Georgia State University
Digital Archive @ GSU

Public Health Theses | Institute of Public Health

7-23-2010

A Comparison of Methods to Assess Practitioner Fidelity in a Parent-Training Program

Ashwini Tiwari
Georgia State University, atiwari1@student.gsu.edu

Follow this and additional works at: http://digitalarchive.gsu.edu/iph_theses

Recommended Citation
Tiwari, Ashwini, "A Comparison of Methods to Assess Practitioner Fidelity in a Parent-Training Program" (2010). *Public Health Theses.* Paper 136.

This Thesis is brought to you for free and open access by the Institute of Public Health at Digital Archive @ GSU. It has been accepted for inclusion in Public Health Theses by an authorized administrator of Digital Archive @ GSU. For more information, please contact digitalarchive@gsu.edu.

ABSTRACT

Ashwini Tiwari
A Comparison of Methods to Assess Practitioner Fidelity in a Parent-Training Program
(Under the direction of Daniel J. Whitaker, Faculty Member)

As evidence-based programs are implemented in real world settings, there is a strong need to effectively and efficiently monitor fidelity, or adherence to a program, in order to maintain the expected effects demonstrated in research settings. The purpose of this study was to compare two methods of assessing fidelity to an evidence-based, parent-training model (SafeCare®) as implemented by community service providers. Specifically, analyses compared fidelity assessed via video versus audio recordings. SafeCare modules often require mobility and high interaction, thus, video recordings may provide a more accurate view of home visitor and family interactions for scoring fidelity. However, videos are more expensive and cumbersome in comparison to audio recordings. Trained coders were randomly assigned to score a video or audio recording of the same session for 25 SafeCare sessions and the codes were compared for agreement. Two types of SafeCare sessions were assessed: assessment and training. Average agreement was somewhat higher for assessment sessions than for training sessions. Average agreement, across all sessions, was higher among items pertaining to SafeCare content than items pertaining to the therapeutic process. Several specific items were identified that are difficult to code via audio recordings. However, more research is needed to determine agreement levels across all SafeCare modules and session types in order to provide insight on the implications for SafeCare's future use of audio and video methods of measuring fidelity.

INDEX WORDS: fidelity, implementation research, parent-training programs, public health

A COMPARISON OF METHODS TO ASSESS PRACTITIONER FIDELITY IN A PARENT-TRAINING PROGRAM

by

ASHWINI TIWARI

HonBSc. UNIVERSITY OF TORONTO

A Thesis Submitted to the Graduate Faculty

Of Georgia State University in Partial Fulfillment

Of the

Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA

A COMPARISON OF METHODS TO ASSESS PRACTITIONER FIDELITY IN A PARENT-TRAINING PROGRAM

By

ASHWINI TIWARI

Approved:

DANIEL J. WHITAKER, Ph.D.
Committee Chair

SHANNON SELF-BROWN, Ph.D.
Committee Member

JOHN R. LUTZKER, Ph.D.
Committee Member

07-23-10
Date

ACKNOWLEDGEMENTS

"O Lord of infinite prowess, my salutations to You from before and from behind. O soul of all, my obeisance to You from all sides indeed. You, who possess limitless might, pervade all; therefore you are all... Therefore, Lord, prostrating my body at Your feet and bowing low, I seek to propitiate You, the ruler of all, and worthy of all praise." Srīmad Bhagavadgīta, 11.40, 11.43

I extend my most sincere 'thank you' to my chairperson, Dr. Daniel Whitaker for guiding me through this thesis process, and allowing me to barge into his office during his lunch breaks, over and over again. Thank you for your undying patience, and irreplaceable mentorship during this period; I would like to acknowledge my committee members, Dr. Shannon Self-Brown and Dr. John Lutzker, and also the staff and faculty at the Center for Healthy Development for providing me with such an enjoyable, unforgettable and valuable experience. I also have to thank all of the Institute of Public Health's faculty and staff for giving me the knowledge and tools to succeed in, and outside of this program. In particular, I have to thank Ms. Courtney Burton for her support, friendship and conversations over the past two years. I especially have to thank Dr. Michael Eriksen, who has been the most influential person throughout my time at IPH, and who has given me limitless opportunities, and continuous votes of confidence in whatever I set out to accomplish.

For personal support: I must thank my Pumi, and my gun-slinging partner-in-crime, Uncle Sham. Who else would have kept me this sane in my times of stress and would have provided me with love, crush-ice, peanut punch, food, and lots of fun?

Arianne, you are my phone support/lifeline and a true friend. Thank you for keeping me entertained and letting me rant to you at a non-stop pace.

To my family, Aunty Deepa, Aunty Indi, Aunty Shanta, Uncle Chris, Uncle Deep, Amrit and Didi Priya: thank you for your support and advice, and for being family. I extend my biggest thank you to my Nanee, for your daily prayers, for putting up with my nonsensical singing and requests for hugs over the phone.

Most importantly, I cannot and will never forget my parents: my Mom/ Bee/Ama/ Her Majesty, Queen Hotness, and my Dad/Dada/Fah-der. In all seriousness, I don't have the right words to give you the most appropriate or adequate recognition. You deserve more than this acknowledgement. I cannot ask for better parents. You are my rocks in life, and my best friends, even when you pester me about getting enough sleep.

Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, College of Health and Human Sciences. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

ASHWINI TIWARI

Signature of Author

Notice to Borrowers Page

All theses deposited in the Georgia State University Library must be used in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this thesis is:

Student's Name: ASHWINI TIWARI

Street Address: 4595 PARKVIEW WALK DRIVE

City, State, and Zip Code: LILBURN, GA, 30047

The Chair of the committee for this thesis is:

Professor's Name: DANIEL J. WHITAKER, Ph.D.

Department: INSTITUTE OF PUBLIC HEALTH

College: HEALTH AND HUMAN SCIENCES

Georgia State University
P.O. Box 3995
Atlanta, Georgia 30302-3995

Users of this thesis who not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulation by signing below. Libraries borrowing this thesis for the use of their patrons are required to see that each user records here the information requested.

NAME OF USER	ADDRESS	DATE	TYPE OF USE (EXAMINATION ONLY OR COPYING)

CURRICULUM VITAE

Ashwini Tiwari
4595 Parkview Walk Drive
Lilburn, GA, 30047
678 924 3103
atiwari1@student.gsu.edu

Education

Georgia State University- Atlanta, GA
Master of Public Health, 2010

University of Toronto at Scarborough-ON, Canada
Honors Bachelor of Science, 2007; Certified Program in Health Studies, 2007

Professional Experience

National SafeCare® Training and Research Center, Atlanta, Georgia
Graduate Research Assistant 2009- Present

Healthcare Georgia Foundation, Atlanta, Georgia
Internship Feb 2009-Aug 2009

Georgia State University
Public Health Institute Student Association (PHISA) Secretary 2008- 2009

Georgia State University
Graduate Research Assistant to the Graduate Coordinator 2008- 2009

University of Toronto at Scarborough
Research Assistant in Anthropology and Health Studies 2006- 2007

Honors and Awards

Emory Center for Injury Control's Summer Student Injury Prevention Award, Public Health Achievement Award, Phi Beta Delta Honor Society for International Scholars, University of Toronto Honors List

Memberships

Board of Trustees: Dharmic Sabha of Georgia 2010-Present
Public Health Institute's Student Association (PHISA) 2008-Present

Publication

Tiwari, A., Nelson, G.D. (2009). *2008 Health Rankings: Georgia and Georgia's Children, Publication #36,* Atlanta, Georgia: Healthcare Georgia Foundation.

TABLE OF CONTENTS

Page
ACKNOWLEDGMENTS....iii

LIST OF TABLES v

CHAPTER I: REVIEW OF THE LITERATURE....1
Purpose of the Study 12
Research Questions 12

CHAPTER II: METHODS AND PROCEDURES 14
Description of Data and Data Source....14
Home Visitor Demographics....15
Study Procedure 15
Study Measures 19
Statistical Analyses....19

CHAPTER III: RESULTS.... 22
Fidelity Checklist Item Classification....22
Descriptive Statistics....23
Reliability Analyses. 23
Primary Analyses....24

CHAPTER IV: DISCUSSION....41
Reliability Analyses....41
Main Analyses....43
Implications for SafeCare Fidelity Monitoring....47
Study Limitations....48
Future Research Aims....49

REFERENCES....52
APPENDIX A. Assessment Fidelity Checklist....59
APPENDIX B. Training Fidelity Checklist....61

List of Tables

Table 1 Number of Videos Stratified by Module and Session Type.....21

Table 2 Number of Videos Included in the Study Stratified by Module and Session Type.....28

Table 3 Frequency Distribution of Video and Audio Rated Recordings by All Coders.....28

Table 4 Summary of Reliability Mean Agreement for All Sessions, by Session Type, by Module, by Method, and by Coder.....29

Table 5 Percent Agreement between Audio and Video Raters for 28 Items on the Assessment Checklist across 7 Health Assessment Sessions.....31

Table 6 Fidelity Classification and Percent Agreement between Audio and Video Raters for 29 Training Fidelity Checklist Items across 17 Recorded Training Sessions.....36

Table 7 Summary of Three Sets of Mean Percentage of Agreement by Session and Module Type.....37

Table 8 Items among Assessment and Training Sessions with Low Audio/ Video Coder Agreement in Comparison to Reliability Scores.....39

Chapter I

REVIEW OF THE LITERATURE

The goal of this study was to compare video and audio recordings as methods of monitoring fidelity for the SafeCare® program. A comprehensive literature review provides insight on the use of these methods among other parent-training programs and how fidelity is typically conceptualized in parent-training programs. This review covers a brief overview of child maltreatment; the need for evidence-based practices; the importance quality assurance methods including coaching and fidelity monitoring; parent-training programs that use video and audio recordings to measure fidelity; an overview of the SafeCare program and how fidelity is monitored, and aspects of fidelity that are measured by this model.

Overview of Child Maltreatment

Child maltreatment (CM) is a well recognized and significant public health problem within the United States (U.S. Department of Health and Human Services, Administration on Children, Youth and Families, 2009), and is associated with a range of adverse outcomes across behavioral, social, cognitive and emotional domains. In 2007, child welfare and protective services (CPS) received approximately 3.5 million referrals for child abuse and neglect and substantiated over 794,000 cases (U.S. Department of Health and Human Services, 2009). However, because not all cases are reported to CPS, the actual number of maltreatment cases is likely to be substantially greater (Leeb, Paulozzi, Melanson, Simon, & Arias, 2008; Sedlak et al., 2010; Whitaker, Lutzker, Self-

Brown, Edwards, 2008). Approximately 80% of perpetrators were parents, with a higher percentage of mothers (39%) than fathers (19%) acting alone (DHHS, 2009). Findings from meta-analyses show that several factors are associated with parents who maltreat their children including part-time employment, low socioeconomic status, prior history of abuse, substance abuse, low familial support, substance abuse, parenthood at a young age, and social isolation (Brown, Cohen, Johnson, & Salzinger, 1998; Schumacher, Slep, & Heyman, 2001).

The Centers for Disease Control and Prevention (CDC) defines CM as "Any act or series of acts of commission or omission by a parent or other caregiver that results in harm, potential for harm, or threat of harm to a child" (Leeb et al., 2008). Common forms of CM include emotional abuse, sexual abuse, physical abuse, neglect, and failure to supervise (Leeb et al., 2008). Abuse is associated with many negative health outcomes, such as child aggression, negative cognitive development and depression (Springer, Sheridan, Kuo, & Carnes, 2003; Swenson, Brown, & J.R. Lutzker, 2007), and physical health outcomes as adults, such as headaches, pelvic, back and abdominal pain (McCauley et al., 1997). Additional evidence suggests that exposure to childhood abuse is related an increased risk of alcohol and drug abuse, smoking, sexually transmitted diseases and also obesity (Diaz, Simantov, & Rickert, 2002; Felitti et al., 1998)

National expenditures for CM total more than $103 billion annually, with over $33 billion in direct costs (i.e., mental health services, hospitalization, law enforcement, child welfare services) and $70 billion in indirect costs (i.e., loss of productivity in society, juvenile delinquency, special education, adult criminal justice system (Wang &

Holton, 2007). However, these estimates only account for costs related to victims and do not include costs for treatment and intervention services for family members and perpetrators, and costs associated with indirect and long term effects of maltreatment on victims, such as risky behavioral practices, and unforeseen physical, mental and social effects (Wang & Holton, 2007). Thus, child maltreatment therefore may lead to significant negative effects on the individual and societal levels, and should be targeted and addressed to reduce negative outcomes.

Expert Recommendations to use Reduce Child Maltreatment

With the long-term health implications (and individual and societal costs of abuse), there is an increasing focus to identify effective intervention strategies. When families are identified through CPS systems and children are not removed, parents are generally referred for 'family preservation services'. Most family preservation services do not currently implement evidence-based protocols (Whitaker et al., 2008), and there is little evidence of their effectiveness (Chaffin, Bonner, & Hill, 2001). However, child welfare agencies are moving toward the implementation of evidence-based programs as part of family preservation and other services (Barth et al., 2005; Whitaker et al. 2008). Evidence-based programs can combine evidence-based practices, or skills, techniques and strategies that can be applied within specific settings for a target population (Fixsen et al, 2002). Several authors have recommended using evidence-based behavioral parent-training programs for reducing and preventing child maltreatment (Barth et al 2005; Chaffin & Friedrich, 2004; Whitaker, Lutzker, & Shelley, 2005), as there is a growing

literature that behavioral Parent-training programs can effectively change parental behaviors and reduce parental risk of CM (Wyatt Kaminski, Valle, Filene, & Boyle, 2008).

Behaviorally based parent-training programs attempt to teach parents effective child management skills (Taylor & Biglan, 1998). Behaviorally based parent-training is based on social learning principles (Chronis, Chacko, Fabiano, Wymbs, & Pelham, 2004) and includes components such as didactic instruction, modeling, skill practice, direct observation of behavior, and differential reinforcement (Serketich & Dumas, 1996). Parents may be trained to minimize neglectful behavior and increase positive interactions with children by using playing techniques, reward systems and positive feedback. In addition, parents may be taught to set and follow clear rules and consequences for their children's behaviors and actions, and to use non-coercive discipline methods. Programs such as Parent Child Interaction Therapy (PCIT) (Timmer, Urquiza, Zebell, & McGrath, 2005), Positive Parenting Program, or Triple P (Sanders, 1999) and SafeCare (Lutzker & Bigelow, 2002) are examples of recognized and effective parent-training programs which focus on the prevention and early intervention of child maltreatment among high-risk parents or among those with substantiated cases of abuse and neglect.

Several studies have shown behaviorally-based parent-training programs can be effective in addressing maltreatment and recidivism rates. For example, a recent randomized control trial of the dissemination of Triple-P was conducted in the Southeast US to assess the effect of the parent-training program on CM (Prinz, Sanders, Shapiro, Whitaker, & Lutzker, 2009). Families among 18 counties were randomly assigned Triple

P services or a control program. Results showed statistically significant reductions across three measured indicators of CM, substantiated cases of CM, child out-of-home placements, and CM injuries, suggesting the positive impact of disseminating Triple P to prevent maltreatment.

In a four-year study by Chaffin et al. (2004), researchers compared the efficiency of PCIT to a PCIT plus enhanced services (EPCIT) and a standard parenting program in preventing recidivism of child physical abuse among parents in the child welfare system. Results from the study show that PCIT was successful in reducing more than 50 percent of the recurrence rate of child physical abuse seen among parents in the standard parenting program (Chaffin et al., 2004). Likewise, in a quasi-experimental evaluation of the SafeCare parent-training program, Lutzker and colleagues (2009) found that families receiving SafeCare had a much lower rate of re-reports compared to families receiving standard family preservation services (Gershater-Molko, Lutzker, & Wesch, 2002). Such findings support the notion that behavioral parent-training programs can effectively reduce child maltreatment risk and reports. Thus, child welfare and mental health systems have been moving toward implementation of such programs.

Importance of Implementing Evidence-based Programs with Fidelity

There are many challenges when implementing evidence-based programs in community settings that must be met to achieve outcomes similar to those found in research studies. One such challenge is to achieve and maintain fidelity to the program model. Fidelity may be described as the extent to which program or intervention delivery

adheres to the procedures of the original model (Durlak & DuPre, 2008). There is clear evidence that program effectiveness is related to fidelity of implementation (Durlak & DuPre, 2008) such that the more a program is implemented as designed, the stronger the program outcomes. Therefore, program effectiveness may be compromised without consistent implementation and monitoring to ensure fidelity (Dane & Schneider, 1998).

Promoting Fidelity through Coaching

Fixsen et al (2002) describes several core, implementation components needed for high quality implementation to achieve desired outcomes. One of those core implementation components focuses on training and implementation support. The training and support model used during implementation can have a large impact on fidelity. Though many training models rely solely on workshops, workshop training and in-class practice alone are typically not be sufficient for sustained implementation with fidelity. Although didactic and workshop training are useful, and necessary to impart knowledge and initial skills to practitioners, several studies have shown that the addition of in-field coaching with corrective feedback to those learning a new practice is critical to achieve implementation with fidelity (Fixsen et al., 2002).

For example, in a meta-analysis of implementation practices in educational settings, Joyce and Showers (2002) found that rates of implementation increased from 5% obtained with workshop only trainings to 95% obtained when workshop training was followed by coaching of teachers in the classroom. Results from health settings also support that workshop training alone is insufficient for successful implementation rates.

In a meta-analysis of continuing medical education strategies assessing physician performance and healthcare outcomes, Davis et al. (1995) reported programs using didactic presentations were mostly ineffective in changing physicians' practices that were the target of continuing medical education (CME). However, delivery methods that included coaching and peer discussion, and practicing skills lead to positive results. Finally, Kelly and Kalichman (2000) focused on disseminating an HIV prevention model to community providers using three strategies: training manuals on implementation, manuals and training workshops, or manuals, workshops and consultation calls. They found that programs using the combination of manuals, workshops and personalized consultation calls were significantly associated with increased adoption of the prevention model. Because coaching (i.e., measuring fidelity and providing feedback) is a key to implementation, it is important to examine the processes by which coaching can be conducted effectively and efficiently. The current research focuses on one part of coaching: monitoring fidelity of program implementation.

Methods of Measuring Fidelity

The literature reflects observation by an independent observer as the 'gold-standard' for fidelity monitoring (Lee et al., 2008; Mowbray, Holter, Teague, & Bybee, 2003). Research indicates that objective observations (i.e., live observations, video recording observations) measuring providers' implementation of a program produce more valid fidelity measurements than self-report measures (Durlak & DuPre, 2008; Mowbray et al., 2003). For example, in a study by Lillehoj, Griffin, and Spoth (2004),

measurements from independent- observer, and provider (self-report) on adherence to a school-based intervention program were compared. Providers reported an inflated level of program adherence across all measures fidelity. Most importantly, providers' reports of fidelity were not associated with youth outcomes in comparison to objective observations of fidelity, which were significantly associated with youth outcomes (Lillehoj, Griffin, & Spoth, 2004).

Directly observing implementation may be ideal, but it is a difficult and costly method of fidelity monitoring. Direct observations are costly in personnel time and equipment, and are especially problematic when services are delivered in home settings, as is the case with most family preservation services (Lee, et al., 2008). In addition, providers may resist to direct observations (Lillehoj et al., 2004). As a result of these disadvantages in directly observing fidelity, many programs may rely on self-report measures to collect ongoing fidelity data or may not collect data at all. However, without reliable and valid collection of fidelity data, program implementation may not occur as planned, and the expected outcomes may not be achieved. A key question, therefore, is how can fidelity data be collected as efficiently as possible?

Parent-training programs have reported a variety of methods of fidelity monitoring for broad implementations. For example, in the Oregon Model Management Training (PMTO) and the Incredible Years, fidelity has been monitored by reviews of videotaped intervention sessions (Forgatch, Patterson, & DeGarmo, 2005; Jones, Daley, Hutchings, Bywater, & Eames, 2007; Webster-Stratton, Reid, & Stoolmiller, 2008). These programs use these videos to allow practitioners and others to assess progress of

trainees towards mastery of skill sets. Webster Stratton (2008) reported that demonstrating successful implementation via videotape is also essential to becoming a certified clinician for the Incredible Years program. Once certification is complete, ongoing fidelity monitoring is conducted by reviewing videotapes, evaluation forms, and attending periodic workshops. In Parent-Child Interaction Therapy (PCIT), training typically includes direct observation of a new PCIT therapist, with real-time coaching providing immediate feedback (Herschell, Calzada, Eyberg, & McNeil, 2002). While this model has been successfully used in university settings to train new PCIT practitioners, most "real-world" implementations of PCIT, in which community practitioners are trained, have relied on phone consultations , which rely on practitioners assessing their own fidelity (Funderburk, Ware, Altshuler, & Chaffin, 2008). Recently, Funderburk and colleagues (2008) have begun to examine the use of telemedicine equipment to conduct remote real-time coaching as a way to promote fidelity. Although there is a range of coaching and fidelity monitoring methods available, little research has examined their comparability. This thesis focuses on examining two methods of measuring implementation fidelity that are currently used within the SafeCare parent-training program.

SafeCare

The SafeCare model is an evidence-based behavioral, parent-training program that focuses on reducing child neglect and abuse among families at high risk of maltreatment (Lutzker & Bigelow, 2002). SafeCare is conducted in the home

environment and consists of three modules: health, home safety and parent child interactions. These modules address aspects of parenting behaviors (Lutzker & Bigelow, 2002), environmental, and healthcare risks, that are associated with child maltreatment (Lutzker & Bigelow, 2002).The PCI module focuses on improving and increasing positive interactions between parents and children. Parents are taught to take care of infants (parent infant interaction) and among toddlers and older children to manage child behavior by using positive interaction and planning skills (planned activities training). The safety module focuses on making the home environment safer and healthier for children through the removal of unsanitary materials, hazards and other harmful objects that can lead to accidents. The health module focuses on having parents assess scenarios and use role playing to identify symptoms of illness, and appropriate treatment options for a particular illness. In conjunction to the modules, problem-solving and counseling skills are also used for occasions when parental problems are not be addressed by SafeCare (Lutzker & Bigelow, 2002).

Each module follows a structured, seven-step process which includes: explaining the rationale for the behavior, demonstration of skills; practice of skills by the parent; observation and data collection of parental behavior by home visitors; positive and corrective feedback from the home visitor, additional parental demonstration of skills; and demonstration of skills to meet mastery criteria (Whitaker et al., 2008). For each module, home visitors are required to first conduct a baseline assessment, followed by several training sessions, and then follow-up assessments to record changes in parental behavior over time.

Current Fidelity Monitoring Used within SafeCare

The National SafeCare® Training and Research Center (NSTRC) was established in 2007 to provide training and research on the SafeCare model ("National SafeCare® Training and Research Center", 2010). All providers who wish to implement SafeCare must agree to the ongoing monitoring of provider's fidelity to ensure appropriate implementation. In the various implementations conducted by NSTRC, several different methods of fidelity monitoring methods have been used, including live observation in which the coach is present at the session, video recording of sessions with post-hoc review, and audio recordings of sessions with post-hoc review. These methods have varied according to the needs and desires of the implementation sites and the available funding.

Because SafeCare is conducted in the home, and the modules may require mobility throughout the home and a high degree of interaction, live observation likely provides the most accurate and complete view of home visitor fidelity. However, as noted, live observations is very costly when considering traveling time to homes and feasibility of scheduling (Self-Brown & Whitaker, 2008). Video recordings of sessions for fidelity monitoring purposes may provide a reasonable proxy for live observation, but video equipment and transfer of files causes considerable expense and inconvenience. For convenience and cost reasons, most SafeCare sites have conducted fidelity monitoring using audio recordings using small inexpensive voice recorders, whose files are easily transferred to coaches. Fidelity monitoring via audio recordings is considerably more

convenient than via video (or live), but it is unclear exactly what, if anything, is lost when using this method. That is, video and audio recordings may or may not be equivalent methods of fidelity monitoring. Understanding the equivalency of these methods provides insight on whether audio recordings capture all aspects of fidelity during session. This comparison may, therefore, provide preliminary data on whether the degree of information that may be lost through audio justifies the use of one method over another when considering costs, resources, and time.

Purpose of this Study

As evidence-based programs are implemented in real world settings, there is a strong need to effectively and efficiently monitor fidelity in order to maintain the expected effects. Few published studies have examined differences methods of fidelity monitoring. This study compares methods of assessing fidelity or adherence within an evidence-based parent-training model (SafeCare®) as implemented in Denver, Colorado. Specifically, analyses compare fidelity assessed via video recordings to fidelity assessed via audio recordings. Results from this study will provide information on aspects of fidelity that may be lost using audio recordings alone. The purpose of this study is to add to the body of research focusing on methods of measuring fidelity among parent-training evidence-based programs.

Research Questions

Specifically, this study will focus on comparing methods of measuring fidelity for

SafeCare by answering the following questions:

(1) What is the level of agreement between video and audio recordings of SafeCare sessions across prescribed items in fidelity checklists?

(2) Will the level of agreement vary by fidelity constructs as summarized in the SafeCare fidelity checklist?

Chapter II
METHODS AND PROCEDURES

This study, (Protocol Number: H10219), was approved by the Georgia State University Institutional Review Board (IRB) in December 2009.

Description of Data and Data Source

For this study, the data were obtained from 31 video recordings made by SafeCare home visitors of live sessions in Denver, Colorado. Video recordings are used for ongoing fidelity monitoring at the Denver implementation site and are stored at the Colorado Judicial Department. Video recordings of family visits were of the three SafeCare modules: home safety, health, and parent child/infant interactions. Of these recordings, 23 were health, 7 were PCI, and 1 was safety. However, three health recordings were mislabeled, missing, or incorrectly recorded respectively. In addition, a combined session of health and safety was also removed from the data. The final set of sessions included 19 health sessions, 6 PCI sessions, and no safety sessions. Recordings from each of the modules varied by session type, which includes assessment, training and end of module. The number of assessment, training and end of module sessions per module-type are described in Table# 1.

Eleven home visitors recorded in these sessions consented to participate in this study. All data used by researchers in this study were delinked from any identifying information about home visitors to ensure privacy of the participants (though they were clearly identifiable on the video tapes, none of the coders had any contact with home

visitors).

Home Visitor Demographics

Demographic information was available for 11 of 11 consenting HVs. Of the HVs, there was 1 African American male and 10 Caucasian females. Ages of participants ranged from 27 years to 52 years. Four participants were below 30 years and one participant was over 50 yrs, six participants were between ages 31-36 years. Of the participants eight obtained their bachelor's degree and three had obtained their masters degree

Study Procedure

Three trained graduate (i.e., Master of Public Health) research assistants from the National SafeCare Training and Research Center (NSTRC) in Atlanta, Georgia were selected to code the recordings obtained from the Denver implementation site (A detailed description of the training process can be found in the section, "Training Process for NSTRC Coders"). Coders are referred to as Coder 1, Coder 2, and Coder 3. In addition, another trained coder from NSTRC was selected as a reliability coder to score recordings from each coder.

All video recordings from the Denver implementation site were de-linked from home-visitors and assigned to a number from 1 to 31 (this was done before any recordings were removed from the data set for the reasons previously described). Because most of the videos were of the health module, videos were then separated in two groups: health videos, and PCI/ safety videos. Using block randomization, videos of a group

were assigned within blocks of six. Randomization to blocks was performed using a random number generator in Microsoft Excel. Each block of six consisted of the six possible permutations for coders to score a video or audio file recording of a particular session. Permutations are shown in Table #2. Coders were randomly assigned to score video or audio recording of the same session within a block. The reliability coder was assigned to score a total of 12 of the sessions, four from each of Coder 1, 2, and 3. Of the four, two were coded via audio and two were coded via video for each coder.

All Coders scored these recordings using two standard SafeCare fidelity checklists used in all implementations: (1) assessment, and, (2) training. Each standard checklist has slightly different items and thus will be analyzed separately.

Training Process for NSTRC Coders

The training process for research coders included a combination of didactic instructions and practice coding for audio recordings of home visitor sessions with SafeCare families. Coders were required to read a fidelity scoring manual and attend weekly meetings to review, discuss and score audio files. The training manual provided fidelity checklists for sessions and detailed instructions for scoring the audio files. Audio recordings covered the three modules of the SafeCare model (i.e., safety, health, parent child/infant interaction). Coders scored at least three audio recordings from each module and at least one assessment, training, and end of module session.

In addition, one coder attended a three-hour coach-training session. Coaches conduct ongoing fidelity monitoring in SafeCare implementations. Approximately 40

hours of training over a 6 month period were needed to achieve the minimal required inter-rater reliability of 85% across sessions. During the training period, weekly meetings were held with NSTRC's research coordinator to discuss and review the audio recordings sessions and the scored checklists. NSTRC trainers trained with these coders for approximately three of these weekly meetings. Initial meetings consisted of coding two to four sessions collectively. Subsequent coding was conducted individually.

Description of Fidelity Checklists

Coders used two of the three generic SafeCare fidelity checklists that are used in routine SafeCare implementation to score all sessions (assessment, training, and end of module checklists). Each of the three checklists contains a different number of items subdivided into categories (See Appendix A for checklists). The assessment fidelity checklists include 28 items divided into 9 categories; the training checklists consist of 29 items in 8 categories; the end of module checklists include 28 items in 7 categories. For all checklists, items are scored on a '+'/ '–' / 'n/a' scale. Definitions for scoring items vary by module and type of session and are found in the coding manual. A '+' score indicates that home visitors used the appropriate technique in a session. A '-' score indicates that home visitors failed to demonstrate a technique. An 'N/A' score is typically used to indicate either that the item was not relevant for the session, or that the coder was uncertain as to whether item was completed.

For this study, coders were instructed to further divide items that would have been coded as N/A into two 2 categories: not applicable (N/A) and not clear (N/C). The

revised N/A score indicates that the item was not relevant to the session. The N/C score indicates that coders were unable to ascertain whether home visitors demonstrated a technique based on limitations due to technology such as audio interference, video camera placement in the home, incomplete recordings.

Classification of Items in Fidelity Checklists

For research and analyses purposes, all items from the assessment and training fidelity checklists were classified into a process or content fidelity category. Fidelity checklists can be found in Appendix A. Process fidelity items indicated areas which involve communication skills and rapport with the family. Content fidelity items were those judged to be the critical content of the SafeCare model, as indicated on the checklist.

Three NSTRC staff and one faculty independently reviewed the items on the assessment, training, and end of module fidelity checklists for classification as content, process, both (an overlap of the two) or unclassified. Items were classified as content (C) or process (P) if there were at least three responses in agreement (e.g. C, C, C, P = C). Any items perceived as an overlap (O) by staff were considered as being both content and process oriented. In cases when responses did not match, the item was not classified (e.g., P, C, P, C= not classified). In addition, when at least 2 responses were in agreement and the remaining response(s) was an overlap, the final classification would take the type of response in agreement (e.g. P, P, O, O = P).

Study Measures

The primary measures to be analyzed in this study were the fidelity scores by audio and video coders. Responses were coded in the following categorical format: 1= "+", 2= "-", 3= "N/A", 4= "N/C". All responses were included in analyses.

Statistical Analyses

Data in this study were entered and analyzed using the Statistical Package for the Social Sciences (SPSS)[R] version 17.0. Descriptive statistics were produced using frequency tables. Frequency tables were obtained to determine the representation of categorical variables in the data. All data were included in the analyses. There were no missing data within the dataset.

Reliability Analyses

Reliability was assessed by computing percent agreement for 11 sessions between the coders' ratings with the reliability coder. The reliability coder scored the same session using the same medium (i.e., video or audio) as the coder. Reliability coefficients were computed for overall sessions, by module (health vs. PCI/PII), by type of session (assessment vs. training), by medium (video vs. audio) and by coder (1, 2, 3). Reliability analyzed across three agreement variables (Agree1, Agree 2, and Agree 3) described below, after which primary analyses are described.

Percent Agreement Analyses

The responses to each of the audio and video coded sessions were compared, and for each session, a set of binary (yes/no) variables were created to represent the agreement between two coders scoring an item for a particular session. Therefore, 28 new variables were created for assessment sessions, and 29 new variables were created for training sessions. New variables (labeled Agree 1) were coded such that: 1= agreement on an item between rater using audio and rater using video and 0= disagreement on an item between audio rater and video rater.

A second set of agreement variables (labeled Agree 2) were also created to examine the level of agreement among coders when the N/A and N/C response options were collapsed into a single category. Collapsing the response option was done to determine if the addition of the N/C option to the original fidelity checklist would affect percent agreement between coders. In addition, collapsing the column would also indicate percent agreement using the unmodified checklists.

In addition, a third set of agreement measures (labeled Agree 3) was created in which agreement was noted only if raters scored an item as occurring (+) or not occurring (-). If either coder scored an item as an N/A or N/C, the item was not coded as an agreement or disagreement. This new set of variables was computed to examine level of agreement between audio and video recordings among items that clearly occurred based on coder scored. Therefore, percent agreement may be higher for this third set of agreement ratings in comparison to the other sets of ratings mentioned above.

To determine the level of agreement by session type, cases were sorted by assessment or training because the two types of sessions had slightly different fidelity checklists. Cases were then sorted by module type: health or PCI//PII. The three measures of agreement (agree 1, and agree 2, agree 3) were computed separately by module type and session type: health assessment sessions (N =7); health training sessions N =12); PCI/PII training sessions (N = 6); and total training sessions (N = 18). No assessment sessions for PCI/PII were included in the analyses; therefore, total assessment sessions were taken for the health module only.

Table #1

Number of Videos Stratified by Module and Session Type

		Module			Total
		Safety	Health	PCI/PII	
Session type	Assessment	0	7	0	7
	Training	0	12	6	18
	End of Module	0	0	0,0	0
Total		0	19	6	25

Chapter III
RESULTS

Fidelity Checklist Item Classification

Classification of assessment, training and end of module fidelity items as process, content or NA can be found in Tables' 5 and 6. Using the four NSTRC staff coding training checklists, there was agreement that 11 of 29 items assessed SafeCare content, 11 assessed process, five items were unclassified and two items were considered as overlap between process and content. Among 28 items from the assessment checklists, 12 items assessed content, 11 assessed process and 5 were not classified.

Descriptive Statistics

Descriptive and frequency distribution for the recorded SafeCare sessions and NSTRC coding are seen in Tables' 2 and 3. Of the 31 sessions received, 25 sessions were included in this study. The number PII/PCI assessment sessions was determined to be too small for inclusion in this study, therefore, only health assessment sessions were included. The majority of sessions were of the health module (76%) and of training sessions (72%). Coders were originally assigned to score similar number of sessions. However, exclusion of some sessions led to coders scoring an uneven number of sessions. Of the 25 recordings used for this study, Coder1 scored seven of these as audio and eight as video. Coder 2 scored nine recordings as audio and six as video. Coder 3 scored 9 recordings as audio and 11 as video.

Reliability Analyses

Reliability Coding

A summary of mean percent agreement for reliability analyses can be found in Table 4. Of the 11 sessions used for reliability analyses, 6 were scored by video. All audio scored sessions were of assessment sessions. Seven sessions were of training sessions, of which, six were video.

Overall Reliability

Agreement across all 11 sessions for Agree 1 averaged 81.15%. Percent agreement overall for Agree 2 was similar to Agree 1 (83.91%) but increased to 93.96% for Agree 3.

Reliability by Session Type

Average agreement was slightly higher for assessment sessions (Agree 1=84.82%; Agree 2= 85.71%) than training sessions (Agree 1= 78.74%; Agree 2= 83.81) for Agree 1 and Agree 2. Percent agreement increased minimally from Agree 1 to Agree 2 for assessment sessions and increased by approximately 5 percent among training sessions. Agreement for Agree 3 increased to 90.58% for assessment sessions and 96.74% for training sessions.

Reliability by Module

Average agreement among health and PCI/PII sessions remained similar to each other across Agree 1, 2 and 3. Agreement increased minimally among health sessions from agree 1 to 2 (Agree1 =81-15; Agree 2= 83.91%) and did not change among PCI PII sessions (81.61%). Agreement for health and PCI/PII module sessions increased by at

least 10 percent for Agree 3.

Reliability by Method of Monitoring

Like reliability by module type, average agreements for audio and video sessions were similar across Agree 1, 2 and 3. Agreement levels increased by approximately 5 percent from Agree 1 to Agree 2 among audio sessions (Agree 1= 80.69%; agree 2= 86.21%) and did not change for video sessions (81.61%). However, both sessions increased by at least 7 percent in Agree 3.

Reliability by Coder

Coder 1 had the highest percent agreement for Agree 1 and 2 (Agree 1 and 2= 87.72 %). Coders 1 and 2 had levels of agreement greater than 85 percent across Agree 1 and 2. However, coder 3 had the lowest level of agreement for these variables (Agree 1= 72.41%; Agree2=77.59), but the highest level of agreement for Agree 3 (95.56%). All coders had levels of agreement greater than 90% for Agree 3.

Primary Analyses- Agreement between Audio and Video Coding

A summary of mean percent agreement across the three agreement variables assessment and training sessions can be found in Table 7.

Assessment Sessions

As noted earlier, all included assessment sessions were of the health module, therefore, no PCI/PII sessions are seen among the results. The overall agreement across the 7 health assessment sessions for Agree 1 averaged to 72.45%. Agreement for Agree

2 was nearly identical (73.47%) to Agree 1, but increased to 91.14% for Agree 3. Thus, it appears that most disagreements between audio and video were due to one of the coders using the N/A or N/C code.

Agreement for individual items of the health assessment sessions is shown in Table 5. Under Agree 1 criteria, full (100%) agreement was obtained for nine items, five of which were content, two of which were process, and two of which were unclassified. Zero percent agreement was seen for three items (i.e., all process: "sits facing client" and "maintains an open posture", "has good eye contact"). Of the 11 total process items, agreement ranged between 0 to 100%. Of the 12 total content items, agreement ranged from 28.6% to 100%. Overall, average agreement among the process items (45.45%) was notably lower than average agreement among content items (80.95%) for Agree 1. Average agreement for process items increased by eight percent for Agree 2, whereas average agreement for content remained the same. Percent agreement for Agree 2 among individual items remained the same, with the exception of “Exchanges an appropriate initial greeting" (a process item). Under Agree 3 agreement criteria, agreement for many process and content items improved dramatically (e.g., “Assessment checklists", "Supplies", "Exchanges an appropriate greeting", "Uses summarizing statements"). Therefore, average agreement for Agree 3 among both process and content increased dramatically (89.17% and 91.69% respectively).

Training Sessions

Table 6 shows agreement for the 18 training sessions, health training sessions (n =

12), and PCI/PII training sessions (n = 6) for the three measures of agreement. Overall, agreement patterns among training sessions closely matched those for assessment sessions. Under Agree 1 criteria, average agreement was 71.46% for all items, 61.62% for process items, and 79.29% for content items. Under Agree 2 criteria, agreement changed minimally, and under Agree 3 criteria, agreement rose to greater than 90% for all items, process items, and content items.

Training sessions for two different modules were coded, health and PCI/PII, and the percent agreements for each module is shown in Table 6. Agreement was somewhat higher for PCI/PII sessions than health sessions for all items (75.86% vs. 69.25% under Agree 1), for process items (66.67% vs. 59.09% under Agree 1), and for content items (83.33% vs. 72.27% under Agree 1). Both health and PCI/PII showed similar patterns of change in agreement across the criteria (i.e., Agree 1, Agree 2, and Agree 3).Specifically, average agreement between audio and video coder rose slightly under Agree 2 and dramatically under Agree 3 for both health and PCI/PII.

Because the pattern of results was similar for health and PCI/PII sessions, the discussion of results of agreement for individual items will focus on agreement across the two modules. Item-level agreements for all 17 training sessions are displayed in the first column of Table 6 (item-level agreements for health and for PCI/PII are displayed in the second and third columns in the same Table). Under Agree 1 criteria, agreement among the 28 individual items of the training sessions ranged from 0% to 100%. Full (100%) agreement was found for eight items, including five process items, and three content items. Zero percent agreement was seen for two process items, "Sits facing client" and

"Maintains an open posture". Average agreement among the process items was notably lower than average agreement among content items for Agree 1. Of the 11 content items, percent agreement ranged from 33% to 100%, with an average agreement of 79.29%. Of the 11 process items, percent agreement ranged from 0 to 100%, with an average agreement of 61.62%. Of the unclassified items, percent agreement ranged from 55.56% to 88.89%.

Under Agree 3 criteria, there were large increases in more than half of all items. Specifically, average agreement for Agree 3 among both process and content increased dramatically (92.02% and 91.52% respectively). Large improvements in Agree 3 among both assessment and training sessions similarly suggest that many discrepancies between audio and video coders were due to the use of N/A or N/C ratings.

Table #2

Number of Videos Included in the Study Stratified by Module and Session Type

		Total		
		Health	PCI/PII	Total
		#	#	#
Session type	Assessment	7	0	7
	Training	12	6	18
Total		19	6	25

Table #3

Frequency Distribution of Video and Audio Rated Recordings by All Coders

Coder I.D.	No. Videos Rated	% of Total Video Ratings	No. Audios Rated	% of Total Audio Ratings	Total Rated Recordings
Coder 1	8	32	7	28	15
Coder 2	6	24	9	36	15
Coder 3	11	44	9	36	21

Table #4

Summary of Reliability Mean Agreement for All Sessions, by Session Type, by Module, by Method, and by Coder

	Agree 1	Agree 2	Agree 3
	% Agreement	% Agreement	% Agreement
Overall	81.15	83.91	93.96
By session type			
Assessment	84.82	85.71	90.58
Training	78.74	83.91	96.74
By Module			
Health	81.15	83.91	93.33
PCI/PII	81.61	61.61	94.87
By Method			
Audio	80.69	86.21	93.46
Video	81.61	81.61	94.44
By Coder			
Coder 1	87.72	87.72	92.86

Coder 2	85.09	85.96	92.71
Coder 3	72.41	77.59	95.56

Table #5

Percent Agreement between Audio and Video Raters for 28 Items on the Assessment Checklist across 7 Health Assessment sessions

Items on Assessment Fidelity Checklist	% Agreement		
	Agree 1*	Agree 2**	Agree 3***
Assessment checklists (c)	71.4	71.4	100.0
Supplies (c)	28.6	28.6	100.0
Other materials (c)	100.0	100.0	-
Exchanges an appropriate initial greeting (p)	42.9	71.4	100.0
States goals for the session (c)	100.0	100.0	100.0
Sits facing client (p)	0	0	-
Communicates empathy, warmth, understanding (p)	100.0	100.0	100.0
Maintains an open posture (p)	0	0	-
Has good eye contact (p)	0	0	-
Uses words/expressions (e.g., "uh-huh") to encourage the parent to talk (p)	85.7	85.7	100.0
Uses open-ended questions (p)	100.0	100.0	100.0
Uses reflecting statements (p)	71.4	71.4	83.3

Uses summarizing statements (p)	57.1	57.1	80
Module overview (c)	85.7	85.7	100.0
Session overview (c)	100.0	100.0	100.0
Explains the purpose of the assessment (c)	100.0	100.0	100.0
Explains the process of the assessment (c)	71.4	71.4	71.4
Assesses the required number of activities, rooms, scenarios (c)	100.0	100.0	100.0
Assesses the required variety of activities, rooms, scenarios (c)	85.7	85.7	85.7
Completes the necessary forms (c)	57.1	57.1	80.0
Provides general, positive feedback about the assessments (n)	85.7	85.7	85.7
Encourages the parent to ask questions and express concerns (p)	57.1	57.1	66.7
Responds to parent questions and concerns (p)	85.7	85.7	83.3
Uses problem solving approaches as appropriate (n)	100.0	100.0	100.0
Summarizes the session (c)	71.4	71.4	71.4
Asks for and answers parent questions (n)	100.0	100.0	100.0
Gives general positive feedback (n)	85.7	85.7	85.7
Schedules meeting date/time for next week	85.7	85.7	100.0

Average Percent Agreement for Process Items (SD)	45.45 (52.22)	57.14 (39.38)	89.17 (12.69)
Average Percent Agreement for Content Items (SD)	72.62 (31.33)	72.62 (31.33)	90.86 (12.48)
Average Percent Agreement for All Items (SD)	72.45 (31.81)	73.47 (31.28)	91.14 (11.29)

C = content, P = process, N = not classified

*Agree 1- Percent agreement calculated based on four possible codes: +. -. N/A, N/C
**Agree 2- Percent Agreement calculated based on 3 possible codes: +, -, N/A (i.e., N/A in Agree 2 is the combined N/A and N/C scores from Agree 1)

Agree 3-Percent agreement calculated based on two codes only: +, -. The N/A and N/C options were not included in this calculation.

Table #6

Fidelity Classification and Percent Agreement between Audio and Video Raters for 29 Training Fidelity Checklist Items across 18 Recorded Training Sessions

Items on Training Fidelity Checklist	% Agreement All Training			% Agreement Health Training			% Agreement PCI /PII Training		
	Agree 1*	Agree 2**	Agree 3***	Agree 1*	Agree 2**	Agree 3***	Agree 1*	Agree 2**	Agree 3***
Checklists-HV Versions (c)	88.9	88.9	100.0	91.7	91.7	100.0	83.3	83.3	100.0
Training materials for families (e.g., Checklists-Parent Versions) (c)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Supplies (c)	33.3	55.6	100.0	25.0.	50.0	100.0	50.0	66.7	100.0
Other materials (c)	83.3	88.9	100.0	75.0.	83.3	100.0	100.0	100.0	-
Exchanges an appropriate initial greeting (p)	66.7	66.7	100.0	58.3	58.3	100.0	83.3	83.3	100.0
States goals for the session (c)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Sits facing client (p)	0	0	-	0	0	-	0	0	-
Communicates empathy, warmth, understanding (p)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Maintains an open posture (p)	0	0	-	0	0	-	0	0	-

Has good eye contact (p)	11.1	11.1	-	16.7	16.7	-	0	0	-
Uses words/expressions (e.g., "uh-huh") to encourage the parent to talk (p)	88.9	88.9	88.9	83.3	83.3	83.3	100.0	100.0	100.0
Uses open-ended questions (p)	100.0	100.0	100.0	100.0	100.0	100.0	66.7	100.0	100.0
Uses reflecting statements (p)	66.7	66.7	75.0	66.7	66.7	72.7	66.7	66.7	80.0
Uses summarizing statements (p)	83.3	83.3	92.3	83.3	83.3	87.5	83.3	83.3	100.0
Conducts assessments as indicated in the outline (c)	83.3	83.3	93.8	83.3	83.3	90.9	83.3	83.3	100.0
Explains the purpose of the assessments (n)	72.2	72.2	80.0	66.7	66.7	70.0	83.3	83.3	100.0
Explains the process of the assessments (n)	88.9	88.9	94.1	91.7	91.7	91.7	83.3	83.3	100.0
Uses the appropriate material (c)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Models steps and behaviors (c)	72.2	77.8	72.2	66.7	66.7	66.7	83.3	83.3	100.0
Has parent practice an appropriate number of times (c)	72.2	72.2	92.9	75.0	75.0	90.0	66.7	66.7	100.0
Balances explain vs. modeling behaviors and steps (c)	61.1	61.1	66.7	50.0	50.0	54.5	83.3	83.3	100.0
Provides general, positive feedback (n)	88.9	88.9	100.0	83.3	83.3	100.0	100.0	100.0	100.0
Provides specific, corrective feedback (c)	77.8	77.8	81.3	83.3	83.3	90.0	66.7	66.7	66.7
Encourages the parent to ask questions and express concerns	94.4	94.4	100.0	91.7	91.7	100.0	100.0	100.0	100.0

Responds to parent questions and concerns (o)	77.8	77.8	100.0	83.3	83.3	100.0	66.7	66.7	100.0
Uses problem solving approaches as appropriate	55.6	55.6	100.0	58.3	58.3	100.0	50.0	50.0	-
Summarizes the session (p)	66.7	72.2	80.0	50.0	58.3	66.7	100.0	100.0	100.0
Gives general positive feedback (n)	66.7	66.7	85.5	58.3	58.3	87.5	83.3	83.3	83.3
Schedules meeting date/time for next week (n)	72.2	72.2	92.9	66.7	66.7	100.0	83.3	83.3	83.3
Average Agreement: Process Items	61.62 (39.31)	62.12 (39.42)	92.02 (9.99)	59.09 (38.09)	59.85 (37.97)	88.78 (13.53)	66.67 (44.10)	66.67 (44.10)	97.50 (7.07)
Average Agreement: Content Items	79.29 (19.89)	82.32 (15.28)	91.52 (12.37)	72.27 (23.30)	80.30 (18.32)	90.19 (15.50)	83.33 (16.67)	86.36 (14.56)	95.00 (11.25)
Average Agreement: All Items	71.46 (28.62)	72.80 (27.51)	92.14 (10.23)	69.25 (28.68)	70.69 (28.62)	90.44 (13.42)	75.86 (30.08)	77.01 (30.02)	95.69 (9.09)

C = Content, P = Process, N = neither

*Agree 1- Percent agreement calculated based on four possible codes: +. -. N/A, N/C
**Agree 2- Percent Agreement calculated based on 3 possible codes: +, -, N/A (i.e., N/A in Agree 2 is the combined N/A and N/C scores from Agree 1)

***Agree 3-Percent agreement calculated based on two codes only: +, -. The N/A and N/C options were not included in this calculation.

Table #7

Summary of Three Sets of Mean Percentage of Agreement by Session and Module Type

Session Type	Agree 1 % agreement	Agree 2 % agreement	Agree 3 % agreement
Health Assessments (note: only health assessments were available)			
All items	72.45	73.47	91.14
Process	45.45	57.14	89.17
Content	80.95	80.95	91.69
All Training sessions			
All items	71.46	72.80	92.14
Process	61.62	62.12	92.02
Content	79.29	82.32	91.52
Health Training sessions			
All items	69.25	70.69	90.44
Process	59.09	59.85	88.78

Content	72.27	80.80	90.19
PCI/PII Training sessions			
All items	75.86	77.01	95.69
Process	66.67	66.67	97.50
Content	83.33	86.26	95.00

Table #8

Items among Assessment and Training Sessions with Low Audio/ Video Coder Agreement in Comparison to Reliability Scores

Session Type	Agree 1		Agree 2		Agree 3	
Assessment Sessions	Audio/Video	Reliability	Audio/Video	Reliability	Audio/Video	Reliability
Supplies	28.6	75.0	28.6	75.0	100.0	100.0
Exchanges an appropriate initial greeting	42.9	50.0	71.4	75.0	100.0	100.0
Uses summarizing statements	57.1	25.0	57.1	75.0	80.0	75.0
Encourages the parent to ask questions and express concerns	57.1	75.0	57.1	75.0	66.7	100.0
Completes the necessary forms	57.1	100.0	57.1	100.0	80.0	100.0
Assessment Checklist	71.4	25.0	71.4	25.0	100.0	100.0
Uses reflecting statements	71.4	75.0	71.4	75.0	83.3	75.0
Explains process of the assessment	71.4	100.0	71.4	100.0	71.4	100.0
Summarizes the session	71.4	100.0	71.4	100.0	71.4	100.0
Training Sessions						
Supplies	33.3	28.6	55.6	42.9	100.0	100.0
Uses problem solving approaches as appropriate	55.6	57.1	55.6	57.1	100.0	-

Balances explain vs. modeling behaviors and steps	61.1	57.1	61.1	57.1	66.7	100.0
Gives general positive feedback	66.7	57.1	66.7	57.1	85.5	66.67
Summarizes the session	66.7	71.4	72.2	85.7	80.0	100.0
Exchanges an appropriate initial greeting	66.7	85.7	66.7	85.7	100.0	100.0
Uses reflecting statements	66.7	100.0	66.7	100.0	75.0	100.0
Schedules meeting date/time for next week	72.2	57.1	72.2	57.1	92.9	66.67
Explains the purpose of the assessments	72.2	71.4	72.2	71.4	80.0	100.0
Models steps and behaviors	72.2	71.4	72.2	71.4	77.8	80.0
Has parents practice an appropriate number of times	72.2	85.7	72.2	85.7	92.9	100.0

Chapter IV
DISCUSSION

The purpose of this research was to compare two methods, the use of audio recordings and the use of video recordings, to rate fidelity to the SafeCare model. Understanding whether the two methods are equivalent and where differences may lie will provide insight on what aspects of fidelity are lost by using audio recordings, which are considerably cheaper and more efficient to use. This information can then be used to determine whether the loss of specific items justifies using one method over another which affect may cost, feasibility, and future fidelity monitoring.

To answer the questions as to whether the two methods are equivalent and what fidelity aspects differ, this research focused on comparing ratings of a trained coder, who scored a session by watching a video, to ratings of another trained coder who scored the same session using only the audio portion of the recording. Additionally, a reliability coder scored a subset of sessions using the same method (video or audio) as the coder to verify overall reliability. Assessing reliability or agreement between coders using the same method is important because one cannot expect cross-method agreement (i.e., audio vs. video) to exceed within-method agreement (audio vs. audio or video vs. video).

Reliability Analyses

Across sessions included in the reliability analyses, and using the most stringent criteria for agreement (Agree 1), overall agreement between coders rating the same session using the same method (i.e., audio only, or video only) was 81.15%. When

criteria for agreement were relaxed to resemble the original SafeCare checklists, average agreement increased minimally to 83.9%. Relaxing the criteria to eliminate all N/A and N/C responses led to a considerable increase in reliability across all sessions. However, eliminating N/A responses is not a realistic scenario for implementation. These trends in increments in average of agreement were seen across modules, method of monitoring, session type and by coder. These results may show that discrepancies in reliability were mainly due in part to coders scoring items as N/C or N/A. Because of this pattern of a slight increase in Agree 2, and a large increase in Agree 3, this discussion will focus on primarily on the Agree 1 results.

Average reliability was fairly consistent across modules, method of monitoring fidelity, and session type. Reliability was somewhat higher among assessment sessions (84.8%) than among training sessions (78.7%). Percent agreement between the reliability coder and the three coders showed some variation with agreement between the reliability coder and Coder 3 being considerably lower than Coders 1 and 2. However, reliability analyses with Coder 3 were limited to only three sessions in comparison to four sessions for Coders 1 and 2, which may bias the results of the reliability analyses.

Overall, inter-rater reliability was not as high as would be desired. This may be due to the fact that the reliability coder was a NSTRC Training specialist while the coders were graduate research assistants, and thus reliability may have been somewhat affected negatively by the different roles of the coders and reliability coder. In addition, coder 3 had the least amount of training of the three coders, which may have led to a decrease in inter-rater reliability between the reliability coder and coder 3, and therefore decreasing

overall reliability. Even so, the overall number of scored recordings for reliability analyses is still low and should be considered when considering the results.

Inspection of item-by-item percent agreement for Agree 1, not presented in the results, shows that large discrepancies were seen among all sessions for common items including "supplies', "other materials", " uses reflecting statements", "use of summarizing statements" and "responds to parent questions and concerns". Low agreement levels for the latter three times were still present even after N/A responses were removed, suggesting that coders may need further training to better understand what to look for when scorings these particular items.

Main Analyses

Results from the main study show that overall percent agreement between video and audio coders for training sessions was 72.45% under the most stringent criteria (Agree 1). In other words, audio coders' ratings of sessions matched those of video coders 72.45% of the time. When the criteria for agreement were loosened, agreements increased as would be expected. The very high rates of agreement for Agree 3, which did not include responses of not applicable or not able to code, suggest that many disagreements were due to specific items being coded either N/A or N/C, presumably by audio coders. As with the reliability analyses, because of the pattern in increments in Agree 2 and 3, this discussion will focus on mostly on the Agree 1 results.

Agreement was considerably lower for process items than content items for both assessment and training sessions. Full disagreement for process items as a function of

method of scoring (audio vs. video) would be expected based on the required visuals needed to code specific items including "sits facing client", "maintains open posture" and "has good eye contact." This was largely the case, except for a 16.7% agreement was found for "has good eye contact" among health recordings. However, removing N/A responses also removed this small agreement. Visual inspection of the raw data showed that two coders scored an N/C for this item, which implies that the video coder was unable to detect this item, possibly from awkward placement of the video recorder at the training session.

Low levels of agreement under Agree 1 and 2 for "supplies" may imply that video coders may have seen the presence of relevant items during a session (e.g., PCI and PII sessions may require the use of toys for play activities; health sessions may require modeling with items such as thermometers). These and other items are sometimes brought by home visitors during a session and can be captured on video recordings, but may not be detected by audio coders unless their use is explicitly mentioned by the home visitor.

Other items showing low agreement cannot be easily explained by a simple lack of visuals, however. For example, low agreement for items including, "Uses summarizing statements", "uses reflecting statements", "uses problem-solving approaches as appropriate" and "has parent practice an appropriate number of times" among all sessions do not require visuals, and may simply be due to coder differences than methods of monitoring fidelity. Agreement for these four items changed differently under the criteria of Agree 3 (e.g., reflecting statements showed only small

improvements, while problem-solving yielded perfect agreement), thus suggesting that there may be different reasons or sources of discrepancies for the items. In an effort to confirm whether these items' scores were low due to coder differences in scoring, Table #8 was created to show a side-by-side comparison of compare audio/video coder agreement to the corresponding reliability agreement for each specific item with low agreement. Results from this table show that some items with low agreement across rating method (i.e., audio vs. video) also had low reliability (e.g., "uses reflecting statements" and "uses summarizes statements"). Thus, the low level agreement across method for these items may reflect general difficulty in obtaining agreement, rather than differences caused by method of monitoring fidelity. Reliability for other items in this table increased dramatically with the removal of the not applicable response in Agree 3.

Within each module, specific items also had consistently low agreement even under Agree 3. Among health assessment sessions, such items included "Explains the process of the assessment" and 'encourages the parent to ask questions and express concerns". Another item among assessment sessions with low agreement included "Completes the necessary forms". Removing the not applicable responses did increase percent agreement over 20%. In this case, video coders may have a greater advantage in which coders can see home visitors score forms in the event that home visitors did not explicitly state their completion of forms. Audio raters would have been more likely to score an N/A or N/C for this item. However, Agree 3 did not remove all disagreement, which suggests that discrepancies still exist between coders when scoring this item.

Among health training items with low agreement across all agreement variables included "models steps and behaviors". Since there was no change in percent agreement in the Agree 3 variable, this indicates that disagreements were due to the audio and video coder disagreeing about whether the modeling of steps and behaviors was adequate or not. That is, one coder thought the home visitor's performance was adequate and the other did not. A similar pattern was found for "balances explain vs. modeling behaviors and steps", a behavior which is directly related to “models steps and behaviors”.

It was somewhat surprising that agreement was higher for PCI/PII than for health; the reverse was expected because PCI/PII training sessions requires more interactions between parent, child and home visitor that may not be captured during audio sessions as compared to health. These results may signify that home visitors were able to translate modeling through audio recordings by explicitly stating or clearly demonstrating their efforts to model a required component of training.

There were also large disagreements for coding the closing sequence for health training sessions. Overall, the results show that removing the N/A option (Agree 3) increased percent agreement by at least 15% for all items. Table 8 shows that reliability among two of the closing sequence items ("provides positive feedback" and "summarizes the session") remained low across the three agreement variables. This is another case, in which coding done via video may show a clear indication as to how a session ended. Audio raters may have been unclear as to whether a recording was stopped deliberately by a home visitor or stopped due to recording device limitations and may therefore have issued an N/A or N/C for an item.

Large coder disagreements were less common among the PCI sessions than health sessions. Among the PCI/PII modules, coders particularly disagreed for a training item "has parents practice an appropriate number of times". However, removing N/A responses led to full agreement for this item. Low agreement occurred across the three agreement variables for "provides specific, corrective feedback", indicating that differences were not due to uncertainty (N/A N/C) and coders clearly differed in their judgments for this variable.

Implications for SafeCare Fidelity Monitoring

Results from this study show that audio coders' ratings agree with many, but not all, of video coders' ratings. Visual inspection of the raw data from study shows that audio coders were more likely to score N/A and N/C than video coders. Differences based on these responses decreased overall agreement for several items. In particular, process items across all sessions had lower levels of agreement under the most restrictive criteria (Agree 1) in comparison to content items. However, once criteria were loosened, coder average agreement across all items, or process or content items exceeded 85%. Using video methods for SafeCare fidelity monitoring should remove the need to score N/A for most checklist items due to technological limitations. Therefore, video coding may allow coders to identify the presence or absence of both process and content checklist items with more accuracy than by audio means. Using video recordings, however, may be more inconvenient for home visitors who may need to transport camera recorders to sessions and set up. Additionally, recorders may become more cumbersome

when sessions require higher levels of interaction and movement around a house for PCI and safety modules respectively. Future research directed towards methods of monitoring fidelity may elucidate the extent of differences in fidelity scores. In the event that video method of monitoring fidelity has a significant impact on fidelity scores, additional research can focus on other forms of technology that may improve video recording quality and efficiency in recording a session.

An alternative option may include maintaining the use of audio recordings to monitor fidelity to the model, but use an additional means of capturing process items that may be lost through audio means. For example, current parent satisfaction surveys given to parents at the end of each module may be expanded to include other questions focusing on process items, or the parent's rapport and alliance with the home visitor. Many of the behaviors assessed by the process items (making eye contact, maintaining an open posture, using reflecting statements) may be important in building rapport between home visitor and parent, and as recipients of training services, parents may be able to report on the communication skills home visitors and their level of rapport.

Study Limitations

Several limitations should be noted within this study. First, a limited number of videotaped sessions were available for coding. As a result, not all of the SafeCare modules could be included in the study, and only the health module could be included in the analyses of assessment sessions. Further research can clarify whether findings would be similar or different for the safety module, and whether PCI and PII modules differ.

In addition, coding of recordings was conducted by graduate research assistants and not by NSTRC training staff, who typically code the sessions. Though GRAs were trained to reliability criteria, scoring of fidelity may differ a bit when it is done for research purposes, as in the current study, versus training purposes where the goal is to provide feedback to the home visitor. Thus, the agreement levels reported here are likely conservative estimates.

In some cases, poor audio and video quality may have affected the scores of fidelity items. For example, audio may have been affected by low microphone settings, interference from other sources within the home, such as television programs, or children. Among video recordings, home visitors may have positioned the recorder such that communication skills (sits facing client, makes good eye contact, etc.) and other interactions were not visible to video coders.

Future Research Aims

This study found some discrepancy in coding agreement between audio coded fidelity scoring and video coded fidelity scoring. Process items were especially difficult to capture in comparison to content items. This finding raises the question of which aspects – process or content – are more important for promoting family behavior change. It has been shown clearly that measures of implementation, such as fidelity relate to more positive outcomes. For example, Durlak & DuPre (2008) reviewed over 500 studies that used various measures of implementation including fidelity, and found a positive relationship between implementation and outcomes. However, it remains unclear what

aspects of implementation and fidelity are most important. It may be that content fidelity is associated more strongly with client/family behavioral change than process fidelity. If so, fidelity monitoring via audio recordings may be adequate in capturing fidelity to the model. However, if process fidelity shows a greater association, these results would indicate that the loss of process items through audio monitoring may negatively affect behavioral outcomes in the future.

Results from this study also suggest that additional training may be required to increase reliability among coders before consistent fidelity scores can be obtained. Future research may examine alternative methods of training coders to enhance skills when scoring SafeCare sessions. In particular, research on feasible technological training services should examine methods of increasing coder practice with pre-scored sessions using both audio and video methods.

Future research looking at a comparison of methods of monitoring fidelity should then include a larger sample size of recordings that includes more proportional numbers of sessions from each module and each session type. Focus should be placed scoring fidelity for safety and PCI/PII modules, in which fidelity may be affected by the use of video or audio methods. Video recordings of PCI/ PII and safety sessions will provide more insight on whether home visitors is modeling steps and interacting with families, and whether home visitors are monitoring rooms for hazard, respectively. Cost-effectiveness analyses should be performed to compare cost differences between video and audio recording methods to the benefits of using video or audio recordings as a

method to monitor fidelity. Results will then provide more accurate insight on the implications for SafeCare's future use of audio and video methods of measuring fidelity.

REFERENCES

Barth, R. P., Landsverk, J., Chamberlain, P., Reid, J. B., Rolls, J. A., Hurlburt, M. S., Farmer, E. M. Z., et al. (2005). Parent-Training Programs in Child Welfare Services: Planning for a More Evidence-Based Approach to Serving Biological Parents. *Research on Social Work Practice*, *15*(5), 353-371. doi:10.1177/1049731505276321

Brown, J., Cohen, P., Johnson, J. G., & Salzinger, S. (1998). A longitudinal analysis of risk factors for child maltreatment: findings of a 17-year prospective study of officially recorded and self-reported child abuse and neglect. *Child Abuse & Neglect*, *22*(11), 1065-1078. doi:10.1016/S0145-2134(98)00087-8

Chaffin, M., Bonner, B. L., & Hill, R. F. (2001). Family preservation and family support programs: child maltreatment outcomes across client risk levels and program types. *Child Abuse & Neglect*, *25*(10), 1269-1289. doi:10.1016/S0145-2134(01)00275-7

Chaffin, M., Silovsky, J. F., Funderburk, B., Valle, L. A., Brestan, E. V., Balachova, T., Jackson, S., et al. (2004). Parent-Child Interaction Therapy With Physically Abusive Parents: Efficacy for Reducing Future Abuse Reports. *Journal of Consulting and Clinical Psychology*, *72*(3), 500-510. doi:10.1037/0022-006X.72.3.500

Chronis, A. M., Chacko, A., Fabiano, G. A., Wymbs, B. T., & Pelham, J. (2004). Enhancements to the Behavioral Parent Training Paradigm for Families of Children with ADHD: Review and Future Directions. *Clinical Child & Family*

Psychology Review, *7*(1), 1-27.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23-45. doi:10.1016/S0272-7358(97)00043-3

Davis, D. A., Thomson, M. A., Oxman, A. D., & Haynes, R. B. (1995). Changing physician performance. A systematic review of the effect of continuing medical education strategies. *JAMA: The Journal of the American Medical Association*, *274*(9), 700-705.

Diaz, A., Simantov, E., & Rickert, V. I. (2002). Effect of Abuse on Health: Results of a National Survey. *Arch Pediatr Adolesc Med*, *156*(8), 811-817. doi:10.1001/archpedi.156.8.811

Durlak, J., & DuPre, E. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology*, *41*(3), 327-350. doi:10.1007/s10464-008-9165-0

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., et al. (1998). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults: The Adverse Childhood Experiences (ACE) Study. *American Journal of Preventive Medicine*, *14*(4), 245-258. doi:10.1016/S0749-3797(98)00017-8

Fixsen, D.l., Naoom, S.F., Blasé, K.A., Friedman, R.M. & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tampa, FL: University of

South Florida, The Louis de la Parte Florida Mental Health Institute, Department of Child & Family Studies.

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. *Behavior therapy*, *36*(1), 3-13.

Funderburk, B. W., Ware, L. M., Altshuler, E., & Chaffin, M. (2008). Use and Feasibility of Telemedicine Technology in the Dissemination of Parent-Child Interaction Therapy. *Child Maltreatment*, *13*(4), 377-382. doi:10.1177/1077559508321483

Gershater-Molko, R. M., Lutzker, J. R., & Wesch, D. (2002). Using recidivism data to evaluate project safecare: teaching bonding, safety, and health care skills to parents. *Child Maltreatment*, *7*(3), 277-285.

Herschell, A. D., Calzada, E. J., Eyberg, S. M., & McNeil, C. B. (2002). Parent-child interaction therapy: New directions in research. *Cognitive and Behavioral Practice*, *9*(1), 9-16. doi:10.1016/S1077-7229(02)80034-7

Jones, K., Daley, D., Hutchings, J., Bywater, T., & Eames, C. (2007). Efficacy of the Incredible Years Basic parent training programme as an early intervention for children with conduct problems and ADHD. *Child: Care, Health and Development*, *33*(6), 749-756. doi:10.1111/j.1365-2214.2007.00747.x

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development*. ASCD.

Kelly, J. A., & Kalichman, S. C. (2002). Behavioral research in HIV/AIDS primary and secondary prevention: recent advances and future directions. *Journal of*

Consulting and Clinical Psychology, *70*(3), 626-639.

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. *Prevention Science*, *9*(3), 215-229. doi:10.1007/s11121-008-0097-6

Leeb, R., Paulozzi, L., Melanson, C., Simon, T., & Arias, I. (2008, January). Child Maltreatment Surveillance: Uniform Definitions for Public Health and Recommended Data Elements Version 1.0. National Center for Injury Prevention and Control, CDC.

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program Provider and Observer Ratings of School-Based Preventive Intervention Implementation: Agreement and Relation to Youth Outcomes. *Health Educ Behav*, *31*(2), 242-257. doi:10.1177/1090198103260514

Lutzker, J., & Bigelow, K. (2002). *Reducing Child Maltreatment: A guidebook for parent services*. New York, NY: The Guilford Press.

McCauley, J., Kern, D. E., Kolodner, K., Dill, L., Schroeder, A. F., DeChant, H. K., Ryden, J., et al. (1997). Clinical characteristics of women with a history of childhood abuse: unhealed wounds. *JAMA: The Journal of the American Medical Association*, *277*(17), 1362-1368.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation*, *24*(3), 315-340. doi:10.1177/109821400302400303

National SafeCare® Training and Research Center. (2010). *Welcome to the National SafeCare® Training and Research Center*. Retrieved April 4, 2010, from http://chhs.gsu.edu/safecare/index.asp

Prinz, R., Sanders, M., Shapiro, C., Whitaker, D., & Lutzker, J. (2009). Population-Based Prevention of Child Maltreatment: The U.S. Triple P System Population Trial. *Prevention Science*, *10*(1), 1-12. doi:10.1007/s11121-009-0123-3

Sanders, M. R. (1999). Triple P-Positive Parenting Program: Towards an Empirically Validated Multilevel Parenting and Family Support Strategy for the Prevention of Behavior and Emotional Problems in Children. *Clinical Child and Family Psychology Review*, *2*(2), 71-90. doi:10.1023/A:1021843613840

Schumacher, J. A., Slep, A. M. S., & Heyman, R. E. (2001). Risk factors for child neglect. *Aggression and Violent Behavior*, *6*(2-3), 231-254. doi:10.1016/S1359-1789(00)00024-0

Sedlak, A., Mettenburg, J., Basena, M., Petta, I., McPherson, K., Green, A., & Li, S. (2010). *Fourth National Incidence Study of Child Abuse and Neglect (NIS-4): Report to Congress*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.

Self-Brown, S., & Whitaker, D. J. (2008). Parent-Focused Child Maltreatment Prevention: Improving Assessment, Intervention, and Dissemination With Technology. *Child Maltreat*, *13*(4), 400-416. doi:10.1177/1077559508320059

Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy*,

27(2), 171-186. doi:10.1016/S0005-7894(96)80013-X

Springer, K., Sheridan, J., Kuo, D., & Carnes, M. (2003). The long-term health outcomes of childhood abuse. *Journal of General Internal Medicine*, *18*(10), 864-870. doi:10.1046/j.1525-1497.2003.20918.x

Swenson, C., Brown, E., & Lutzker, J. (2007). Issues of Maltreatment and Abuse. In *Personality Disorders in Childhood and Adolescence* (pp. 229-295). New Jersey:: John Wiley & Sons, Inc.

Taylor, T. K., & Biglan, A. (1998). Behavioral Family Interventions for Improving Child-rearing: A Review of the Literature for Clinicians and Policy Makers. *Clinical Child and Family Psychology Review*, *1*(1), 41-60. doi:10.1023/A:1021848315541

Timmer, S. G., Urquiza, A. J., Zebell, N. M., & McGrath, J. M. (2005). Parent-Child Interaction Therapy: Application to maltreating parent-child dyads. *Child Abuse & Neglect*, *29*(7), 825-842. doi:10.1016/j.chiabu.2005.01.003

U.S. Department of Health and Human Services. (2009). *Child Malreatment 2007*. Washington, DC:U.S.: Government Printing Office.

Wang, C., & Holton, J. (2007). *Total Estimated Cost of Child Abuse and Neglect in the United States*. Prevent Child Abuse America.

Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *Journal of child psychology and psychiatry, and allied disciplines*, *49*(5), 471-488. doi:10.1111/j.1469-

7610.2007.01861.x

Whitaker, D.J., Lutzker, J.R., Self-Brown, S., Edwards, A. (2008). Implementing an Evidence-Based Program for the Prevention of Child Maltreatment: The SafeCare Program. *Emotional & Behavioral Disorders in Youth*, 55-62.

Whitaker, D. J., Lutzker, J. R., & Shelley, G. A. (2005). Child Maltreatment Prevention Priorities at the Centers for Disease Control and Prevention. *Child Maltreat*, *10*(3), 245-259. doi:10.1177/1077559505274674

Wyatt Kaminski, J., Valle, L., Filene, J., & Boyle, C. (2008). A Meta-analytic Review of Components Associated with Parent Training Program Effectiveness. *Journal of Abnormal Child Psychology*, *36*(4), 567-589. doi:10.1007/s10802-007-9201-9

APPENDIX A: Assessment Fidelity Checklist

Home Visitor Fidelity Checklist: Assessment

Home Visitation Staff _____ Session Date _____ Family #_____
Coach_____ Module_____ In-person or Recorded?_____

Has materials ready				
Assessment checklists	+	-	n/a	n/c
Supplies	+	-	n/a	n/c
Other materials	+	-	n/a	n/c
Opens the session				
Exchanges an appropriate initial greeting	+	-	n/a	n/c
States goals for the session	+	-	n/a	n/c
Demonstrates appropriate demeanor				
Sits facing client	+	-	n/a	n/c
Communicates empathy, warmth, understanding	+	-	n/a	n/c
Maintains an open posture	+	-	n/a	n/c
Has good eye contact	+	-	n/a	n/c
Uses active listening techniques				
Uses words/expressions (e.g., "uh-huh") to encourage the parent to talk	+	-	n/a	n/c
Uses open-ended questions	+	-	n/a	n/c
Uses reflecting statements	+	-	n/a	n/c
Uses summarizing statements	+	-	n/a	n/c
Gives overviews				
Module overview	+	-	n/a	n/c
Session overview	+	-	n/a	n/c
Explains the assessment				
Explains the purpose of the assessment	+	-	n/a	n/c
Explains the process of the assessment	+	-	n/a	n/c
Completes the assessment				
Assesses the required number of activities/rooms/scenarios	+	-	n/a	n/c
Assesses the required variety of activities/rooms/scenarios	+	-	n/a	n/c
Completes the necessary forms	+	-	n/a	n/c
Provides general, positive feedback about the assessments	+	-	n/a	n/c
Addresses issues that arise during the session				
Encourages the parent to ask questions and express concerns	+	-	n/a	n/c
Responds to parent questions and concerns	+	-	n/a	n/c
Uses problem solving approaches as appropriate	+	-	n/a	n/c

Follows an appropriate closing sequence				
Summarizes the session	+	-	n/a	n/c
Asks for and answers parent questions	+	-	n/a	n/c
Gives general positive feedback	+	-	n/a	n/c
Schedules meeting date/time for next week	+	-	n/a	n/c
Items scored +				
Total items scored + or -				

Percent correct = Items scored + / Total items scored

____ %

APPENDIX B Training Fidelity Checklist

Home Visitor Fidelity Checklist: Training

Home Visitation Staff _____ Session Date _____ Family #_____
Coach_____ Module_____ In-person or Recorded?_____

Has materials ready				
Checklists-HV Versions	+	-	n/a	n/c
Training materials for families (e.g., Checklists-Parent Versions)	+	-	n/a	n/c
Supplies	+	-	n/a	n/c
Other materials	+	-	n/a	n/c
Opens the session				
Exchanges an appropriate initial greeting	+	-	n/a	n/c
States goals for the session	+	-	n/a	n/c
Demonstrates appropriate demeanor				
Sits facing client	+	-	n/a	n/c
Communicates empathy, warmth, understanding	+	-	n/a	n/c
Maintains an open posture	+	-	n/a	n/c
Has good eye contact	+	-	n/a	n/c
Uses active listening techniques				
Uses words/expressions (e.g., "uh-huh") to encourage the parent to talk	+	-	n/a	n/c
Uses open-ended questions	+	-	n/a	n/c
Uses reflecting statements	+	-	n/a	n/c
Uses summarizing statements	+	-	n/a	n/c
Conducts assessments as needed				
Conducts assessments as indicated in the Outline	+	-	n/a	n/c
Explains the purpose of the assessments	+	-	n/a	n/c
Explains the process of the assessments	+	-	n/a	n/c
Trains the parent				
Uses the appropriate material (SICC-P and scenarios, HAPI-P, PAT-P, Infant Observation Worksheet-P) to train the parent	+	-	n/a	n/c
Models steps and behaviors	+	-	n/a	n/c
Has parent practice an appropriate number of times	+	-	n/a	n/c
Balances explain vs. modeling behaviors and steps	+	-	n/a	n/c
Provides general, positive feedback	+	-	n/a	n/c
Provides specific, corrective feedback	+	-	n/a	n/c
Addresses issues that arise during the session				
Encourages the parent to ask questions and express concerns	+	-	n/a	n/c
Responds to parent questions and concerns	+	-	n/a	n/c

Uses problem solving approaches as appropriate	+	-	n/a	n/c
Follows an appropriate closing sequence				
Summarizes the session	+	-	n/a	n/c
Gives general positive feedback	+	-	n/a	n/c
Schedules meeting date/time for next week	+	-	n/a	n/c
Items scored +				
Percent correct = Total items scored + or -				