of Data from the Original Study

Date of Request 9/12/07

**Request for Use of Data from the Study
"Acceptance and Change in Marital Therapy"**

1. Name and degree Arlene Cruz M.A. (Psy.D. doctoral student)
2. Professional affiliation Pepperdine University
3. Address 6100 Center Dr., Los Angeles, CA 90045
4. E-mail and phone number arlene_06@yahoo.com (818)282-3075
5. Advisor (if a student) Kathleen Eldridge, Ph.D.
6. Advisor's contact information Kathleen.Eldridge@pepperdine.edu 310-506-8559
7. Type of Project (Students only)
 - a. Dissertation
 - b. Master's Thesis
 - c. Undergraduate Honor's Thesis
 - d. Other (describe)

8. Brief Description of Research Project

TITLE: Concordance Between Therapists' Self-Reports and Neutral Observers' Ratings of Adherence to Marital Therapy

The current dissertation topic sets out to explore couple therapists' self-perceived adherence to two marital treatments, TBCT or IBCT, by comparing their ratings with those of neutral observers. Using the original data from the Christensen et al. (2004) study, two methods were used to gather observer ratings of adherence, and one measure was used to collect therapist self-reports.

The research questions that will be investigated in this study are:

1. What is the concordance between therapist self-reports and undergraduate student global observer ratings of adherence?
 - a. What is the correlation between therapists and raters on Change interventions?
 - b. What is the correlation between therapists and raters on Acceptance interventions?
2. What is the concordance between therapist self-reports and graduate student detailed observer ratings of adherence?
 - a. What is the correlation between therapists and raters on Behavior Exchange?
 - b. What is the correlation between therapists and raters on Communication Training?
 - c. What is the correlation between therapists and raters on Problem Solving Training?
 - d. What is the correlation between therapists and raters on Empathic Joining?
 - e. What is the correlation between therapists and raters on Unified Detachment?
 - f. What is the correlation between therapists and raters on Tolerance?

9. Description of Data Needed for Research Project

- Adherence ratings from graduate students who used the detailed *Behavioral Couple Therapy Rating Manual (BCTRM)*.

Date of Request 9/12/07

- Adherence ratings from undergraduate students who used the *Couple Therapist Rating Scale (CTRS)*.
- Therapist self-ratings on adherence to treatment from their responses on the *Sessions Ratings by Therapist (SRT)* questionnaire. Specifically, data from the checklist of the specific treatment interventions that the therapist used: Three of the items represent TBCT interventions (Behavior Exchange, Communication Training, and Problem Solving) and three of the items represent IBCT interventions (Emphatic Joining, Unified Detachment, and Tolerance Intervention).

10. Approximate Time Line for Research Project

October 2007: dissertation preliminary exam
November 2007: data analysis begins
July 2008: dissertation final oral exam

11. Proposed authorship

Dissertation: sole author Arlene Cruz
If published: Eldridge, Cruz and Christensen

Decision _____
 Request Approved Contingent on approval from
 Request Denied Because Pepperdine's IRB.

Signature Andrew Christensen Date 10/12/07
Andrew Christensen, PI

APPENDIX C
Copy of Measure BCTRM

Behavioral Couple Therapy Rating Manual
4/4/94

Table of Contents

1. Setting and Following Agenda (p.6)
2. Ordinary Conversation (p.6)
3. Assessing Collaborative Set (p.7)
4. Inducing Collaborative Set (p.7)
5. Behavior Exchange (p.7)
6. Praising Change (p.8)
7. Sex Therapy (p.8)
8. Companionship (p.8)
9. Problem Solving (p.9)
10. Problems as Differences (p.9)
11. Reasons for Partners' Differences (p.9)
12. Cognitive Interventions (p.10)
13. Genogram (p.10)
14. Reframing (p.10)
15. Soft Disclosures (p.11)
16. Communication Training (p.12)
17. Talking about Interaction Theme as an It (p.12)
18. Circular Questioning (p.13)
19. Preparing for Slip-ups and Lapses (p.13)
20. Positive Features of Negative Behavior (p.14)
21. Restraint of Change (& Other Strategic) (p.14)
22. In-session Rehearsal of Negative Behavior (p.14)
23. Instructing to Fake Negative Behavior at home (p.15)
24. Self-care (p.15)
25. Explicit Guidance (p.15)
26. Homework Assigned (p.15)
27. Homework Reviewed (p.16)
28. Generalization and Maintenance (p.16)

Introduction to Raters

The purpose of this study is to describe as accurately as possible what the therapist does during the sessions of couple therapy you will be coding. Because many of the interventions described in this manual could be used in both the therapies being compared, it is important that you listen and code each item carefully, based on what you actually hear rather than based on your guess about the type of therapy. Here are a few guidelines (adapted from the CSPRS Raters Manual) to help you rate the sessions.

Rate Therapist Behavior

All items are designed to measure therapist behavior. It is important to distinguish the therapist's behavior from the client's behavior in response to the therapist. The rater should attempt to rate the therapist behavior, not the client response to that behavior. In rating therapist behavior, the rater should consider what the therapist attempted to do, not whether those attempts met with success or failure.

Rate Extensiveness, Not Quality

The items are designed to measure the extent to which the therapists' engage in the behaviors being measured, rather than the quality with which those behaviors are performed. Although extensiveness is not totally independent of the quality of therapist behavior, the rater should not consider the quality of the therapist behavior per se when rating the items.

Frequency versus Intensity

Most of the items require the rater to rate how extensively (or thoroughly) the therapist behavior occurred. In order to determine the extent to which a therapist behavior occurred the rater must consider BOTH the frequency with which that behavior occurred during the session and the intensity with which that behavior was engaged in when it did occur. (Intensity means the therapist's concentration of effort or focus on the intervention.)

Items vary with regard to how relevant frequency and intensity are in determining how that item should be rated and there are no fixed rules for determining the importance of each concept. The relative weighing of these two concepts depends not only on which item is being rated, but also on which specific techniques the therapist uses to accomplish the strategy or goal stated in the item. For example, Instructing to Fake Negative Behavior at Home is an item for which intensity is more relevant than frequency.

This intervention may take comparatively little time within the session; however, as long as it is discussed directly with the couple it should receive a high rating. The less directly it is discussed the lower the rating it should be. On the other hand, Ordinary Conversation is an example of an item whose rating is based entirely on frequency. The more the therapist engages in ordinary conversation, the higher the rating should be.

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

Avoid Haloed Ratings

These items were designed for the purpose of describing therapist's behavior in the session. In order to use the scale correctly, it is essential that the rater rates what she/he hears, NOT what she/he thinks OUGHT to have occurred. The rater must be sure to apply the same standards for rating an item regardless of:

- 1) what type of therapy the rater thinks she/he is rating;
- 2) what other behaviors the therapist engaged in during the session;
- 3) what ratings were given to other items;
- 4) how skilled the rater believes the therapist to be in a particular modality;
- 5) how much the rater likes the therapist;
- 6) whether the rater thinks the behavior being rated is a good thing to do or a bad thing to do.

Rating Conjunctive Relationships

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

Use of Guidelines

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

Use of Examples

Whenever possible, examples have been included to illustrate how to rate therapist behavior. These examples, however, are only guidelines for how to rate an item. Often the example will only state that therapist behavior similar to the example merits a rating greater than a “1”. This is because the examples are of brief interchanges whereas the rater must consider the entire session when rating an item. The examples are a better guide to the kinds of behaviors and the intensity with which they should occur, than they are to the frequency with which behaviors should occur.

The manual includes reference to “low”, “medium” and “high” ratings in discussions of how examples should be rated. Because the rater must consider the entire session and not just a discrete incident or period of time in deciding the exact rating, these suggested ratings should not be considered fixed. In general, however, a low rating corresponds to 2, medium rating to 3 or 4, and high rating to 5. The manual explicitly states when the rater should assign a rating of 1. A low rating does not refer to a 1.

Making Distinctions

Because the items vary in terms of breadth of coverage, the same therapist behaviors which are appropriately rated in one item, may also be rated in another item.

Conversely, the rater is often required to make fine distinctions between therapist behaviors which are similar yet should be rated distinctly. Some items measure therapist behaviors which are similar and which may covary, but yet are distinct. The rater should be careful to rate them distinctly (i.e., in rating each item, the rater should consider the extent to which the behavior specified in that item occurred and should not consider other similar behaviors).

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

The Raters Manual also contains an “Important Distinctions” section within the entry for some items. This section contains information regarding how the item is similar to and different from other items. These “Important Distinctions” are not the only important similarities or differences that need to be attended to- don’t rely on “Important Distinction” sections to point out all of the important similarities and differences which exist.

Specific Instances Required for Rating

In order to give a rating greater than a “1”, the rater must hear a specific example of the therapist behavior being rated. The rater should be careful not to rate behavior as having occurred if she/he thinks it probably occurred but cannot think of a specific example.

Substantiating Ratings

The starting point for rating each item in the scale is 1, “not at all”. Give a rating higher than a 1 only if there is an example of the therapist behavior specified in the item. This is particularly difficult to do when rating the facilitative conditions items where the rater may be tempted to assign an average rating unless the therapist’s behavior was remarkable either by its absence or abundance. **DO NOT DO THIS.** The rater must be able to substantiate the rating she/he assigns to every item.

In particular, a high rating for facilitative items should be reserved for instances in which the therapist makes verbal statements that communicate rapport, warmth, etc. For example, a session characterized by frequent therapist statements such as, “*I really appreciate the risks you both have been willing to take to talk about such a sensitive topic with me,*” would receive a higher rating of rapport than a session in which the rapport is evidenced only through non-verbal actions such as the session seeming to flow smoothly without any obvious rifts. In other words, raters should substantiate ratings for facilitative items with verbal statements rather than solely non-verbal indications of facilitative conditions.

Overlap between Current versus Prior Sessions

Often an issue that was discussed in an earlier session is implicitly or explicitly referred to in the session being rated. For example, the client may seem to know what the therapist means when referring to communication training (because the couple must have learned it in a previous session). However, if communication training is mentioned only passing without the therapist conducting communication training in the current session, communication training should not be rated. Discussions, which took place in an earlier session, should not be considered in determining a rating given to the current session.

Instructions to Raters

1. RATE EVERY ITEM.
2. READ CRITERIA FOR ITEMS EACH TIME THAT THEY ARE RATED.
3. ATTEND TO MANUAL NOTES.
4. LISTEN BEFORE RATING.
5. TAKE NOTES.
6. FILL OUT CODE SHEETS CLEARLY AND CORRECTLY.

NOTE: There will be some therapist behavior that is not described by any item in this manual. One common example of this are seeking questions by the therapist: If the couple came in having had a fight during the week and the therapist simply asked, “*What happened?*” that statement need not be coded. Typically, the therapist will follow-up information seeking questions with interventions that you will be able to code under items in the manual.

1. Setting and Following Agenda.

Therapist worked with the clients to formulate and follow a specific agenda for the session.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Setting an agenda may include generating items to be discussed, choosing which of the items will be discussed, determining the order in which items are discussed, and allotting time to be spent on discussing each item.

Following the agenda includes therapist comments that remind the couple of the agenda and keep the discussion focused in order to cover items on the agenda. Sometimes the agenda must be revised and such therapist comments should also be rated here.

There are two aspects to consider when rating this item: 1) did the therapist work with the clients to set a specific agenda for the session? 2) did the therapist work with the clients to follow the agenda during the session?

2. Ordinary Conversation.

The therapist talked with the client about topics that seemed more likely ordinary conversation than therapy AND that cannot be classified under any other item.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

For example, the client and therapist may have talked about the weather, some recent news event, movies or a book, some place that they all have visited, etc., but in no way does the therapist tie the discussion topic to the client’s feelings, thoughts, or actions, currently or in the past. This item should not be rated higher than 1 unless the therapist in no way uses the conversation for assessment or intervention. Before rating this item, the rater should thoroughly check to rule out other items that might better describe the client and therapist’s interactions.

3. Assessing Collaborative Set.

Therapist asked questions in order to determine the extent to which each partner viewed himself or herself as the cause of some of the problems in the relationship and was willing to assume responsibility to make changes in his or her behavior to improve the relationship.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Important Distinction. This item differs from Item #4 Inducing Collaborative Set. In Inducing Collaborative Set, the therapist tries to get partners to act collaboratively despite how they feel. In Assessing Collaborative Set, the therapist simply asks questions to determine how each person views his or her role in causing problems.

4. Inducing Collaborative Set.

Therapist actively encouraged partners to work together collaboratively (i.e., changing his/her own behavior to improve the relationship without waiting for the other to change first), and/or reinforced positive client behavior which reflects an effort to behave collaboratively.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Inducing collaborative set can include the therapist presenting a model in which both partners accept responsibility for their own actions that contribute to marital distress, and the therapist persuading the couple to act collaboratively regardless of how they feel. Induction of collaborative set may sometimes have a “preachy” or “hard sell” tone as the therapist strongly tries to persuade each partner to make changes.

Important Distinction. Item #4 Induce Collaborative Set differs from Item #3 Assess Collaborative Set. The crucial aspect of Induce Collaborative Set is that the therapist actively asks the couple to adopt a particular orientation to therapy (focus on own role in creating problems and on changes he or she can independently make to improve the relationship). Whereas for Assess Collaborative Set, the therapist does not ask the couple to adopt a collaborative set but rather determines the extent to which the couple is or is not already collaborative.

Note: A rating of 4 or 5 should be reserved for when the therapist is actively persuading the couple to adopt a collaborative set, rather than solely presenting the model.

5. Behavior Exchange.

Therapist initiated and/or facilitated discussion of things each partner could independently do to improve spouse’s satisfaction with the relationship.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The therapist encouraged partners to make changes in order to increase marital satisfaction by:

- 1) generating lists of behaviors likely to please the spouse, OR
- 2) discussing hypothetical attempts to increase partners' marital satisfaction, OR
- 3) discussing past efforts to promote marital satisfaction through increases in pleasing behavior, OR
- 4) giving direct advice or suggestions about changes either partner should make to increase the other's satisfaction, OR
- 5) teaching parenting skills (e.g., how to get your kid to go to bed, or time out procedures).

Important Distinctions. When the therapist suggests or advises one or both partners to make changes in order to increase marital satisfaction AND the therapist does not make these suggestions in the context of formal problem solving, the therapist's behaviors should be rated as Item # Behavior Exchange. In other words, when the therapist helped the couple resolve some problem or difficulty by asking questions, proposing alternatives, etc., without using a specific format, this is rated as Item #5 Behavior Exchange rather than Item #9 Problem Solving.

6. Praising Change.

Therapist praised the couple's efforts at making changes by summarizing what worked, commenting on how hard they are working, how differently the interaction went because of their hard work, etc.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

7. Sex Therapy.

Therapist helped the couple improve sexual dysfunctions or dissatisfactions (e.g., used techniques such as sensate focus).

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Therapist helped the couple work on sexual problems: sexual dysfunctions (i.e., impotence, premature ejaculation, orgasmic dysfunction) and/or sexual dissatisfaction (e.g., different preferences regarding sexual activity or frequency, sexual boredom). The therapist may have developed activities designed to reduce fear of failure or pressure to engage in sexual activity. For example, the therapist may have used specific sex therapy techniques such as sensate focus (mutual, non-goal-oriented sensual interaction between the partners).

8. Companionship.

Therapist initiated/facilitated discussion of enjoyable activities that the couple could or has participated in together.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

9. Problem Solving.

Therapist taught or initiated practice in using a specific format for solving interpersonal conflicts.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The problem solving format includes defining the problem, brainstorming possible solutions, discussing the costs and benefits of various solutions, and coming to an explicit agreement. The therapist's teaching role involves didactic instruction, behavior rehearsal, and providing feedback.

10. Problems as Differences.

Therapist reformulated the problem either as deriving from a difference between the partners, OR as a vicious cycle resulting from each partner's attempt to solve the problem that their differences create.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The therapist pointed out how each one's behavior is reasonable and understandable given its place in the vicious cycle. A session could receive a rating of up to 5 if the therapist discussed problems either in terms of deriving from a difference between the couple, or as a vicious cycle that results from efforts to solve the problem; the therapist does not have to do both in order to receive a high rating.

Important Distinction. Item #10 Problems as Differences may occur with Item #11 Reasons for Partner Differences. The important aspect for Item #10 Problems as Differences is that the therapist emphasizes that the couple's problem is a result of how they ineffectively handle their differences as opposed to emphasizing the reasons for those differences. Item #11 Reasons for Partner Differences, however, should be rated when the therapist helps the couple understand the reasons for the differences, not the reasons for the problem.

11. Reasons for Partner Differences.

Therapist explored reasons why partners might differ regarding preferences for intimacy, time alone, need for reassurance, ways of showing affection, etc.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

These reasons should involve family history, factors in the current environment, or culture (sex roles, ethnic differences, or religious differences).

12. Cognitive Interventions.

The therapist led the couple to examine evidence for interpretations of or attributions about each other's behavior or to examine whether expectations about each other or marriage were reasonable.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The therapist challenged, through Socratic questioning, the logic or reasonableness of the client's interpretations, attributions, or expectations of each other. In the following example, the wife was complaining that the husband had not taken initiative nor followed through with helping one of their children with a school assignment. She attributes his inaction to a lack of interest in the children.

T: Mike, if it isn't just a lack of interest, as she is interpreting it, what is it?

H: No, I am interested. For example, I've been appalled at how little they know about what is happening in the world and I've been trying to read them some things from the newspaper or talk over things I hear on the news. It's just that assignment that he had to do was just not something I felt, I just felt incompetent.

T: So Gloria, I want to go back to your initial mis-guess about what's going on with him about why he doesn't get engaged more. Your original thought was, "He just doesn't care about the kids. He doesn't care about what is going on with them in school." And Mike just said that no I am interested and I have evidence that I am interested: I've been trying to think about how to increase their exposure to current events. If you had that different understanding, how would that make things different for you? How might this feel different to you?

13. Genogram.

Therapist asked each partner about their families of origin to create a structural diagram showing how patterns are transmitted intergenerationally and how past events such as death, illness, great success or immigration have influenced current patterns.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

14. Reframing.

The therapist reinterpreted one partner's negative behavior in a more positive light.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

For example (J & M, 1979, p. 144), "In the following excerpt, the couple is discussing the husband's tendency to conceal certain things from his wife; here they are discussing a bounced check which the husband intercepted before the wife discovered it.

W: You can't accept responsibility for your behavior. Whenever you do something wrong, you lie, deceive me. I can't stand your dishonesty.

T: It seems like her approval is very important to you (to husband). You care so much about what she thinks that you can't get yourself to tell her when you screw something up.

Here the therapist chooses to interpret the husband's behavior as indicating that he cares very much about his wife's opinion of him, a much more positive, and not any less accurate, outlook than the wife's perspective which attributes the husband's behavior to the trait of "dishonesty".

Important Distinction. Reframing should be rated only when the therapist reinterprets behavior, not emotions. If the therapist relabels emotions in a more positive light, that should be rated under Item #15 Soft Disclosures.

15. Soft Disclosures.

When clients were blaming, hostile, contemptuous (or expressing other strongly negative emotion), the therapist solicited partner disclosure of "soft" feelings and thoughts (e.g., fear, sadness, insecurity) and/or reinterpreted hard emotions in terms of their underlying softer emotions.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The therapist attempted to heighten the client's expression of her/his softer emotions or thoughts instead of the harder emotions expressed when attacking or blaming. To do this, the therapist may have solicited partner disclosure by helping the client to recognize and express softer thoughts or feelings that:

- 1) the client is unaware of; OR
- 2) the client is aware of but not expressing; OR
- 3) the client is expressing nonverbally but not verbally.

The therapist may either say what the client is feeling for the client or encourage the client to voice the softer emotions him or herself; either therapist behavior should be coded here.

NOTE: This item should not be rated higher than a 3 unless the therapist paid particular attention to helping the client express "soft" emotions. To give a rating higher than a 3 the

therapist must not only help the client express thoughts and feelings, but, in particular, help the spouse express vulnerability, sadness, disappointment, etc., likely to draw the couple together.

Important Distinction.

Soft Disclosure can be confused with two other items, Item #14 Reframing and Item #16 Communication Training. The important distinction between reframing and soft disclosure is the targeted behavior that is relabeled in a more positive light. Rate soft disclosure when the therapist relabels hard emotions in terms of their more primary softer emotions. Rate Item # 14 reframing when the therapist relabels overt behavior in a more positive light.

Soft disclosure should also be discriminated from Item #16 Communication Training. Although the therapist using communication training may ask the couple to talk about feelings, the therapist uses a specific format in order to increase the couple's skill in communicating; whereas in soft disclosure the therapist does not use a specific format, but instead seeks to articulate the softer emotions likely to draw the couple together.

16. Communication Training.

Therapist taught or initiated practice of active listening or expressive communication skills.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Communication training involves didactic instruction (e.g., modeling use of a specific format), behavior, rehearsal, and feedback from the therapist. Feedback is the provision of information to a couple regarding some aspect of their interaction; modeling (coaching) is instructing or demonstrating alternative responses; behavioral rehearsal is practice of new communication skills. Communication training may target any of the following: helping partners to listen more effectively and demonstrate understanding of each other; validating each other; teaching how to express positive and negative feelings; teaching how to express caring, appreciation, affection, and how to give compliments and praise; or teaching assertiveness skills. The essential element of communication training is that it is done in a teaching, didactic manner. The therapist's intervention need not be formal, but should definitely include feedback and rehearsal in order to be coded as communication training.

Communication training can occur in conjunction with other interventions. For example, while having the couple discuss the outcome of BE homework, the therapist may instruct and give feedback about the way partners describe their feelings about what the other did to please them. Or the therapist may comment during problem-solving training, "*Joe, when you repeatedly interrupt Mary as she tries to paraphrase what she heard your issue to be, it seems to be de-railing her. Try to wait until she is completely finished before you tell her what she isn't understanding about what you said.*" In these examples, communication training should be rated in addition to the other interventions (BE, Homework review, Problem-Solving Training). If the therapist asked the couple to practice communication skills at home, this should be rated both as communication training and as homework assigned.

17. Talking about an Interaction Theme as an “It”.

Therapist engaged partners in a general discussion of an interaction theme or issue without a focus on what could be done to change it, and without explicitly trying to teach expressive communication skills.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Therapist helped partners talk about the problem as something they share, rather than something that one does to the other. Said differently, the therapist tries to develop a descriptive rather than blaming account of the couple’s troubling interaction pattern. The therapist may do this in a variety of ways. The therapist may have helped each partner elaborate and articulate his/her particular feelings, thoughts, and actions in the theme. The therapist may have helped the couple identify the mutual traps. Humor or “short hand” labels to describe an interaction sequence may be used in order to help the couple gain a different perspective. These discussions could, but do not necessarily, involve:

- a) upcoming events, where the event is relevant to the interaction theme; or
- b) recent incidents, where a recent positive or negative incident was relevant to an interaction theme.

Important Distinction. When an interaction pattern is defined as the problem to be solved within the problem solving format, the therapist’s behavior should be rated under Item #9 Problem Solving rather than Item #17 Talking about an Interaction Theme as an “It”.

Similarly, when the therapist focuses on “reciprocal causation”, that is how what each does is in part caused by the other, but also focuses discussion on what partners can do to change this interaction pattern, this should not be coded as Interaction Theme as an “It”. Instead, when the therapist identifies reciprocal causation and asks the couple to consider changing, you should consider whether the therapist’s intervention is more appropriately rated as items Inducing Collaborative Set, Behavior Exchange, or Communication Training. For example, if the therapist said, “*when he does x, you do y. As soon as you do y, he does more of x. I want you both to take a minute to think about what you should do to make this go differently*”, and then the therapist went on to help each identify ways to change, this would be coded as Inducing Collaborative Set (focus on each changing own behavior in a slightly preachy “should” way) and as Behavior Exchange (changes to improve the other’s satisfaction).

18. Circular Questioning.

Therapist invited client(s) to describe the partner’s relationship with a third family member.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Rather than (or in addition to) asking the client directly about a conflict he or she experiences with a family member, the therapist invited the spouse to describe what he or she has observed. For example, the therapist might ask the husband, “*How does you mother-in-law see this conflict*”

between your wife and your son? When your wife disciplines your son, what does her mother do? How does your son then respond to his grandmother?"

19. Preparing for Slip-ups and Lapses.

Even during success with change efforts, therapist alerted the couple to the likelihood that “slip-ups” or “lapses” will occur.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

For this item to be rated highly, the therapist must have communicated that the couple cannot count on change by, for example, helping the couple prepare for the lack of change or discussing how the couple can have a good relationship while the problem occurs and as they try to recover from a slip-up. In other words, high ratings should be reserved for therapist interventions that clearly propose acceptance of lack of change and coping with lack of change.

It’s important to note that preparing for slip-ups and lapses should only be rated when the therapist intervention is future oriented or is a reminder of having predicted some problem would occur, rather than solely providing a rationale for change/progress being unsteady as a way to control damage after a slip-up.

20. Positive Features of Negative Behavior.

Therapist discussed or engaged couple in a discussion of the positive features of one or both partner’s negative behavior.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Therapist highlighted how what one or both partner’s view as negative actually serves an important use in the relationship. For example, the therapist might say, *“You, Mr. Brown, like to spend money and you, Mrs. Brown, like to save money. Even though this gives rise to a lot of conflict, your problems would be even worse if you were both the same; in your old age you would either be in debt from spending beyond your means or have savings but not have enjoyed yourselves. There is a real benefit of having both qualities in a marriage.”*

21. Restraint of Change (and Other Strategic Interventions).

Therapist suggested that couple should NOT change because change might be harmful or have a negative impact. Therapist may appear to be arguing against what is a “positive” change or to be playing devil’s advocate.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Strategic interventions are sometimes used in the context of client resistance to change: the therapist intervenes to create some contrasting position that pushes the client toward change. The therapist may instruct the couple not to change some troubling behavior with the intention of freeing the couple TO change.

22. In-session Rehearsal of Negative Behavior.

Therapist attempted to increase one or both spouse's ability to tolerate the other's upsetting behavior.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Therapist requested one member of the couple to role-play negative behavior in the session as a means of discovering feelings, thoughts, and actions as well as partner's reactions.

23. Instructing Couple to Fake Negative Behavior at Home.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Therapist asked one member of the couple to fake some negative behavior during the coming week by doing the negative behavior when they don't really feel it. Therapist explained the purpose of such faking to both partners.

24. Self-care.

Therapist encouraged couple to explore self-care possibilities, particularly, but not exclusively, those he or she can use when the partner does engage in negative behavior.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

25. Explicit Guidance.

The therapist directed or guided the session in an explicit way

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The rater should not rate how explicit the guidance was on any particular occasion. Raters should consider the extent to which the therapist explicitly controlled the direction of the session. The therapist might accomplish this by initiating a significant change in content or shift in focus of the session or by maintaining the focus on topics which she/he wants to discuss. If no guidance was provided OR if the guidance that was provided was not explicit, this item should be rated 1.

26. Homework Assigned.

Therapist developed or helped the couple develop homework.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Homework is a specific assignment which the client is to engage in (but not necessarily complete) before the next session. Rate this item low if the therapist off-handedly suggested, in order to bring the discussion to an end, that the clients engage in some behavior between sessions. Rate low to medium if the therapist asked the couple to do something between sessions but did not attempt to make the assignment more specific. Do not rate this item higher than a 4 unless the therapist helps the couple anticipate and resolve difficulties they might have in performing a homework assignment.

27. Homework reviewed.

Therapist paid attention to homework previously assigned to the couple.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

Homework refers to one or more specific assignments given by the therapist for the couple to complete between sessions. A high rating should be given only if the therapist attempted to use the couple's experiences with the homework as a basis for further discussion in the session.

Regardless of whether the clients completed the homework, the therapist can use the clients' experiences with the assignment as a basis for discussion (e.g., "*Were you able to attempt the homework? If not, what happened to prevent you from trying it?*"). In other words, this item should be rated independently of whether the couple completed or even attempted the homework; a rating of up to 5 can be given in such cases.

28. Generalization and Maintenance.

Therapist fostered the couples' ability to continue to apply skills or ideas learned in therapy to improve the relationship when problems arise in the future.

1	2	3	4	5
not at all	some	moderately	considerably	extensively

The therapist initiated discussion of how what the couple has learned in therapy can be continued outside the session or after therapy has stopped. A high rating should be given when the therapist thoroughly plans how the couple can continue to use what they have learned in therapy outside the session or after therapy has ended. For example, the therapist may introduce

the idea of state of the relationship meetings in which the couple agree to meet at a specific time to function as their own therapist after therapy.

Important Distinction. Item #28 Generalization and Maintenance is different from Item #19 Preparing for Slip-ups and Lapses in that Generalization and Maintenance has to do with how the couple will maintain change, whereas Slip-ups and Lapses the focus is on accepting a lack of change.

APPENDIX D
Copy of Measure CTRS

Appendix A:
Couple Therapist Rating Scale

Introduction to Rating

The purpose of this project is to accurately describe therapist behavior in couple therapy. It is expected that therapists have their own unique style that is expressed with every client. In addition, it is expected that therapists may modify their own unique style according to the personality and needs of each client. Here are some guidelines for you to follow while coding the therapy sessions (adapted from the CSPRS and BCT Rating Manuals).

Each session will last approximately 50 minutes. An overall rating will be made at the end of the session that represents the entire session. Each code will be rated on a 9 point scale with one end representing the behavior not happening at all and the other end representing the behavior occurring a lot. The coder makes a judgment based on the extensiveness of the given behavior relative to other therapists. Therefore, it is necessary that the coder get a sense of the typical frequency of the given behavior in therapy sessions. In order to gain this sense, coders will participate in a training period in which they practice the coding system with a series of therapy sessions.

The codes are explained below. Although some are distinct from each other, many are not mutually exclusive; therapist behavior may be an example of more than one code at a time. In making ratings, coders should consider the extensiveness of the behavior in question. The extensiveness can be a combination of the frequency and intensity of the behavior displayed.

Coders should focus primarily on therapist behavior, although some of the codes take into account what the couple is doing. The coder is also permitted and encouraged to replay any portions of the session necessary to make an accurate final rating of the overall session. The following guidelines should ensure the accuracy of your ratings.

Rate Therapist Behavior

The codes reflect therapist behavior only. Therefore, it is necessary to rate only therapist behavior, not client behavior. It is important to make the distinction between the therapist behavior and the client's response to the therapist. The client's response or the success or failure of what the therapist attempts to do is not considered in the code. The coder should only consider what the therapist attempted to do.

Rate Extensiveness, Not Quality

The codes are meant to reflect the extent to which the therapist engaged in the given behavior, not the quality with which the coder thinks those behaviors are performed. Although extensiveness and quality are not completely independent, the coder should not consider quality of the therapist behavior per se when making a rating.

Frequency vs. Intensity

To rate the extensiveness of therapist behavior, it is important to consider two things: frequency and intensity. Frequency is the number of times the therapist engaged in the behavior. Intensity is the amount of concentration, effort, or focus the therapist places in the intervention when it occurs.

The importance of frequency and intensity in making a rating will depend on the behavior in question. Some behaviors take little time within the session but may vary in the intensity with which the therapist engages in them. A less explicit behavior is usually considered less intense. No fixed rules exist for determining the equivalence of a behavior done intensely for a short period of time versus a behavior not done intensely but done frequently. It is up to the coder to weigh the frequency and intensity in the given situation to make a rating.

Avoid Haloed Ratings

Haloed ratings are ratings based on what the coder thinks OUGHT to have happened and should be avoided. Instead, the coder should rate what is actually heard. The coder should rate what is heard, not what should have occurred, regardless of:

- 1) what other behaviors the therapist has engaged in during the session;
- 2) what ratings were given to other items;
- 3) how skilled the coder believes the therapist is;
- 4) how much the coder likes the therapist;
- 5) whether the coder thinks the behavior being rated is a good or bad thing to do.

Use of Guidelines

The descriptions of behavior included in this manual are not meant to encompass all possible behaviors, and should be considered guidelines and not rules. Coders are expected to use their best judgment when rating all behavior including behavior not explicitly outlined in this manual.

Specific Instances Required for Rating

The starting point for each code is “1”, not at all. In order to give a rating greater than “1”, the coder must hear a specific example of an item under the code being rated. It is important to avoid rating behavior as occurring if the coder thinks it probably did occur but can not think of an actual example.

Overlap Between Current Versus Prior Sessions

Occasionally, an issue that was discussed in a previous session is referred to in the present session. However, if the issue is mentioned in passing without the therapist engaging in a specific behavior again in the current session, the behavior from the previous session should not be rated in the current session. Discussions that occurred in an earlier session should not be considered when determining a rating for the current session.

Trained versus Non-trained codes

While there are 14 items on the coding sheet, we will be training only on the first four items. Descriptions of these items appear below. You should rate all of the remaining 10 items for every session, but we will only touch briefly on these in our training. However, all of the above suggestions still apply to these remaining items as well.

Additional Instructions

1. Listen carefully to the entire session.
2. Take notes if necessary.
3. Attend to manual instructions.
4. Read the criteria for the codes each time they are rated.
5. Always rate every code.
6. Fill out the coding sheets clearly and correctly.

anger, intolerance) or therapist reinterprets negative emotions in terms of the underlying painful emotions.

7. Therapist encourages a nonblaming, descriptive discussion of an interaction theme or problematic issue.
8. Therapist acknowledges the probability of lack of change, slip-ups, or lapses in the future.
9. Therapist points out the positive features of partner's negative behavior.
10. Therapist encourages increased tolerance of negative behavior by rehearsal of negative behavior or instruction to fake negative behavior at home.
11. Therapist encourages exploration of ways partners can be more self-reliant and get needs met outside the relationship, particularly when a partner engages in negative behavior.

1	2	3	4	5	6	7	8	9
does not display/promote understanding/acceptance							extensively displays/promotes understanding/acceptance	

4. Homework

1. Therapist directly or indirectly assigns homework assignments for the couple to complete in between sessions.
2. Therapist reviews progress on previously assigned homework or reviews reactions to the assignment.

1	2	3	4	5	6	7	8	9
homework not assigned/reviewed at all							homework extensively assigned/reviewed	

APPENDIX E
Copy of Measure SRT

Couple ID _____
Session number: _____

Date of session: _____

Session Ratings by Therapist

Fill in the bubbles of all that apply:

1. Couple called me since the last session. Reason for call was (please circle one):
a) scheduling
b) emergency
c) other
If emergency, briefly describe:

2. Couple was late by _____ minutes.

3. Couple failed to show for a session since the last session I had with them.

4. Husband failed to complete homework assignment for this session.

5. Wife failed to complete homework assignment for this session.

6. Treatment procedures which I used in this session (fill in all that you used):

- | | |
|---|---|
| <input type="radio"/> Behavior Exchange | <input type="radio"/> Empathic Joining |
| <input type="radio"/> Communication Training | <input type="radio"/> Unified Detachment |
| <input type="radio"/> Problem Solving Training | <input type="radio"/> Tolerance Intervention |
| <input type="radio"/> Discussed a recent conflict in detail | <input type="radio"/> Discussed an upcoming event |

7. I was adherent to the treatment procedures (ICT or TBCT)

Not adherent
(included strategies from
alternative treatments)

Somewhat adherent

Extremely adherent
(included only specified
treatment strategies)

8. How effective do you believe you were as a therapist in this session?

Not effective

Somewhat effective

Extremely effective

9. How beneficial do you believe this treatment session was to the couple?

Not beneficial

Somewhat beneficial

Extremely beneficial