

DETERMINING INTERVENTION FIDELITY FROM CHRONOLOGICAL FIELD
NOTES

Jo Ann Dowell

A dissertation submitted to the faculty of the University of North Carolina at
Chapel Hill in partial fulfillment of the requirement for the degree of Doctoral in the
School of Nursing of

Chapel Hill
2011

Approved by:

Linda S. Beeber, PhD, FAAN Chair

Sheila Santacroce, PhD

Marcia Van Riper, PhD

Sarah E. Bledsoe, PhD

Todd Schwartz, Dr.PH

©2011
Jo Ann Dowell
ALL RIGHTS RESERVED

ABSTRACT

Determining Intervention Fidelity from Chronological Field Notes (Under the direction of Linda Beeber)

Intervention fidelity (IF) is a relatively recent methodological consideration in nursing research that refers to the adherent and competent behaviors by the interventionist in the delivery of the intervention as planned (Santacroce, Maccarelli, & Grey, 2004; Stein, Sargent, & Rafaels, 2007). Intervention fidelity is a major contributor to the internal validity of a study. When an intervention is delivered as prescribed, the inference that the outcome is caused by the intervention is supported. If intervention fidelity is low, the expected differences between intervention group and control group may not occur. Historically, researchers have used various recording devices [audio or video tapes] to gather data on intervention fidelity. However, concerns about privacy or disruption of the intervention as a result of taping may prevent this method from being used. Field notes written at the time of the delivery may be a solution. However, field notes have not been extensively tested as an alternative device. The purpose of this pilot study was to explore an alternative method of determining intervention fidelity using chronological field notes (CFNs) (Stein, et al., 2007; Waltz, Addis, Koerner, & Jacobson, 1993). Method: A secondary analysis using data from an intervention study that used chronological field notes was completed. A randomly selected sample of chronological field notes (N=181) written by interventionist nurses (N=20) were coded using the Collaborative Study Psychotherapy Rating Scale (CSPRS), a validated fidelity instrument used in similar studies. Results: The interrater reliability (IRR) on the CSPRS was excellent on adherence (IRR .95 and Cronbach alpha .97) and for the new competence scale, IRR .74 and Cronbach alpha .91. The results on the CSPRS were

comparable to a study (Hill, 1992) that used traditional methods of transcribed audiotapes. This exploratory study provided support that an alternative methodological approach to determining intervention fidelity is feasible. Further research with a larger sample is needed to conclude with greater certainty that intervention fidelity can be determined using field notes.

Keywords: intervention fidelity, adherence, competence

ACKNOWLEDGEMENTS

One of the most interesting and exciting aspects of this study was having a network of a well--renowned experts in nursing science and supporters to guide me throughout the process. To the casual observer, a doctoral dissertation may appear to be solitary work. However, to complete a project of this magnitude requires a network of experts and supporters, and I am indebted so many people. Although part of the work of the dissertation is in isolation, there was this envelope of protection that was provided by so many people in the School of Nursing and others.

Dr. Linda Beeber, PhD, PMHCNS, FAAN who chaired my committee, inspired me to pursue the integrity of intervention studies. I am most grateful to my mentor and my strongest advocate, Dr. Beeber for her tireless mentoring and allowing me to explore new avenues of research related to a new methodological approach in determining the fidelity of an intervention. It is the endless hours sitting in her office reviewing and discussing from the time the idea was conceptualized to the final written document that probably felt to Dr. Beeber as a never ending process. However, this direction from an expert nursing research scientist who was willing to share her expertise has afforded me the knowledge toward developing a successful program of research.

I am most especially grateful to Dr. Martha Keever, who has provided me with continuous editorial support throughout my study. Although I would often call on Dr. Keever for her editorial support, it was often not her editorial that was received. Instead, it often was encouragement and support. It was Dr. Keever who gave me the fortitude that I have the ability to complete this program. To her I owe tremendous amount gratitude.

I am most humble to have committee members with enormous amount of expertise in nursing research to guide me throughout this journey with continuous support. Especially, I would like to thank Dr. Todd Schwartz for helping me to understand different ways of establishing the analyses for this study.

I am truly grateful to a friend and colleague, Brandy Mechling for her patience and willingness to learn about the role of the rater. Brandy's participation as a volunteer second rater in this study contributed to the completion my research. I truly thank you for all of your contribution to this body of knowledge.

I also am most grateful for my family especially my mother, my daughter, Hawanya, my sisters and brothers for their support and extraordinary courage to be patience with me when I was too involved in the work to participate in family gatherings throughout this process.

To Ginny Lewis and all of Dr. Beeber's Team past and present, I appreciate your support and continuous assistance in obtaining the tools that I needed to complete this work. It was their enduring patience and assistance with a smile that made my day a success although I may have been doubtful.

Last, I would like to acknowledge Dr. Clara E. Hill, Professor of Psychology at the University of Maryland at Baltimore for permission to use the Collaborative Study Psychotherapy Rating Scale and her continuous emails in explaining the function of this reliable instrument.

TABLE OF CONTENTS

LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
CHAPTERS	
I. INTRODUCTION.....	1
Aims.....	2
Conceptual Framework.....	5
Operational Definitions.....	7
II. LITERATURE REVIEW.....	9
Fidelity and Internal Validity.....	10
History of Intervention Fidelity in Psychotherapy Research.....	15
Intervention Fidelity Studies in Nursing Research.....	24
Methods of determining fidelity.....	31
Limitations of Existing Systems of Measuring Fidelity.....	40
III. METHODOLOGY.....	45
Design.....	45
Sample.....	47
Methods.....	50
Procedure.....	53
Rater Guidelines.....	54
Data Preparation and ongoing management.....	58
Procedure for rating the CFNs.....	58

	Analysis.....	59
	Limitations.....	64
IV.	RESULTS.....	65
	Characteristics of the intervention nurses.....	65
	Results by the Aims.....	67
	Aim 1.....	67
	Aim 2.....	76
	Aim 3.....	79
	Aim 4.....	80
	Summary of results by Hypotheses.....	84
V.	DISCUSSION.....	87
	Rating from different observers.....	88
	Summary of the findings.....	89
	Limitations.....	93
	IMPLICATIONS for Nursing Research.....	94
VI.	Appendix A Collaborative Study Psychotherapy Rating Scale (IPT).....	96
VII.	Appendix B Collaborative Study Psychotherapy Rating Scale (CBT).....	106
VIII.	Appendix C Rater Applied Performance Scale.....	117
IX.	Appendix D Intervention Fidelity: The rater’s manual.....	118
X.	Appendix E Program Evaluation.....	120
XI.	REFERENCES.....	121

LIST OF TABLES

Tables

II.1 Characteristics of Intervention research and IF.....	24
II.2 Characteristics of nursing intervention research and IF.....	31
III.1 Nurses, geographic location and setting type.....	47
IV.1 Nurses' age, years of experience, length of time on the project and location of the delivery.....	66
IV.2 Reliability with the master coder for each subscale total (MDS 1) in Trial 1.....	71
IV.3 An example of a grid of the raters' and master coder's ratings used for training purpose.....	72
IV.4 Reliability with mater coder (MDS 2) in Trial 2	74
IV.5 Inter-rater reliability (MDS 1) in Trial 1.....	74
IV.6 Inter-rater reliability (MDS 2) in Trial 2.....	75
IV.7 Level of agreement over time between the master coder and raters.....	75
IV.8 Level of agreement between raters over time.....	75
IV.9 Means and Std. Deviation.....	77
IV.10 Inter-rater reliability for IPT and CBT for the overall sample.....	78
IV.11 Internal Consistency between the raters and for the first 10 field notes	78
IV.12 Comparison from the study of subscale IPT & CBT on adherence with NIMH	79
IV.13 Inter-rater reliability and Cronbach alpha on IPT competence.....	80
IV.14 Correlations between adherence and competence on IPT and CBT plus between prescribed behaviors (IPT) and proscribed behaviors (CBT).....	81
IV.15 Factors means and standard deviations.....	81

IV.16 Bivariate correlation of the predictors on adherence and competence with IPT.....	82
IV.17 Linear regression predicting greater adherence (Models 1 & 2 and competence (Models 3 & 4).....	83
IV.18 Multiple regressions for all factors predicting greater adherence and competence for IPT.....	84
V.1 CSPRS CBT items measuring strategies that the nurses were trained to do.....	92

LIST OF FIGURES

Figure

I.1 Conceptual Framework Model for the study.....	7
II.1 Dimensions of a broader fidelity concept.....	17
III.1 Determining the sample of nurses, sessions and cases.....	49

ABBREVIATIONS

IF	Intervention Fidelity
CFN	Chronological Field Notes
CSPRS	Collaborative Study Psychotherapy Rating Scale
IRR	Inter-rater reliability
IPT	Interpersonal Psychotherapy
CBT	Cognitive Behavior Therapy
CM	Clinical Management
FC	Facilitative Conditions
ED	Explicit Directiveness
NIMH	National Institute of Mental Health
TDCRP	Treatment of Depression Collaborative Research Program
RCT	Randomized Clinical Trials

CHAPTER 1

INTRODUCTION

Intervention fidelity (IF) is defined as the adherent and competent behaviors by the interventionist nurse in the delivery of an intervention as planned (Santacroce, et al., 2004; Stein, et al., 2007). Intervention fidelity is a major contributor to the internal validity of an intervention study or the extent to which one can infer that an outcome occurred as a result of an intervention instead of uncontrolled, extraneous factors (Cook & Campbell, 1979; Nezu & Nezu, 2008). The domains of IF are *adherence*, defined as “the extent to which the prescribed (*desirable*) elements of an intervention have been delivered and proscribed (*prohibited*) elements avoided”, (Santacroce *et al.*, 2004, p. 63; Waltz *et al.*, 1993, p.630) and *competence*, defined as the extent to which the interventionist “has displayed behaviors that typically engage participants in the intervention and affect outcomes in the desired direction” (Najavits & Weiss, 1994, p.679; Santacroce, *et al.*, 2004, p. 63).

Traditionally, researchers have used observers, audio tapes and/or video recordings, or interventionists’ self-reports to gather data on IF (Moncher, 1991). The “gold standard” for IF requires the use of a reliable measure applied to a randomly-selected sample of taped intervention sessions that are stratified on relevant dimensions and rated by trained raters who demonstrate consistent reliability (Santacroce, et al., 2004). However, this method may be problematic for certain settings and types of research where audio or video taping will interfere with the delivery of the intervention. For example, the use of recording devices in

intervention studies that require the revelation of sensitive issues may limit individuals from sharing this information. In addition, traditional IF procedures that use observational methods require labor-intensive work and high cost to complete.

An alternative device, field notes written at the time of delivery of the intervention, may be another basis for IF measurements, especially when sensitive topics or trust are issues. The proposed study explored the use of data retrieved from a standardized form of field documentation (chronological field notes [CFN]) using trained raters and a reliable measure of IF.

AIMS:

The overall aim of this study was to explore whether or not an alternative type of data collection method (chronological field notes [CFN]) could be used as the basis for the valid and reliable assessment of intervention fidelity (IF). To test the hypothesis we proposed the following specific aims:

AIM 1. Describe the process used to establish and maintain the reliability of the raters on the Collaborative Study Psychotherapy Rating Scale (CSPRS) applied to CFNs:

1a. Establish a master set of ratings by an expert intervention researcher for the CSPRS using CFNs that will not be used in the study.

1b. Develop a raters' manual for determining the CSPRS item ratings.

1c. Develop a training program for raters.

1d. Train two raters using the materials developed in 1a-c.

1e. Evaluate the outcomes of the training program.

1f. Determine each rater's level of agreement with the master coder ratings and with one another at completion of the training. We hypothesize that:

H1f1. Raters who receive comprehensive training will achieve adequate agreement with the master coder; that is, the level of agreement will reach a reliability coefficient of 0.9.

H1f2. Raters who receive comprehensive training will achieve an IRR with one another of greater than 0.9.

1g. Determine each rater's level of agreement with the master coder and with one another over time.

H1g1. Raters will achieve a reliability coefficient of 0.9 with the master coder and an IRR of 0.9 on a master data set at a checkpoint conducted midway through the data.

AIM 2. Explore how the CSPRS ratings derived from CFNs function as a measure of the IF domain of adherence. Using the CSPRS adherence scale specifically:

2a: Compute IRR between the two raters and internal consistency on adherence for each of the theoretically-derived two subscales: interpersonal psychotherapy (IPT) and cognitive behavior therapy (CBT).

2b. Compare IRR and internal consistency of each CSPRS subscale on adherence using CFNs with the scores derived from observational transcribed data from the original National Institute of Mental Health (NIMH) study. We hypothesized:

H2b IRR and Cronbach's alpha on adherence ratings for each of the two subscales (IPT and CBT) will be equal to the NIMH published scores for IRR (IPT - .78, CBT - .90,) and alphas (IPT - .90, CBT - .94).

Aim 3. Explore how the CSPRS ratings derived from CFNs function as a measure of the IF domain of competence. Specifically, using a newly-added competence scale on the CSPRS:

3a. Compute IRR for the two raters and internal consistency on competence for each of the two subscales (IPT and CBT).

3b. Compare the internal consistency on competence for each of the five CSPRS subscales to established standards. Specifically:

H3a. IRR on competence for each of the two subscales (IPT and CBT) on competence will achieve a reliability coefficient of equal to or greater than 0.80.

H3b. Each of the two CSPRS subscales on competence will achieve a Cronbach's alpha of .80.

Aim 4: To explore whether CSPRS ratings derived from CFNs reflect the expected relationships among nurses' adherence, competence and relevant personal and contextual factors. Specifically:

H4a. Adherence (score on IPT subscale) and competence (score on IPT subscale) will have a strong positive correlation.

H4b. IPT scores (prescribed behaviors) and CBT scores (proscribed behaviors) will have a strong negative correlation.

H4c. Greater age, years of experience, and length of time on project will predict greater adherence.

H4d. Greater age, years of experience, and length of time on project will predict greater competence.

H4e. Type of setting (urban, suburban and rural) in which the nurses delivered the intervention will be differentially associated with adherence.

The science of IF is essential and significant for the validity and reliability of nursing intervention research. The accomplishment of these aims will provide information on the reliability of chronological field notes as a source of data for determining IF. This study will move nursing forward toward establishing preliminary evidence that CFNs can be used as an alternative when other devices, such as tape or video recordings, are not feasible.

Conceptual Framework

The proposed framework for this study is based on the work of Waltz et al. (1993) on the behaviors that affect adherence and competence of the therapist (unique and essential, essential but not unique, acceptable but not necessary, and proscribed); the work of Barber et al. (1996, 2001) relating to interpersonal psychotherapy skills (IPT; e.g., warmth, empathy, congruence, and genuineness) by the therapist; and the work of Hollon et al. (1984) relating to preparation and engagement by the therapist.

Waltz's Work on Adherence and Competence

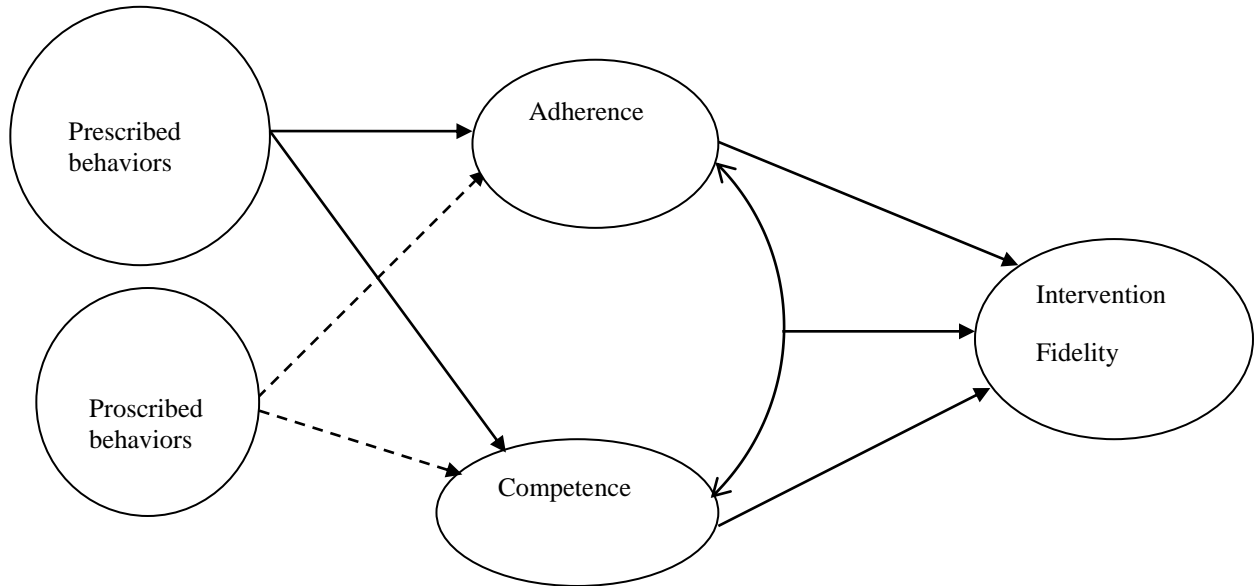
The conceptual framework uses Waltz et al.'s (1993) definition of adherence and competence. Waltz defined *adherence* as the extent to which the therapist applies the intervention per treatment protocol and avoids procedures proscribed by the manual, and *competence* as the level of skill, or the extent to which the interventionist displayed behaviors that typically engage the participants in the intervention and has an effect on the outcomes in the desired direction. Adherence includes prescribed behaviors (essential and unique, essential but not unique, and acceptable but not necessary) that are specific to the treatment intervention. In addition, adherence includes avoidance of proscribed behaviors. Proscribed behaviors are not a part of the specific treatment intervention. Further, if they are present but not accounted for they may or may not have an effect on the intervention outcome, which

could confuse interpretation of the results. Waltz's definition of competence also accounts for contextual elements, i.e. the relevant aspects of the therapeutic frame of reference. This may include the a) participants' degree of impairment, b) problems that manifested, c) life situations and stressors of participants, and d) degree of improvement already achieved and appropriate sensitivity to the timing of intervention within a therapy session (Waltz, et al., 1993). For example, one treatment may require the therapist to be warm and empathic in delivering the treatment, whereas another treatment would not (Waltz, et al., 1993).

Waltz et al. made three recommendations for IF research. These are: 1) interventionist competence in an IF research study must be defined relative to the protocol manual; 2) that adherence measures should include four types of therapist behaviors (essential and unique to a particular treatment modality, essential but not unique, acceptable but not necessary, and behaviors that are proscribed); and 3) that therapists must establish their competence through independent measurement and verification, rather than investigators assuming that experience level and training are equivalent to the skill level (Waltz, et al., 1993).

For this study, adherence and competence and the indicators of each as defined by Waltz will be the focus. The model (Figure 1.1) explains IF. Intervention fidelity is proposed to be determined by the adherence and competence of the interventionists. The extent to which proscribed behaviors are present and proscribed behaviors are absent determine the adherence and competence of the nurses. The broken lines indicate that there may or may not be a relationship between proscribed behaviors and adherence/competence of nurses.

Figure I.1 Conceptual Framework Model for the Study



Operational Definitions

In this study, the following concepts and operational definitions will be used.

Adherence:

- *Conceptual Definition:* Adherence is defined the extent to which the prescribed elements of an intervention have been delivered and proscribed elements are avoided (Santacrose, et al., 2004; Waltz, et al., 1993).
- *Operational Definition:* In this study, adherence will be represented by two ordinal ratings representing the degree to which interventionist nurses' chronological field notes reflect: prescribed behaviors measured by the IPT sub-scale of CSPRS and proscribed behaviors as measured by the CBT sub-scales of the CSPRS.

Competence:

- *Conceptual Definition:* Competence is defined as the extent to which the interventionist has displayed behaviors that typically engage the participants in the intervention and affect outcomes in the desired direction (Najavits & Weiss, 1994; Santacrose, et al., 2004)
- *Operational Definition:* In this study, competence will be represented by an ordinal rating representing the degree to which interventionist nurses' chronological field notes reflect their skill level in the delivering the treatment intervention and engaging the participants using prescribed behaviors.

CHAPTER 2

LITERATURE REVIEW

The purpose of this chapter is to provide a synthesis of the literature on IF that includes the definition of IF, the relationship of IF to internal validity, the history of IF in psychotherapy and nursing research, and methods of determining intervention fidelity. The literature that was reviewed included a wide array of research on IF, especially within the past ten years. Several different disciplines were represented in the literature review, including public health, psychology, psychiatry, education, medicine, nursing, and social work. The literature on IF showed considerable variation in the methods of determining IF including face-to-face observations or the use of different devices such as audiotapes, videotapes, process notes, transcripts, or field notes. In addition, multiple measures of IF have been used as well.

Definition of Fidelity

Fidelity derives from the Latin word *fidēlis*, meaning “faithful or loyal” (“Merriam Webster's Collegiate Dictionary,” 2002). Through intervention fidelity (IF), or the adherent and competent behaviors by the interventionist nurse in the delivery of an intervention as planned (Santacrose, et al., 2004; Stein, et al., 2007), researchers may collect evidence to support and verify a theoretically-based, predicted causal link between an intervention and the outcome (Bellg, et al., 2004; Moncher & Prinz, 1991; Resnick, Inguito, et al., 2009; Waltz, et al., 1993). Intervention fidelity has two core components: adherence and competence (Stein, et al., 2007).

Adherence has been defined as the therapist applying the prescribed ingredients of the treatment intervention and avoiding proscribed behaviors (Santacrose, et al., 2004; Stein, et al., 2007; Waltz, et al., 1993). There are interrelated conditions that affect the adherence of the interventionist such as the interventionist's ability to provide some degree of congruence, quality of delivery, and appropriateness of the dose (Barber, Stratt, Halperin, & Connolly, 2001). Congruence relates to the degree to which the interventionist must come to grips with the inevitable restrictions that adherence to the protocol requires and maintain strategies conducive to the experimental condition (Weissman, Rounsaville, & Chevron, 1982). The interventionist's lack of adherence to the prescribed protocol may lead to a diminished quality of the delivery, producing a compromised outcome.

Competence, the second core component of IF, is defined as the extent to which the interventionist has displayed behaviors that typically engage participants in the intervention and affect outcomes in desired direction (Najavits & Weiss, 1994). The literature is limited on measures for competence of psychotherapists; however, Carroll, K. et al. (2000) reported in their article that competence was measured by how well the therapist performs. Waltz et al. (1993) moved away from the general therapeutic definition and focused instead on the contextual definition of competence. These contextual factors that may influence the interventionist's competency are the client difficulty, the stage of therapy, or the client's characteristics, such as the presence or absence of collateral problems. Competency is defined relative to the intervention manual (Waltz *et al.*).

Fidelity and internal validity

Internal validity is the degree to which the findings of a study can be attributed to the study intervention as opposed to other factors, i.e., that the experimental treatment

(independent variable), rather than uncontrolled, extraneous factors, is responsible for observed effects (Cook & Campbell, 1979; Nezu & Nezu, 2008). These extraneous factors may include general threats to internal validity relating to the design of the intervention, its delivery, the manipulation process, the number of individuals participating in the study, and the behaviors of the interventionists. The threats to internal validity relating to the design of the study are a) history, the effect that events that are external to the study design have on the participants; b) selection, the effects of preexisting differences between the groups; c) maturation, the developmental changes that take place within a participant over the course of the intervention study; d) testing effects, the effect from participants' learning from the instruments and repeated responses to the questions; e) instrumentation, the effect of changes in data collection procedures or measurement tools over the course of the study; and f) mortality, the differential loss of participants from intervention and control groups (Kerlinger & Lee, 2000). In a scientific experimental intervention, randomized clinical trials (RCTs), the researcher maintains internal validity through the procedures and the processes built into the study's design (Horner, Rew, & Torres, 2006) including the systematic delivery of the intervention (Spillane, et al., 2007)

IF, through its dimensions of adherence and competence, is a significant contributor to internal validity as well as the effect size and statistical power of a study (Bruckenthal & Broderick, 2007; Resnick, Inguito, et al., 2005; Spillane, et al., 2007), external validity, construct validity, and statistical conclusion validity (Calsyn, 2000; Moncher & Prinz, 1991; Nezu & Nezu, 2008). Cook and Campbell (1979) explained that statistical conclusion validity focuses on the appropriate use of various statistical analytic procedures to determine

the actual co-varying relationship between a given independent variable (IV) and a given dependent variable (DV).

Methods for establishing reliable delivery are central to the integrity of any randomized trial. Researchers have revealed that the more complex an intervention is, the more likely there is a negative effect on the delivery of the intervention (Calsyn, 2000; Horner, et al., 2006). The risk of variation may occur when the intervention is offered multiple times. Using multiple interventionists also increases the risk of deviation from the original study design (Calsyn, 2000). In addition, intervention studies that have multiple sessions may be at risk for variations in the delivery (Kerns & Prinz, 2002). For example, an interventionist may make minor adjustments to the protocol based on assessments from earlier sessions (Bellg, et al., 2004; Horner, et al., 2006). Further, as intervention studies become more complex, there is the risk of what is known as interventionist “*drift*.” This means that unplanned changes take place as the interventionist drifts from the protocol manual (Bellg, et al., 2004; Horner, et al., 2006; Nigg, Allegrante, & Ory, 2002; Santacrose, et al., 2004). Even though, there are some adjustments made during most intervention studies, it is this subsequent adjustment that may lead to further deviation from the original design. The more changes that occur during the intervention research, the more likely there will be deviation from the original design of the research study, which ultimately affects the validity of the study. To summarize, the researchers’ strategies to maintain the integrity of the intervention delivery across all sessions contributes to the confidence that the study findings occurred because of the intervention (Horner, et al., 2006). Interventionist training and ongoing supervision can be a means to maintain a high level of IF and thus support internal validity in RCT; however, the provider (interventionist) who lacks competence in the

intervention may fail to be adherent. This action affects the outcome and the integrity of the intervention.

To determine whether the independent variable (treatment) is responsible for changes in outcomes, manipulation checks such as IF monitoring should be completed (Brandt, Kirsch, Lewis, & Casey, 2004; Waltz, et al., 1993). Manipulation checks are necessary to provide information to support assumptions about ensuring internal validity (Mowbray, Holter, Teague, & Bybee, 2003). For the researcher, providing manipulation checks means IF monitoring, which includes examining both adherence and competence of the interventionist during the intervention delivery. Waltz (1993) concluded that manipulation checks are essential to infer conclusions about the treatment intervention effects or the absence of an effect. In addition conducting manipulation checks such as IF monitoring throughout the study may provide information about the strengths and weakness of the interventionists and help researchers know whether extra training is needed to be able to infer a treatment effect.

Multiple interventionists can also increase the risk of deviation from the original study design (Calsyn, 2000). The increase in the number of research team members with different responsibilities may pose a threat to the validity (Whitmer, Sweeney, Slivjak, Sumner, & Barsevick, 2005). This means that the more people there are implementing the intervention, the more likely there will be some deviation from the original intended treatment. Studies have shown that intense training of the team members and ongoing supervision may decrease the threat to some degree (Mowbray, et al., 2003; Resnick, Inguito, et al., 2005).

IF is also determined by the behavior of the interventionists. Literature on the behavior of interventionists in psychotherapy has described four behaviors: a) unique and essential, b) essential but not unique, c) acceptable but not necessary, and d) proscribed or prohibited (Waltz, et al., 1993). For the interventionists to maintain these behaviors there needs to be a close adherence to the protocol manual, which may provide instructions or guidelines in implementing the treatment intervention. Unique and essential behaviors by the interventionist include techniques that are specific to the treatment intervention. Essential but not unique behaviors relate to the techniques that are essential to the intervention but are not necessarily unique (Waltz, et al., 1993). Waltz et al. report that these behaviors may also be present in a comparison treatment group. Interventionists may also behave in ways that are acceptable but not necessary or included in the intervention protocol. For example, an interventionist may disclose personal information or incorporate motivational interviewing techniques during delivery of an intervention that is not intended to be included.

The interventionist may perform behaviors that are not prescribed in protocol manual and that may or may not have a negative effect on the outcome of the intervention or intervention fidelity. There are behaviors that are prohibited and are not a part of the experimental condition. Researchers identify these behaviors as proscribed, which are in violation of the protocol manual.

In summary, to control for threats to internal validity, the investigator must ask: a) was the therapy being studied carried out by the interventionist providers as prescribed by the treatment manual and (b) was it implemented in a competent manner? These questions address the therapist's adherence and competence.

History of Intervention Fidelity in Psychotherapy Research

Historically, the concept of intervention fidelity (IF) was identified in the 1970s and addressed in social and behavioral studies. The original guidelines of IF were initiated to further enhance internal validity (Hill, O'Grady, & Elkin, 1992; Moncher & Prinz, 1991). Prior to this date the term IF was used interchangeably with treatment integrity, but now IF has a distinct meaning of its own. The earlier definition of treatment integrity referred to the degree to which a treatment condition is implemented as intended over the duration of the study (Moncher & Prinz, 1991).

The demand for accountability in psychotherapy influenced the need for intervention fidelity (Moncher & Prinz, 1991). Despite the efforts by researchers, IF was not adequately addressed (Moncher & Prinz, 1991). Researchers' concerns over validity of controlled clinical trials of psychotherapy were the focus of what constitutes the best method for empirically validating treatments (Ablon & Jones, 2002; Mowbray, et al., 2003; Santacrose, et al., 2004; Stein, et al., 2007). There were debates as to efficacy versus effectiveness being preferable for empirically validating treatments (Ablon & Jones, 2002). Efficacy studies emphasize the importance of being able to draw a causal inference between the treatment provided and the outcomes (Ablon & Jones, 2002; Leventhal & Friedman, 2004). Efficacy studies typically assign participants randomly to different treatment conditions whose parameters are closely controlled. An underlying assumption of these studies is that the interventions are the cause of the observed changes that constitute the study outcomes. Effectiveness studies, or testing an intervention under real world conditions, emphasize external validity (Ablon & Jones, 2002). External validity for an RCT refers to the generalization beyond the particular providers and patients participating in the study to other

populations, settings, and variables (Nezu & Nezu, 2008). Participants in effectiveness studies may be selected without using exclusion criteria. These studies do not have a set number of sessions or a treatment manual.

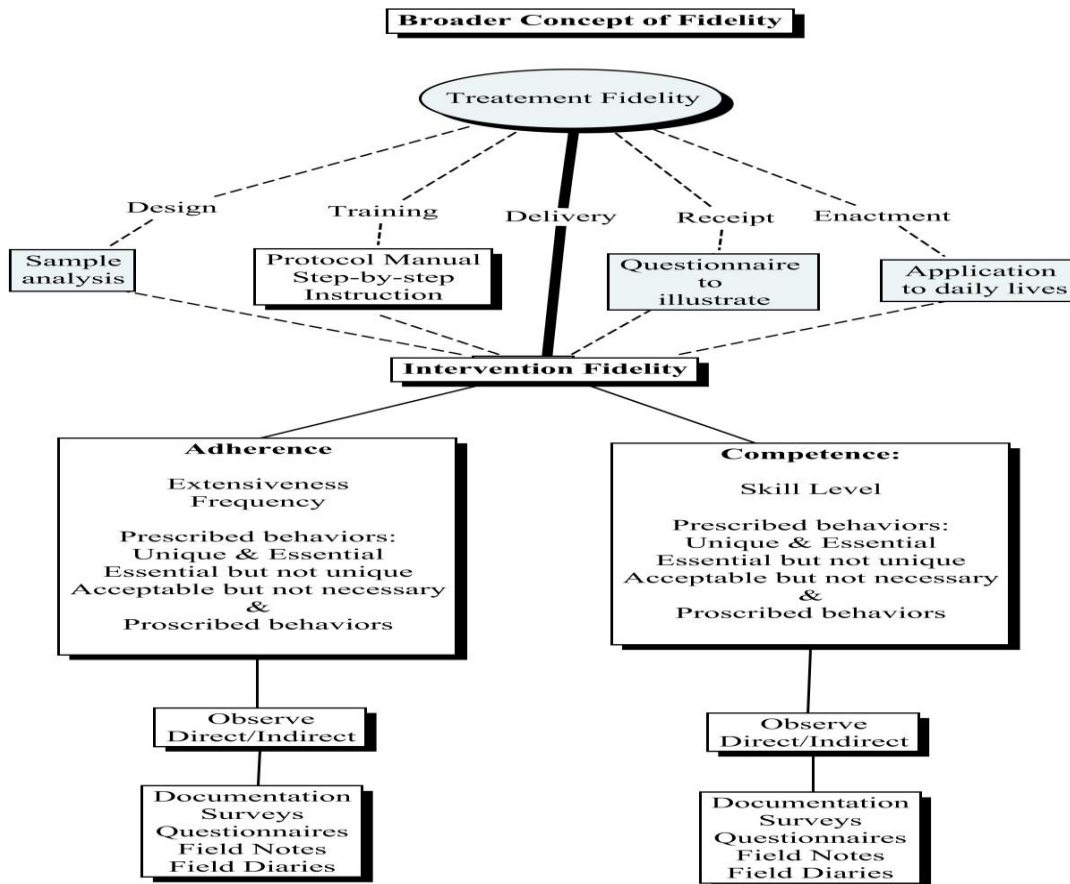
Concerns over the validity of assumptions underlying clinical trials research led to the development of the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP), which focused their attention on constructing a methodologically sound RCT comparing different forms of psychological therapy for depression (Ablon & Jones, 2002). These researchers emphasized the utilization of program manuals for training and supervising the intervention staff to monitor program quality and performance in order to ensure fidelity (Borrelli, et al., 2005; Mowbray, et al., 2003). Although studies have implemented parts of treatment integrity, such as design, there were limited specific guidelines for reporting the process in which IF was monitored (Borrelli, et al., 2005).

Treatment fidelity involves the degree to which treatment condition is implemented as intended (Moncher & Prinz, 1991). During the earlier years researchers expanded the definition of treatment fidelity to include five accepted dimensions: (1) the design of the study, (2) the technique with the administration of the intervention, (3) how it was received, (4) whether it was given as intended per protocol, and (5) “enactment,” or whether the individual can apply the intervention to everyday life (Moncher & Prinz, 1991; Peterson, Homer, & Wonderlich, 1982). The Behavior Change Consortium (BCC), established as a branch of the National Institutes of Health (NIH), consisted of scientific researchers from multiple disciplines and set the standards for fidelity of behavior change intervention research with a set of recommendations published in 1999. These included the dimensions

that had been previously identified, i.e., (1) the study design, (2) training of the interventionists, (3) delivery [IF], (4) how received, and (5) enactment (Ory, Jordan, & Bazzarre, 2002; Resnick, Inguito, et al., 2005).

An organized model illustrates how the concept of IF has been conceptualized over time in Figure II.1. The dimensions of this model show how treatment fidelity includes multiple components that include *adherence* and *competence* of therapist in maintaining internal validity. Treatment fidelity is especially relevant to the delivery process, in which the focus is the behaviors of the providers or the ways in which the providers carry out the experimental condition. As previously mentioned, the dimensions of delivery that the proposed study will address are IF and its domains of *adherence and competence*.

Figure II.1 Dimensions of a broader fidelity concept



To further understand the link between BCC guidelines and intervention fidelity, a full explanation will be provided. First, treatment fidelity as it relates to the study design ensures the study can adequately test its hypotheses in relation to the theory and clinical processes (Resnick, Bellg, et al., 2005). Secondly, training and supervision of the therapist indicate that the providers of the intervention have received adequate information and were monitored in the delivery of the intervention (Resnick, Bellg, et al., 2005). The third guideline, delivery, is associated with the process of monitoring interventionists or behavioral performances that take place as the intervention is given to the participants. For example, questions about delivery ask under what circumstance the delivery was implemented. Monitoring assures that the interventionist nurses delivered the intervention. The fourth guideline is that the participants received the treatment and the participants' behavior changed. The fifth and final guideline is related to enactment, which refers to the participant's ability to perform the behavioral skills and cognitive strategies in real life settings. The BCC concluded that ongoing work is needed to quantify treatment fidelity rather than assuming that it is achieved at an acceptable level because of the study design (Resnick, Bellg, et al., 2005; Resnick, Inguito, et al., 2005). The consensus from the BCC group was that IF, that is, intervention delivery, should be evaluated by direct or indirect observation. The BCC recommended use of checklists to measure delivery (Resnick, Bellg, et al., 2005; Resnick, Inguito, et al., 2005). Additionally, a measure of the ability of the interventionist to follow the protocol was previously recommended as a way of quantifying IF (Kazdin, Mazurick, & Siegel, 1994). In the psychotherapy literature, the focus was on treatment fidelity in outcome studies which related to the assurance that any therapy intervention under investigation was actually implemented in a competent matter in

accordance to the procedural elements as delineated in a treatment manual (Nezu & Nezu, 2008).

A specific group of studies have reported implementing IF procedures in the delivery of a psychotherapy treatment (Table II.1). Lee et al. (2008) used an innovative technology, e.g., a web-based monitoring system, to assess fidelity of a school-based intervention. The Early Risers “Skills for Success” program was an early-age preventive intervention for children at risk for serious conduct problems that was delivered over 16 sessions of psychotherapy intervention to 28 independent elementary schools (by site visits to the schools and homes). Fidelity was measured by using an on-line monitoring system for documentation. The web-based fidelity system was developed to monitor the implementation and to collect data on adherence. The therapists were trained using the Early Riser protocol manual. The measurement technology featured a fidelity monitoring system that required weekly reporting to assess exposure and adherence. Quality of delivery was assessed by using five telephone surveys of the interventionists. No measure of competence was used. The study showed that interventionists had higher adherence to the prescribed protocol around the fixed components of the intervention as opposed to the variable components of the intervention.

Martino et al.’s (2008) goal was to determine the extent to which the clinicians in an addiction treatment program implemented different therapies. The dimensions studied were adherence and competence of 35 therapists from five outpatient community programs who delivered either a three-session adaptation of motivational enhancement therapy or an equivalent number of drugs counseling sessions (usual care) to 461 clients within a National Institute on Drug Abuse Clinical Trial Network. All of the therapists were trained with

expert-led intensive workshops, audio taped practice cases, and protocol manuals. Each therapist's adherence and competence were evaluated using the Independent Tape Rater Scale (Ball et al. , 2002), adapted from the Yale Adherence Competence Scale (YAC-S) (Carroll, et al., 2000). Fundamental and advanced motivational strategies included monitoring to prohibit behaviors of the interventionists. Using interclass correlations (ICC), the investigators analyzed adherence and competence between treatment conditions (Motivation enhance therapy [MET] and Counseling-as-usual [CAU]), program sites, and therapists. The findings were that each of the first two dimensions (adherence and competence) was significantly and positively related to client change (negative drug screens). Inconsistent strategies were significantly and negatively correlated with client change. Adherence and competence were more variably related to other client outcomes (such as treatment retention and abstinence). The IF measures were limited to the YAC-S items; therefore other aspects of the performance by the therapists were mentioned in this study, for example, proscribed behaviors. Proscribed behaviors may be strength or a problem for the treatment outcomes.

Hogue, Liddle, Becker, & Johnson-Leckrone (2002) examined IF in a multidimensional family prevention, a family-based prevention counseling model for adolescents at risk for substance abuse and behavior problems. Monitoring fidelity included using the Therapist Behavior Rating Scale-Version 2 (TBRS-2) to capture adherence and differentiate between Multidimensional Family Prevention (MDFP), Multidimensional Family Therapy (MDFT), and Cognitive-Behavioral Therapy (CBT). Fourteen counselors were trained using didactic seminars and treatment manuals. Fidelity (adherence and dose) was determined from a pool of randomly-selected videotapes of sessions by 14 counselors

who provided family therapy to 43 adolescents and families (Hogue, Liddle, Becker, & Johnson-Leckrone, 2002). Nine raters were selected from undergraduate students in advanced psychology courses and one graduate student. The videotapes were coded by trained raters using TBRS-2 (Hogue, et al., 2002). The two dimensions of TBRS include intervention domain and intervention techniques. The raters were trained in group format for 90 minutes per week over a 4-month period to reach an adequate pre-study reliability (ICC > .65 for most study items). No other report was indicated for the raters' reliability. Analysis of variance (ANOVA) comparing average session length revealed no significance among the conditions: MDFP: $M = 51.1$ ($SD = 24.1$); MDFT: $M = 52.8$ ($SD = 17.9$); CBT: $M = 45.9$ ($SD = 11.1$); $F(2,194) = p = 1.13$. The Levene statistic = 7.97, $p < .001$ showed that there was variability in length for MDFP and MDFT greater than that in CBT. A MANOVA was used to conduct a multivariate analysis in which all levels of the dependent variable were measured on the same scale. The investigators found overall that the MDFP and MDFT were identical to one another and predictably different from CBT (2005). The major IF findings were that MDFP demonstrated adherence to its core intervention principles in comparison to MDFT and in contrast to CBT across two dimensions of model fidelity [intervention parameters focused on dose in minutes and intervention techniques focused on adherence].

Carroll, K. et al (1998) study focused on a manual-based psychotherapy process ratings systems, the Collaborative Study Psychotherapy Rating Scale (CSPRS) as method of assessing treatment integrity and discriminability. A randomized clinical trial of psychotherapy and pharmacotherapy was done using 121 ambulatory patients who were labeled as cocaine abusers. The patients were randomized to one or four treatments which were delivered in individual sessions over 12 weeks of an abstinence initiation trial. These

included 1) desipramine in combination with CBT; 2) desipramine in combination with clinical management (CM), a supportive psychotherapy control condition; 3) CBT plus placebo, and 4) CM plus placebo. All treatment sessions were videotaped for supervision and assessment. The authors evaluated discriminability of the two forms of psychotherapy and adherence to manual guidelines using independent observers who rated the session videotapes. The observers used the CBT and CM subscales from the CSPRS to rate adherence. The therapists also completed the CB checklist after each session. Both the therapists and raters received didactic training.

To analyze the outcomes, the authors computed psychometrics on the adapted rating scale and compared to those reported for the CSPRS (Hill et al, 1992). The coefficient alphas for the CBT subscale, CM and nonspecific factors were 0.90, 0.76, and 0.58, respectively. IRRs for the subscales were computed using ICC (CBT-.78; CM-.85, and nonspecific scale-.57). For the item discriminability, those on the CM subscale were higher than those on the CBT subscale. Results indicated earlier sessions were more reflected higher level of adherence to the treatment manuals (earlier- θ =.78 versus later sessions- θ =.55). Cognitive behavioral sessions were rated as significantly higher than clinical management sessions on nonspecific factors (CBT .31 versus CM.13). There was a gradual reduction across treatment weeks in the number of cognitive-behavioral interventions delivered (linearity $F= 48.8$, $p < 0.001$).

A study in progress by Wahab et al. (2008) is focusing on the delivery of a motivational interviewing intervention to increase screening for colon cancer. In a longitudinal study, 524 people have been randomized to three study groups: Group 1, the usual care group, which is colon cancer screening; Group 2, tailored health counseling; and

Group 3, motivational interviewing by telephone. The intervention fidelity study had focused on the delivery of the intervention from Group 3, motivational interviewing by telephone. Three interventionists were trained in motivational interviewing according to a manual. Delivery is being measured by using the Motivational Interviewing Treatment Integrity scale (MITI). The authors have reported that supervision is essential to fidelity but have not specifically reported how interventionist fidelity is being measured.

In summary, there was variability among the IF outcome studies, the level of adherence and the accuracy of reporting on adherence. Although all five of the studies showed that there was high adherence to the prescribed protocol manual, with one exception, the results were limited to the individual intervention studies. This seemed to indicate that as long as the interventionist is trained to a particular intervention or clinical trial and receives ongoing supervision, the adherence level will probably be high. However, none of these adherence measures of fidelity can be generalized to other studies, except Marino, who used the YACS. Although YACS is a generalized system measure that has been used in other studies, it is not specific to an intervention. Three studies reported measures of competence; however, only two revealed a positive relationship to the outcome of the intervention. The one fidelity study that measured adherence and competence of the therapist showed no main effects for competence on any outcome variable. The problem may be that three of the intervention studies have been tested in real world practice settings and implemented under broader, less controlled conditions. One of the fidelity studies did not provide any specifics on how the intervention fidelity was measured. Despite discussion of fidelity, the studies reviewed above illustrate that only a few studies have actually included measures of fidelity and reported the results. Adherence has been measured much more frequently than

competence. However, both are necessary for the delivery of an intervention. Martino et al. (2008) showed that high fidelity by the therapist correlated positively with the intervention outcomes. This study used the YAC scale which has items measuring both adherence and competence. Adherence alone strongly contributes to a high fidelity level, and future studies will benefit from separating competence and understanding its unique contribution to fidelity.

Table II.1 Characteristics of intervention research and IF

CITATIONS	INTERVENTIONISTS	FIDELITY STRATEGIES	MEASURES	CONCEPTS/ SCALE DIMENSION	LENGTH OF INTERVENTION (DOSE)	DETAILED DESCRIPTION OF THE INTERVENTION	FIDELITY OUTCOME
Psychiatric/psychology intervention studies							
Lee <i>et al.</i> , 2008	Family advocate	Web-based system	Web survey	Exposure Adherence delivery	16 sessions	Early Risers "Skills for Success" program	Exposure-0.43-0.87 Adherence-0.40-0.98 Quality of implementation-0.65-0.86
Martino <i>et al.</i> , 2008	therapists	Audio tape	YACS	Adherence Competence	3 sessions	Community program: Motivational enhancement therapy	Fundamental MI skills Adherence- Competence- $p < .001$ Advanced MI skills Adherence Competence $p < .001$ MI inconsistency Adherence Competence
Hogue <i>et al.</i> , 2005	Counselors [3 masters level & 1 doctoral]	Video tapes	Therapist Behavior Rating Scale	Adherence	25 sessions	MDFP-family-based intervention	
Carroll, K. <i>et al.</i> , 1998	Doctorial level therapists	Video tapes Self-report & observers	CSPRS	Adherence	Individual sessions over 12 weeks	Treatment of cocaine abusers	IRR-CBT-.78; CM-.85; Non-specific-.57/ F= 48.8, $p < 0.001$
Social Work intervention study							
Wahab <i>et al.</i> , 2008	Health counselors	Hand written note/audio tape	MI	Adherence	6 sessions	Motivational interviewing	All 3 interventionists score an average of 100% MI-adherent

Intervention Fidelity Studies in Nursing Research

For this section, the literature search consisted of IF information from empirical research studies. The search included databases of PubMed, CINAL, Google Scholar, PsycInfo, and ISI web. The terms used for the search included *intervention fidelity research in nursing, intervention fidelity, intervention fidelity in nursing, randomized clinical trials, adherence of interventionist, competence of interventionist, adherence of psychotherapist, competence of psychotherapist, and nursing intervention research*. An overview of the literature search included a wide array of research on intervention fidelity research. Intervention fidelity (IF) is relatively new to nursing research. Experimental nursing

interventions often are flexible, dynamic, and individualized to be sensitive to the unique characteristics of the participants. Nursing studies have addressed empirical support showing the extent to which the intervention was implemented as planned (Horner, et al., 2006; Resnick, Bellg, et al., 2005; Santacroce, et al., 2004; Stein, et al., 2007). Eight studies have reported implementing fidelity procedures in the delivery of nursing intervention.

A Res-Care intervention study was a randomized controlled trial that involved 12 nursing homes in which six were exposed to the Res-Care intervention and six received the control condition (Resnick, Galik, et al., 2009). The restorative care intervention in nursing home settings focused on teaching the nursing assistants how to motivate residents to engage in restorative activities and incorporate them into routine daily care. Res-Care Nurses (RCN) that were trained based on a training manual provided restorative care education to 256 nurse's assistants (NA) for 20 hours per week during the 12 month duration of the intervention. The dimensions of fidelity focused on design, training, delivery, receipt, and enactment. Fidelity was measured using a restorative care documentation flow sheet (Resnick, Galik, et al., 2009). Treatment fidelity was met if manual-based training was successfully provided by each of the five RCNs. Delivery fidelity was measured by the NAs' provision of restorative care (Res-Care). NAs, who provided 60% (SD = 29%) of all care interactions at four months and 65% (SD = 25%) at 12 months, were consistent with the Res-Care training. The investigators measured fidelity related to receipt (e.g., NAs learned material) with a pre-test and post-test. A total of 222 out of 256 NAs completed the sessions. Enactment was measured by the completion of flow sheets; the NAs enacted 70.5 (SD = 56.2) minutes daily of restorative care to the resident participants in the study.

Erien et al. (2008) delivered a 12-week telephone intervention designed to improve medication adherence in 99 participants with HIV infection. The purpose was to determine the level of fidelity to each of the specified end points (number of intervention sessions, session duration, number of telephone attempts per session, the length of time between sessions, sessions with multiple interventions, and attribution from intervention protocol) in the delivery of intervention. Although there was no report about the number of nurses involved, nurse interventionists were trained to deliver the intervention according to a protocol. Fidelity was determined from randomly selected audio-recorded sessions, field notes, and tracking forms. Delivery was the primary focus for fidelity monitoring. The project director reviewed each intervention session for fidelity to the protocol and provided feedback to the nurses. The findings, focused on fidelity of the specified endpoints, revealed that the nurses provided on average 8.1 sessions (SD = 4.07) per participant. Forty-three (43%) of the participants did receive all 12 sessions. Seventy-seven of the participants received at least the first five sessions. Twenty-one participants dropped out of the treatment before completion, of which two received none of the 12 intervention sessions. A fourth of the participants received more than one intervention topic delivered during a single call. The delivery of the intervention lasted an average of 11.3 minutes (median = 10.0; range = 4.7-26.7). Even though the delivery occurred within the expected time, more than half of the participants received less than the planned treatment. Women in this study were more likely to double up the interventions ($p = .036$). A marginally significant difference ($p = .075$) was noted in the number of sessions received between whites and non-whites, with whites (M = 8.8, SD = 3.9) receiving more sessions compared to non-whites (M = 7.4, SD = 4.2). The

race-by-income effect was observed for the mean number of attempts to contact the participants ($p = 0.44$). The overall outcome of the intervention was less effective.

Minnick et al. (2008) delivered an intervention that included the use of nurse coaches (NC) to assist primary care practices in adopting a new model of patient care called Virtual Integrated Practice (VIP). The main focus of VIP is better care management of the adult Type 2 diabetic, including a tool box approach for primary care practices. The three practices that participated were two internal medicine practices (seven physicians) and a family practice (three physicians). Three nurse coaches with Master's degrees were trained as nurse coaches and on the VIP protocol. The dimensions measured were delivery and receipt. Delivery fidelity was determined using a Nurse Coach Process Marker record developed by the investigator, a 38-item self-report of the coaching activities associated with NC domains, as well as interviews from physicians and staff. The investigators found that the NC from two of the practices completed 94.7% of tasks. NC from the third practice completed 71.1% of tasks. The investigators reported verification from the physician and staff interviews. Receipt fidelity was determined by the lead physician and staff interviews. Each NC succeeded in facilitating the adoption of VIP processes in the assigned practice. The limitation in this fidelity study appears to be lack of information on adherence or competence of the nurses. Although the method for assessing intervention fidelity captured the activities and general time input of the NCs, there is limited information on the quality of the delivery and its impact on the outcomes for diabetic patients.

A nursing intervention study was developed to help low-income families improve lifestyle and weight-related health indicators in their overweight children. The study used structured field notes and audio tapes for collecting fidelity data (Tyler & Horner, 2008).

Two advanced practice nurses were trained in motivational interviewing. Fidelity and consistency in the delivery of the intervention were determined by a structured field-note guide and randomly-selected intervention visits that were audiotape recorded. Field notes were completed following the sessions. Fidelity was measured by the principal investigator and a research assistant reviewing and coding the audiotape recordings. The researchers review of field notes and audiotape helped to ensure that nurse interventionists adhered to the intervention. The investigator gave a descriptive analysis from the field notes and the audiotapes of the parent-child-provider interactions and reported that they had an effect on families making changes to improve eating patterns and physical activity. The collaborating negotiation intervention overall outcome provided the families with a plan in managing their child's obesity. Although the investigator's intent was to report the interventionist nurses' adherence, there was no report on the outcome of adherence or competency of the interventionist nurses.

Pbert et al. (2006) delivered a smoking cessation intervention to improve abstinence rates among adolescents interested in quitting. This one-on-one nurse-delivered smoking cessation intervention focused on 37 schools and 571 adolescents. School nurses were trained in the implementation of the intervention using an intervention manual. Fidelity implementation was determined by adherence and dose. The measure of fidelity included self-report questionnaires from the students and the schools. The intervention effects on participants' smoking status at six weeks and three months were significant ($p < 0.001$). After adjusting for school, students in the intervention had odds of quitting that were seven times greater than the students in the control school (OR = 7.3; 95% CI 3.4, 15.6). A

stronger intervention effect was observed at three months (OR = 5.9; 95% CI 3.6, 9.6). The investigators reported high fidelity by the nurses.

Although the previous nursing intervention studies did not include competence of the interventionists, investigators from two different intervention studies included both adherence and competence of the interventionist nurses in the fidelity measures for their studies. Brandt et al. (2004) evaluated the strength and integrity of a pilot study of a behavioral intervention to assist mothers with breast cancer and their children in the home. The intervention study involved eight families and included five sessions over a 10-week period. The intervention investigator included a scripted protocol. There was no report as to how the interventionist was trained in the pilot intervention study. The primary dimensions of fidelity monitoring included purity as a measure of integrity or adherence to the scripted protocol, dose (number and duration of sessions), and competence (skill level). Adherence to the scripted protocol was obtained by using a checklist of expected behaviors to review two audio-recorded intervention sessions for each family intervention session. Dosage was determined by tracking the number of the sessions and completed homework assignments. The last dimension, competency, was determined by a self-report measure, the Working Alliance Inventory (WAI). The investigators found that no errors were found in the delivery of the content as outlined in the protocol; however, skill-building omissions by the interventionists were detected in the majority of audited sessions (Brandt, et al., 2004). Although the skill-building was not explicitly scripted in the protocol, it was written as an overall guideline for the interventionists to use. With regard to dosage, all eight of the families completed all the sessions; however, only six families completed all of the homework assignments. One of the eight families completed 80% of the assignments and

another family completed 60% of the assignments. For the competence of the interventionist, descriptive statistics of WAI were calculated to determine the mutuality of goals between the mother and the interventionist. Using the Wilcoxon signed ranks test, the results of mother and interventionist pairs showed no statistically significant differences, which means that the mother and the interventionist had comparable perceptions about the working alliance formed with each other.

Bruckenthal et al. (2007) assessed IF in a pilot study of nurse practitioner-delivered pain Coping Skills Training (CST). This pilot intervention study included six patients who received ten sessions of training. The nurse practitioner received two days of training using a CST treatment manual, role playing, evaluation, and feedback. Fidelity dimensions were determined by adherence to the manual-based protocol and enactment. The measure of fidelity included a comprehensive checklist that provided information on delivery style and interpersonal skills. Adherence by the NP as measured by the interpersonal skills used during the delivery was determined by audiotapes, which were analyzed by two raters. The raters' inter-rater reliability was 0.71, using the *k* statistics. Enactment was determined by a post-treatment questionnaire given to the recipients concerning their ability to manage pain from their osteoarthritis. Nurses' adherence to the training points of the manual was 86%. The raters reported an overall effectiveness of the nurses in the delivery on a scale of 1= poor to 5 = excellent, with a final mean score of 4.4. The second rating of the nurses' interpersonal skills by the raters gave a mean score of 4.5. Recipients reported a mean rating of 4.2 out of 5, indicating overall satisfaction with the treatment; 4.0 for ability to manage their osteoarthritis; and 4.8 in the nurses' ability to be helpful. Table II.2 provides further information on the previously mentioned fidelity monitoring research.

Table II.2 Characteristics of nursing intervention research and IF

CITATIONS	INTERVENTIONISTS	FIDELITY MONITORING DEVICES	MEASURES	CONCEPTS/ SCALE DIMENSION	LENGTH OF INTERVENTION (DOSE)	DETAILED DESCRIPTION OF THE INTERVENTION	FIDELITY OUTCOME
Nursing and Field Notes							
Resnick <i>et al.</i> 2009	RCN	Flow sheets	Training Delivery Receipt Enactment	Training Delivery Receipt enactment	12 months	Educational program	
Erien & Sereika, 2006	Nurse	Audio taping & field notes	Intervention tracking form	Delivery Duration Time elapsed	12-session	Medication-taking behaviors of HIV patients	Intervention session delivery 8.1 (SD= 4.07)
Minnick <i>et al.</i> 2008	Nurses	NC Process Markers record notes	Interviews to verify records	Adherence Delivery Receipt	22 weeks	Nurse Coach Quality improvement intervention	Authors report 94% adherence rate
Tyler & Horner, 2008	School nurses	Structure field notes & random selected audio	Audio used to verify field notes Code sheet	Adherence	12 weeks with booster on 25 th week-follow for 37 weeks	“Collaborative negotiation intervention” with (CHeWS)	Investigator report s adherence
Nursing and Other/ Audio/Video Devices Only							
Pbert <i>et.al.</i> 2006	School nurses	Training Questionnaire	Questionnaire Self-report	Adherence	4 sessions	Smoking cessation intervention	High Level of Fidelity
Brandt <i>et al.</i> 2004	Nurses	Audio	WAI	Adherence competence dose	5 sessions	Cognitive-behavioral approach to improve the quality of m-c relationships & child’s adjustment to their mothers’ breast CA	Competence- 0.95=client 0.87=Interventionist
Bruckenthal & Broderick, 2007	Nurse practitioners	checklist audio	NP-Log	Adherence competence	10 sessions	Coping skills training for chronic pain	Inter-reliability 0.71 using k statistics 86% adherence of NP

In spite of the eight studies in which there were attempts to measure adherence of the interventionists, there continues to be limited information on the measures of these dimensions. Further, there are only two nursing intervention studies that included the dimension of competence in the monitoring of fidelity. Competence of the interventionist provides information on the skill level and whether the interventionist preformed in a competent matter. It is essential to measure the competence of the interventionist for IF. As noted before, fidelity should include the adherence of the interventionist to the prescribed elements of the intervention and the competence, skill level, or performance of the interventionist to assure that delivery has occurred.

Methods of determining fidelity

To determine the IF of a treatment intervention, data must be generated to support the assumption that the manipulation of the independent variable occurred as planned (Moncher

& Prinz, 1991; Stein, et al., 2007). Hence, the data support that the theoretically predicted causal link between an intervention and behavioral change on outcomes in the participants has occurred (Bellg, et al., 2004; Moncher & Prinz, 1991; Resnick, Inguito, et al., 2005). Several structures have been developed to create these data: protocol manuals, which are used to determine consistent preparation and, later, the degree of interventionists' adherence to the protocol; methods to obtain data about the delivery directly and indirectly; and instrumentation to measure various dimensions of fidelity. This section will discuss each of these aspects.

To understand the essential features in providing instruction with randomized clinical trials of psychosocial intervention, the researcher needs to have some form of written documentation explaining how the therapy is to be conducted (Carroll, et al., 2000). If no documentation were to exist, the source of the intervention effect would remain vulnerable to alternative interpretations. Researchers have met this need through the development of intervention protocols that are laid out in manuals to provide the basis for training interventionists, guidelines as to how the intervention is to be conducted, and, also, the means by which to measure how well adherence has occurred. For example, Bruckenthal et al. (2007) reported that the protocol manual was utilized at two levels: both to train the nurses and to provide a guide for measurement of delivery.

Although there have been improvements in the past twenty years with use of protocol manuals and supervised training, investigators have had a greater interest in conducting manipulation checks (Stein, et al., 2007). These manipulation checks are ways of monitoring the behaviors of the interventionist at periodic intervals to assess whether the treatment has been successfully delivered. Delivery success is evaluated on the extent of the therapists'

adherence to the treatment manual and the degree to which their performance was competent (Waltz, et al., 1993).

Traditionally, studies have used direct and indirect observation and interventionists' self-report to gather data on the fidelity of the intervention delivery (Resnick, Bellg, et al., 2005; Resnick, Inguito, et al., 2005). Direct observation is a method in which an independent rater reviews the interventionist providers as they deliver the intervention and records notes on all the activities that takes place. Indirect observation may include the use of recording devices such as video recorder, audio tape, or the interventionist's hand-written notes. Another method of determining IF may be the interventionist therapist's self-report.

Direct observations have been viewed as an unbiased source of information which is thought to be more closely associated with the actual outcomes (Dane & Schneider, 1998; Lee, et al., 2008). Direct observation was used in the earliest psychotherapy studies in only a few cases, possibly because direct observation of interventions may not be feasible. Direct observation has the potential to affect the intervention delivery process, and observation does not "capture" the delivery process for repeated viewing and teaching or for repeated assessment to ascertain the reliability of a direct rater. Observed behaviors directly rely on supervisors or live raters using a fidelity scale to measure the activities between the interventionists and the participants in a controlled environment (Chevron & Rounsaville, 1983; Clarke, 1998). Interventions that involve sensitive material and trust issues for the participants may make direct observation an impediment to the delivery of the intervention. Additionally, a therapist who is aware of the monitoring activities or knows when he or she is being watched may not adhere as well to the intervention protocol (Moncher & Prinz, 1991). Consequently, indirect observations have been used much more frequently. Sources for

indirect observations have included audio or video taping or the field notes of the interventionists. The current gold standard consists of the method to which IF was conducted. This includes random selection of the sample, treatment condition, and ongoing reliability of the raters (Santacroce, et al., 2004). Although there are several methods, the gold standard for IF is audio or audiovisual tapes and analysis for prescribed and proscribed behaviors of the interventionists (Stein, et al., 2007). In addition, observer ratings using multiple raters are a part of the gold standard as well (Carroll, et al., 2000).

In recent years researchers have attempted to use audiotapes and videotapes and the resulting transcripts as sources of data on IF (Bellg, et al., 2004; Resnick, Bellg, et al., 2005). Both methods require a labor-intensive coding process and the cost can be high, especially when multiple raters are used (Lee, et al., 2008). In fact the cost and time in using video recordings have forced researchers not to include these in their research. Cost and time to generate the technological training to use the equipment in different environments is a likely reason that studies have not routinely included audio or video tapings (Stein, et al., 2007). Taping can also be problematic in and of itself, since it can be distracting as well as affected by background noise. Although taping has decreased the negative effect of direct observation on the intervention delivery, there continue to be concerns over recordings relative to sensitive material.

Several research studies have used audio taping and checklists to record behaviors of the interventionists. These studies are shown in Table 2.1 and 2.2 (Brandt, et al., 2004; Bruckenthal & Broderick, 2007; Martino, Ball, Nich, Frankforter, & Carroll, 2008; Wahab, Menon, & Szalacha, 2008). Most of the studies focused on the interventionists' adherence to the protocols. In six of the studies, interventionists' field notes were used to supplement the

audio/video tapes. Several of the studies indicated problems related to the use of audio or video recording. These issues were the time to gather data with audio or video and the coding of the information gathered in this way. Even though researchers have proposed that audio tapes and video tapes provide the most information, there has been limited usage of them because of the challenges associated with collecting the data with these sources (Bruckenthal & Broderick, 2007; Wahab, et al., 2008). Audio recorders often may result in interruption of the data collection in order to manage the equipment or requests by the participants to stop the recordings, which may result in loss of time or information. Video recordings share these problems and in addition may require another individual to participate in the collection of data to manage the equipment. Researchers have suggested that the fear of being monitored and the invasion of participants' privacy may have a negative effect on these kinds of studies and hence the outcomes (Tyler & Horner, 2008).

Self-reported data from the interventionist that uses an evaluation questionnaire designed for the specific intervention study has been another source of collecting data on intervention fidelity. Self-reports have consisted of questionnaires that require the individual to respond to specific questions using a forced response scale. An interventionist who uses these evaluation questionnaires, however, may over-report fidelity outcomes (Carroll, et al., 2000). Researchers have used participants and interventionist therapist self-report instruments to capture interventionist adherence and the competence of the interventionist (Minnick, Catrambone, Halstead, Rothschild, & Lapidus, 2008; Pbert, et al., 2006). A study that measured adherence using a self-reported 14-item component checklist from the participants also is problematic because it does not take into account the other activities that may have occurred.

Validated questionnaires have been developed to organize data from raters who indirectly observe intervention delivery via audio or videotapes. Scientists have tried to measure two subconcepts, adherence and competence of the interventionists. The National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP) developed The Collaborative Study for Psychotherapy Rating Scale-Form 6 (CSPRS-6) (Hollon et al, 1984). This instrument has 96-items that use a 7-point Likert-type response scale to measure adherence. The CSPRS-6 has been shown to be a reliable general measure of interventionist psychotherapist adherence behaviors, with a Cronbach alpha of .90 overall. The CSPRS-6 has been tested in a variety of empirical studies (Hill, et al., 1992; Luty, S., et al., 2007). A study was conducted to explore adherence of therapists to behaviors specified in cognitive-behavior therapy, interpersonal psychotherapy, and clinical management manuals (Hill, et al., 1992). Using CSPRS, the researchers found that the therapists exhibited behaviors specific to the individual therapy and also that the therapists used techniques from each other's approach to the therapy. Few of the approaches were associated with clinical management. Another study aimed to compare the efficacy of interpersonal psychotherapy and cognitive-behavioral therapy in people receiving out-patient treatment for depression and explore the responses in severe depression (MADRS score >30) and in melancholic depression (Luty, S. , et al., 2007). The analysis using the CSPRS revealed that the therapists adhered to the treatment protocol and the sessions were classified correctly 100% of the time, with more than 90% of these having strict adherence to protocol.

The CSPRS is designed to measure the extent to which 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 with the CSPRS. The CSPRS includes (a) modality-specific scales (CBT, IPT, and CM), (b) three tangential modality scales, and (c) two non-modality specific scales (FC and ED). These are: I) CBT, containing 28 items in six subscales: cognitive rationale (3 items), assessing cognitive processes (5 items), evaluating and changing beliefs (7 items), behavioral focus (4 items), homework (3 items), and collaborative structure (6 items) and (b) two tangential cognitive-behavior scales- alternative cognitive strategies (2) and operant approaches (2); [items-19-21;54-80]; II) IPT 28 items in seven subscales: interpersonal rationale (3), focus on feelings (4), assessing interpersonal relationships and tendencies (5), assisting changes in interpersonal functioning (4), role transitions (4), interpersonal disputes (4), and interpersonal deficits (4) and (b) tangential interpersonal behaviors (4); [items:25-53 (tangential interpersonal therapy items-22-24, 47)]; III) clinical management scale (CMS) (20) with subscales: Pharmacology rationale (3), symptoms/illness focus (4), medication effects expected and achieved (4), medication regime (6) and side effects (3); [items 81-101]; and IV) Facilitative conditions (FC) (8) and Explicit directiveness (ED) (4). The interrater reliabilities for the CSPRS were CBT (.88); IPT (.78); CM (.80); FC (.47); and ED (.58) for the ICC random effect. The ICC fixed effect was CBT (.92); IPT (.82); CM (.92); FC (.58); and ED (.73). The overall internal consistency (α) was CBT (.79); IPT (.86); CM (.69); FC (.79); and ED (.50) (Hill, et al., 1992).

Another instrument that has been used and validated is the Yale Adherence and Competence Scale (YACS) (Carroll, 1996; Carroll, et al., 2000). This scale was developed to assess interventionist adherence and competence/skill level in behavioral research that targets people with addiction disorders. Carroll (2000) reported reliability using an ICC that

ranged from 0.80 to 0.95 for adherence ratings and from 0.71 to 0.97 for competence. It has been determined that the six subscales are highly reliable in measuring adherence and competence of therapists. An article reported on a systematic review of measures of fidelity in motivational enhancement as they relate to adherence and competence measures (Madson & Campbell, 2006). The authors reported psychometric properties of the YACS ($p^2 = .82$). Although Carroll et al. (2000) found construct validity of YACS through confirmatory factor analysis and the goodness-fit indices for adherence (.92 to .99), Madson & Campbell found evidence of convergent and discriminant validity when compared with the Motivational Interviewing Supervision and Training Scale (MISTS). Santacroce et al. (2004) explored the use of a modification of this instrument, the Yale Adherence and Competence Scale for Nursing (YACS-N), in nursing research. Both the CSPRS-6 and the YACS instruments have been applied in a variety of studies (Luty, S., et al., 2007; Markowitz, Spielman, Scarvalone, & Perry, 2000; Martino, et al., 2008). In addition, there have been other instruments: the Penn Adherence-Competence Scale for Supportive-Expressive therapy (PACS-SE) (Barber & Critis-Christoph, 1996), a 45-item rating scale; the Therapist Competency/Adherence Scale Behavioral family management (Miklowitz & Goldstein, 1990), a thirteen item seven-point Likert-like scale; the Fidelity of Implementation Rating System (FIMP) (Knutson *et al.*, 2003); and the Cognitive Therapy Adherence and Competence Scale (CTACS), consisting of 25 items with five sections. These previously mentioned instruments have measured adherence and competence with supportive therapy, cognitive therapy, and other therapies. Although there have been several empirical studies using these scales, a modified version of the CSPRS developed from the NIMH-TDCRP study will be used for this study to measure adherence and competence.

One alternative method that researchers have used to monitor IF is field notes of the interventionists (Davis, et al., 2002; Moncher & Prinz, 1991; Shadish, 2002; Tessaro, Campbell, & Benedict, 2002). Field notes written within reasonable time after the intervention may provide a solution to some the problems associated with audio or video recordings because they do not interrupt the intervention (Sanjek, 1990). Field notes can serve as a systematic data collection that includes a report of events which occurred during what was observed or asked (Roper & Shapira, 2000). An interventionist can be a participant-as-observer, one who actually participates in the ongoing activities in the research setting (Hesse-Biber & Leavy, 2006). The interventionist has a direct view of the participants and their environment and can become an account of the interventionists' minute-to-minute experiences with the participants (Roper & Shapira, 2000). Historically, field notes are recorded raw data from observation, conversations, feelings, and interpretations by an interventionist (Roper & Shapira, 2000). Sanjek (1990) contends that field notes are statements constituting an objective, uncontested world of interpretations, both indigenous and scientific. Anthropologists and ethnographers have provided scientific information from field work using field notes for years (Sanjek, 1990). Although field notes have been traditionally used in scientific research over the years, the technique has not been transferred to IF research. However, field notes may be very useful in providing information about the adherence of the interventionist to the prescribed protocol manual. In research on sensitive topics where recording or direct observation is not feasible, field notes may provide the only opportunity to obtain data on how closely the interventionists actually adhere to the prescribed protocol (Erlen & Sereika, 2006; Horner, et al., 2006) by providing a thick

description of the setting, the participant, and the interventionist (Hesse-Biber & Leavy, 2006).

Limitations of Existing Systems of Measuring Fidelity

Direct observational measures are thought to be the most accurate type of measure in behavioral science. However, the interventionist who knows treatment activity is monitored may have a reactive response (Addis & Waltz, 2002; Moncher & Prinz, 1991; Waltz, et al., 1993). A reactive response means that the therapist may conduct the intervention activity differently in response to being observed. Indirect observation using different devices have been an alternative to face-to-face observation.

Indirect observation may include audio/video recordings. Although most of the studies have used these various electronic devices for collecting the data [Table II.1 and Table II.2], there appear to be problems associated with using these devices. In the case of indirect observation, the recording of the intervention may pose limitations. For example, certain cultures are reluctant to have their family life photographed or their words recorded (Tyler & Horner, 2008). Also, if the intervention involves the exploration of sensitive issues, a participant may ask that the recording device be stopped. This impedes the accuracy of the observation, and may have an effect on the delivery of the intervention as well. An additional problem with video or audio recordings is the cost and complex logistics of their use, such as transcription and coding. Finally, in some situations such as delivery of an intervention in the home, maintenance of privacy and elimination of background noise or other environmental distractions are difficult to achieve without the use of noise-filtering equipment and multiple microphones. Although digital recorders plus lavalier microphones and tabletop microphones are the current recommendation, Tyler et al. (2008) reported in

their study that the audiotapes were of poor quality due to the low, soft voices of the participants, background noise from extended family members in the room, or other activities in the clinic environment. Where tape recordings have not been feasible for reasons related to privacy, technology, or funding, some researchers have shown that alternate devices such as field notes can be a substitute (Crits-Christoph, Connolly, & Shaffer, 1999; Erlen & Sereika, 2006; Tyler & Horner, 2008). There has been less information on the comparison of IF measurement based on audio or video recordings and documentation based on methods such as the use of chronological field notes.

Although structured field notes are systematic data collection, i.e., the notes are a reflection of salient categories and behavior patterns of interest to a particular study, there are limitations (Roper & Shapira, 2000). Recorded observations or impressions intended to describe events, recollections, thoughts, and feelings are subjective information and are an individual's personal account of the activities that transpired (Sanjek, 1990). Structured field notes are either recorded during the delivery of an intervention or at a later time. Problems arise because information is missed or too much time has lapsed between collecting the data and writing the notes.

Field notes contain large amounts of data that require several hours of analysis. There are few checks of validity with field notes, that is, the degree to which the scientific observations recorded in field notes actually record what they purport to describe (Sanjek, 1990). Despite these shortcomings, the large volume of field notes collected over time in a single observation is a rich data source that may provide information on prescribed and proscribed behaviors of the interventionist nurse during the delivery of an intervention.

When audio- or video-recording is used, raters are trained to reliably code the interactions of the interventionists' behaviors. One of the methodological problems of this approach is the rater's bias specific to recordings (Hill, O'Grady, & Price, 1988). There is limited research on raters' bias specific to field notes. Raters who rate video-recordings may be subject to bias due to interpreting scale items differently or over time or due to developing a relationship with the client (Hill, et al., 1992; Hill, et al., 1988). Bias may also arise from changes in rater sensitivity over time, differences in raters' qualifications, and raters' training. A study exploring raters' characteristics revealed that raters' fatigue, a waning of effort, and sensitivity affect their rating process (Hill et al., 1988). Waltz et al. (1993) recommended that the raters be required to be sufficiently experienced and sophisticated to understand the implications of the contextual variables described in the protocol manual. They further suggested that while a rater who is skilled in the given treatment may be appropriate, raters should have no affiliation with the treatment intervention. Another limiting factor mentioned in the literature was information on the raters and inter-rater reliability. The characteristics of the raters, such as experience or relationship to the study, may contribute to the effects of the intervention outcomes. Training the raters for reliability is thus essential for consistency (Forgatch, Patterson, & Degarmo, 2005). Multiple raters and rater's reliability are essential to treatment fidelity. The limited literature on how researchers have managed the raters suggests that future research is needed on the use of multiple raters and their inter-rater reliability. Although a single rater can be trained in the rating process, typically one rater provides less reliability than multiple raters. However, Hollon et al. (1988) determined that only two raters are necessary to establish reliability levels when using CSPRS (Hill, et al., 1992). Hill (1988) suggested that there are other factors associated with

rater error, such as the characteristic of the rater, consistency, and rater's agreement. The length and complexity of the measure increases the risk for rater error. This means a longer measure is more at risk for rater error as opposed to a measure with fewer items.

Another challenge to the measurement of fidelity has been the lack of a clear definition of the concepts comprising fidelity (Dusenbury, Brannigan, Falco, & Hansen, 2003; Shadish, 2002). For example, adherence and competence have been measured within the same scale with inadequate separation of the two concepts. For example, one study used one instrument to measure both adherence and competence (Martino, et al., 2008). Another study reporting fidelity outcome used a valid instrument, the WAI, to measure mutual alliance. Mutual alliance in the literature has been associated with the competence of the interventionist (Brandt, et al., 2004). The second and fourth recommendations of Waltz et al. (1993) suggest that researchers need to establish competence through independent measurements and formulate different frameworks for adherence and competence. The literature has revealed researchers' belief that adherence is related to quantity, as shown with the checklist approach. The checklist provides information on the number times treatment intervention take place. For example, if ten randomly selected treatment sessions were evaluated, two raters would use a checklist to determine the number of times the therapist applies the treatment and inter-rater reliability would be assessed (Bruckenthal & Broderick, 2007). Research scientists believe that competence is related to quality, a measure of performance. For example, several studies measured the competence of the therapist by how well the therapist delivered the intervention (Bruckenthal & Broderick, 2007; Hogue, Dauber, et al., 2008; Martino, et al., 2008).

To summarize, investigators have attempted to use the standard devices (audio or video tapings); however, the researchers who used them included only a selected number to conduct IF measurements, probably as a means to manage the cost of IF monitoring. Additionally, researchers have found the use of these devices to be problematic due to the coding process and the cost of training coders on the recording devices and direct observation. There is limited literature on CFN. On the whole, there needs to be further research that strengthens the measurement of fidelity. Improvement in healthcare is a priority in United States. Nurse-led clinical trials or intervention studies are making significant contributions to the achievement of these goals. For the future of clinical trials and intervention research, especially in nursing for improvement in the quality of care, rigorous implementation of IF is critical to ensure the validity of these RCT and the transfer of knowledge from the research setting to the community (Whitmer, et al., 2005).

CHAPTER 3

METHODOLOGY

Design

The study design was a secondary analysis of existing chronological field notes from the primary study, “*Reducing Depressive Symptoms in Low-Income Mothers*” (NIMH-RO1 MH065524), known as the “HILDA” study. For the purpose of the proposed study, the focus was on intervention delivery, which included the dimensions of adherence and competence of the interventionists. Intervention fidelity was assessed from chronological field notes (CFN) using the Collaborative Study of Psychotherapy Research Scale-Form 6 (CSPRS-6), a measure that has been used in several studies of IF (Carroll, 1996; Hill, et al., 1992). Two subscales of the CSPRS were to examine the primary study questions, the Interpersonal Psychotherapy (IPT) and Cognitive Behavior Therapy (CBT) subscales. The remaining subscales were for descriptive purposes. The proposed study was limited to examination of the experimental (intervention) group only. In the primary intervention study, there were no detailed field notes from the control group.

The main aim of the HILDA study was to test the efficacy of an in-home delivered intervention for low-income mothers with depressive symptoms. Participating mothers were randomized to one of two treatment conditions: an experimental condition or an attention-control condition. The experimental condition consisted of in-home delivery of an adapted version of Interpersonal Psychotherapy (IPT) to decrease depressive symptoms and increase self-efficacy of low-income mothers with depressive symptoms. The control condition

consisted of a registered nurse who established a relationship with the mother and provided health information and teaching every week.

Interpersonal psychotherapy (IPT) is a time-limited treatment for depression that has three phases: an initial phase (1-3 sessions), middle phase (4-13 sessions), and terminal phase (3 sessions) (Markowitz, 2004). For the HILDA study, 16 sessions over 19 weeks was used to meet the requirements of a modified IPT treatment intervention. The study interventionists were Psychiatric-Mental Health Advanced Practice Nurses (PMHNs). The HILDA protocol manual provided specific instruction for the nurses for the application of the modified IPT that included the following essential activities: (a) establishing a therapeutic relationship with the mother; b) assessment of the mother's view of the depressive symptoms using a visual aid, the "assessment circle"; (c) clinical assessment of the presence and severity of depressive symptoms according to the Diagnostic and Statistical Manual of Mental Disorders (APA, 2000) using a clinical instrument, the Depressive Symptoms Inventory (DSI); (d) assessment of the life issues and social support using a clinical instrument, the Mother's Life Issues and Support Inventory; (e) assessment of the mother's parenting efficacy using a clinical instrument, the Mother-Child Observation Guide; (f) establishment of a contract; (g) face-to-face IPT sessions, followed by (h) 4-5 telephone booster sessions; and (i) a face-to-face termination session. These essential activities, were specified to occur over the sixteen sessions with (b)-(f) occurring in the initial sessions of the intervention. During the first visit, the nurse was to complete the assessments described above. On every subsequent visit, the assessment circle was repeated. The second visit included the mutual formulation of the problem, a contract for work, and formalization of the mutual alliance. The remaining visits were the psychotherapy work between the nurse and

the mother, ending in a formal termination. The DSI was repeated twice more at 6 and 12 weeks.

The geographic locations for the primary intervention research were New York and North Carolina. To determine whether nurses visited mothers who lived in urban, suburban or rural areas, the population census for 2000, (US Census Bureau, 2000), was used for the catchment area served by each Early Head Start program. Onondaga County, New York with Syracuse as a major city has a population 458,336. The second location was Guilford County, North Carolina with Greensboro as a major city has a population of 421,048. Greater Buncombe County, North Carolina with Ashville as a major city the third location has a population of 206,270. Orange County that included Chapel Hill has a population 115,536. The last location included a rural area, Wayne County with Goldsboro as a major city has a population of 113,332. Table III.1 represent the geographic location and number of nurses for whom CFNs were used for each area served.

Table III.1: Nurses, geographic location and setting type

Nurses	Geographic Location	Setting Type
3	Onondaga County (Syracuse), New York	Urban
3	Greater Buncombe County North Carolina	Suburban
4	Orange County, North Carolina	Suburban
4	Guilford County (Greensboro), North Carolina	Urban
6	Wayne County, North Carolina	Rural
Total	20	

Sample

The sample of field notes for the proposed secondary analysis was drawn from 1,737 chronological field notes written between 2003 and 2009 by 27 nurses in the primary intervention study. A chronological field note (CFN) is a structured, linear account of the activities that occurred from beginning to end in each contact between the HILDA nurse and

the mother. The structured field notes were written at the end of every session that the nurse conducted with a participating mother. The researcher and co-investigator of reviewed the notes with the nurses within one week and discussed them during weekly supervision in order to facilitate the nurses' memory of the events that transpired during the delivery of the intervention. When missing information was discovered, the field notes were rewritten to identify missed activities. The unit of analysis was the entire set of notes written over the course of the intervention about a particular study participant ("mother") assigned to the nurse. The set of notes written by the interventionist nurse were assumed to represent the record of the intervention delivery.

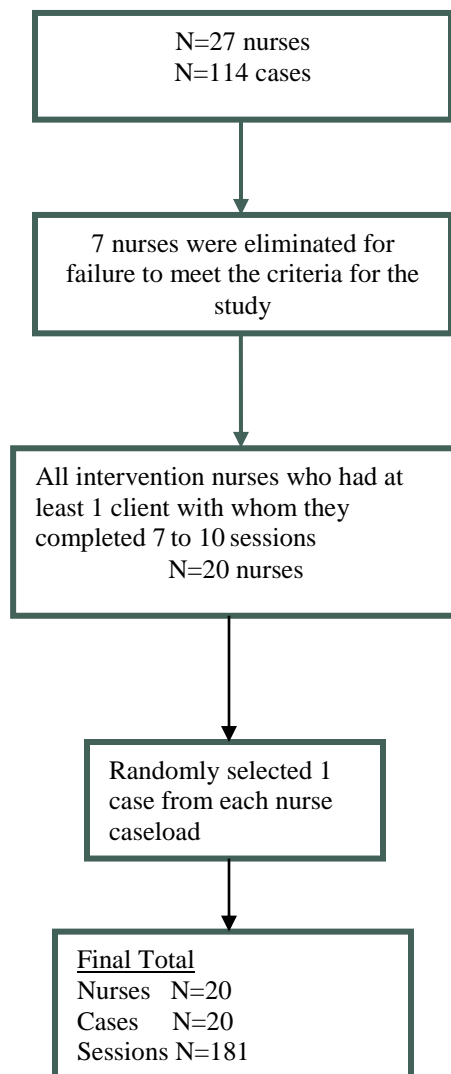
The procedure for random selection of sets of notes for assessment was determined by the nurse's case load and, within that caseload, the number of sessions completed by the nurse. Over the life of the study, nurses varied as to the length of time they participated as interventionists. As a consequence, nurses' caseloads ranged from a total of 1 to a total of 30 mothers. While the protocol manual for the experimental condition called for a total of 10 face-to-face visits, 4-5 phone calls and a termination to be completed per participant, due to cancellations, missed appointments, and early dropouts, the nurses varied on the number of visits, completing between one and 16 visits with their assigned mothers. A mother who completed at least seven face-to-face sessions was determined to have received a minimum dose of the intervention. To be included in the analysis of IF, nurses had to complete a total of at least seven face-to-face sessions with at least one mother. For nurses who completed at least seven face-to-face sessions with more than one mother, one set of notes was randomly selected from their total caseload. Random selection was used to increase the probability that

the set of field notes chosen for analysis was representative of the nurse's overall work with her assigned mothers.

Using these criteria, the final sample included 181 CFNs out of the original 1,737 CFNs written by 20 interventionist nurses representing 20 cases (mothers) (approximately 10% of all the CFNs). Figure III.1 further provides an illustration of how the nurses, sessions, and cases were randomly selected for inclusion in this study.

Approval for this secondary analysis was obtained from the University of North Carolina at Chapel Hill Institutional Review Board (IRB).

Figure III.1: Determining the sample of nurses, sessions, and cases.



Methods

The two main concepts to be considered in the proposed analysis of IF are nurses' *adherence* and *competence*.

Adherence

Adherence is defined as the degree to which the prescribed elements of the intervention have been delivered and proscribed elements avoided (Santacroce, et al., 2004; Waltz, et al., 1993). Waltz et al. (1993) recommended that measures to assess adherence should include four types of items: therapist behaviors that are unique to the treatment modality and essential; behaviors that are essential to the treatment, but that may not be unique to it; behaviors that are acceptable but may not necessary for the specified treatment therefore will not be prohibited or unique. Finally, the fourth type of item measures behaviors that are proscribed. Proscribed behaviors are prohibited for the specified treatment. Without knowing the occurrence of these behaviors, one would be unable to determine whether the protocol manual has been violated (Waltz, et al., 1993).

Competence

Competence encompasses the extent to which the interventionist nurses have displayed behaviors that typically engage the participants in the intervention and affect outcomes in the desired direction (Najavits & Weiss, 1994; Santacroce, et al., 2004). Although competence has shown moderate effects in some studies and no effect in others, the competence of the interventionist must be judged on the skill, appropriateness, and timing of the intervention, as well as the degree of responsiveness to the mothers' behaviors (Hogue, Dauber, et al., 2008; Hogue, Henderson, et al., 2008). Competence should be grounded in a thorough knowledge about these mothers and the therapeutic context (Hogue, Henderson, et

al., 2008; Waltz, et al., 1993). For IPT, nurse competence included: demonstration of the skills considered necessary to establish mutual alliance, establishment and maintenance of clear boundaries, establishment of a time-limited relationship, flexibility, respect for the mother's culture and home environment, and conveyance of acceptance (Beeber, 2004; Hill, et al., 1992; Waltz, et al., 1993). For this study, competence was represented by the interventionist displaying behaviors that engaged the mothers in the modified version of interpersonal psychotherapy intervention. To accommodate the immediate environment of the mothers, the interventionist needed to be flexible and adaptable in the delivery of the treatment intervention. Carroll, C. (2007) identifies adaptability to the local conditions, in the real-world environment as an essential component.

Instrument

To test the hypothesis that CFNs can be instrumental in determining IF, the Collaborative Study Psychotherapy Rating Scales - Form 6 (CSPRS-6) was used to evaluate the CFNs on prescribed and proscribed behaviors by the interventionist nurses.

The Collaborative Study Psychotherapy Rating Scale - Form 6

The Collaborative Study Psychotherapy Rating Scale - Form 6 (CSPRS-6) (Hollon et al, personal communication, 1984) is a 96-item instrument purported to measure intervention adherence that was developed for use in the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program (TDCRP) (Appendix B). The CSPRS consists of descriptive statements with a 7-point Likert-type response scale anchored by 4 descriptors: "not at all," "some discussion," "considerable discussion," and "extensive discussion." The subscales of the CSPRS are interpersonal psychotherapy (IPT-28 items), cognitive behavior therapy (CBT-28 items), tangential interpersonal psychotherapy, clinical

management (CM), facilitative conditions (FC), and explicit directiveness (ED). For this study, IPT and CBT were used. The internal consistency (α) for each subscale were as followed: CBT = .94; IPT = .90; CM = .84; FC = .89; and ED = .70 (Evans, Piasecki, Kriss, & Hollon, 1984; Hill, et al., 1988). Items were rated using the original 7-point rating scale. For the purpose of this study, the prescribed behaviors were represented by the CSPRS Interpersonal Psychotherapy subscale (IPT). This subscale addresses the essential elements of IPT: assessing problems (i.e., recent grief, role disputes, role transition, or interpersonal deficits), making an inventory of the symptoms, focusing on the relationship with another individual (role disputes), and focusing on social relationships (interpersonal deficits). Proscribed behaviors were represented by the CSPRS Cognitive Behavioral Therapy subscale (CBT). Nurses were not trained to do CBT and therefore, were not expected to demonstrate CBT as part of the primary psychotherapy intervention.

For the purposes of this study, the CSPRS was expanded to include a dimensional measure of competency that explored the quality of the interventionists' delivery of IPT. This was achieved by attaching a second 7-point Likert-like scale for competence using four anchors (very poor, acceptable, good, and excellent). Each of the subscales was measured on quantity (adherence) and quality (competence).

Preliminary Work

The doctoral student investigator for the proposed study conducted several small projects to establish the feasibility of intervention fidelity using chronological field notes. These projects focused on the preliminary development of a fidelity audit and testing an instrument to determine fidelity from field notes. A small study project that included audio taping of the nurses' delivery of the experimental condition during the HILDA Project

provided information on the nurses' difficulties in adhering to the protocol manual when audio recorders are used. A second project included the use of a reliable instrument to determine fidelity from field notes that was presented as a student poster presentation at the Society of Nursing Research conference. The student investigator selected two sets of three sessions from two interventionist nurses. The field notes were analyzed line-by-line and segment-by-segment and reviewed against the instrument (Yale Adherence and Competence Scale for Nurses [YACS-N],) (Santacroce, et al., 2004). The limitations were that the instrument had not been used in other studies and only four items from the scale were used, which limited the generalizability of the study.

To further determine feasibility for the proposed study, the student investigator wanted to establish the time frame for determining fidelity from the chronological field notes. The student investigator selected three interventionist nurses and one case from each. Field notes with 16 sessions for each case were analyzed line-by-line and segment-by-segment and reviewed using the four subscales (assessment, general support, goals of treatment, and clinical management) of the original version of YACS. The results were that data collection for 48 sessions or chronological field notes with three participants could be completed over a six-hour period. To summarize, the author completed extensive preliminary work in preparation for the proposed project.

Procedure

The chronological field notes that were used in the proposed study were entered electronically into a Microsoft Access software file at the University of North Carolina at Chapel Hill from 2004 to 2008. The program output was designed specifically for the HILDA Project. Once the records were sampled, the selected notes were extracted from the

general set of chronological notes. The sampled field notes were recoded with new identifying numbers to eliminate any potential personal information that identified either the interventionists or the participating mothers in the original study. The field notes were re-entered under the new identification codes into ATLAS.ti 6.1.7 software (ATLAS.TI, 2009). A computer-generated table was maintained to link the primary study nurses and mothers to the secondary analysis identification code numbers. This list was maintained in a locked file at the University of North Carolina at Chapel Hill, School of Nursing. Only the investigator for this study, research staff, and the dissertation chair had access to this file.

Rater guidelines

Phase one of the procedures was focused on identification of a master coder and creation of master code sets, creating a manual and a training program to prepare the raters, training, establishment of agreement between raters and the master codes and determination of inter-rater reliability.

Step 1 Selection of a master coder

The author's dissertation chair, who conducted the study from which the CFNs were drawn, served as the master coder. The master coder's responsibility was to establish a master dataset from which pre-study agreement with the master codes and inter-rater reliability for the two raters was established and repeated at one checkpoint midway through the data coding phase of the study.

Step 2 Establishment of a master dataset

The student investigator selected at random four sets of CFNs (40 notes) that were not used in the sample for the fidelity study. These four master data sets were rated by the master coder using the CSPRS-6 in order to establish the master codes. These master codes

were used to establish reliability between the master coder and each of the raters and to establish IRR between the two raters. Additionally, at another checkpoint the master coder monitored internal consistency for each of the raters for this study.

Step 3 Development of raters manual

A modified version of the CSPRS manual originated by Hollon and colleagues for the NIMH Collaborative study, the Intervention Fidelity Raters' Manual (IFRM) was developed to train the raters and to foster reliability and consistency of the rating for this study. The manual provided step-by-step instructions on how the raters should rate interventionist adherence and competence using the Collaborative Study Program for Psychotherapy Rating Scale -Version 6 (CSPRS-6) described earlier. The manual provided a detailed description with examples of the items and how each item should be distinguished from other items that were closely related.

Step 4 Raters' selection and training

The literature supports that measurements of adherence and competence are improved and bias is reduced by having multiple independent raters (Hill, et al., 1992). Due to limited resources, this study used two raters. Researchers have demonstrated validity with the use of two raters (Barber, Crits-Christoph, & Luborsky, 1996; Chan, O'Neill, McKenzie, Love, & Kissane, 2004; Shapiro & Startup, 1992; Symes, Remington, Brown, & Hastings, 2006).

Because of the complexity of rating, it was deemed essential that the raters be experienced clinicians. Raters' clinical experience, age, motivation, theoretical orientation, gender and similarity between raters' attitudes may potentially contribute to raters' bias (Hill, et al., 1988). Waltz et al. (1993) recommended that the qualifications and training of raters to collect data must be similar to the knowledge of the interventionists. Therefore, careful

selection and screening of the raters was done in order to select at least one rater with experience in psychiatric-mental health and research. The final raters consisted of an advanced psychiatric mental health clinical nurse specialist and the author who has participated in a psychiatric mental health courses and was an advanced pediatric and family nurse practitioner. Neither rater participated in the primary study.

Procedures for the raters' training focused on how to apply the CSPRS items and submit the data results electronically. The training was planned to take place over four weeks for a total of 40 hours. Researchers have indicated that 40-50 hours are needed for training the raters in order to establish knowledge and reliability (Forgatch, Degarmo, & Beldavs, 2005; Wilson & Crisanti, 2009).

The raters were instructed to monitor their responses and make notes in a journal. This procedure was done to enhance the validity. Researchers have found that when raters take notes, the accuracy of the ratings is improved because it reminds the raters of the information which is relevant and keeps the raters focused (Hill, et al., 1992).

Evaluation of the raters was done with the Rater Applied Performance Scale (RAPS) that measured the effectiveness of the training and how well the raters used the electronic data management system (Appendix D).

Step 5 Establishing agreements between the master coder and the raters plus inter-rater reliability

First, each rater was required to establish a reliability score of .9 with the first set (10 CFNs) of master codes. If .9 was difficult to obtain with one or both raters, more training using CFNs and CSPRS was planned. Another set (10 CFNs) of master codes prepared by the master coder was included during the training. Once each rater reached the required ICC

.9 for agreement with the master coder, the raters began the process of establishing inter-rater reliability (IRR) with each other. The two raters' results were calculated using the intra-class correlation coefficient to establish IRR. Initially, the raters had to establish an intra-class correlation of 0.9 or greater. While .9 was the goal, it was understood that this high correlation might need to be adjusted to reflect the complexity of coding across multiple session notes. The plan was that if the raters failed to establish IRR of ≥ 0.9 during training, the master coder would retrain the raters by reviewing discrepancies on particular items until agreement between the master coder and the raters was reached. Following that additional examples of that item were planned to be reviewed to determine if there are continued disagreements between raters.

Midway through data collection, the plan was to have the master coder select a checkpoint at which to repeat the coding procedure on a set of master codes. The goal was to have each rater establish a score of .9 with the master codes and at least IRR of $\geq .8$ with each other. If the raters fell below .8, the plan was to retrain the rater with the master coder.

The two raters met weekly by way of face-to-face meetings or the Skype Web site (Skype, Inc.; <http://www.skype.com>) to review the technical aspects of the rating process and discuss problems. Skype allows individuals to have audio and visual communication using video cameras. This approach was used to enhance the consistency of the raters over the duration of the study. All the item scores for each the raters and the master scorer were entered for Hypotheses 1. For Hypotheses 2 and 3 the item scores were entered into the analyses. The subscale total scores were entered for Hypothesis 4. All data scores were electronically submitted or mailed and entered into statistical analysis software, Statistical Package for Social Science (SPSS) version 19 database.

To establish reliability between the raters, the Intraclass Correlation Coefficient (ICC) for multiple raters was used. Intraclass correlation coefficient (ICC) (r) has different versions, depending on the various assumptions made. ICC computes the correlation among multiple observations of the same variable, that is within a class of variable, as distinguished from Pearson's correlation (r^2), which is usually between different variables (Streiner & Norman, 2005). Streiner & Norman (2003) discussed another version (ICC-C, 2) in which all patients are evaluated by all raters, looking at the *consistency* among the raters rather than *absolute* agreement, where 1 represents the reliability of a single rater. Intraclass correlation coefficient (ICC) takes into account the differences due to the raters (Howell, 2002). An ICC of .40 to .60 is considered fair, .60 to .75 as good, and over .75 as excellent (Fleiss, J.L., 1981). In this study, the plan was to achieve an ICC of 0.9 agreements or greater between the two raters, thus reflecting an excellent level of agreement. Once the initial inter-rater reliability was established for the two raters, the raters were to begin rating the CFNs.

Data preparation and ongoing management

Data preparations included the use of password-protected files that were stored on the University of North Carolina at Chapel Hill School of Nursing server. Continuous review by the investigator assured that data from the rater was being entered accurately on a weekly basis.

Procedures for Rating the CFNs

Data collection occurred in three steps:

Step 1. Raters read one full set of chronological field notes about the sessions between nurses and mothers.

Step 2. Upon reading all sessions in the set, the raters evaluated each nurse's prescribed and proscribed behaviors as reflected in the notes using the CSPRS-6. The raters completed every item on the scale for each therapy session.

Step 3. The raters submitted the scores from each item for each of the subscale to the primary investigator for this study.

Analysis

After data collection was completed and entered electronically, the investigator completed the analysis. In outcome studies, IF may have an effect on statistical power, which suggests that if IF is high, then the probability of behavior change will also be high (Moncher & Prinz, 1991; Sanchez, et al., 2007; Santacroce, et al., 2004). Although behavior changes may occur with low IF, such a situation may suggest that the treatment intervention was not related to the outcomes but other causes may have been. The proposed analyses were: 1) to compute the IRR on adherence and competence for the IPT and CBT subscales from the CSPRS; 2) Establish inter-rater reliability for the entire sample, 3) to establish internal consistency (Cronbach's alpha) on adherence and competence for IPT and CBT subscales; 4) to compare the IRR and internal consistency on adherence for IPT and CBT subscales to the IRR and internal consistency from the original NIMH study; 5) to explore how the CSPRS ratings derived from the CFNs function as a measure of the IF domains of adherence and competence, and 6) explore whether the CSPRS ratings from CFNs reflect the expected relationships among nurses' adherence, competence and relevant personal and contextual factors.

The subscale scores for both adherence and competence from the IPT and CBT subscale of the CSPRS-6 instrument were entered into Statistical Package for Social Science

(SPSS, version 19) and descriptive statistics were computed. Details on the additional analyses are discussed in the following steps.

Step 1. Establishment of pre-study inter-rater reliability

In order to establish inter-rater reliability between the two raters an intra-class correlation coefficient (ICC) was computed to provide an estimate of inter-rater reliability for each of the two subscales from the CSPRS instrument. The ICC measures the proportion of the total variance of a variable that is accounted for by the case clustering (group membership) (Cohen, Cohen, West, & Aiken, 2003; Fleiss, Joseph L. & Cohen, 1973). The intra-class correlation coefficient is a descriptive statistic that can be used when quantitative measurements are made on units that are organized into groups. It shows how the units in the same groups correlate. The statistical application of intra-class correlation coefficient produces a measure of both agreement and consistency (Streiner & Norman, 2005).

Equation III.1: Equation for absolute agreement in determining IRR using ICC

$$ICC2(A, k) = \frac{MS_{subjects} - MS_{error}}{MS_{subjects} + (MS_{observers} - MS_{error})/n}$$

To monitor the raters for consistency across the duration of the study, ICC was computed at one checkpoint midway through the data collection by the master coder. ICC is sensitive to systemic error in addition to differences in association and is particularly useful for small samples. The two raters were expected to achieve consistency with test-retest reliability scores of 0.90.

Step 2. Establish inter-rater reliability for the entire sample

To address hypotheses 2a and 3a, regarding the inter-rater reliability for the two raters on adherence and competency for each of the two sub-scales of the CSPRS, the intra-class correlation coefficient was computed for the entire sample of the CFNs. Equation III.1 was

used to assess multiple ratings for each item by computing an average of the items for each rater for these observations.

Step 3. Establish Cronbach's coefficient alpha

To address hypotheses 2b and 3b, Cronbach's coefficient alpha (α), a reliability coefficient that indicates how much the item scores on the subscales for the two raters measure the same underlying dimension was computed to determine internal consistency on adherence and competency for each of the two subscales from the CSPRS. Cronbach's coefficient alpha represents the homogeneity of the items for each of the two CSPRS subscales.

Equation III.2 Cronbach's alpha (k--items, σ^2 --the variance of the observed total test scores) to determine internal consistency for the CSPRS-items

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_{Y_i}^2}{\sigma_x^2} \right)$$

α =alpha Σ =sum

Step 4. Comparison between adherence scores and the NIMH study

To address hypothesis H2b, IRR and internal consistency on adherence for each of the two CSPRS subscales using CFN-based data was compared to the same scores in the original NIMH study that were derived from audiotaped and transcribed (traditional format) data.

Step 5. Comparison of CSPRS competence scores with established standards

To address hypothesis H3a & b, IRR and internal consistency on competence for each of the two CSPRS subscales using CFN-based data was compared to established psychometric standards.

Step 6. Explore the relationship between adherence, competence, and relevant personal and contextual factors.

To address hypothesis 4, Pearson Product-Moment correlation coefficient (r) was used to measure the degree of the relationship between nurses' adherence and competence on the IPT and CBT subscales from the CSPRS. Correlation coefficients indicate both the magnitude and direction of the relationship. Pearson Product-Moment correlation coefficient (r) can be positive or negative, meaning ranging between -1 and +1 (Howell, 2002). Stronger relationships move r closer to +1 or -1, while a value of zero indicates the total absence of a relationship, as long as the relationship is captured by a straight line (Spicer, 2005).

A multiple regression analysis was computed to assess further the magnitude of the relationships of selected characteristics of the nurses and adherence and competence as measured by two subscales of the CSPRS-6, IPT [representing prescribed behaviors] and CBT [representing proscribed behaviors]. This analysis also explored the associations of the CSPRS ratings on CFNs with the nurses' characteristics that would be theorized to affect adherence and competence. Multiple regression analysis provides a set of coefficients or weights that minimize the amount of variable error (Kerlinger & Lee, 2000). The main purpose is to correlate multiple independent variables with a single variable. To compute the dependent variable (DV) as a function of the independent variables (IV) we used an *ordinary least squares* analysis (OLS) (Cohen, et al., 2003). *Ordinary least squares* regression pursues the minimal sum of squared differences between the observed and the predicted squares of the dependent variable. One key assumption of this model is that the residuals are normally distributed and exhibit homoscedasticity (Cohen, et al., 2003). The assumption is

that there is a consistent and efficient estimate of each true regression coefficient of the population. A problem may arise if the assumptions are violated; however, there are several possible approaches that may be used in this case, according to Cohen et al. (2003). First, the analysis may build terms into the OLS model so that the relationship between the DV and IV is adequately represented. Residual analyses can be checked for normality of distribution. Second, the analysis may delete outlying observations or transform the data. Third, analysis may consider an alternative to OLS to make the estimate more robust to the specific problem that has been identified (Cohen, et al., 2003).

The IPT adherence and competence scores was entered into separate regression models as dependent variables with age, years of experience and length of time on the project as independent variables (Equation 3.4). Years of experience was operationalized as the experience as a nurse. Time on the project was operationalized as the total time of the nurse had participated on the project from the time hired until the first session of the note used in the analysis. We computed a correlation between IPT_{adherence} and IPT_{competence} as an unadjusted association. The variables were entered as a continuous data. Lastly, a set of model for adjusted association between types of setting (rural, urban, and suburban) in which the nurses delivered the intervention was computed using IPT adherence and IPT competence as dependent variables. In summary the analysis that included computing correlations and regressions models will have illustrated the fit with the primary question about reliability and validity of IF rating from CFN.

Equation III.3 The equation for the models

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Y=dependent variable
X=independent variable
 β =Slope

Limitations

The proposed study used one intervention study to test fidelity, which limits the generalizability of this method to other intervention studies. Only the intervention group was used for the fidelity study. There was no comparable study for the quality (competence) of the nurses. In addition, the method included only two raters, a relatively small sample and having the investigator as a rater posed a limitation. Despite these limitations, this proposed study may lay the foundation for a larger research study focusing on extracting fidelity information from narrative field notes of intervention research. Applying a variety of procedures to monitor the delivery of a complex intervention research can be a complicated process. The methods must be based on the availability of raters and financial resources. Using available field notes as an alternative to taping or transcribing may offer additional ways to monitor fidelity. To summarize, the methodological approach enhances the science by adding knowledge about the performance of the CFN using a validated instrument. The results can provide information on the future of CFN as a source for fidelity monitoring.

CHAPTER 4

RESULTS

Sample

Data were obtained from a sample of 181 chronological field notes from 20 mothers written by 20 interventionist nurses. One case [i.e. mother] from each nurse was used. The nurses' notes were comprised of 7 to 10 sessions lasting one to 1.5 hours.

Characteristics of the Intervention Nurses

The mean age for the nurses was 54 years [standard deviation (SD) 6.63]. Nurses' experience ranged from <1 year to 35 years (195.81 months). The length of time on the project varied between six to 72 months. Table IV.1 presents a descriptive data about the interventionist nurses.

Table IV.1 Nurses, age, years of experience, length of time on the project and location of delivery

Nurses	Age	Years of experience	Total time on the project by days	Urban = 3 Suburban = 2 Rural = 1
N001	61 years	35 years	144 days	3
N002	58 years	27years	493 days	3
N003	59 years	11 year	17 days	1
N004	58 years	26 years	383 days	2
N005	57 years	33 years	126 days	3
N006	50 years	24 years	34 days	3
N007	61 years	10 years	295 days	2
N008	51 years	18 years	3 days	1
N009	57 years	12 years	86 days	3
N010	56 years	24 years	434 days	2
N011	50 years	18 years	216 days	1
N012	52 years	20 years	29 days	3
N013	43 years	1 year	759 days	2
N014	61 years	11years	84 days	1
N015	49 years	5 years	1783 days	3
N016	55 years	23 years	50 days	2
N017	57 years	33 years	916 days	2
N018	58 years	8 years	315 days	2
N019	53 years	17 years	18 days	1
N020	36 years	0 years*	127 days	1

*the nurse was a recent graduate

The mothers who were the nurses' cases resided in different geographic locations that could have contributed to variations in how the intervention was delivered. For example, delivering the intervention in areas that was less populated required long drives by the nurse, whereas delivery in urban areas required travel through densely-populated neighborhoods that were often blighted and crime-ridden. These challenges may have affected the fidelity of the intervention delivery.

To capture whether that variation contributed to intervention fidelity, the geographic location of the nurse's case selected for analysis was represented according to the size of the population in which each mother lived. These locations were coded for analysis as urban, suburban or rural depending on the population in the 2000 census. Onondaga-County

(Syracuse), NY with a population of 458,336 and Guilford County (Greensboro), NC with a population of 421,048 were the largest areas and each had a mid-size city. These two cities were coded as *urban*. Greater Buncombe County and the town of Asheville, NC had a population of 206,270. These sites and Orange County, NC that included the town of Chapel Hill with a population of 115,536 were coded as *suburban*. Wayne County, NC with a population 113,332 was coded *rural*. We established dummy codes to represent the three types of settings.

Results by Aims

Aim 1 describes the process used to establish and maintain the reliability of the raters on the CSPRS-6 applied to CFN. The following includes a description of the findings for each of the sub-aims:

1a. Master set

The master data set (MDS) was a sample of field notes that was used to train the coders and verify that they had reached agreement with the master coder and with each other prior to coding the field notes. An expert with an extensive history in intervention research coded the notes using the CSPR IPT and CBT subscales. Four randomly-selected sets of 10 field notes ($n = 40$), from four different intervention nurses were used for the master codes. Data set 1 was used in trial 1 of the training session. Data set 2 was used during trial 2. Neither data set 3 nor 4 were used in the study. The master coder used the coding manual developed by the investigator. The master coder rated the 40 field notes over an eight-day period. During the creation of the master codes, the master coder consulted with the investigator for clarification about items.

1b. Intervention fidelity raters' manual

A manual was created for this study to train multiple raters on the use of C SPR scale with field notes. The intervention fidelity raters' manual (IFRM) was a modification of the original raters' manual by Evans et al., (1984) used to train multiple raters on the use of the C SPR scale. The IFRM provided directions for coding the C SPRS-6 and examples of the items. The manual was also used by the master coder during creation of the master codes. The title page and table of content pages appear in Appendix C.

1c. Training program

A comprehensive training program was developed to train two individuals on the rating of field notes using the C SPR scale. The program consisted of an overview of the original study that generated the field notes, the review of the nurse's role as an interventionist, the essential elements of the intervention, and the analysis of a field note. As a part of the training the two raters rated one field note.

1d. Training the raters

The raters received forty-five hours of a comprehensive training program over 16 weeks with face-to-face, Skype video communication, and telephone conference calls using the coding manual (IFRM). After reading and discussing the codes in the IFRM, the raters scored the C SPRS-6 IPT and CBT adherence and competence subscales on two sets of field notes that had been coded by the master coder. Scores were compared with the master coder. The scores that appeared to be greater than two points different were reevaluated. Several telephone conference calls between the master coder and the raters to discuss the items were required. The differences between the raters and the master coder were resolved with

clarification. All of the training between the two raters was audio recorded to provide data for later analysis.

1e. Training evaluation

Evaluation of the outcomes from the comprehensive training program was developed as means of determining its effect on the process of determining intervention fidelity. The evaluation included detailed information on whether the teaching strategies were appropriate and met the needs of the raters. The evaluation is included in Appendix D. Raters reported that their needs were met by the training program.

1f. Establishing raters' agreement with the master codes and with each other (Hypothesis 1f1)

At the completion of the training and prior to coding the field notes for the study, each rater's level of agreement with the master codes and with one another was tested using two sets of the master codes. To determine each rater's level of agreement with the master codes and each other, the reliability of the raters was computed using Strout and Fleiss (1979) intra-class correlation coefficient [ICC].

ICC is a reliability index that computes a ratio of the variance of interest over the sum of the variance plus error. There are several versions that provide different results. A two-way random model was selected for this study computing: 1) absolute agreement for measurements that are averages based on k independent measurements on randomly selected objects; and 2) the degree of consistency for measurements that are averages of k independent measurements on randomly selected objects (McGraw & Wong, 1996). Absolute agreement takes the total score variance as its denominator whereas consistency does not (McGraw & Wong, 1996). The variability between the raters was of interest

therefore; the absolute agreement was used. However, if this source of variability was deemed to be non-essential then we used consistency (McGraw & Wong, 1996). The ICC in SPSS/PASW generates an output that includes both the single measure and an average measure. The single measure is the computed outcomes from the judges' individual subjects and item scores whereas average measurement compute average ratings for k judges, the average score for k -subjects and k -items (McGraw & Wong, 1996; Shrout & Fleiss, 1979). The average measure approach was used for the purpose of comparing to the Hill et al. 1992 study. The version of ICC selected for this study was based on the decision to compute the differences between the raters' mean ratings relevant to the reliability between raters for each of the subjects (Equation IV.1).

Equation IV.1

$$ICC2(A, k) = \frac{MS_{subjects} - MS_{error}}{MS_{subjects} + (MS_{observers} - MS_{error})/n}$$

Using Statistical Package for Social Science-Predictive Analysis Software (SPSS-PASW version 18), the scores from the rated master datasets by the master coder and each of the raters were entered. For this study the adequacy of the initial intraclass correlation [ICC] was set at $\geq .90$. The initial high reliability score was used because it was anticipated that over the course of the coding, these correlations would decrease. The formula for calculating reliability scores for the first master data set (MDS1) between the master coder and the raters plus inter-rater reliability is presented in in Equation IV.1.

The post-training Cronbach alpha scores were $> .90$ for the first set of master codes. Reliability of both raters for MDS 1 is presented in Table IV.2. Rater 1 had a Cronbach alpha of $> .90$. The ICC for each of the CSPRS subscales was as follows: $> .90$ for Interpersonal Psychotherapy-Adherence (IPTa; prescribed behavior), Interpersonal

Psychotherapy-Competence (IPTc; prescribed), Cognitive Behavioral Therapy-Adherence (CBTA; proscribed behavior) and for Cognitive Behavioral Therapy-Competence (CBTc; proscribed behaviors). Rater 2 had a Cronbach alpha of $>.90$. The ICC was $>.90$ for IPTa, IPTc, CBTA and CBTc. Another set of master codes were used to confirm our findings of reliability and to establish a $.90$ for all the items.

Table IV.2 Reliability with the master coder for each subscale total (MDS1) in Trial 1

	R1	IPTa	IPTc	CBTa	CBTc	R2	IPTa	IPTc	CBTa	CBTc
Cronbach Alpha		.93	.91	.95	.91		.96	.94	.96	.90
ICC		.91	.91	.95	.90		.96	.94	.95	.90

ICC=intraclass correlation coefficient
 IPTa=interpersonal psychotherapy-adherence
 IPTc=interpersonal psychotherapy-competence
 CBTA=cognitive behavior therapy-adherence
 CBTc=cognitive behavior therapy-competence

In establishing reliability with the master coder, it was determined that the raters having difference of two points or greater on an item from the master code warranted discussion, clarification and retraining on the correct codes with the master coder. To address the discrepancies between raters and the master coder, the investigator documented and displayed discrepancies on a grid. The grid was a way of visualizing our problem items during the training (Table IV.3) and focusing discussion on the problem items that required clarification and agreement between the raters and the master coder.

Table IV.3. An example of a grid of the raters' and master coder's ratings used for training purposes

Item number	IPTa					
	Post Trained			Post retrained		
	MC	R1	R2	MC	R1	R2
1	7	4	7	7	7	7
2	3	2	7	3	4	2
3	6	4	6	6	6	6
4	7	5	7	7	7	7
5	6	3	6	6	5	6
6	6	6	7	6	6	6
7	5	5	6	5	6	6
8	5	3	5	5	5	4
9	6	5	6	6	7	6
10	6	5	6	6	7	6
11	6	5	7	6	7	5
12	2	1	2	2	1	3
13	7	7	7	7	7	6
14	5	5	5	5	5	5
15	3	5	4	3	5	3
16	5	5	5	5	6	5
17	3	6	3	3	3	3
18	3	2	3	3	2	3
19	1	2	3	1	1	2
20	2	2	3	2	2	2
21	1	4	7	1	2	1
22	3	2	7	3	3	3
23	2	4	2	2	3	5
24	3	2	6	3	3	3
25	1	1	1	1	1	1
26	5	7	7	5	7	7
27	4	5	7	4	5	3
28	3	4	6	3	7	2

Some of the problems were related to clarity of a selected number of items. For example, one item in the IPT subscale, [“did the nurse attempt to use the therapeutic relationship to help the mother understand how she relates to others or as a model for the mother to use in developing satisfying relationships outside of the therapy?”], was confusing for the raters. The item refers to whether the interventionist nurse used her relationship with the mother as model for developing a therapeutic interpersonal relationship with an important person in her life. Telephone conference calls with the master coder that included a discussion how to identify that behavior of the nurse from the chronological field notes and until an agreement was met between the raters and the master coder. Table IV.3 provides the scores after the initial training and post retraining scores.

Another interesting example emerged in the adherence rating for the Interpersonal psychotherapy subscale on items 15, 23, 26, 27, and 28 during the training. According to the Collaborative Study for Psychotherapy Raters Scale version 6 (CSPRS-6) raters’ manual, the nurses’ behaviors were given a rating that reflected their actions across all of the sessions with the mothers. Some of the nurses’ behaviors were in response to the mother’s action; the CSPRS items were worded in a way that implied that the nurse would initiate all actions. Considerable discussion was needed to resolve the issue so that the raters could determine that the nurse met the primary objective of the item either by initiating it or by nurse facilitating the behavior being measured.

One item (#12 on the IPTC scale) required a change in the master code on the competence subscale. The item (did the interventionist nurse help the mother to consider ways of developing new relationships?) posed variation in the raters and master coder scores.

With much discussion and clarification of the item, an agreement was reached between the master coder and raters that resulted in a change in the master coder item score from 3 to 4.

When all discrepancies in the first set of codes were resolved a second master data set (MDS2) was rated by both raters who were able to obtain adequate agreement with the master coder; that is Cronbach alpha > .90 and ICC > .90. Thus, hypothesis 1f1 was supported. (See Table IV.4)

Table IV.4 Reliability with the master coder (MDS2) in Trial 2

	R1	IPTa	IPTc	CBTa	CBTc	R2	IPTa	IPTc	CBTa	CBTc
Alpha		.94	.93	.94	.94		.95	.92	.97	.95
ICC		.93	.92	.93	.94		.95	.92	.97	.95

ICC=intraclass correlation coefficient
 IPTa=interpersonal psychotherapy-adherence
 IPTc=interpersonal psychotherapy-competence
 CBTa=cognitive behavior therapy-adherence
 CBTc=cognitive behavior therapy-competence

Inter-rater' reliability (Hypothesis 1f2)

The raters also needed to establish reliability with each other. Inter-rater reliability (IRR) helps to establish the extent of consensus between the two raters using CSPRS instrument in rating field notes. For this study we wanted to determine if raters who received comprehensive training could achieve an adequate IRR using the CSPRS with field notes. To establish initial inter-rater reliability both raters rated the master codes twice and obtained ICC scores as illustrated in Tables IV.5 and IV.6 on the IPT and CBT subscales. The final inter-rater correlation coefficient score between the raters were > .90 on adherence and > .80 on competence. The Cronbach alpha scores were .90 on adherence and > .80 on competence for the MDS2.

Table IV.5 Inter-rater reliability (MDS1) for Trial 1

	IPTa	IPTc	CBTa	CBTc
Alpha	.89	.88	.92	.85
ICC	.88	.88	.91	.85

Table IV.6 Inter-rater reliability (MDS-2) for Trial 2

	IPTa	IPTc	CBTa	CBTc
Alpha	.90	.83	.95	.89
ICC	.90	.83	.94	.89

In summary, for this study competence was added to the CSPRS instrument to measure the function on the field notes. With extensive training the pre-study competence reliability for the IRR remained > .80. Although the pre-study ICC scores for IPT and CBT on competence did not meet the original goal of .9, the scores revealed an adequate level (Cicchetti, 1994). Thus, Hypothesis 1f2 was supported.

H11g1. Level of agreement with the master codes over time

To determine whether the coders stayed reliable with the master coder and each other overtime, we repeated the rating of one set of the master codes (MDS1) at the mid-point during data collection. An unrated copy of the codes was mailed to the raters who returned them in one week. The IRR using the ICC estimates are expected to be lower compare to the initial reliability scores. The initial IRR were [IPTa-.88; IPTc-.88; CBTa-.91; and CBTc-.85]. In comparison at the midpoint, the scores were [IPTa-.92; IPTc .83; CBTa-.91; and CBTc-.82]. Tables IV.7 and IV.8 provides the results the MDS1 reliability scores.

Equation IV.2 ICC to determine level of agreement of the subscale totals for the raters

Table IV.7 Level of agreement over time between MC and raters

	R1	IPTa	IPTc	CBTa	CBTc	R2	IPTa	IPTc	CBTa	CBTc
Alpha		.95	.95	.94	.84		.96	.92	.97	.87
ICC		.95	.95	.94	.84		.96	.92	.97	.87

Table IV.8 Level of agreement between the raters over time

	IPTa	IPTc	CBTa	CBTc
Alpha	.92	.83	.91	.82
ICC	.92	.83	.91	.82

Raters' agreement with the master coder was > 0.9 for adherence and IPTc; however, competence for CBT was lower 0.84 and 0.87. The IRR scores were $>.9$ for both IPTa and CBTa whereas IPTc and CBTc were $>.8$ at the mid-point of data collection. In summary, the level of agreement remained excellent on adherence and the level of agreement remained acceptable on competence midway through data collection. Thus, Hypothesis 1h1 was supported.

Aim 2 CSPRS Ratings of the Chronological Field Notes (CFN)

Analyses to achieve AIMS 2-4 were conducted using measurements that were made after the raters had completed training and demonstrated an acceptable level of reliability. We computed inter-reliability (IRR) between the two raters and internal consistency on adherence for each of the theoretically-derived subscales: interpersonal psychotherapy (IPT) and cognitive behavior therapy (CBT). Other subscales in the CSPRS (clinical management (CM), facilitative conditions (FC) and explicit directiveness (ED) were not included in this study.

The statistical analysis used SPSS/PASW for computations. Descriptive analysis included the means, standard deviation and variance for each IPTa, IPTc, CBTa and CBTc (Table IV.10). The means across items for each rater were used as the unit of analysis. A high score indicates a high level of inter-rater consistency on the average of all 56 ratings per subject (1,120) for IPT and 64 ratings per subject (1,280) for CBT.

Adherence. The extent that the nurses reported in the field notes the prescribed elements of the intervention which resulted in an average mean score of 131.92 with standard deviation of 97.23 for all of the nurses (N=20) for IPT (Table IV.10). The nurses'

application of the proscribed elements of the intervention resulted in average mean score of 110.67 with a standard deviation of 32.01 for CBT.

Competence. In this sample, the extent to which the interventionist nurse displayed engagement behaviors with the participant confirmed their competence in the delivery of the intervention protocol (Najavits & Weiss, 1994; Santacroce, et al., 2004). With the reliability analysis, the total mean score for IPTc was 145.5 with standard deviation of 36.77. The CBTc total mean score was 167.7 with a standard deviation of 36.27.

Table IV.9 Means and Std. Deviation

	Mean	SD
IPTa	131.9	97.2
IPTc	145.5	36.7
CBTa	110.6	32
CBTc	167.7	36.3

Inter-rater reliability analysis on adherence

The overall reliability and Cronbach alpha scores are indicated in Table IV.10. One of the aims of this study was to compare the CSPRS scores derived from field notes to those that were derived from audiotapes/transcriptions. In the NIMH study and in the Hill et al, 1992 study, the ICC average measure reliability (the mean of all ratings is the unit of analysis) that is, the mean of all the raters was used and thus, was used in this study. This included the average across items and across raters. For this study the ICC average measures for IPT and CBT were comparable to the NIMH scores and slightly higher than the Hill study. The Cronbach coefficient alpha for IPT and CBT were comparable to the NIMH scores as well. Cicchetti (1994) reported that the reliability of a given assessment instrument whether in terms of a coefficient alpha, test-retest, interrater or temporal reliability coefficient is distinguished by the level of clinical meaningfulness. Therefore, the guidelines

take into consideration the broader context of all reliability measures (e.g. interrater reliability) (Cicchetti, 1994). As recommended by Cicchetti (1994), the ICC guidelines state that, when the reliability coefficient is below .40 the clinical significance is poor; however levels between .40 and .59 is fair, the clinical significance is good between .60 and .74 and between .75 and 1.00 is excellent (Cicchetti, 1994). Using Cicchetti’s guidelines in this study, the reliability coefficient on IPTa and CBTa had a level of clinical significance that fell into the “excellent” range whereas the level of clinical significance on IPTc and CBTc was good.

Table IV.10 Inter-rater reliability for IPT and CBT

	Cronbach Alpha	Intraclass Correlation
IPTa	.97	.95
IPTc	.91	.74
CBTa	.98	.96
CBTc	.90	.70

Internal Consistency

In the analysis of reliability, internal consistency is used to measure the consistency of the raters on the items within the test. For this study internal consistency of the subscales (IPT and CBT) was determined with coefficient alpha using a combined scores for the two raters (Nunnally, 1967). To establish internal consistency among the raters, the fifth week was determined as the mid-point of the data collection for establishing internal consistency. Once the raters finished the tenth nurse, a Cronbach alpha was computed for the first ten field notes. Table IV.11 provides the results. Raters were able to achieve a Cronbach alpha coefficient of > 0.9 for adherence; however, competence was lower 0.88 and 0.83.

Table IV.11 Internal consistency between the raters on the first ten field notes

	IPTa	IPTc	CBTa	CBTc
Cronbach alphas	.96	.88	.95	.83

Comparison between the IRR with NIMH (Hypothesis 2b)

We then compared the IRR and internal consistency of each CFN-derived CSPRS adherence subscale with the scores derived from audiotaped transcribed data from the original National Institute of Mental Health (NIMH) study. We hypothesized that:

H2b IRR and Cronbach's alpha on adherence ratings for each of the two subscales would be equal to the NIMH published scores for IRR (IPT .78 and CBT .90) and alphas (IPT .90 and CBT .94). The reliability and consistency of the CFN scores were equal or superior to those derived from the audiotapes/transcriptions. Table IV.13 shows the results of these comparisons on adherence.

Table IV.12 Comparison from the study of subscale IPT & CBT on adherence with NIMH

		IPT	CBT
Scores from this study	IRR	.95	.96
	Alpha	.97	.98
Scores from the NIMH study	IRR	.78	.90
	Alpha	.90	.94

Competence was not measured on the original NIMH study and therefore, only the adherence scores were compared.

Aim 3. CFNs function as measure of competence

To explore how the CSPRS ratings derived from CFNs functioned as a measure of the IF domain of competence, the newly-added competence scale on the CSPRS was evaluated. Specifically, we computed IRR for the two raters and internal consistency on competence for the IPT and CBT subscales.

H3a. IRR on competence for each of the two subscales (IPT and CBT) was not successful in achieving a reliability coefficient of equal to or greater than .80. The

scores were less than .80 as represented in Table IV.14; therefore, hypothesis 3a was not supported.

Table IV.13 IRR and Cronbach alpha on competence for IPT and CBT

	IRR	Cronbach alpha
IPT	.74	.91
CBT	.70	.89

3b. Compare the internal consistency on competence for each of the two CSPRS subscales to established standards. Once the raters finished the tenth nurse, a Cronbach alpha was computed for the first ten field notes. Specifically:

H3b. Each of the CSPRS subscales on competence will achieve a Cronbach’s alpha of .80. Table IV.14 represent the Cronbach alpha scores. Thus, hypothesis 3b was supported.

Aim 4 CSPRS ratings and relationships among the major variables

We then explored whether CSPRS ratings derived from CFNs reflected the expected relationships among nurses’ adherence, competence and relevant personal and contextual factors by correlating these factors with the IPT adherence and competence subscales.

Hypothesis 4a: Relationship of Adherence and Competence

To establish a correlation between adherence and competence, a Pearson product moment coefficient correlation was computed on adherence for both IPT and CBT subscales using the average of the raters’ scores. There was a strong positive correlation between adherence and competence for IPT and CBT suggesting that in this sample, nurses who were adherent to the IPT elements of the protocol also were adherent to CBT elements as well.

Tables -IV.15 shows these results.

Table IV.14 Correlations between adherence and competence plus between prescribed (IPT) and proscribed (CBT) behaviors

Correlation among adherence and competence for IPT and CBT			
	IPTa		CBTa
IPTc	.541*	CBTc	.840**

Correlation between IPT (prescribed behaviors) and CBT (proscribed behaviors) on adherence and competence			
	IPTa		IPTc
CBTa	.619**	CBTc	.966**

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

Hypothesis 4b. Relationship between prescribed and proscribed behaviors

A correlation analysis was computed to determine if proscribed behaviors (CBT) would be negatively correlated with prescribed behaviors (IPT). There was a strong positive correlation between prescribed behaviors and proscribed behaviors with significance on adherence and competence. The nurse, who adhered to the prescribed behaviors, also demonstrated proscribed behaviors. Table IV.15 displays these data.

Factors associated with adherence and competence of nurses

Nurses' age, years of experience, total time on the project from time hired to the first session of the note and the geographical location where the nurse delivered the intervention were factors that were hypothesized to be related to adherence and competence. Table IV. 16 present descriptive data on these variables.

IV.15 Factors, means and standard deviations

Factors	Mean	Std. Deviation
Age (years)	54.10	6.357
Years of experience	205.2	130.5
Length time on the project (days)	315.6	428.3

Bivariate correlations of these predictors were computed to evaluate the nature of the relationships between these factors and IPT adherence and competence. There was no

correlation between IPT adherence and competence and the factors; however there was a correlation between age and years of experience. Table IV.17 present these correlations.

Table IV. 16. Bivariate correlation of the predictors on adherence and competence with IPTa

Predictors	Correlation between each predictor and adherence							
	IPTa	IPTc	A	YE	TOP	Urban	Sub	R
IPTa								
Age (A)	.205	.188						
Years of experience (YE)	.377	.157	.525*					
Time on project (TOP)	.344	.242	-.163	-.122				
Urban	-.126	.035	.090	.359	.122			
Suburban (Sub)	.290	-.023	.157	.052	.237	-.584*		
Rural (R)	-.171	-.013	-.257	-.428	-.373	(-.538*-IPTc)	-.480	

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

H4c. Greater age, years of experience, and length of time on project will predict greater adherence. Nurses' age, years of experience, length of time on the project and the geographical location where the nurse delivered the intervention were entered into two separate regression models of main effects using the adherence and competence scores as the dependent variable in each. The linear combination of independent variables was non-significantly associated with IPT adherence, ($R^2 = .299$ with an adjusted $R^2 = .168$, $F(3, 16) = 2.28$, $p = .119$) as shown in Table IV.19 Model 1. Despite not being statistically significant, it was noteworthy that the model explained nearly 30% of the variance in IPT adherence. The linear combination of the independent variables was non-significantly associated with CBT adherence (model 2) as shown in Table IV.19. Hypothesis 4c was not supported in that the linear combination of factors was non-significantly associated with the nurses' adherence to the intervention protocol.

Table IV.17. Linear regression predicting greater adherences (Models 1 & 2) and competence (Models 3&4)

Nurses' Factors	Model	Model	Model	Model
	1	2	3	4
	(IPTa)	(CBTa)	(IPTc)	(CBTc)
	β	β	β	β
Age	.065	.415	.185	.145
Years of experience	.392	.040	.094	.078
Time On Project (in days)	.402	.314	.284	.264
R^2	.299	.244	.118	.091

H4d. Greater age, years of experience, and length of time on project would be associated with greater competence. The linear combination of factors was not significantly associated with IPT competence ($R^2=.118$ with an adjusted $R^2 =-.047$, $F(3, 16) =.717$, $p = .556$) as shown in Table IV.19, Model 3 or CBT competence (Table IV.19, Model 4).

H4e. To determine if the type of geographic location of the mother affected nurses' adherence, dummy codes were established for each type of area and entered into the regression. Model 1 included age, years of experience, time on project, urban, suburban and rural predicting adherence. Model 2 included age, years of experience, time on project, urban, suburban and rural areas predicting competence. The linear combination of factors in Model 1 was non-significantly associated with adherence for IPT (Table IV.20); $R^2 .430$ and adjusted $R^2 .226$; $F(5, 14) 2.11$, $p = .124$ (Table IV.20). The second linear combination of factors (Model 2) was not significantly associated with competence for IPT (Table IV.20), $R^2 .172$; adjusted $R^2 -.123$; $F(5, 14) .582$, $p = .713$. The model factors that included geographic location was not significantly associated with adherence or competence. Hypothesis 4e was not supported.

Table IV.18. Multiple regression models for all the factors predicting greater adherence and competence for IPT

Nurses' Factors	Model 1 IPTa		Model 2 IPTc	
	β	p value	β	p value
Age	.037	.880	.222	.462
Yrsexperience	.593	.044*	.209	.528
Time on project (in days)	.508	.048*	.414	.164
Geographic location				
suburban		.164		.899
rural		.137		.445
R ²	.430 [±]	.124	.172	.713

*. Correlation is significant at the 0.05 level (2-tailed).

± Non-significant

Summary of the results by Hypotheses

Hypothesis 1f1. Using the ICC average measure the raters who received comprehensive training achieved adequate level of agreement with the master coder, that is; the level of agreement was reached with a reliability coefficient > 0.9 and Cronbach alpha > 0.9 .

Hypothesis 1f2. After the raters reached reliability with the master coder on both data sets, the raters' scores were computed to establish inter-rater reliability with each other. The two raters achieved a level of agreement with reliability coefficient $>.9$ for IPT and CBT on adherence as compared to $>.8$ for IPT and CBT on competence. Although the IRR scores were not met for our original goal set by the study, the $>.8$ was considered an acceptable level of reliability.

Hypothesis 1g. Midway through the data collection the raters maintained internal consistency with a Cronbach alpha $>.9$ for IPT and CBT on adherence; however, IPT and CBT on competence were $>.8$. There was minimal drift between the training and during the data collection.

Hypothesis 2b. The scores from the study IRR and Cronbach alpha were comparable to the NIMH scores which included IRR (IPT - .78, CBT - .90,) and alphas (IPT - .90, CBT - .94).

Hypothesis 3a. The IRR on competence was $\geq .70$. The result was below acceptable levels for established instruments but was acceptable for a new measure.

Hypothesis 3b. The Cronbach alpha on competence was > 0.90 . This coefficient alpha tells us that the instrument has the potential of being a reliable measure for competence.

Hypothesis 4a. The hypothesis that there would be a strong positive significant correlation between adherence and competence for IPT and CBT for the total scores was supported.

Hypothesis 4b. This hypothesis was not supported. There were positive correlations between IPT and CBT on adherence and competence as opposed to the original expectation of a negative correlation between prescribed (IPT) and proscribed (CBT) behaviors. Nurses who delivered the intervention included elements of both IPT and CBT.

Hypothesis 4c. The linear regression model of age, years of experience as a nurse, and length of time on the project was not a predictor of adherence to the intervention protocol manual (R^2 of .299 with p value .119).

Hypothesis 4d. The linear combination of the nurses' characteristics did not predict a significant amount of variance on competence, ($R^2=.118$ with an adjusted $R^2 =-.047$, $F(3, 16).717$, $p = .556$) for IPT.

Hypothesis 4e. Type of geographic setting where the nurses delivered the intervention was non-significantly associated with nurse therapist adherence or competence

controlling for age, experience and length of time on the project. For Model 1, the beta coefficient for the suburban area was .349 (p value = .164) controlling for age, years of experience, length of time on the project. The beta coefficient for rural area was .450 (p value = .137). For the Model 2 the addition of location did not increase the association of the independent variables with competence (Table 20, Model 2). Despite not reaching statistical significance, it was noteworthy that the model explained nearly 43% of the variance in IPT adherence when geographic location was added to the previous model.

CHAPTER 5

DISCUSSION

The primary purpose of the study was to explore whether intervention fidelity could be determined from chronological field notes (CFN), a documentation method that was less intrusive and less often used than observational methods (audio or video taping or direct observation). Taking the theoretical approach that two key representations of fidelity are adherence and competence (Waltz et al., 1993), the study examined whether the (CFNs) provided adequate data from which to determine whether the intervention nurses displayed essential and unique behaviors with skillfulness (Waltz, et al., 1993). The essential element of the treatment intervention being evaluated was IPT. Prohibited (proscribed) behaviors were those associated with CBT.

The study produced preliminary data that CFNs could be used as data within a defined measurement structure that included notes with sufficient documentation to determine nurses' activities, a reliable and valid instrument (the CSPRS) for coding the data and multiple raters trained in a replicable, manualized method who had achieved adequate reliability with master codes and each other. Each step of the study documented the development and implementation of these steps. Within this structure, the CFN-derived adherence data from the CSPRS on showed comparability to that derived using traditional methods in a previous study. However, CFN data revealed that the interventionist nurses also displayed proscribed (CBT) behaviors.

As reliability is defined through error the greater the error the more unreliable the measure and vice-versa (Streiner & Norman, 2005). Reliability, as defined by adequate psychometric indicators, was established after extensive training with a master coder and two raters. The raters with comprehensive training did achieve adequate level of agreement with the master coder to reach a reliability of .90. The comprehensive training that included 45 hours of face-to-face discussion for the items on the IPT and CBT subscales was effective in providing the knowledge needed to produce an acceptable level of skill in the raters. This included back and forth questions and answers to clarify items and instructions in the manual. Raters used examples from the manual frequently. There were frequent practice sessions over the course of 17 weeks. The second stage included the establishment of reliability with the master coder and between raters. With multiple telephone conference calls and meeting with master coder to retrain on the master codes, the raters and master coder were able to reach a level of agreement.

Rating from different observers

During the training the raters kept a journal of their thoughts. There were several issues that became apparent associated with competence of the nurses. One was related to the raters' sense that despite some degree of subjectivity, the comprehensive training provided the guidelines to evaluate the behaviors of the nurses objectively. The second issue that occurred frequently in the journal was the measure of competence. Although the nurse might not have completed the prescribed treatment intervention as planned, both raters agreed that this deviation did not seem to indicate the lack of competence. The skills that the nurse displayed during the delivery of the treatment seemed to be more important.

A few of the nurses failed to explore issues which the mother expressed. Raters found this to be challenging when the mother began to open up with significant feelings and nurse did not explore further. For example, the nurses who presented in unorganized way during the intervention delivery missed an opportunity to explore important issues expressed by the participants.

Another issue emerged where the nurse who wrote long field notes was not necessarily more adherent to the prescribed treatment intervention. A few nurses documented a large amount of conversation between themselves and the mother without goals or plans for treatment. As this was an essential prescribed element in the primary intervention study, a field note such as this would be rated as non-adherent.

Although the raters demonstrated reliability in evaluating the nurse therapist's report of the intervention delivery, the raters were not infallible. For example, the raters may have missed detecting some of the interventions because the CSPRS did not include all possible components. Raters were not privy to the interaction between the mother and the nurse; thus, a nurse who failed to record the core elements of the intervention may have been rated low in adherence and competence.

Summary of the findings

After the raters reached reliability with the master coder on both data sets, the raters' scores were computed to establish inter-rater reliability with each other. The two raters achieved a level of agreement with reliability coefficient $>.9$ for IPTa and CBTa as opposed to $>.8$ for IPTc and CBTc. Although the IRR scores were not as high as those sought in the study aims, the strength of the agreement ($>.8$) can be considered to be an excellent level (Cicchetti, 1994).

The raters were also able to maintain a reliability coefficient of .90 with the master coder after ten weeks of data collection on adherence for both subscales, IPT and CBT. However, the reliability of the newly-created competence scale remained at $>.80$.

The reliability ratings were consistent throughout the study on adherence and competence but the scores were from $>.7$ to $.8$, lower than our intended score of $.9$. There may be several problems associated with measuring competence from CFNs using the scale created for this study. Some of the possible reasons may be 1) the sample may not be adequate for the newly-created competence scale and 2) the items from CSPR instrument may not adequately measure competence. Although the pre-study reliability on competence did not meet our original goals, the reliability over all for the study sample provided a good level of clinical significance. A larger sample is needed to provide enough data in determining intervention fidelity that includes both adherence and competence. The reliability for the CSPR derived from the CFN was comparable to the original NIMH scores.

The inter-rater reliability was $.74$ on competence for IPT with a Cronbach alpha $.91$. ICC scores were $.70$ on competence for CBT with Cronbach alpha $.90$. These scores were lower than the pre-study reliability for IPT ($.83$ with alpha $.83$) and for CBT ($.89$ with Cronbach alpha $.89$). Although the CFN may not be used as measure of competence, there is a level of agreement that was good. The results shows that more research is needed to determine the function of field notes on competence.

This study provided the information on the relationship between the two core concepts of intervention fidelity by the positive, significant correlation of adherence and competence for both IPT and CBT subscales. The results supported that this critical relationship could be detected using CFNs.

Although we hypothesized that there would be a negative correlation among IPT and CBT, instead, the relationship was a strong positive significance correlation. This correlation between the prescribed and proscribed behaviors indicated that there was an overlap in the nurses' behaviors that may have also indicated overlap in the two treatment modalities. Approximately the first third of the questions in the CSPRS CBT subscale addresses maintenance items that a therapist would apply with either IPT or CBT. For example, one of the CBT items asked whether the nurse worked collaboratively with the mother to formulate and follow a specific agenda for the session. The intervention protocol for the primary intervention study trained nurses to carry out other functions essential to engaging and maintaining a therapeutic relationship with the mothers. These behaviors were addressed in the CBT subscale. A few other CBT items that were included in the nurses' training are represented in Table V.1. Again, these were foundational maintenance behaviors that were generic to IPT as well as CBT, but were only asked in the CBT subscale. Therefore, the prescribed and proscribed behaviors, as defined by the CSPRS, together may have been needed to deliver the intervention. These results illustrate that it is important to measure more than required behaviors in an intervention fidelity study.

While this explanation is strong, these results need to be taken into consideration when reporting the overall results of the primary study. If the IPT operationalized in the primary study intervention contained elements that, according to the CSPRS ratings, were CBT in nature, the replicability of the outcomes may depend on incorporating these elements into the intervention. However, the pilot nature of this study and limited sample size requires caution in this interpretation. Research with a larger sample is needed to explore this issue further.

Although the primary intervention study used a modified version of IPT as the theoretical foundation, the CSPR scale may not have captured the full components of the modified version to determine fidelity. CSPRS measures the adherence of the therapist who applies behaviors consistent with IPT but not the modified version of IPT. These findings provide information on the nurse applying both prescribed and proscribed behaviors as it applies to this particular intervention.

Table V.1 CSPRS CBT items measuring strategies that the nurses were trained to do

Item number	CBT items
1	Did the interventionist nurse work collaboratively with the mother to formulate and follow a specific agenda for the session?
2	Did the nurse review previously assigned homework with the mother?
3	Did the interventionist nurse actively attempt to engage the mother in working <u>together</u> to explore the therapeutic issues?
4	Did the interventionist nurse encourage the mother's independence from the nurse in dealing with her problems?
23	Did the interventionist nurse attempt to teach the mother skills (e.g., assertiveness, social skills, task relevant skills) in the session?
24	Did the interventionist nurse or mother develop one or more specific assignments for the mother to engage in prior to the next session?
26	Did the interventionist nurse work with the mother to schedule or structure one or more specific activities for the purpose of increasing the likelihood that the mother will initiate or follow through on those activities?

(Evans, et al., 1984)

Nurses' age, years of experience, total time on the project from time hired to the time of the CFN used in the study and the geographical location where the nurse delivered the intervention were not statistically relevant to adherence or competence. The small sample size and limited opportunity to sample among nurses may have prevented the relationships among these conceptually relevant factors to be detected. In addition, factors that may have been relevant were not tested, e.g. investment in training and prior community intervention experience. However, despite not reaching statistical significance, the model was associated with 43%, and two independent variables (age and years of experience) were significant

contributors indicating that there are some meaningful relationships which a larger sample or more raters might reveal.

In summary, the data provided some preliminary support for the hypotheses that CFNs could function as an alternate to traditional observational methods for determining intervention fidelity. The methods that were utilized in this study were comparable to Hill *et al.* (1992) study in which the methods were similar using multiple raters and the same instrument as the larger studies. Therefore, the similarity of this study and Hill's study supports that CFNs are a reasonable alternate. A new competence scale added to the CSPRS showed initial feasibility in that the raters were able to establish adequate interrater reliability. Further work, however, is needed to develop it.

Limitations of the study

The CSPRS was designed to be used while a study is being conducted, not retrospectively. However, because this study compared CFN data to observational data used in gold standard fidelity studies, the measure was used retrospectively. Therefore, the retrospective use of the CSPRS is a limitation.

Along with the previous limitations mentioned in Chapter 3, the sample, raters, raters' training process for this study were limited. In spite of raters' comprehensive training using a rater's manual, the training that they received on rating CFNs was conducted for the first time in this study. The sole sources of data were the perceptions of the nurses through their accounts in the CFNs and these accounts may have been biased and/or incomplete. These findings do not challenge that direct observation is more accurate and should remain the gold standard. However, in spite of the limitations, the data support that CFNs offer a unique alternative.

In this pilot study the reliability scores were comparable to Hill's study of raters' reliability scores and the NIMH scores. Intra-class correlation is used to measure inter-rater reliability for two or more raters; however, due to the small sample and number of raters, reliability scores may have been inflated as a function of unstable variability (Bartko, 1976). Future studies should use a larger sample and more raters with random pairing of raters to determine interrater reliabilities.

Another limitation with regards to the scale used for competence, is that the lowest value for scale was 1= very poor. However, we cannot measure competence on something that is missing; the scale needs a non-applicable (N/A) end point that would represent competence more accurately. For example, in further research with the use of the concept, competence, the Likert-like scale should include 0=N/A - 1=very poor - 2 - 3 - 4=acceptable, 5-6 and 7= excellent. Finally, further analysis could be computed on an item-by-item basis, rather than on the subscale totals to shed light on precisely where discrepancies between raters may have occurred.

Implications for nursing research

Nurse researchers find themselves faced with the need to make adjustments in determining fidelity to accommodate the specific characteristics of the intervention or the participants. This study explored whether field notes provided enough information about the environmental factors and the behaviors of the nurse therapist to determine adherence and competence in regard to an intervention protocol. In nursing intervention research studies, the process of rating intervention fidelity has posed difficulties that have prevented the measurement of fidelity. This study has shown that at least one alternative method has the

potential as an addition to the traditional methods that may not be possible to use in some studies.

Improvements that are needed in future studies include: 1) the use of three or more raters based on sample size requirements for estimating ICC (Bonett, 2002); 2) a larger sample of subjects; 3) a random selection of field notes based on all sessions; 4) separate ratings for each of the sessions; 5) determine intervention fidelity using both the experimental and control group for comparison; and 6) explore the characteristics of the therapist nurses. Using the experimental and the control group for comparison would test whether there were diffusions of the intervention into the control group. For example, with further study, we could use the same subscales (IPT and CBT) to explore whether these dimensions were absent in the control group, an indicator of intervention fidelity. Intervention fidelity needs to be addressed in a comprehensive manner that includes training manuals and extensive training time that needs to be considered in the planning of a study. Finally, this study has shown interesting preliminary data on the function of field notes in determining intervention fidelity that can be pursued through further research.

APPENDIX A

COLLABORATIVE STUDY PSYCHOTHERAPY RATING SCALE (CSPRS)

IPT Subscale

*Directions: Read the chronological field notes for all the interventionist nurses for this study. You are to read each note that was related to the interventionist nurse's case [each mother]. Read carefully each field note and respond to each item on this questionnaire appropriately. There should only be one questionnaire used for each case **[not each field note]**.*

1. INTERPERSONAL RATIONALE FOR DEPRESSION: Did the interventionist nurse relate the mother's depression or specific depressive symptoms to difficulties in the mother's interpersonal relationships or role expectations?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

2. INTERPERSONAL THERAPY RATIONALE: Did the interventionist nurse provide a rationale which emphasized that working on understanding and changing the mother's interpersonal relationships and/or social roles would alleviate her depression?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

3. RELATING INTERPERSONAL CHANGE TO THERAPY: Did the interventionist nurse relate changes in the mother's interpersonal relationships or role expectations to the emphasis in therapy on understanding and changing the mother's interpersonal functioning or role expectations?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

4. INTERPERSONAL FOCUS: Did the interventionist nurse focus on the mother's interpersonal relationships or role expectations?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

5. PATTERNS IN RELATIONSHIPS: Did the interventionist nurse draw parallels or point out patterns in two or more of the mother's relationships for the purpose of helping the mother understand how she functions in interpersonal relationships?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

6. COMMUNICATION PATTERNS: Did the interventionist nurse explore with the mother how the mother's way of communicating affects her interpersonal relationships?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

7. INTERPERSONAL EXPECTATIONS: Did the interventionist nurse help the mother explore her expectations and the expectations of another person regarding their relationship?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

8. RELATIONSHIP SATISFACTION: Did the interventionist nurse review with the mother the satisfying and unsatisfying aspects of the mother's past or present important interpersonal relationship (s)?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

9. CHANGES DESIRED IN RELATIONSHIP (S): Did the interventionist nurse explore with the mother what changes the mother would like to see in an important relationship or in role expectations within a relationship?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

10. CONSIDER OPTIONS FOR INTERPERSONAL CHANGE: Did the interventionist nurse encourage the mother to consider a broad range of potential options for dealing with an interpersonal problem (or role expectation issue within a relationship)?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

11. WAYS OF CHANGING RELATIONSHIPS: Did the interventionist nurse help the mother to consider ways in which the mother can bring about desired changes in her interpersonal relationships (or role expectations in those relationships)?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

12. DEVELOPING NEW RELATIONSHIPS: Did the interventionist nurse help the mother to consider ways of developing new relationships?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

13. FOCUS ON INTERPERSONAL DISPUTE: Did the interventionist nurse focus on the mother's current overt or covert dispute with an important other person?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

14. CONSEQUENCE OF DISPARATE EXPECTATIONS: Did the interventionist nurse help the mother to relate interpersonal disputes to disparate expectations and values held by the mother and an important other person?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

15. RELATING DISPUTED RELATIONSHIP TO OTHER RELATIONSHIPS: Did the interventionist nurse attempt to help the mother see similarities or differences in a currently disputed relationship and other relationships (past or present) in the mother's life?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

16. MAINTAINING THE DISPUTE: Did the interventionist nurse and mother discuss processes through which the mother's interpersonal dispute is perpetuated?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

17. FOCUS ON ROLE TRANSITION: Did the interventionist nurse focus on the mother's interpersonal difficulties in current transition from one role to another?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

18. EXPLORING POSITIVE ASPECTS OF NEW ROLE: Did the interventionist nurse help the mother to explore the positive aspects of (including opportunities offered by) a new role?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

19. EVALUATE LOSSES IN ROLE CHANGE: Did the interventionist nurse help the mother to evaluate what she lost as part of her role change?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

20. PREPARATION FOR NEW ROLE: Did the interventionist nurse help or encourage the mother to prepare for the new role (e.g., by developing a social support system, necessary social skills, and a sense of mastery regarding the demands of the new role)?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

21. FOCUS ON INTERPERSONAL DEFICITS: (Defined as : Social isolation resulting from a significant lack of healthy important relationships or significant social unfulfillment that is present across relationships due to the mother's inability to get needs met from her relationships.) Did the interventionist nurse focus on the mother's interpersonal deficits?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

22. UNDERSTAND PAST DIFFICULTIES: Did the interventionist nurse help the mother to explore past difficulties in the mother's interpersonal relationships and to understand how these difficulties arose or how to avoid them in the future?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

23. USING THERAPEUTIC RELATIONSHIP AS A MODEL: Did the interventionist nurse attempt to use the therapeutic relationship to help the mother understand how she relates to others or as a model for the mother to use in developing satisfying relationships outside of therapy?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

24. BEST ASPECTS OF PRIOR RELATIONSHIPS: Did the interventionist nurse help the mother to explore the best aspects of the mother's prior relationships as a means of providing a model for the development of satisfying new relationships?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

25. FEELINGS ABOUT LOSS/DECEASED: Did the interventionist nurse help the mother to identify and explore feelings that the mother had about a deceased person when she was alive, or feelings that the mother now has about the loss or the deceased person?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

26. EXPLORATION OF FEELINGS: Did the interventionist nurse help the mother to explore her feelings related to an interpersonal relationship or role expectation?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

27. ACKNOWLEDGEMENT OF AFFECT: Did the interventionist nurse attempt to help the mother acknowledge affect that she was not expressing or of which she was unaware?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

28. ACCEPTANCE OF AFFECT: Did the nurse encourage the mother to accept feelings of which the mother is aware but which are painful or uncomfortable?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -----	2 ---	3 ---	4 ----	5 ----	6 -----	7
Very poor		Acceptable		Good		Excellent

[Extracted from CSPRS-6 with permission from M. Hill, PhD, University of Maryland, Department of Psychiatry
 Evans, M., Piasecki, J., Kriss, M., & Hollon (1984). Rater's Manual for the Collaborative Study Psychotherapy Rating Scale –
 Form 6. University of Minnesota and St. Paul-Ramsey Medical Center.]

APPENDIX B

COLLABORATIVE STUDY PSYCHOTHERAPY RATING SCEALE
CBT Subscale

*Directions: Read the chronological field notes for all the interventionist nurses for this study. You are to read each note that was related to the interventionist nurse's case [each mother]. Read carefully each field note and respond to each item on this questionnaire appropriately. There should only be one questionnaire used for each case **[not each field note]**.*

1. **SETTING AND FOLLOWING AGENDA:** Did the interventionist nurse work collaboratively with the mother to formulate and follow a specific agenda for the session?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		thoroughly

COMPETENCE/SKILL LEVEL

1 -----	2 -----	3 -----	4 --	5 -----	6 -----	7
Very poor		Acceptable		Good		Excellent

2. **HOMEWORK REVIEWED:** Did the nurse review previously assigned?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Did not Or none assigned		Some attention		Considerable attention		Thoroughly

COMPETENCE/SKILL LEVEL

1 -----	2 -----	3 -----	4 -----	5 -----	6 -----	7
Very poor		Acceptable		Good		Excellent

3. COLLABORATION: Did the interventionist nurse actively attempt to engage the mother in working together to explore therapeutic issues?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Nurse made no attempt to involve the mother in working <u>together</u>		Nurse occasionally attempted to involve the mother in working <u>together</u>		Nurse frequently attempted to involve the mother in working <u>together</u>		Throughout the session nurse actively solicited the mother's involvement in working <u>together</u>

COMPETENCE/SKILL LEVEL

1 -----	2 -----	3 -----	4 -----	5 -----	6 -----	7
Very poor		Acceptable		Good		Excellent

4. ENCOURAGES INDEPENDENCE: Did the interventionist nurse encourage the mother's independence from the nurse in dealing with her problems?

ADHERENCE LEVEL

1	2	3	4	5	6	7
No Encouragement of the mother's independence		Some encouragement of the mother's independence		Much encouragement of the mother's independence		Extensively encouragement of the mother's independence from the nurse

COMPETENCE/SKILL LEVEL

1 -----	2 -----	3 -----	4 -----	5 -----	6 -----	7
Very poor		Acceptable		Good		Excellent

5. SPECIFIC EXAMPLES: Did the interventionist nurse urge the mother to give concrete, specific examples of beliefs or events?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -----	2 -----	3 -----	4 -----	5 -----	6 -----	7
Very poor		Acceptable		Good		Excellent

6. RELATIONSHIP OF THOUGHTS AND FEELINGS: Did the interventionist nurse encourage the mother to relate affective states that the mother had experienced (or will experience in the future) to the mother's ongoing thoughts?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -	2 -	3 -	4 -	5	6	7
Very poor		Acceptable		Good		Excellent

7. COGNITIVE THERAPY RATIONALE: Did the interventionist nurse provide a rationale which emphasized the importance of evaluating the accuracy of the mother's beliefs and changing inaccurate beliefs in order to alleviate the mother's depression?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some discussion		Considerable discussion		Extensive discussion

COMPETENCE/SKILL LEVEL

1 -	2 --	3 -	4 -	5 -	6 --	7
Very poor		Acceptable		Good		Excellent

8. RELATE IMPROVEMENT TO COGNITIVE CHANGE: Did the interventionist nurse relate improvement that has occurred in the mother's depressive symptoms or related problems to changes in the mother's beliefs?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 --	3 -	4 -	5 --	6	7
Very poor		Acceptable		Good		Excellent

9. REPORTING COGNITIONS: Did the interventionist nurse ask the mother to report specific thoughts (as verbatim as possible) that the mother experienced either in the session or in a situation which occurred prior to the session?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Infrequent requests		Several requests		Frequent requests for specific thoughts

COMPETENCE/SKILL LEVEL

1 --	2 --	3 -	4 -	5 --	6 -	7
Very poor		Acceptable		Good		Excellent

10. EXPLORING PERSONAL MEANING: Did the interventionist nurse probe for beliefs related to a thought the mother reported in order to explore the personal meaning associated with the mother's initial thoughts?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some exploration of the mother's personal meaning system		Considerable exploration of the mother's personal meaning system		Extensive exploration of the mother's personal meaning system which includes a discussion of the impact of those related beliefs on the mother's affect

COMPETENCE/SKILL LEVEL

1 --	2 --	3 -	4 -	5 --	6 -	7
Very poor		Acceptable		Good		Excellent

11. RECOGNIZING CONGITIVE ERRORS: Did the interventionist nurse help the mother to identify specific types of cognitive distortions or errors (e.g., all-or-none thinking, overgeneralization) that were present in the mother's thinking?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 -	3 -	4 -	5 -	6 -	7
Very poor		Acceptable		Good		Excellent

12. IDENTIFYING UNDERLYING ASSUMPTIONS: Did the interventionist nurse explore with the mother a general belief that underlies many of the mother's specific negative thoughts and beliefs?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some mention of underlying assumption (s)		Considerable discussion of mother's underlying assumption (s)		Extensive discussion of mother's underlying assumption (s)

COMPETENCE/SKILL LEVEL

1 --	2 --	3 --	4 -	5 -	6 -	7
Very poor		Acceptable		Good		Excellent

13. CHALLENGING OF BELIEFS: Did the interventionist nurse encourage the mother to view her thoughts as beliefs which may or may not be true rather than as established facts?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 ---	2 -	3 -	4 -	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

14. EXAMINE AVAILABLE EVIDENCE: Did the interventionist nurse help the mother to use currently available evidence or information (including the mother's prior experiences) to test the validity of the mother's beliefs?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 ---	2 --	3 --	4 --	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

15. TESTING BELIEFS PROSPECTIVELY: Did the interventionist nurse encourage the mother to engage in specific behaviors for the purpose of testing the validity of her beliefs or make explicit predictions about external events so that the outcomes of those events could serve as tests of those predictions or review outcome of a previously devised prospective test or review outcome of a previously devised prospective test?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 -	3 -	4 -	5 -	6 --	7
Very poor		Acceptable		Good		Excellent

16. SEARCHING FOR ALTERNATIVE EXPLANATIONS: Did the interventionist nurse help the mother to consider alternative explanations for events besides the mother's initial explanations for those events?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 --	3 -	4 ---	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

17. REALISTIC CONSEQUENCES: Did the interventionist nurse work with the mother to determine what the realistic consequences would be if the mother's belief proved to be true?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 ---	2 --	3 --	4 --	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

18. ADAPTIVE/FUNCTIONAL VALUE OF BELIEFS: Did the interventionist nurse guide the mother to consider whether or not maintaining a specific belief is adaptive for the mother (regardless of whether or not it accurate)?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 -	3 -	4 -	5 --	6 ---	7
Very poor		Acceptable		Good		Excellent

19. DIDACTIC PERSUASION: Did the interventionist nurse use didactic persuasion to urge the mother to change her belief (s)?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 -	3 --	4 -	5 -	6 -	7
Very poor		Acceptable		Good		Excellent

20. SUBSTITUTING POSITIVE THOUGHTS: Did the interventionist nurse encourage the mother to substitute a more positive belief for another (whether or not the substitute belief was more accurate or realistic), solely because the mother would feel better if she thought another way?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 ---	3 ---	4 ---	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

21. PRACTICING “RATIONAL RESPONSES”: Did the interventionist nurse and mother practice possible rational responses to the mother’s negative thoughts or beliefs?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 -	3 -	4 --	5 --	6 ---	7
Very poor		Acceptable		Good		Excellent

22. PLANNING/PRACTICING ALTERNATIVE BEHAVIORS: Did the interventionist nurse work with the mother to plan or to practice alternative overt behaviors for the mother to utilize outside of therapy?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 --	3 --	4 --	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

23. SKILLS TRAINING: Did the interventionist nurse attempt to teach the mother skills (e.g., assertiveness, social skills, task relevant skills) in the session?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 -	3	4 --	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

24. HOMEWORK ASSIGNED: Did the interventionist nurse or mother develop one or more specific assignments for the mother to engage in prior to the next session?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		Some attempt to develop homework		Considerable attempt to develop homework		Extensive attempt to develop homework

COMPETENCE/SKILL LEVEL

1 ---	2 ---	3 ---	4 ---	5 --	6 ---	7
Very poor		Acceptable		Good		Excellent

25. INCREASING PLEASURE AND MASTERY: Did the interventionist nurse encourage the mother to engage in activities which would be pleasurable to the mother or from which the mother would obtain a sense of accomplishment?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 ---	2 ---	3 ---	4 --	5 --	6 -	7
Very poor		Acceptable		Good		Excellent

26. SCHEDULING/STRUCTURING ACTIVITIES: Did the interventionist nurse work with the mother to schedule or structure one or more specific activities for the purpose of increasing the likelihood that the mother will initiate or follow through on those activities?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 -	3 --	4 --	5 -	6 --	7
Very poor		Acceptable		Good		Excellent

27. SELF-MONITORING: Did the interventionist nurse encourage the mother to record feelings, activities, or events between sessions or review the mother's records of feelings, activities, or events?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 --	3 ---	4 --	5 -	6 --	7
Very poor		Acceptable		Good		Excellent

28. RECORDING THOUGHTS: Did the interventionist nurse encourage the mother to record thoughts between sessions or review the mother's records of her thoughts?

ADHERENCE LEVEL

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 --	3 ---	4 --	5 -	6 ---	7
Very poor		Acceptable		Good		Excellent

29. MANIPULATING BEHAVIOR VIA CUES OR CONSEQUENCES: Did the interventionist nurse help the mother to arrange for cues (i.e., stimulus control) or consequences behaviors in order to manipulate the occurrence of those behaviors?

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 ----	2 --	3 --	4 -	5 --	6 ---	7
Very poor		Acceptable		Good		Excellent

30. ATTEMPTS TO NEGOTIATE THERAPY CONTENT: Did the interventionist nurse negotiate with the mother assignments, changes in direction, or major emphases of the session in a way that gave the mother opportunity to have input?

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 --	2 -	3 --	4 --	5 --	6 --	7
Very poor		Acceptable		Good		Excellent

31. EXPLANATION FOR INTERVENTIONIST'S DIRECTION: Did the interventionist nurse explain to the mother the interventionist's reasons for pursuing a particular topic in the session?

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 ---	2 --	3 ----	4 -	5 --	6 ---	7
Very poor		Acceptable		Good		Excellent

32. SUMMARIZING: Did the interventionist nurse summarize or encourage the mother to summarize key issues discussed either in a previous session or in the current session?

1	2	3	4	5	6	7
Not at all		some		considerably		extensively

COMPETENCE/SKILL LEVEL

1 -	2 --	3 -	4 --	5 ---	6 --	7
Very poor		Acceptable		Good		Excellent

[Extracted from CSPRS-6 with permission from M. Hill, PhD, University of Maryland, Department of Psychiatry
 Evans, M., Piasecki, J., Kriss, M., & Hollon (1984). Rater's Manual for the Collaborative Study Psychotherapy Rating Scale –
 Form 6. University of Minnesota and St. Paul-Ramsey Medical Center.]

APPENDIX C

RATER APPLIED PERFORMANCE SCALE (RAPS)

Checkpoint number	①	②	③
RATER ID	_____		

Review based on: audiotape videotape **field notes** live review

1. Adherence: Adherence to rating scale content and item sequence.
NA unsatisfactory fair good excellent
2. Follow-up: Use of follow-up questions to obtain more information.
NA unsatisfactory fair good excellent
3. Clarification: Use of questions and re-phrasing of what has already been said to clarify ambiguous information:
NA unsatisfactory fair good excellent
4. Neutrality: Use of open-ended questions. Avoiding leading questions:
NA unsatisfactory fair good excellent
5. Accuracy of scoring: The rater scores items based on anchor descriptions and other (e.g., manual) guidelines. Ratings are based on positive information obtained.
NA unsatisfactory fair good excellent

Lipsitz, J. et al. (2004). The rater applied performance scale: development and reliability. Psychiatry Research 127, 147-155.

APPENDIX D

INTERVENTION FIDELITY: The raters' manual

[Two page copy]

For a pilot intervention fidelity research study:

“Determining Intervention Fidelity from Chronological Field Notes”

*Developed by JADowell, Phdc as a part of the dissertation proposal:
A modified version of Evans, M., Piasecki, J., Kriss, M., & Hollon (1984). Rater's Manual
for the Collaborative Study Psychotherapy Rating Scale University of Minnesota and St.
Paul-Ramsey Medical Center.]*

Table of Contents from the Raters' Manual

I.	-----	Purpose
II.	-----	Key concepts
III.	-----	Brief overview of the primary intervention study
IV.	-----	Raters responsibilities
V.	-----	Electronic rating
VI.	-----	Rating adherence and Competence behaviors
VII.	-----	Step 1
VIII.	-----	Step 2
IX.	-----	Step 3
X.	-----	Submitting the scores
XI.	-----	Appendix A
XII.	-----	Appendix B
XIII.	-----	Appendix C
XIV.	-----	Appendix D

Appendix E

Program Evaluation Form

This program evaluation instrument is designed to give input on how well you thought this program provided you with the skills and knowledge to complete your responsibilities as a rater in this study. Do not put your name on the form. Your response is anonymous. We encourage you to be frank and honest in your evaluation.

Please indicate your answers on the computerized answer sheet. Comments should be made on the General Evaluation Form.

Disclosures

Prior to the start of this Raters' Training Program offering you were informed in writing and/or verbally about whether there was or was not any conflict of interest for training instruction.

1. Do you feel that this training activity was unbiased?

General Training Information

2. Training was consistent with my expectation.
3. Training challenged me and motivated me to learn.
4. Teaching strategies used were appropriate to the training content.
5. Facilities chosen for the training were appropriate and conducive to learning.
6. The training objectives contributed to the overall purpose of the raters' training.
7. Overall evaluation for this raters' training program.

Training Objectives

Upon completion of this training, the participants will be able to:

8. Systematically, perform the duties of a rater.
9. Provide a rationale for their decisions.
10. Record appropriate response as per instructed with this study.

REFERENCES

- Ablon, J., & Jones, E. (2002). Validity of Controlled Clinical Trials of Psychotherapy: Findings From the NIMH Treatment of Depression Collaborative Research Program. *Am J Psychiatry, 159*(5), 775-783. doi: 10.1176/appi.ajp.159.5.775
- Addis, M., & Waltz, M. (2002). Implicit and Untested Assumptions About the Role of Psychotherapy Treatment Manuals in Evidence- Based Mental Health Practice. *Clinical Psychology: Science and Practice, 9*(4), 421-424.
- ATLAS.TI. (2009). ATLAS.TI version 6.05-The Qualitative Data Analysis Software. Germany: ATLAS.TI.
- Barber, J., & Critis-Christoph, P. (1996). Development of a Therapist Adherence/Competence Rating Scale for Supportive-Expressive Dynamic Psychotherapy: A Preliminary Report. *Psychotherapy Research, 6*(2), 81 - 94.
- Barber, J., 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. doi: 10.1037/0022-006x.64.3.619
- Barber, J., Stratt, R., Halperin, G., & Connolly, M. (2001). Supportive Techniques: Are They Found in Different Therapies? *J Psychother Pract Res, 10*(3), 165-172.
- Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*(5), 762-765. doi: 10.1037/0033-2909.83.5.762
- Beeber, L. (2004). HILDA PROJECT: Nurse Intervention Manual RO1 MH065524-01A2. University of North Carolina at Chapel Hill: National Institutes of Health, National Institute of Mental Health.
- Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D., Ory, M., et al. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol, 23*(5), 443-451. doi: 10.1037/0278-6133.23.5.4432004-18051-001 [pii]
- Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in medicine, 21*(9), 1331-1335. doi: 10.1002/sim.1108
- Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A., Czajkowski, S., Breger, R., et al. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73*(5), 852-860. doi: 10.1037/0022-006x.73.5.852
- Brandt, P., Kirsch, S., Lewis, F., & Casey, S. (2004). Assessing the strength and integrity of an intervention. *Oncology Nursing Forum, 31*(4), 833-837.

- Bruckenthal, P., & Broderick, J. (2007). Assessing treatment fidelity in pilot studies assist in designing clinical trials: an illustration from a nurse practitioner community-based intervention for pain. *ANS Adv Nurs Sci*, 30(1), E72-84. doi: 00012272-200701000-00015 [pii]
- Calsyn, R. (2000). A Checklist for Critiquing Treatment Fidelity Studies. [10.1023/A:1010109205676]. *Mental Health Services Research*, 2(2), 107-113.
- Carroll, K. (1996). The Technology model: An introduction to psychotherapy research in substance abuse. *Training Series*: Yale University Psychotherapy Development Center.
- Carroll, K., Nich, C., Sifry, R., Nuro, K., Frankforter, T., Ball, S., et al. (2000). A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence*, 57(3), 225-238.
- Chan, E., O'Neill, I., McKenzie, M., Love, A., & Kissane, D. (2004). What works for therapists conducting family meetings: treatment integrity in family-focused grief therapy during palliative care and bereavement. *J Pain Symptom Manage*, 27(6), 502-512. doi: 10.1016/j.jpainsymman.2003.10.008S0885392404001009 [pii]
- Chevron, E. S., & Rounsaville, B. J. (1983). Evaluating the clinical skills of psychotherapists. A comparison of techniques. *Arch Gen Psychiatry*, 40(10), 1129-1132.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290. doi: 10.1037/1040-3590.6.4.284
- Clarke, G. (1998). Intervention Fidelity in the Psychosocial Prevention and Treatment of Adolescent Depression. *Journal Of Prevention And Intervention In The Community*, 17, 19-34.
- Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Cook, T., & Campbell, D. (1979). *Quasi-Experimentation Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.
- Crits-Christoph, P., Connolly, M., & Shaffer, C. (1999). Reliability and base rates of interpersonal themes in narratives from psychotherapy sessions. *Journal of Clinical Psychology*, 55(10), 1227-1242.
- Dane, A., & Schneider, B. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev*, 18(1), 23-45. doi: S0272-7358(97)00043-3 [pii]

- Davis, M., Baranowski, T., Hughes, M., Warneke, C., de Moor, C., & Mullis, R. (2002). Using Children as Change Agents to Increase Fruit and Vegetable Consumption Among Lower-Income African American Parents. In A. Steckler & L. Linnan (Eds.), *Process Evaluation for Public Health Interventions and Research* (First ed., pp. 249-267). San Francisco, CA: Jossey-Bass.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings (Vol. 18, pp. 237-256): Oxford Univ Press.
- Erlen, J., & Sereika, S. (2006). Fidelity to a 12-Week Structured Medication Adherence Intervention in Patients With HIV. [Article].
- Evans, M., Piasecki, J., Kriss, M., & Hollon, S. (1984). *Rater's Manual: For the Collaborative Study Psychotherapy Rating Scale-form 6*. University of Minnesota and St. Paul-Ramsey Medical Center, Minnesota.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. . New York: John Wiley.
- Fleiss, J. L., & Cohen, J. (1973). The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement, 33*(3), 613-619. doi: 10.1177/001316447303300309
- Forgatch, M. S., Degarmo, D. S., & Beldavs, Z. G. (2005). An efficacious theory-based intervention for stepfamilies. *Behav Ther, 36*(4), 357-365.
- Forgatch, M. S., Patterson, G. R., & Degarmo, D. S. (2005). Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behav Ther, 36*(1), 3-13.
- Hesse-Biber, S., & Leavy, P. (2006). *The Practice of Qualitative Research*. Thousand Oaks: Sage Publications.
- Hill, C., O'Grady, K., & Elkin, I. (1992). Applying the Collaborative Study Psychotherapy Rating Scale to rate therapist adherence in cognitive-behavior therapy, interpersonal therapy, and clinical management. *Journal of Consulting and Clinical Psychology, 60*(1), 73-79. doi: 10.1037/0022-006x.60.1.73
- Hill, C., O'Grady, K., & Price, P. (1988). A method for investigating sources of rater bias. *Journal of Counseling Psychology, 35*(3), 346-350. doi: 10.1037/0022-0167.35.3.346
- Hogue, A., Dauber, S., Chinchilla, P., Fried, A., Henderson, C., Inclan, J., et al. (2008). Assessing fidelity in individual and family therapy for adolescent substance abuse. *Journal of Substance Abuse Treatment, 35*(2), 137-147. doi: DOI: 10.1016/j.jsat.2007.09.002

- Hogue, A., Henderson, C., Dauber, S., Barajas, P., Fried, A., & Liddle, H. (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. doi: 10.1037/0022-006x.76.4.544
- Hogue, A., Liddle, H., Becker, D., & Johnson-Leckrone, J. (2002). Family-based prevention counseling for high-risk young adolescents: Immediate outcomes. *Journal of Community Psychology, 30*(1), 1-22.
- Horner, S., Rew, L., & Torres, R. (2006). Enhancing intervention fidelity: a means of strengthening study impact. *Journal for Specialists in Pediatric Nursing, 11*(2), 80-89.
- Howell, D. (2002). *Statistical Methods for Psychology*. Duxbury United States: Thompson Learning.
- Kazdin, A. E., Mazurick, J. L., & Siegel, T. C. (1994). Treatment outcome among children with externalizing disorder who terminate prematurely versus those who complete psychotherapy. *J Am Acad Child Adolesc Psychiatry, 33*(4), 549-557.
- Kerlinger, F., & Lee, H. (2000). *Foundations of Behavioral Research*. Wadsworth United States: Thompson Learning.
- Kerns, S. E., & Prinz, R. J. (2002). Critical issues in the prevention of violence-related behavior in youth. *Clin Child Fam Psychol Rev, 5*(2), 133-160.
- Lee, C., August, G., Realmuto, G., Horowitz, J., Bloomquist, M., & Klimes-Dougan, B. (2008). Fidelity at a distance: assessing implementation fidelity of the Early Risers prevention program in a going-to-scale intervention trial. *Prev Sci, 9*(3), 215-229. doi: 10.1007/s11121-008-0097-6
- Leventhal, H., & Friedman, M. (2004). Does Establishing Fidelity of Treatment Help in Understanding Treatment Efficacy? Comment on Bellg et al. (2004). *Health Psychology, 23*(5), 452-456. doi: 10.1037/0278-6133.23.5.452
- Luty, S., Carter, J., McKenzie, J., Rae, A., Frampton, C., Mulder, R., et al. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *The British Journal of Psychiatry, 190*(6), 496-502. doi: 10.1192/bjp.bp.106.024729
- Luty, S., Carter, J., McKenzie, J., Rae, A., Frampton, C., Mulder, R., et al. (2007). Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *Br J Psychiatry, 190*, 496-502. doi: 10.1192/bjp.bp.106.024729 [pii]

- Madson, M. B., & Campbell, T. C. (2006). Measures of fidelity in motivational enhancement: a systematic review. *J Subst Abuse Treat*, *31*(1), 67-73. doi: S0740-5472(06)00089-4 [pii]10.1016/j.jsat.2006.03.010
- Markowitz, J. (2004). Interpersonal psychotherapy: principles and applications. *World Psychiatry*, *3*(3), 136-139.
- Markowitz, J., Spielman, L., Scarvalone, P., & Perry, S. (2000). Psychotherapy Adherence of Therapists Treating HIV-Positive Patients With Depressive Symptoms. *J Psychother Pract Res*, *9*(2), 75-80.
- Martino, S., Ball, S., Nich, C., Frankforter, T., & Carroll, K. (2008). Community program therapist adherence and competence in motivational enhancement therapy. *Drug and Alcohol Dependence*, *96*(1-2), 37-48. doi: DOI: 10.1016/j.drugalcdep.2008.01.020
- McGraw, K. O., & Wong, S. P. (1996). Forming Inferences About some Intraclass Correlation Coefficients. *Psychological Methods*, *1*(1), 30-46.
- Merriam Webster's Collegiate Dictionary. (2002) (10th Edition ed.). Springfield, Massachusetts: Merriam-Webster, Incorporated.
- Miklowitz, D. J., & Goldstein, M. J. (1990). Behavioral family treatment for patients with bipolar affective disorder. *Behav Modif*, *14*(4), 457-489.
- Minnick, A., Catrambone, C., Halstead, L., Rothschild, S., & Lapidus, S. (2008). A nurse coach quality improvement intervention: feasibility and treatment fidelity. *West J Nurs Res*, *30*(6), 690-703. doi: 0193945907311321 [pii]10.1177/0193945907311321
- Moncher, F., & Prinz, R. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247-266. doi: Doi: 10.1016/0272-7358(91)90103-2
- Mowbray, C., Holter, M., Teague, G., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation*, *24*(3), 315-340. doi: 10.1177/109821400302400303
- Najavits, L. M., & Weiss, R. D. (1994). Variations in therapist effectiveness in the treatment of patients with substance use disorders: an empirical review. *Addiction*, *89*(6), 679-688.
- Nezu, A., & Nezu, C. (2008). *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions*. New York: Oxford University Press.
- Nigg, C., Allegrante, J., & Ory, M. (2002). Theory-comparison and multiple-behavior research: common themes advancing health behavior research. *Health Educ. Res.*, *17*(5), 670-679. doi: 10.1093/her/17.5.670

- Nunnally, J. (1967). *Psychometric Theory*. New York: McGraw-Hill Book Company.
- Ory, M., Jordan, P., & Bazzarre, T. (2002). The Behavior Change Consortium: setting the stage for a new century of health behavior-change research. *Health Educ. Res.*, 17(5), 500-511. doi: 10.1093/her/17.5.500
- Pbert, L., Osganian, S. K., Gorak, D., Druker, S., Reed, G., O'Neill, K. M., et al. (2006). A school nurse-delivered adolescent smoking cessation intervention: a randomized controlled trial. *Prev Med*, 43(4), 312-320. doi: S0091-7435(06)00159-9 [pii]10.1016/j.ypmed.2006.04.006
- Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, 15, 477-492.
- Resnick, B., Bellg, A., Borrelli, B., DeFrancesco, C., Breger, R., Hecht, J., et al. (2005). Examples of Implementation and Evaluation of Treatment Fidelity in the BCC Studies: Where We Are and Where We Need to Go. [Article]. *Annals of Behavioral Medicine*, 29, 46-54. doi: 10.1207/s15324796abm2902s_8
- Resnick, B., Galik, E., Pretzer-Aboff, I., Gruber-Baldini, A., Russ, K., Cayo, J., et al. (2009). Treatment fidelity in nursing home research: the Res-Care Intervention Study. *Research in Gerontological Nursing*, 2(1), 30-38.
- Resnick, B., Inguito, P., Orwig, D., Yahiro, J., Hawkes, W., Werner, M., et al. (2005). Treatment fidelity in behavior change research: A case example. *Nurs Res*, 54, 139 - 143.
- Resnick, B., Inguito, P., Orwig, D., Yahiro, J., Hawkes, W., Werner, M., et al. (2009). Treatment Fidelity in Behavior Change Research: A Case Example. [Miscellaneous Article].
- Roper, J., & Shapira, J. (2000). *Ethnography In Nursing Research*. Thousand Oaks: Sage Publications, Inc.
- Sanchez, V., Steckler, A., Nitirat, P., Hallfors, D., Cho, H., & Brodish, P. (2007). Fidelity of implementation in a treatment effectiveness trial of Reconnecting Youth. *Health Educ. Res.*, 22(1), 95-107. doi: 10.1093/her/cyl052
- Sanjek, R. (1990). *Fieldnotes: The Makings of Anthropology*. Ithaca: Cornell University Press.
- Santacroce, S., Maccarelli, L., & Grey, M. (2004). Intervention Fidelity. *Nursing Research January/February*, 53(1), 63-66.
- Shadish, W. (2002). Revisiting Field Experimentation: Field Notes for the Future. *Psychological Methods*, 7(1), 3-18.

- Shapiro, D., & Startup, M. (1992). Measuring Therapist Adherence in Exploratory Psychotherapy. *Psychotherapy Research*, 2(3), 193 - 203.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, 86(2), 420-428.
- Spicer, J. (2005). *Making Sense of Multivariate Data Analysis*. Thousand Oaks: Sage Publications.
- Spillane, V., Byrne, M., Byrne, M., Leathem, C., O'Malley, M., & Cupples, M. (2007). Monitoring treatment fidelity in a randomized controlled trial of a complex intervention. *Journal of Advanced Nursing*, 60(3), 343-352.
- Stein, K., Sargent, J., & Rafaels, N. (2007). Intervention Research: Establishing Fidelity of the Independent Variable in Nursing Clinical Trials. [Article].
- Streiner, D., & Norman, G. (2005). *Health Measurement Scales: A practical guide to their development and use* (Third ed.). Oxford: Oxford University Press.
- Symes, M., Remington, B., Brown, T., & Hastings, R. (2006). Early intensive behavioral intervention for children with autism: Therapists' perspectives on achieving procedural fidelity. *Research in Developmental Disabilities*, 27(1), 30-42. doi: DOI: 10.1016/j.ridd.2004.07.007
- Tessaro, I., Campbell, M., & Benedict, S. (2002). Health works for women. In A. Steckler & L. Linnan (Eds.), *Process Evaluation for Public Health Interventions and Research* (First ed., pp. 184-202). San Francisco, CA: Jossey-Bass.
- Tyler, D., & Horner, S. (2008). Collaborating With Low-Income Families and Their Overweight Children to Improve Weight-Related Behaviors: An Intervention Process Evaluation. [Article]. *Journal for Specialists in Pediatric Nursing*, 13(4), 263-274. doi: 10.1111/j.1744-6155.2008.00167.x
- US Census Bureau. (2000). Your Gateway to Census 2000, from <http://www.census.gov/>
- Wahab, S., Menon, U., & Szalacha, L. (2008). Motivational interviewing and colorectal cancer screening: A peek from the inside out. *Patient Education and Counseling*, 72(2), 210-217. doi: DOI: 10.1016/j.pec.2008.03.023
- Waltz, J., Addis, M., Koerner, K., & Jacobson, N. (1993). Testing the integrity of a psychotherapy protocol: Assessment of adherence and competence. *Journal of Consulting and Clinical Psychology*, 61(4), 620-630. doi: 10.1037/0022-006x.61.4.620

- Weissman, M., Rounsaville, B., & Chevron, E. (1982). Training psychotherapists to participate in psychotherapy outcome studies. *American Journal of Psychiatry*, *139*, 1442-1446.
- Whitmer, K., Sweeney, C., Slivjak, A., Sumner, C., & Barsevick, A. (2005). Strategies for Maintaining Integrity of a Behavioral Intervention. *West J Nurs Res*, *27*(3), 338-345. doi: 10.1177/0193945904270087
- Wilson, D., & Crisanti, A. (2009). Psychometric Properties of the Dual-Disorder Treatment Fidelity Scale: Inter-Rater Reliability and Concurrent Validity. [10.1007/s10597-008-9167-2]. *Community Mental Health Journal*, *45*(3), 171-178.