A BRIEF PREVENTIVE MODEL OF PARENT-CHILD INTERACTION THERAPY FOR OUT-OF-HOME CARE

Ryan Egan

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Department of Psychology

Central Michigan University
Mount Pleasant, Michigan
December 2017

ABSTRACT

A BRIEF PREVENTIVE MODEL OF PARENT-CHILD INTERACTION THERAPY FOR OUT-OF-HOME CARE

by Ryan Egan

Children placed out-of-home are at a higher risk for psychological and behavioral problems than other children (Pecora, White, Jackson, & Wiggins, 2009). Unfortunately, trainings to prepare foster parents for their challenging roles often lack scientific evidence (Dorsey et al., 2008). The current study piloted a brief parent training program based on an evidence-based intervention model for children whose parents have voluntarily and temporarily placed them with volunteer host families through the Safe Families for Children program. A combination of analytical methods, including reliable change indices and mean difference statistics, were used to evaluate two different formats (online and in-person training) featuring the same content. Sixty-eight families from Safe Families for Children enrolled for either the online or in-person training format; thirty families initiated the training. Both training formats were followed by five weekly group consultation sessions (offered twice a week) after the initial trainings. The primary client outcomes of this evaluation included caregiver-rated improvements in family functioning and caregivers' skill and knowledge uptake, while the primary implementation outcomes of this evaluation included acceptability of the program, feasibility of the program, and adoption and/or usage of the program. Although conclusions are constrained by small sample, tests of individual trainee change suggest that the intervention improved child pro-social behavior and parent stress in many cases, and acceptability was high amongst all respondents. Implementation challenges included barriers to participation (e.g., scheduling

conflicts, competing priorities), high attrition, low turnout, and low engagement with the material. Considerations for future implementations are discussed.

TABLE OF CONTENTS

LIST OF TABLES.....v

I. INTRODUCTION.....1
Out-Of-Home Placement Risk.....1
Preparing Out-Of-Home Caregivers.....2
Parent-Child Interaction Therapy.....4
Implementation.....5

II. METHOD.....10
Participants.....10
Training.....12
Measures.....13
Parent Stress Index Short Form (PSI-SF).....13
Psychosocial Strengths Inventory for Children and Adolescents (PSICA).....13
Strengths and Difficulties Questionnaire (SDQ).....14
Dyadic Parent-Child Interaction Coding System – Fourth Edition (DPICS-IV).....14
Knowledge of Behavior Principles as Applied to Children – Revised (KBPAC-R).....15
Child-Directed Interaction (CDI) and Parent-Directed Interaction (PDI) Quizzes.....16
Abbreviated Acceptability Rating Profile (AARP).....16
Treatment Evaluation Inventory – Short Form (TEI-SF).....17
Barriers to Treatment Participation Scale (BTPS).....18
Other Indicators of Implementation Outcomes.....18
Data Collection Procedure.....19

III. RESULTS.....22
Data Cleaning.....23
Client Outcomes.....24
Implementation Outcomes.....30
Acceptability.....31
Feasibility/Adoption.....35

IV. DISCUSSION.....41
Client Outcomes.....41
Implementation Outcomes.....44
Strengths and Limitations.....49
General Conclusions.....50

REFERENCES.....51

LIST OF TABLES

TABLE PAGE

1. Percentages of Demographic Characteristics of Trainees 11

2. Additional Demographic Characteristics of Trainees.....11

3. Visualization of Data Collection at All Time Points 21

4. Means of Client Outcomes, by Time Point.....25

5. Reliable Change of Raw Scores, by Measure and by Caregiver Number 26

6. Average Mean Item-Rating for Acceptability Forms, by Training Group 32

7. Item-Level Descriptives of AARP.....33

8. Item-Level Descriptives of TEI-SF 34

9. Descriptives of Video Module Viewing, by Module and Device Type 37

10. Selected Item-Level Descriptives of BTPS, in Order of Highest to Lowest Mean Value.39

CHAPTER I

INTRODUCTION

Children placed out-of-home are at risk for a variety of negative outcomes (Pecora et al., 2009), and often show impairment in multiple domains, including emotional, behavioral, social, and neurological development (Leve et al., 2012). Despite this increased psychosocial risk, parent training for hosting caregivers often does not incorporate evidence-based parenting techniques that may help mitigate this risk (Dorsey et al., 2008). Institutional barriers such as cost and time commitment make it difficult to implement many evidence-based parenting interventions (Festinger & Baker, 2013). Thus, some researchers have tested briefer models of treatment, such as Parent-Child Interaction Therapy (PCIT), with some success (McNeil, Herschell, Gurwitch, & Clemens-Mowrer, 2005; Mersky, Topitzes, Grant-Savela, Brondino, & McNeil, 2016). The current study piloted a novel abbreviated PCIT training, measured changes in family functioning and caregiver skill, and assessed implementation outcomes that included trainee acceptability, adoption of practices, and feasibility of the program.

Out-of-Home Placement Risk

Converging evidence from multiple large-scale studies shows that foster children are at significantly higher risk for behavioral and emotional disorders (Pecora et al., 2009). Estimates of the prevalence and intensity of these psychological difficulties vary, with as many as 50% of displaced children showing clinically significant levels of behavioral or emotional dysfunction (Burns et al., 2004; Garland et al., 2001). These disturbances in functioning have also been shown to persist: Adults with a history of foster care placement have poorer educational, financial, physical health, and psychological outcomes relative to their peers (Jackson Foster et

al., 2015; Pecora et al., 2006; Pecora et al., 2005). Even children without initial risk for mental health problems show increased prevalence of psychological disorders after over a year of placement out-of-home (Newton, Litrownik, & Landsverk, 2000).

One particularly problematic aspect of the mental health of displaced children is the prevalence of externalizing behaviors (e.g., aggression, rule-breaking, acting out, etc.). Externalizing behaviors are more common in foster children (Chamberlain et al., 2008; Clausen, Landsverk, Ganger, Chadwick, & Litrownik, 1998) and are associated with increased placement disruption (Chamberlain et al., 2006; Fisher, Burraston, & Pears, 2005; Leathers, Spielfogel, Gleeson, & Rolock, 2012). High levels of placement disruptions are associated with the exacerbation of existing behavior problems (Newton et al., 2000). Furthermore, externalizing behaviors often persist unless treated (Campbell & Ewing, 1990; Daly, Buchanan, Dasch, Eichen, & Lenhart, 2012; Shaw, 2013) and can be made worse by negative patterns of interactions with others (Moffitt, 1993) and parental inconsistency (Campbell, 1995). Foster care thus has the potential to be particularly iatrogenic to children with externalizing behaviors, leading to what one researcher has called "serial evictions" for these children (Barber & Delfabbro, 2006). Disruptive behavior begets instability, which begets more disruptive behavior, and the cycle continues.

Preparing Out-of-Home Caregivers

Unfortunately, out-of-home caregivers are often ill-equipped to deal with challenging behavior. Foster parents have reported that they do not receive enough training in parenting skills (Hampson & Tavormina, 1980; Hochstadt, Jaudes, Zimo, & Schachter, 1987; Spielfogel, Leathers, Christian, & McMeel, 2011), feel incapable of dealing with problem behaviors

(Cooper, Peterson, & Meier, 1987), and experience significant levels of stress in response to these behaviors (Timmer, Sedlar, & Urquiza, 2004; Vanschoonlandt, Vanderfaeillie, Van Holen, De Maeyer, & Robberechts, 2013). Parent training may be a good solution to this problem, but unfortunately, existing trainings for out-of-home caregivers have very little scientific support (Dorsey et al., 2008; Festinger & Baker, 2013; Leve et al., 2012; Rork & McNeil, 2011). The most common parent trainings focus on education about the child welfare system and what to expect as a prospective caregiver, rather than focusing on parenting skills (Dorsey et al., 2008). The studies that do evaluate parent skills-training commonly have methodological limitations, including a lack of control or comparison group, single-method subjective outcome measures, few instances of child outcome data, small sample sizes, lack of randomization to training conditions, and attrition (Festinger & Baker, 2013; Leve et al., 2012; Rork & McNeil, 2011; Solomon, Niec, & Schoonover, 2017).

There are some interventions that show promise for improving parent skills and child behaviors in the child welfare context. Behaviorally based parent management training has strong support for improving externalizing behaviors in children (Eyberg, Nelson, & Boggs, 2008; Kazdin, 2005; Maughan, Christiansen, Jenson, Olympia, & Clark, 2005), and principles from this approach have been adapted to multiple out-of-home parent training programs. Parent management training uses principles of reinforcement to modify problematic behaviors in an incremental and consistent manner (Kazdin, Holland, Crowley, & Breton, 1997). Outcomes of these trainings include parent reports of improved child behavior (Chamberlain & Mihalic, 1998; Chamberlain et al., 2008; McNeil et al., 2005; Mersky et al., 2016; Price, Chamberlain, Landsverk, & Reid, 2009; Timmer et al., 2006), better placement stability (Fisher et al., 2005; Price et al., 2008), increased use of positive parenting skills and effective discipline

(Chamberlain et al., 2008; Linares, Montalto, Li, & Oza, 2006), lower parent stress (Fisher & Stoolmiller, 2008; Timmer et al., 2006), lower child cortisol levels (Fisher, Gunnar, Chamberlain, & Reid, 2000), and lower parent abuse potential (Timmer et al., 2006). However, many of these trainings require multiple weeks of in-person attendance, practice, and phone consultation (e.g., Chamberlain et al., 2008), which is potentially at odds with some of the most pervasive barriers in parent training in the child welfare system: funding and participant commitment (Festinger & Baker, 2013).

Parent-Child Interaction Therapy

Parent-Child Interaction Therapy (PCIT; Zisser & Eyberg, 2010) is a behavioral parent training program that shows great potential for addressing common treatment barriers. PCIT draws from both behavioral theory and attachment theory (Eyberg, 1988) to increase the warmth of the interactions between parent and child while also increasing the parent's control over negative behaviors. In the first stage of treatment, therapists focus on increasing the parent's use of positive attention and praise to increase a child's prosocial behaviors while ignoring inappropriate behavior (McNeil & Hembree-Kigin, 2010a). In the second stage of treatment, therapists focus on teaching parents how to give effective directions and judiciously apply non-coercive discipline for remaining unacceptable behaviors (McNeil & Hembree-Kigin, 2010b). Most of the weekly treatment hour focuses on parents practicing the skills in the clinic with their own child; this allows the therapist to provide immediate feedback via "coaching" (often done via a radio earpiece while the therapist watches from another room). This type of in-clinic, *in vivo* practice is what makes PCIT unique from many other parenting interventions and has been associated with the highest treatment effect sizes among behavioral parent training studies

(Kaminski, Valle, Filene, & Boyle, 2008). The live-coaching part of PCIT has been successfully used by itself without other treatment components to improve child behavior (Shanley & Niec, 2010).

Research shows wide support for PCIT in general, but also in particular for the child welfare context. Meta-analyses examining outcomes of PCIT have revealed large multi-method effect-sizes in the reduction of externalizing behaviors (Thomas & Zimmer-Gembeck, 2007; Ward, Theule, & Cheung, 2016) and parent stress (Cooley, Veldorale-Griffin, Petren, & Mullis, 2014). PCIT also shows a high degree of portability, achieving similar clinical outcomes in non-university based settings (Lanier et al., 2011; Lyon & Budd, 2010; Self-Brown et al., 2012). Recently, PCIT has been extended to child welfare populations, showing efficacy in reducing problem behaviors, keeping parents in treatment, and reducing child abuse potential (Chaffin et al., 2004; Chaffin et al., 2009). In a sample of 75 foster children, treatment completion rates and child behavior improvements were found to be no different from the sample of 98 non-foster families referred for behavior problems (Timmer et al., 2006). Additionally, two briefer PCIT trainings have been tested with foster parents, with both models showing improved parent ratings of child behavior (McNeil et al., 2005; Mersky et al., 2016). In addition to their promising clinical outcomes, these two briefer models have broken exciting new ground with regard to their innovations in implementation.

Implementation

Given the aforementioned scarcity of evidence for out-of-home parent trainings, an examination of the factors that might facilitate uptake of more evidence-based practices is needed. Although little research has been done on the process of adoption of evidence-based

practices in child welfare agencies (Horwitz, Chamberlain, Landsverk, & Mullican, 2010), existing models of implementation science may be helpful for evaluating existing training models. Proctor et al. (2011) proposed a taxonomy for understanding the key aspects of implementation that defined eight outcomes: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability. These outcomes were defined through a review of existing implementation research. They all describe different results of attempts to disseminate new practices and are distinct from client outcomes (which refer to actual symptom improvement at the client level). This vocabulary allows for a more thorough consideration of the factors that help bring a lab-tested intervention into the community.

In light of this implementation taxonomy, the most recent brief PCIT trainings by McNeil et al. (2005) and Mersky et al. (2016) have broken exciting new ground with regard to feasibility, implementation cost, appropriateness, and fidelity. With regard to feasibility and implementation cost, the intensive 2-day format is easier to attend than a multi-week format. Given the prevalence of attrition in other PCIT research (e.g., Timmer et al., 2006), ease of attendance makes for a compelling reason to have a truncated delivery model. Additionally, the group format is more resource-efficient (Barkley, 1987; Brightman, Baker, Clark, & Ambrose, 1982). With regard to appropriateness, the group format of these trainings may address some specific needs among out-of-home caregivers. Out-of-home caregivers have reported feeling isolated from other parents (Cavazzi, Guilfoyle, & Sims, 2010) and have expressed a desire for increased social and administrative support (Brown, 2008; MacGregor, Rodger, Cummings, & Leschied, 2006). A group format may improve perceived support. With regard to fidelity, one aspect of fidelity is identifying the different components of a treatment (or dosage level) that can work without the complete treatment package. Mersky et al. (2016) provided a test of training

dosage for two brief formats, including one group that received training and phone consultation only, and one group that received training, consultation, and a booster session with additional consultation. Understanding dosage allows for more equitable allocation of services.

As both research groups noted, PCIT training can still be further modified to maximize implementation outcomes within the child welfare context. First, the scope of the previous trainings was still quite large, and therefore difficult to implement with more limited resources. Both trainings required two 8-hour days of time commitment for families, which may still be prohibitive for many hosting families. The Mersky et al. (2016) implementation required individual weekly phone consultation with over 100 families, which would require large staffing and time demands for any implementation of this model. Second, in-person attendance adds another barrier that is common for parent training: transportation (Boggs et al., 2005; Harwood & Eyberg, 2006). Families often do not have a way to get to the clinic, or they are simply too far away (Owens et al., 2002). Alternative content delivery formats (e.g., self-guided materials, web-based content, mobile applications, video) may circumvent this barrier. Third, previous trainings did not measure acceptability. McNeil et al. (2005) used a measure of satisfaction, but Proctor et al. (2011) noted that satisfaction is a less specific construct that "generally refers to the general service experience" (p. 67) and may not relate to perceptions of the actual intervention itself.

Finally, there is an additional innovation that has been suggested McNeil et al. (2005) that could address multiple implementation outcomes: a prevention model. Behavioral parent training has strong support in the prevention of conduct problems (Daly et al., 2012; Powell, Lochman, & Boxmeyer, 2007), and it also has several implementation advantages. First, treating the antecedents of behavior problems can be more cost effective than treating a more entrenched behavior problem if risk is appropriately assessed and targeted (Foster & Jones, 2006). Second, a

prevention approach would increase penetrance by being available to all prospective caregivers, not just those dealing with clinically significant behavior issues. Third, a prevention approach may be a better fit with the current state of out-of-home caregiver training, which is part of the key implementation outcome of appropriateness. As noted by Festinger and Baker (2013), the most common forms of out-of-home caregiver training are pre-service, and the contents of these pre-service trainings are not rooted in evidence-based practices. PCIT, on the other hand, is an evidence-based practice that could be offered pre-service without compromising its essential ingredients. Finally, prospective caregivers may find it more acceptable to feel prepared for their placements before those children in the home. Given the aforementioned relationship between caregivers' lack of confidence and subsequent placement instability, feeling competent sooner rather than later could have a buffering effect on this relationship.

The current study tested a brief pre-service, out-of-home caregiver training based on PCIT, guided by previous studies and the preceding discussion of what is needed on the implementation front. Two client outcomes of the training were measured: a) caregiver-rated improvements in family functioning, b) caregivers' skill and knowledge uptake. Three implementation outcomes of this evaluation included a) acceptability of the program, b) feasibility of the program, and c) adoption and/or usage of the program. Differences in these outcomes across the two formats (online vs. in-person) were also examined. With client outcomes, family functioning was assessed with measures of child behavior and caregiver stress while caregiver skill and knowledge was assessed with a skill-based task and a written test. With implementation outcomes, acceptability was assessed with established acceptability questionnaires, whereas the other outcomes were assessed with various behavioral indicators

(e.g., turnout, attrition, homework completion, etc.) and a questionnaire about barriers to participation.

CHAPTER II

METHOD

Participants

Sixty-eight caregivers from the Safe Families program in Chicago enrolled to take part in one of two different formats of the training (online or in-person). Thirty caregivers initiated the training, with seventeen in the online group and thirteen in the in-person group. Seven families from the initial thirty participated in a second in-person booster training. To participate in the training, participants were advised to a) have their own children between 2.5–7 years old, b) agree to participate in all parts of the training, and c) agree to accept a child 2.5–7 years old into their home within three months of the initial training. Safe Families stakeholders identified qualified families and provided contact information to the research team. Researchers then sent each caregiver an orientation packet with information about the upcoming training with pre-training evaluation forms (including the measures discussed below), which were then sent back to the researchers. Caregivers indicated which format (online or in-person) they wanted to do with their agency leaders, who then communicated this preference to the researchers.

Demographic data about the thirty participants are presented in Tables 1 and 2. Participants were mostly female, white, middle-aged, high socio-economic status (SES), well-educated, and had a small-to-medium-sized family with two caregivers. Participants had an average of two years of membership with Safe Families, ranging from one to three years of membership. Thus, all participants were fairly new to the program. Participants varied widely in their experience hosting displaced children, ranging from zero to thirty hosted children in the past three years. The majority of training participants (66.67%) had hosted at least one child.

Across all demographic variables, χ^2 analyses and independent-samples *t*-tests were conducted to ascertain the degree of heterogeneity across training formats (in-person vs. online). No variables showed significant differences between the thirteen in-person trainees and the seventeen online trainees (p = 0.24 – 0.87).

Table 1. Percentages of Demographic Characteristics of Trainees

Variable	Percentage of Sample
Gender	
Female	83.3%
Male	16.7%
Ethnicity	
Caucasian	86.7%
American Indian	3.3%
Asian	6.7%
Annual household income	
25k-50k	3.3%
50k-75k	10.0%
75k or higher	83.3%
Education level	
Some college	6.7%
Bachelor's Degree	60.0%
Master's Degree or higher	33.3%

Note. One participant did not want to report their ethnicity.

Table 2. Additional Demographic Characteristics of Trainees

Variable	*M*	*SD*	Minimum	Maximum
Age	44.77	10.61	27	65
Number of Caregivers	2.03	0.32	1	3
Household size	3.5	1.36	1	7
Years of membership with SF	2	0.87	1	3
Age of children in the family	6.32	3.47	2.63	15
Total displaced children hosted	5.83	7.5	0	30

Note. Three participants had no children, and thus did not report on age. These data only include the individuals who returned pre-training packets (and not the five additional respondents who participated in the barriers questionnaire at the end).

Training

The training program, called the "Better Families Together" program (BFT), was based on a brief model of PCIT, PCIT-Selective Prevention (Acevedo-Polakovich et al., 2014; Niec et al., 2014). For both formats (online and in-person), training content included basic behavioral principles, attachment, and all essential parenting techniques in PCIT. Content for the in-person format was presented first in didactic sessions, and then extensive practice through role-play, games, and actual in vivo practice with volunteering children from their families. Content for the online format was presented via individual video-recordings of all the didactic sessions from the in-person training, and online learning experiences were utilized to provide an analog of the practice received in the in-person format. The in-person training lasted about six hours, while viewing all of the modules for the online component and completing the associated exercises took approximately two-thirds that time. After initial training took place, the consultation phase began. Families were offered two hour-long group phone consultations each week for a period of five weeks to consolidate the use of their skills at home. Both formats were offered the same amount of consultation.

An in-person booster training of the Better Families Together program took place approximately three months after the initial in-person BFT training was offered. The content of this six-hour training included a didactic review of child-led play skills and parent-led discipline skills, troubleshooting for the aforementioned skills and how to practice them at home, further practice of the skills through role-play, and an introduction to basic engagement strategies to help other caregivers use these skills. This booster training was followed by an additional five-week period of phone consultation.

Measures

All measures described in this section were used by Safe Families to help evaluate the program. The research team at Central Michigan University was granted access to these measures. Two client outcomes of the training were measured: a) caregiver-rated improvements in family functioning, b) caregivers' skill and knowledge uptake. Three implementation outcomes included a) acceptability of the program, b) feasibility of the program, and c) adoption and/or usage of the program. Both client outcomes and the first two implementation outcomes (acceptability and feasibility) were assessed with specific measures. The first three measures assessed caregiver-rated family functioning.

Parent Stress Index Short Form (PSI-SF)

The PSI-SF (Abidin, 2012) is a 36-item parent-report measure of the stress level of the respondent due to different aspects of being a parent. The Total Stress scale of the PSI-SF is highly correlated ($r = 0.94$) with the full PSI score, with test retest reliability at 0.84 over 6 months. Parents' full PSI scores have been significantly correlated with externalizing behaviors in their children (Eyberg, Boggs, & Rodriguez, 1993). The PSI-SF was used to assess whether caregivers' stress declined subsequent to the training.

Psychosocial Strengths Inventory for Children and Adolescents (PSICA)

The PSICA (Niec, Peer, Courrégé, & Solomon, 2016) is a 36-item parent-report measure of child psychosocial competence, including behaviors such as listening quickly to adults, remaining calm when challenged, and sharing. Each item describes a behavior and has the parent rate 1) the frequency with which this behavior occurs (ranging from 1 – *never* to 7 – *always*) and 2) whether the parent is satisfied with this behavior ("yes" or "no"). The PSICA demonstrates

good internal consistency (α = 0.96), is positively correlated with measures of emotion regulation and the parent-child relationship, and is negatively correlated with measures of conduct problems and hyperactivity (Niec et al., 2016). The PSICA was used to measure the caregivers' biological child's behavior both before and after training to judge whether the training results in more prosocial behavior.

Strengths and Difficulties Questionnaire (SDQ)

The SDQ (Goodman, 1997) is a 25-item parent-report measure of positive and negative behaviors, with a five factor structure that encompasses emotional, conduct, hyperactivity-inattention, peer, and prosocial domains (Goodman, 2001). Each item is a statement about the child's behavior, and the parent rates how true it is for the child on a 3-point Likert scale (*not true, somewhat true,* and *certainly true)*. Among parent raters across 48 studies, the SDQ shows good internal consistency, (α = 0.69-0.87; mean = 0.81) and test-retest reliability (r = 0.72-0.86; mean = 0.76) (Stone, Otten, Engels, Vermulst, & Janssens, 2010). Additionally, high scores on the SDQ predict a higher probability for existing psychiatric diagnosis from physicians (Goodman, 2001). The SDQ was used as a broadband measure of problematic behaviors in the caregivers' biological child both before and after training to judge whether the training results in any reduction in these behaviors.

The next three measures assessed caregiver learning.

Dyadic Parent-Child Interaction Coding System – Fourth Edition (DPICS-IV)

The DPICS-IV (Eyberg, Nelson, Ginn, Bhuiyan, and Boggs, 2013) is a behavioral observation coding system that is typically used before, during, and at completion of PCIT. It categorizes and totals the verbalizations spoken between parents and a child during play

situations (e.g., praise, behavior descriptions, etc.) and is a useful tool for assessing caregivers' skill acquisition over time. Some of these code categories help to identify skills that caregivers should use a lot (called "Do-Skills") and behaviors that caregivers should aim to use less ("Don't Behaviors"). Interrater reliability for the DPICS-IV categories for verbalizations is adequate, ranging from 0.59 to 0.85 (Eyberg et al., 2013). The DPICS-IV was used to evaluate how well caregivers could identify the skills they were taught during the training, as they were using DPICS-IV rules to attempt to code a sample of a scripted parent-child interaction on video. The parent in the video spoke every 10 seconds, and his verbalizations were provided by an off-screen assistant reading from a randomized list of verbalizations (weighted toward more Do-Skills). Verbalizations were selected for relative ease of coding and ease of applicability to a wide range of play situations. After caregivers watched and coded this video, intra-class correlation was used to measure how closely their coding tallies matched the correct tally.

Knowledge of Behavior Principles as Applied to Children – Revised (KBPAC-R)

The KBPAC-R (Furtkamp, Giffort, & Schiers, 1982) is a 10-item multiple-choice test about basic behavioral principles and common childhood learning scenarios. It is based off of a longer version, the KBPAC (O'Dell, Tarler-Benlolo, & Flynn, 1979), and has two parallel forms A and B with different items and equal length. Item-total correlations from the original 50 items were used to select the shorter sets of items. Internal consistency of the KBPAC-R parallel forms (α = 0.74 and 0.77) are comparable to the original (α = 0.86), and the adjusted correlation of the parallel forms is 0.87 (Furtkamp et al., 1982). The original KBPAC shows sensitivity between different levels of expertise with child behavior in professionals (McLoughlin, 1985) and is correlated with parent-rated behavior problems in their children (Tiano & McNeil, 2014). The

KBPAC-R was used to measure caregivers' knowledge of behavior principles both before and after training to judge whether the training provides a greater understanding of applied behavior principles.

Child-Directed Interaction (CDI) and Parent-Directed Interaction (PDI) Quizzes

During the first in-person training, two multiple-choice quizzes were designed by the trainers to promote engagement with the material and briefly check for participant understanding. The five-item CDI quiz pertained only to the first phase of PCIT (e.g., "select two toys that are appropriate for special play time" or "which of the following is not a behavior you should ignore?"), and the ten-item PDI quiz pertained only to the second phase of PCIT (e.g., "what are characteristics of effective commands?" or "what should you do immediately after giving an effective command?"). The quizzes were not intended as formal knowledge assessments or as primary outcomes, but they may be useful as informal indicators of learning for some of the content.

The next two measures assessed caregiver-rated acceptability of the training.

Abbreviated Acceptability Rating Profile (AARP)

The AARP (Tarnowski & Simonian, 1992) is an 8-item measure of treatment acceptability for child mental health treatment that was designed in response to concerns of the time-intensiveness and higher reading level of other acceptability measures (Finn & Sladeczek, 2001). Each item is a statement about some aspect of the treatment's acceptability (e.g., how likable the treatment is, whether it would have bad side effects), and respondents rate how true each statement is on a 6-point Likert scale (ranging from 1 – *strongly disagree* to 6 – *strongly agree*). It has good internal consistency (α = 0.97) and split-half reliability (r = 0.95), and was

less time-intensive and more readable than the other most common short acceptability forms (Tarnowski & Simonian, 1992). Factor analysis revealed a unitary factor of acceptability for the AARP, and it discriminates between different psychosocial interventions and medical interventions for children with behavior problems (Tarnowski & Simonian, 1992). Some of the wording for this measure was altered to make it appropriate for a preventive program for out-of-home caregivers. The AARP was used to assess the degree to which caregivers found this training acceptable.

Treatment Evaluation Inventory – Short Form (TEI-SF)

The TEI-SF (Kelley, Heffer, Gresham, & Elliott, 1989) is a 9-item parent-report measure of satisfaction with a parent training program. The TEI-SF is a distilled version of a longer acceptability instrument, the Treatment Evaluation Inventory (Kazdin, 1980), and was reduced to nine items via factor analysis. The two factors were labeled Acceptability (i.e., how much the treatment was rated positively) and Discomfort (i.e., extent of aversiveness of the treatment) (Kelley et al., 1989). Each item on the TEI-SF is a statement about some aspect of the treatment's acceptability (e.g., beliefs about efficacy, expectations about side effects), and parents rate how true each statement is using a 5-point Likert scale (ranging from 1 – *strongly disagree* to 5 – *strongly agree*). The TEI has good internal consistency (α = 0.85) and showed discriminant validity in a sample of adults rating the acceptability of different doses of punishment in behavioral treatments (Kelley et al., 1989). Some of the wording for this measure was altered to make it appropriate for a preventive program for out-of-home caregivers. The TEI was also used to assess the degree to which caregivers find this training acceptable, in addition to

the AARP. There is no consensus about the suitability of either the AARP or the TEI-SF (Finn & Sladeczek, 2001), and they have slightly different factor structures and content.

The final measure assessed an aspect of caregiver-rated feasibility.

Barriers to Treatment Participation Scale (BTPS)

The BTPS (Kazdin et al., 1997) is a 44-item parent-report measure of barriers to participating in a psychosocial treatment program. Each item is a statement about an aspect of the program (e.g., "the program was in conflict with another one of my activities," "I lost interest in the program"), and parents rate how much each statement was problematic for their participation using a 5-point Likert scale (ranging from 1 – *never a problem* to 5 – *very often a problem*). The BTPS has good internal consistency ($\alpha = 0.85$) and was shown to predict treatment drop-out and lack of participation beyond what is typically predicted by other demographic indicators (Kazdin et al., 1997). Some of the wording for this measure was altered to make it appropriate for a preventive program for out-of-home caregivers. The BTPS was used to assess the extent and nature of barriers to participation, which can be conceptualized as an aspect of feasibility (Proctor et al., 2011). It was administered at the 6-month follow-up after the initial in-person training.

Other Indicators of Implementation Outcomes

Additional observational data, collected as the implementation of the program unfolded, measured the extent to which participants used the techniques and content of the program. These data may be considered part of the implementation outcomes referred to as adoption and/or feasibility (Proctor et al., 2011), and they included consultation attendance (i.e., the sum of attended consultations for each caregiver), homework completion (i.e., the sum of times the

caregiver completed “special time” practice play sessions at home), “play count” of the online modules (i.e., how many times the didactic videos were watched for the online version), and completion rates of online modules (i.e., what proportion of each video was watched). All of these data were simply tallies of the named behavior, or in the case of the completion rates of online modules, they were average proportions of completed video content (e.g., participants watched an average of 85% of this video before stopping it or closing the window in their browser).

Data Collection Procedure

Data were collected from participants at three time points. At Time 1, 30 participants of the original 68 who signed up returned a packet containing all pre-training questionnaires, including demographic information, family functioning (SDQ, PSICA, and PSI), and caregiver knowledge (KBPAC-R). Of these initial thirty, thirteen participants chose the in-person training and seventeen chose the online training. Six participants of these thirty did not complete the child behavior measures (e.g., SDQ, PSICA) because they did not have a child of their own (despite the inclusion criterion to have a child); they were welcomed to participate anyway to maximize accessibility and penetrance. Immediately after the in-person training (before consultation), trainees submitted their acceptability measures anonymously (AARP, TEI-SF). Online participants submitted these acceptability measures whenever they completed the online modules. At Time 2 (three months after Time 1), seven participants attended the booster training and submitted another packet of materials, including family functioning (SDQ, PSICA, PSI) and caregiver knowledge (KBPAC-R) before the day’s training activities. They also submitted acceptability measures immediately after the booster training (AARP, TEI-SF). At Time 3 (six

months after Time 1), a follow-up phone survey was conducted with twelve participants. Five of these twelve were part of the initial 34 individuals who did not return an initial questionnaire packet or participate in any of the training. Due to the lack of uniformity of data collection, a visualization of the three times points is presented in Table 3 to clarify the nature of the data (see below).

Table 3. Visualization of Data Collection at All Time Points

	Time 1 Before training (0 months)	Time 2 Booster training (3 months)	Time 3 Final Phone Survey (6 months)
In-person group			
Caregiver 1*	X	X	X
Caregiver 2*	X	X	X
Caregiver 3	X	X	X
Caregiver 4	X	X	X
Caregiver 5*	X		X
Caregiver 6	X		X
Caregiver 7	X		
Caregiver 8*	X		
Caregiver 9	X		
Caregiver 10	X		
Caregiver 11	X		
Caregiver 12	X		
Caregiver 13	X		
Online group			
Caregiver 14	X	X	
Caregiver 15*	X	X	
Caregiver 16*	X	X	
Caregiver 17	X		X
Caregiver 18	X		
Caregiver 19	X		
Caregiver 20	X		
Caregiver 21	X		
Caregiver 22	X		
Caregiver 23	X		
Caregiver 24	X		
Caregiver 25	X		
Caregiver 26	X		
Caregiver 27	X		
Caregiver 28	X		
Caregiver 29	X		
Caregiver 30	X		

Note. Caregivers with an asterix had incomplete or absent child forms (e.g., SDQ, PSICA) because they did not have a child in their home at the time of the training.

CHAPTER III

RESULTS

As is evident from Table 3, there are limitations to the data that warrant some attention. Of the initial thirty participants who completed pre-training packets, only ten completed the measures a second time, and only four completed them at all three time points. Furthermore, half of repeat-respondents did not report child data. Consequently, several analytical constraints are noted from the outset. First, analyses of "outcomes" necessarily focused on individual reliable change as a function of the reliability of the measure used and the effect size observed by each individual caregiver for each individual measure (Reliable Change Index [RCI]; Jacobson & Truax, 1991). Mean-based statistics were reserved for Time 1 group comparisons, since there were not sufficient numbers to do any within-subjects comparisons of means for any measures. These mean-based statistics were primarily used to determine if there were any pre-training differences between groups. Second, perceptions of causation regarding the outcomes of the training should be regarded cautiously due to inconsistent participation within conditions. For example, the online participants often did not complete, or initiate, the online training modules. Furthermore, it is difficult to separate the effects of the booster training from that of the other trainings (online and in-person) because individuals from both the online and in-person groups elected to attend. Researchers were disinclined to turn away caregivers, and compromised research design priorities in favor of service to the collaborating agency. Third, a mean-level and item-level descriptive analysis of client and implementation outcomes will be utilized to draw conclusions about the implementation and how future efforts might proceed.

Data Cleaning

Typically, identifying outliers is important for protecting the integrity of any mean-level statistical tests, since they are sensitive to extreme values (Keppel & Wickens, 2004). Because mean-level statistics were not emphasized in the current study's analytic strategy, outlier identification took on a different purpose: identifying extreme or erroneous values that may not be explained by the study's parameters and variables, and thus might not be interpretable. Additionally, outliers may identify particularly unusual participants who may not resemble the average participant in important ways, and thus their response to intervention may be idiosyncratic.

A three-standard-deviation cut-off rule was used to identify potential outliers (Keppel & Wickens, 2004). Any value that was three standard deviations above or below the mean was considered a potential outlier (but not necessarily removed; see below). This rule was selected primarily because techniques that depend on normally distributed data, like the Outlier Labeling Rule (Hoaglin & Iglewicz, 1987; Hoaglin, Iglewicz, & Tukey, 1986), were inappropriate for this dataset. Visual inspection of the questionnaire data collected revealed many non-normal distributions, many of which remained non-normal after multiple transformation attempts. Only three outliers in the data were identified by the 3-SD rule: one participant attended four consultation calls (most did not attend any), one participant hosted thirty displaced children, and one family had three caregivers instead of the typical two. Because these variables were not critical for any analyses, and because of the small sample likely resulting in an expected pattern of "overdispersion" in the data (Keppel & Wickens, 2004), these cases were included.

Client Outcomes

Client outcomes were measured at three time points: right before beginning the training (either online or in-person), right before the booster training (three months after the initial training), and six months after the initial training ended. These outcomes included family functioning (PSICA, SDQ, and PSI) and caregiver knowledge (KBPAC-R). Coding ability was only assessed at in-person contact times (during the first in-person training and the booster training). Descriptives will be reviewed first, followed by reliable change analyses.

Descriptives for all variables of interest (both family functioning and caregiver knowledge) are shown in Table 4, across all three time points. The table note cautions against interpreting any changes in means across time points because the groups are not uniform (i.e., different individuals made up these groups). For this reason, scores from Time 2 and Time 3 will only be emphasized in the proceeding RCI analyses. At Time 1, the SDQ and PSI scores were in the average range according to their respective normative samples, and the PSICA score was less than one standard deviation above the original sample's mean (M = 186.37 and SD = 35.14; Courrégé, White, Niec, Solomon, & Peer, 2015). From this, it is safe to assume that this sample of children was sub-clinical regarding broadband measures of psychological functioning. With the KBPAC-R, caregivers scored above chance with a mean score of 5.07 out of 10 (random responding would result in a mean score of 2.5 out of 10). Coding ICCs during the coding activity during the training were good (0.76), just shy of the minimum criterion for coding agreement in the DPICS-IV manual (Eyberg, Nelson, Ginn, Bhuiyan, & Boggs, 2013). It is worth reiterating that the coding video contained carefully spaced verbalizations selected for their ease of recognition, so it may not be representative of their coding ability with a more real-

world interaction. Nevertheless, the high ICC average is still impressive considering caregivers' lack of explicit training in DPICS-IV rules, which are extensive.

Table 4. Means of Client Outcomes, by Time Point

Measure	Time 1			Time 2			Time 3		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
SDQ 0-4	12	11.17	5.18	2	10	7.07	0	-	-
SDQ 4-10	11	12.32	4.68	2	7.5	0.71	3	10.67	5.77
PSICA - F	23	174.48	24.45	4	193.0	11.94	4	200.75	10.53
PSICA - S	23	22.83	7.02	3	23.0	8.89	4	28.0	5.23
PSI	24	81.0	12.15	7	73.71	20.59	7	67.43	16.06
KBPAC	30	5.07	2.21	7	7.0	1.63	7	7.29	2.21
Coding ICC	13	0.76	0.19	7	0.74	0.15	-	-	-

Note: SDQ has two age groups, 0-4 and 4-10. PSICA has Frequency and Satisfaction scales. Across time points, groups are not composed of the same participants, so mean changes cannot be interpreted in a meaningful way. Completing KBPAC did not require having children in a certain age range to report on, which is why all 30 packets had scorable KBPAC data. One participant at time 2 completed the Intensity but not the Satisfaction scale of the PSICA. Coding ICC was assessed *during* the training, whereas all other variables in this table were assessed *before* the training.

All RCI analyses are shown in Table 5, with all available raw scores of the outcome variables of interest accompanied by asterisks indicating significant differences. For the only measure of broadband psychopathology, the SDQ, RCI analyses revealed very little change. Five caregivers had more than one SDQ score, with one having all time points, two having times 1 and 3, and two having times 1 and 2. None of the raw scores changed significantly, across any time point comparison. In most cases, change only occurred by one or two raw score points, and even the largest (and still non-significant) change with Caregiver 4 went in a direction that suggested more behavior problems, and not less. SDQ scores were largely in the average range or slightly above average in the first place, so it is possible that any lack of change observed with

Table 5. Reliable Change of Raw Scores, by Measure and by Caregiver Number

Measure	Time 1 Raw Score	Time 2 Raw Score	Time 3 Raw Score
SDQ			
C3	3	5	4
C4	4	7	-
C14	7	8	-
C6	13	-	14
C17	15	-	14
PSICA			
C3	196	204	206
C4	197*	196	213*
C14	208	196	-
C6	173*	-	191*
C17	178*		193*
PSI			
C1	74*	63	52*
C2	54	37	42
C3	90	91	86
C4	95*	82	72*
C14	56**	74**	-
C5	78	-	62
C6	88*	-	62*
C17	76	-	79
KBPAC			
C1	8	7	8
C2	3*	4	7*
C3	5*	9*	9
C4	2*	8*	3
C14	2*	7*	-
C15	7	8	-
C16	4	6	-
C5	9	-	10
C6	9	-	7
C17	7	-	7

Note: C = Caregiver. *Reliable Change was significant, and in the desired direction. **Reliable Change was significant, and in the non-desired direction. PSICA is only frequency score. C14 and C17 reported not completing any training, so any changes are likely not due to participation in the program.

the SDQ scores was due to floor effects (i.e., no decrease in symptoms was observed because of too few symptoms to begin with).

For the measure of pro-social behavior, the PSICA, a promising pattern of improvement emerged. The same five caregivers had more than one PSICA score, with a slightly different time point breakdown: two had all time points, one had times 1 and 2, and two had times 1 and 3. Caregivers 4, 6, and 17 all had significant reliable change in the desired direction between time 1 and time 3, with Caregiver 3's raw score improvement being four raw score points shy of being statistically reliable. No significant changes were observed between times 1 and 2. Caregiver 14 saw no significant improvement, which is notable because of her reported lack of completion of any training components (she signed up for the online training but did not start it). However, this lack of participation is also true of Caregiver 17, who reported a significant improvement on the PSICA (although it was the smallest magnitude of the group). Without mean statistics, and without more careful experimental control of the conditions, it is difficult to make strong conclusions about the change on the PSICA. Nevertheless, all individuals who participated in the training showed either trending or significant improvement on the PSICA by time 3, suggesting a possible benefit of this training for their child's caregiver-rated pro-social behaviors. Two of the three caregivers reporting the significant improvements on the PSICA attended the booster training at time 2, making it difficult to ascertain the resultant effects of the initial training versus the booster training.

For the measure of parent-stress, the PSI, there was another intriguing pattern of improvement. Eight caregivers had more than one PSI score, with four having all three time points, one having times 1 and 2, and three having times 1 and 3. Caregivers 1, 4, and 6 saw significant improvements in parent stress, while Caregivers 2, 3, and 5 saw non-significant

improvements. All significant improvements were noted between times 1 and 3. Caregivers 14 and 17, who did not participate in the training, reported higher parent stress, with Caregiver 14's change being clinically significant. Once again, mean statistics would render any conclusions more robust, particularly with an even split of participants with reliable improvement and non-reliable improvement. Taken together with the slight deterioration reported by trainees who did not complete training, these data may indicate another promising effect of the training on parent stress. Once again, two of the three caregivers who reported significant improvements on the PSI also participated in the booster training, adding some uncertainty to the effect of each training.

Turning toward caregiver learning outcomes, knowledge of behavior principles (the KBPAC-R) showed some improvement when initial scores were low, but otherwise did not change significantly. Ten caregivers had more than one KBPAC-R score, with four having all three time points, three having times 1 and 2, and three having times 1 and 3. It is worth noting again that a score of 2.5 is expected for random responding, as it is a forced-choice test of ten items with four answer choices for each item. Caregivers 2, 4, and 14 started near or below this threshold of random responding, and all of them showed significant improvement between time 1 and a later time point. However, Caregiver 4 puzzlingly showed a backslide toward random responding after showing a significant improvement between times 1 and 2. Caregiver 3 showed significant improvement from time 1 to time 2, and then remained the same at time 3. All other caregivers showed insignificant changes, often responding identically between testings. This is not surprising, given that it is a short knowledge test with a few fixed answers and not a more expansive rating scale of observed child behavior (like the other measures). It is possible that memory of previous responses affected performance on subsequent administrations, keeping both correct responses and incorrect responses consistent. Regardless, a pattern of three

improvements (discounting the backslide of Caregiver 4) versus seven insignificant changes is hardly suggestive in any particular direction, especially when it seems plausible that regression toward the mean was a factor for low scorers.

Two other outcomes were considered as possible indicators of caregiver learning, despite the impossibility of reliable change analysis with either of them: DPICS coding intraclass correlations (ICC) from the coding activity and quiz scores during the first training. First, as shown in Table 5, several caregivers participated in both the initial training and the booster training, and participated in an identical coding activity twice. This allowed for a comparison of their coding agreement with the answer key at two different time points. Reliable change was not calculable because the DPICS coding ICC is, in itself, a calculated statistic, and lacks the psychometric characteristics necessary to calculate the RCI (Jacobson & Truax, 1991). Comparing the ICC's between time 1 and time 2, one score appears to improve, two scores appear to remain the same, and one score appears to decline. Thus, no strong conclusions appear possible on coding ability as it changed over time.

Quiz data collected during the initial in-person training provided a small indication of some caregiver learning, prompting caregivers to answer simple questions about PCIT-specific knowledge. On the Child-Directed Interaction quiz, caregivers scored an average of 4.54 out of a possible 5, which suggests they knew the PRIDE skills, knew how to allocate their selective attention regarding child behavior, and knew how to set up special time effectively. On the Parent-Directed Interaction quiz, caregivers scored an average of 7.77 out of 10, with only three questions answered correctly less than 80% of the time (the hardest question was answered correctly 61.5% of the time, and included a partially correct and popular response as a distractor). This suggests that they knew the rules of effective commands, examples of effective

commands, and the rules and script of the time-out sequence. Taken together, the quiz results are a simple confirmation of caregiver engagement with the material at a basic level (during the in-person training).

Overall, all participants who participated in the training showed some degree of improvement on the PSICA and PSI (if they completed it), with some of these changes being statistically reliable. All significant changes in these measures occurred between time 1 and time 3. Either the booster training, or simply maturity effects, seemed necessary to manifest the significant changes reported by the few caregivers who had scores from multiple time points. Amount of consultation attendance or homework completion did not seem to correspond with extent of reliable change in the measures reported above. Several cases showed either no change or negative change when they did not participate in any part of the training, but non-significant positive change also occurred with several caregivers who did participate. It may be the case that the measures selected were not appropriately sensitive to the changes that a non-referred sample of caregivers would experience, and the measures of knowledge and skill were not sufficiently difficult or pertinent to the training content to show meaningful change. Regardless, a larger sample and more stratified training groups would clarify any measured change.

Implementation Outcomes

There were three primary implementation outcomes from this study: acceptability, feasibility, and adoption. To review these terms, acceptability has been defined as the perception among those with direct experience with an intervention that it was "agreeable, palatable, or satisfactory" (Proctor et al., 2011, p. 67). Feasibility refers to how well a program can be put to use within an agency, and any potential barriers to its use (Proctor et al., 2011). Adoption refers

to trying or actually using what was learned from a program (Proctor et al., 2011). Acceptability was assessed using two modified acceptability questionnaires, the AARP and the TEI-SF. Feasibility was primarily assessed with the BTPS and simple turnout/attrition rates. Other observed outcomes pertained to aspects of both feasibility and adoption, including consultation attendance, homework completion, "play count" of the online modules, and completion rates of online modules. The content of the questionnaire responses proved particularly useful in judging exactly what was both acceptable and problematic about the training experience. Thus, a more fine-grained item-level analysis is included after a more descriptive, quantitative account.

Acceptability

Overall, acceptability of the program based on both forms was very high. Table 6 shows that the average rating for each form, across different participant groups, was overwhelmingly positive. If the mean rating for each group was rounded to the nearest anchor-point on the Likert scale, all four groups (first in-person training, online only, booster training, post-training) indicated that they "strongly agree" with all acceptability statements on the AARP. ANOVA revealed no significant differences between groups on the AARP score ($p = 0.70$). Similarly, the TEI-SF showed a high average rating, although the rounded average of the online group was "agree" instead of "strongly agree." Again, ANOVA revealed no significant differences on the TEI-SF between online and in-person groups ($p = 0.92$). The reason for slight differences in AARP and TEI-SF confidence (average of "strongly agree" vs. "agree) may become apparent in the item-level descriptives.

Table 6. Average Mean Item-Rating for Acceptability Forms, by Training Group

Measure	Time 1, In-person			Online Only			Time 2, Booster			Time 3, Post-training		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
AARP	12	5.74	0.38	3	5.71	0.40	8	5.57	0.61	6	5.87	0.27
TEI-SF	12	4.58	0.37	3	4.41	0.39	8	4.50	0.51	6	4.57	0.51

Note: Maximum AARP rating = 6 (strongly agree); maximum TEI-SF rating = 5 (strongly agree). Means displayed in this table represent the averages of all participants' average ratings, not their total ratings. One additional participant who did not fill out the other forms participated in a part of the training at time 2, and submitted an acceptability form.

The AARP item-level responses were mostly homogeneous, with very few items having significant variability. Table 7 shows how similarly each item was treated by respondents. Every single item was generally strongly agreed with by all respondents, with very small standard deviations indicating very little variability in these ratings. The "minimum" column indicates that there was occasional dissent with the majority opinion on some items, most notably item 5 ("This program would not have bad side effects for children"). Only one individual marked "slightly disagree" for this item, but otherwise had a positive impression of the program and rated all other items very favorably. The lack of variability comports with research on the AARP indicating its unitary factor structure (Tarnowski & Simonian, 1992).

The TEI-SF item-level responses, on the other hand, revealed more instances of dissent, but still indicated broad acceptability of the program (Table 8). While most items resulted in general agreement from respondents with little variability, items 3, 6, and 7 were slightly more controversial. Item 3 ("I believe that it would be acceptable to use this program without children's consent") was mostly agreed with, but one caregiver said "disagree" and two said "neutral." Item 6 ("I believe children will experience discomfort when I use the techniques with them") was reverse-coded for the total score, since agreement with the item would *not* be

Table 7. Item-level Descriptives of AARP

Item wording	Rounded response (mean)	*SD*	Minimum
1. This is an acceptable program for children who are in SF	Strongly Agree (5.65)	0.57	Slightly Agree
2. The program should be effective in helping children with their behaviors	Strongly Agree (5.65)	0.49	Agree
3. Some children in SF can benefit from this program	Strongly Agree (5.70)	0.47	Agree
4. I am glad to have started this program	Strongly Agree (5.87)	0.34	Agree
5. This program would not have bad side effects for children	Strongly Agree (5.57)	0.79	Slightly Disagree
6. I like this program	Strongly Agree (5.74)	0.54	Slightly Agree
7. This program is a good way to help children before their disruptive behaviors become severe	Strongly Agree (5.65)	0.49	Agree
8. Overall, this program would help children	Strongly Agree (5.70)	0.47	Agree

Note: $n = 23$; only responses from first in-person, online, and booster training were included because post-training individuals overlapped considerably with other groups and do not represent distinct data. SF = Safe Families. Rounded response is the closest anchor point based on rounding up or down from the average response to the item.

indicative of acceptability. While the average response was disagreement, eight caregivers were "neutral" and two "agreed." Item 7 ("I believe this program is likely to result in permanent improvement") was mostly agreed with, but one caregiver was "neutral." The slightly larger disagreement with some of the TEI-SF items compared to the AARP items may be because of the two-factor structure of the TEI-SF, with some items measuring "Acceptability" and some

Table 8. Item-level Descriptives of TEI-SF

Item wording	Rounded response (mean)	*SD*	Minimum
1. I find this program to be an acceptable way of dealing with a child's problem behavior	Strongly Agree (4.65)	0.49	Agree
2. I would be willing to use this program to improve my relationship with children in SF	Strongly Agree (4.78)	0.42	Agree
3. I believe that it would be acceptable to use this program without children's consent	Agree (4.48)	0.85	Disagree
4. I like the skills used in this program	Strongly Agree (4.78)	0.42	Agree
5. I believe this program is likely to be effective	Strongly Agree (4.70)	0.47	Agree
6. I believe children will experience discomfort when I use the techniques with them	Disagree (2.30)	0.93	Agree*
7. I believe this program is likely to result in permanent improvement	Agree (4.39)	0.58	Neutral
8. I believe it would be acceptable to use this program with parents who have placed their children through SF	Agree (4.48)	0.51	Agree
9. Overall I have a positive reaction to this program	Strongly Agree (4.83)	0.39	Agree

Note: $n = 23$; only responses from first in-person, online, and booster training were included because post-training individuals overlapped considerably with other groups and do not represent distinct data. SF = Safe Families. Rounded response is the closest anchor point based on rounding up or down from the average response to the item. *Item 6 is a reverse coded item, so it was more useful to display the maximum value instead of the minimum value.

measuring "Discomfort" (Kelley et al., 1989). Indeed, all of the items with slight disagreement in the sample did not have obvious semantic analogues on the AARP.

In sum, caregivers whose responses are available indicated a high degree of acceptability across different training groups, and only a few voiced concerns about permanence of the effects, consent of the children, and discomfort for children when using BFT techniques. Responses were mostly limited to individuals who attended the in-person trainings (including the booster training); only a few caregivers in the online group completed the program and returned an acceptability evaluation. Thus, the sample reviewed here may not be indicative of other caregivers whose opinions were not heard.

Feasibility/Adoption

Predictably, the feasibility and adoption outcomes are more mixed than the acceptability outcomes. Furthermore, they were fragmented in terms of which data mapped onto which outcomes. As an example, the BTPS measured barriers to participation in the program (i.e., feasibility) but also the perceived relevance of the program for Safe Families caregivers (i.e., appropriateness). A flexible application of the terminology of Proctor et al. (2011) will be necessary here.

It is safe to say that overall participation and usage of the program was the largest challenge for the BFT program. From the outset, many caregivers signed up but did not participate. There were 32 prospective trainees for the first in-person training, and 13 came (41%). For the online training, there were 36 prospective trainees, 17 of which returned the introductory packets (47%), and 3 of which returned evaluations for the completed online course (8%). During the weeks that phone consultation was offered for trainees, only 8 out of the 30 trainees made use of the consultation times (27% of all 30 trainees), and among these 8, the mean number of attended consultations was 2 (out of the 5 possible times). When trainees did

participate in consultation, reported home practice was low: trainees completed the recommended 5-minute "special time" home-practice session an average of 2.5 out of 7 days a week (skewed by a few participants who practiced more frequently; the modal number was 0). The lack of engagement indicated by these numbers may have contributed to lower reliable change in client outcomes.

From the aforementioned participation rates, it may be tempting to conclude that the online format had much worse participation than the in-person format, but play counts of the online videos tell a different story. Table 9 shows the play counts and mean percentage of the video that was watched for each online module. Across six different videos, there was a total of 162 viewings, ranging from 43 viewings for the first module to 14 for the last module. The overall mean percentage of video watched was 78%, meaning that most online participants who started videos watched the majority of the video. "Giving Effective Commands" had the largest "watched percentage," meaning that caregivers who started this video watched a greater portion of it than other videos. The play count for all modules shows a reliable taper from the first video onward, meaning that caregivers varied in their commitment to finishing the series after starting the first module. Unfortunately, it is not possible to tell from these data whether each play count represents a unique user watching the video anew, or simply a repeat viewing from the same participant. It is likely that there were some repeat viewings, given the play count for the first video is higher than the total number of individuals who participated in the training. Regardless, it seems likely that our assumption (based on returned forms) that only three caregivers completed any online content is too strict. More likely, the online content was used by many more caregivers than just the ones with whom the researchers had significant contact.

Table 9. Descriptives of Video Module Viewing, by Module and Device Type

Category of viewing	Play count	Mean % watched
By Module		
1. Why are good parent-child relationships important?	43	75
2. Building Healthy Parent-Child Relationships	41	67
3. Applying Child-Centered Skills	24	86
4. Using Effective Discipline Techniques	21	78
5. Giving Effective Commands	19	91
6. Applying Effective Discipline Techniques	14	88
By Device		
Desktop	112	72
Tablet	31	86
Mobile Device	19	81

Note: Total number of unique viewers is unknown; data only reflects raw watch count.

The BTPS was the only formal questionnaire associated with this set of implementation outcomes, and the total scores revealed a general lack of perceived barriers to participating in the BFT program. For the twelve caregivers who completed the BTPS at the 6-month follow-up survey, their average rating for the 44 barrier statements on the questionnaire was 1.33 (recall that ratings on the BTPS ranged from 1 – *never a problem* to 5 – *very often a problem*). While only twelve caregivers responded to the 6-month follow-up questionnaire (and thus provided responses for the BTPS), no differences were observed in average barrier ratings between the seven trainees and the five non-participants (p = .85). Nevertheless, responses may not be representative of all of the other participants who did not respond. The overwhelming modal response for all 44 questions was "1 – *never a problem*," and it was typical that only a handful of barriers were endorsed by any one participant. Thus, an item-level analysis was necessary to observe meaningful patterns.

Table 10 shows the more frequently endorsed items of the BTPS, with the most frequent items pertaining to time and scheduling conflicts for caregivers. The BTPS has four qualitative categories for each item, including "stressors/obstacles that compete with treatment" (20 total), "perceived relevance of treatment" (8 total), "treatment demands/issues" (9 total), and "relationship to the therapist" (7 total; in this case, relationship to the trainers). The five items with the highest average rating pertained to the stressors/obstacles category, and most of the selected items in table 10 pertain to competing priorities and time constraints (including items from different categories, e.g., "The program took too long to finish"). In particular, consultation calls were a challenging aspect of the program for these caregivers, taking the top two ranks for highest rating.

It is encouraging that "stressors/obstacles" was the most common category to endorse because the other categories pertained more to the content and service aspects of the training (which are more within the trainers' control). For example, the seven "relationship to the therapist" items only received three endorsements greater than "never" (e.g., "the trainers did not do enough consultation calls," "I had to give too much personal information to the trainers," "the trainers made me feel uncomfortable"), to the exclusion of items indicating more overt disliking or perceived lack of support. Additionally, the only "treatment demands" items that were endorsed frequently pertained to time commitment (e.g., "the program took too long to finish," "the program was more work than I expected), and not to problems with willingness to participate in the first place (e.g., "I disagreed with the program content," "my child refused to participate"). The "perceived relevance" items were moderately rated, with several items included in the selected list in table 10, (e.g., "I lost interest," "it did not seem as important as time went on"), and several items that were never endorsed (e.g., "the program did not seem to

Table 10. Selected Item-level Descriptives of BTPS, in Order of Highest to Lowest Mean Value

Item wording	Rounded response (mean)	*SD*	Frequency of >1	Category
16. I was not available during consultation times	Sometimes (2.92)	1.83	7	Obstacles
4. Scheduling of consultation calls	Seldom (2.25)	1.86	4	Obstacles
6. The program was in conflict with another of my activities	Seldom (2)	1.54	5	Obstacles
34. I did not have time for the work	Seldom (2)	1.65	4	Obstacles
38. Getting time away from my child so I could attend the training	Seldom (2)	1.60	4	Obstacles
42. My job got in the way of completing the program	Seldom (1.83)	1.53	3	Obstacles
14. During the course of the program I experienced a lot of stress in my life	Seldom (1.75)	1.36	3	Obstacles
21. I felt the program did not seem as important as time went on	Seldom (1.67)	1.07	4	Relevance
20. The program added another stressor	Seldom (1.58)	1.00	4	Obstacles
22. I felt this program was more work than expected	Seldom (1.58)	1.24	3	Demands
5. The program took too long to finish	Never (1.42)	1.16	2	Demands
31. The in-person training days were inconvenient	Never (1.42)	1.16	2	Obstacles
39. Figuring out how to call in for the consultation was too difficult	Never (1.42)	1	2	Obstacles
2. Getting to the training site was too difficult	Never (1.42)	1	2	Obstacles
15. I lost interest in the program	Never (1.42)	1	2	Relevance
44. I had trouble with other children at home which made it hard to do the program	Never (1.33)	0.89	2	Obstacles
11. The program was not what I expected	Never (1.25)	0.62	2	Relevance

Note: $n = 12$. Twenty-seven other items excluded for infrequency of endorsement. Frequency of >1 indicates raw number of responses greater than "never a problem (1)." Categories abbreviated; see text.

be working," "my child has new or different problems now"). In general, however, it is important to note that across all items, endorsements of any kind were rare (see table 10, "frequency of >1" column).

In summary, the feasibility/adoption outcomes were a puzzling combination of generally low perceived barrier ratings juxtaposed against what appeared to be low turnout, high attrition, and low participation in post-training activities (consultation, home practice). However, high usage rates of the online content were encouraging. It is difficult to make sense of a group of caregivers who identified very few barriers to participating and still did not participate. Six participants offered their own explanations for their lack of engagement after providing responses on the BTPS. Mostly, these caregivers either 1) noticed the requirement for having a child in the home *after* they signed up or 2) missed important deadlines after signing up. These reasons may be part of a larger implementation challenge that will be elaborated upon further in the discussion.

CHAPTER IV

DISCUSSION

The current study identified several promising preliminary outcomes of a brief out-of-home caregiver training, and also provided insight about aspects of implementation. Both client outcomes and implementation outcomes are considered below.

Client Outcomes

The most encouraging client outcome from the study was the fact that all participants with repeat-responses on measures of child prosocial behavior and parent stress reported changes in the desired direction by the 6-month follow-up. If these increased pro-social behavior and reduced parent stress findings are replicable, then the BFT program may address many of the challenges of out-of-home care that are discussed in the introduction. Caregivers would have access to skills that quantifiably improve child pro-social behavior and some of the parent-related anxieties about out-of-home care. Ergo, two significant risk factors in the etiology of placement instability would be abated.

There are substantial caveats to these forecasts. First, many trainees did not provide post-training or follow-up data. There is simply no way to know from the incomplete data whether the overall pattern of change from the training conforms to this smaller sample of trainees. Second, within this smaller sample, not all trainees reported statistically reliable change on the most successful outcomes. The strength of one's conclusions should rest upon the statistical rigor of the findings, and therefore caution is warranted even with the available single-subject data. Third, all measured results were likely multi-determined from different components of the training (first in-person training, booster training, consultation, online modules, etc.). This makes

it difficult to know exactly which components contributed to the reported changes. There was also no obvious pattern between consultation and reliable change on client outcomes, which would normally be expected for a program traditionally associated with at-home practice. Thus, while the RCI findings may foreshadow the kinds of results that could be expected with a larger-scale implementation of the program, they are only considered an encouraging suggestion.

Client outcomes that did not change significantly included knowledge of behavior principles (KBPAC-R), coding ability, and child negative behavior (SDQ). There may be many reasons for the lack of significant results from these measures, but once again, limiting factors from the design and nature of the data make interpretation a challenge. For coding ability, it may not be entirely accurate to say that the training did not improve this ability. There was no baseline assessment of coding ability, so there may have been some initial increase in coding ability that did not *continue* to increase between Time 1 and Time 2. After all, participants did have relatively high ICC's with the answer key, so they were performing quite well relative to an objective standard (and without any explicit instruction on coding rules). The knowledge of behavior principles measure (KBPAC-R) was a very brief measure (10 items), and may not have been sensitive enough to capture the small changes in knowledge that may result from a prevention program. Furthermore, the training content did not provide explicit instruction on principles of behavior, which may have been needed to produce measurable change. A general implication of these null findings may be that client outcome measures for prevention programs should be highly specific to program content because of their often-abbreviated nature.

The null findings from the SDQ expand upon this implication. The inconclusive results from the SDQ may be because it is primarily sensitive to changes in problematic behaviors, which are not expected to be substantial in a preventative sample. Indeed, four out of the five

subscales on the SDQ pertain to problematic behavior, and the pro-social scale is omitted when calculating the total problem score. Thus, when a caregiver reports an already low frequency of problem behaviors, the SDQ score is not likely to decrease further with intervention (i.e., a floor effect). Compare this to the PSICA, which only measures the frequency of desired pro-social behaviors, and was one of two measures that showed some indication of positive change from the training. This underscores the importance of selecting instruments that are sensitive to the kinds of changes one could expect *a priori* from a given prevention program.

The improvements in parent stress on the PSI may inspire similar assessment considerations in future prevention program evaluation. The appraisal that out-of-home caregivers often felt ill-equipped for their roles (Cooper et al., 1987; Hampson & Tavormina, 1980; Hochstadt et al., 1987; Spielfogel et al., 2011) was one of the primary motivators for the development of the BFT program from the outset: if caregiver stress can be reduced, then placement disruption and coercive parenting may also be reduced. Because the PSI does not measure child psychopathology per se, but rather caregiver perceptions about their role and its associated challenges, it may be another excellent candidate for evaluating prevention programs.

To summarize, the conclusions that flow from client outcomes are twofold. First, preliminary evidence suggests that child pro-social behavior and parent stress may be improved with an abbreviated, pre-service parent training based on PCIT. Second, future prevention programs should consider assessment measures that are content-specific and strength-focused (rather than symptom-focused).

Implementation Outcomes

High regard for the program was the strongest implementation outcome of the current study. Acceptability was high for all participants who provided responses, with very little dissent or differences between forms. The TEI-SF showed slightly more variability between participants than the AARP, which may mean that the TEI-SF is descriptively superior in this context. Similar caveats that apply to the client outcomes also apply to acceptability: there may have been wider dissent if all participants responded. However, there was reason to doubt a "silent majority" of detractors: all participants in the study were part of a tightly-knit and supportive community, and interacted warmly with one another and the trainers during consultation and training. Given this general climate of warmth and openness, it seems likely that significant concerns would have been communicated to the researchers if they existed. One drawback of the anonymity of the acceptability forms was that it could not be linked with other existing data from the participants, and therefore no predictive analysis on acceptability ratings was possible (i.e., which demographic or client outcome factors were related to positive perceptions of the program). On the other hand, anonymity may be essential to obtain honest acceptability ratings.

With regard to all other implementation outcomes, it is necessary to be more integrative because no single outcome appears to explain the observed pattern of engagement. One theme that emerges from this integration of outcomes is that efforts to make the training more accessible were not sufficient to elicit participation. One might expect that the online materials, relatively short trainings, low barrier ratings, lack of monetary cost to trainees or the agency, and high acceptability ratings would have resulted in high engagement. This is inconsistent with the reality of low turnout, high attrition, and low participation in consultation and home practice. Even the higher-than-expected rates of video watching showed a negative taper from Module 1

to Module 6, suggesting that attrition was high even amongst the online trainees who did not return questionnaire materials. One may be tempted to point to the barriers ratings to explain the lack of consultation attendance (the highest rated items involved not having enough consultation times or scheduling conflicts surrounding consultation), but even this is difficult to reconcile with the fact that consultation times were structured around the trainees' requests for specific weekend and evening times (and modifications of these times subsequent to time change requests still did not seem to affect participation). If the training appeared to be well-liked and easy to do, what accounts for these outcomes?

Perhaps the simplest reason that any given implementation attempt stumbles is that it lacks a guiding theory. The current study made use of the terminological framework of Proctor et al. (2011) in 1) analyzing recent implementation advances in preventive parenting groups, and 2) prioritizing and defining new outcomes of interest. However, this was largely where the use of implementation theory stopped. There are many implementation models focused solely on the *logistical* and *practical* strategies involved with integrating evidence-based practice into new settings; a review by Tabak, Khoong, Chambers, and Brownson (2013) discussed 61 such models. Another review by Chambers (2014) organizes some of these strategic models by level of analysis (e.g., individual helper, organization, community, policy, etc.), flexibility, and utility for research efforts. Without a guiding theory, the current study may have lacked a structured plan for managing all of the messy challenges of implementation. Unfortunately, such a shortcoming is reportedly common in preventive parenting programs, as described by a recent review by Olofsson, Skoog, and Tillfors (2016): "…even though different ecological models were often referred to in the introductions, they did not inform the evaluations… neither the conceptual terminology nor the ecological system has been fully investigated" (p. 77). Using a

previously tested and unified implementation strategy may improve future implementation outcomes, or at least add to our understanding of how these strategies work.

One example of a neglected aspect of implementation strategy in the current study was agency readiness. Agency readiness refers to how motivated and prepared the different stakeholders and leaders in an organization are before implementing a new program (Olofsson et al., 2016). In one study of a parenting program implementation by Romney, Israel, and Zlatevski (2014), four out of the six participating agencies completed a self-assessment of how prepared they were to execute all of the different procedures and logistical aspects of an evidence-based parenting program. Subsequent parent completion rates of the program were much higher in the four agencies that first conducted a formal assessment of their readiness. In the current study, it is possible that a more formal assessment of agency readiness may have identified key areas of increased vigilance and management for the training team and the agency that could have contributed to increased caregiver completion rates.

Effective communication between stakeholders is related to agency readiness and represents another factor in implementation success that could have been more carefully considered. First and foremost, identification of stakeholders comes before establishing regular communication between them. For the current study, there was one agency leader who managed all of the logistics of the trainings. Beyond this individual, the researchers did not make additional efforts to identify stakeholders at other levels of the organization (e.g., the caregivers themselves, leaders amongst the caregivers, funding sources, other organizational leadership, etc.). Indeed, as the program unfolded, it was revealed that several of the caregivers in the training held important administrative positions in Safe Families, and may have been willing to handle important logistical tasks and dissemination of information. The value of having more

than one influential individual talking about the program and providing reminders and encouragement about it during agency gatherings was not sufficiently recognized in the current study. It may have also been helpful for answering questions and concerns in a timely fashion, which has been referred to as "rapid-cycle problem-solving" and has been shown to improve the outcomes of even the best-planned implementation efforts (Akin et al., 2014).

Returning to the individual level of implementation strategy, it is possible that engagement with a program cannot be predicted by its ease of use, its acceptability, or even its potential to buffer against stress, but rather its *value*. In other words, the researchers may have focused too much on reducing reasons to *disengage* from the training, rather than increasing reasons to *engage* with the training. One way that this has been accomplished in PCIT implementation research is through a "local needs assessment," or an assessment of the kinds of challenges that are faced by the prospective trainees with their population of youths (Beveridge et al., 2015). This "knowledge utilization" approach places higher value on the prospective trainees' perspectives regarding exactly what is needed, and can increase the priority of the training. One might object to this explanation by citing that trainees from the current study did not strongly endorse the "perceived relevance" items on the barriers questionnaire, but the same was true of all categories of barriers, and the "perceived relevance" items as a whole were still moderately rated in comparison to the other categories. Furthermore, one wonders whether a program with higher perceived value would have still resulted in endorsements of time and scheduling conflicts.

Another trainee-level reason for lack of engagement is that the pre-service nature of the training may have removed a possible motivator for a lot of out-of-home caregivers: a sense of "urgency." Much of the findings about parent stress discussed in the introduction pertained to

caregivers who were asked to reflect upon the challenges of having *been* an out-of-home caregiver, rather than what it might be like to be one. Thus, it could be that imagined challenges might be less motivating than experienced ones. None of the trainees in the current study were faced with the pressure of a challenging child in their home, and may not have felt a need to seek help. Indeed, an interesting anecdotal observation from the from the researchers is that some of the participants who participated the most reliably had mild behavioral concerns with their own biological children, and not concerns about their role as a prospective out-of-home caregiver. In sum, a paradox of prevention programs is apparent, whereby the very advantage of a pre-service format (e.g., it is completed before things become challenging) is also what makes it seem unimportant from the perspective of the trainees.

The words of Olofsson et al. (2016) nicely summarize the implementation outcomes of this study: "it seems as if it is not enough to deem a program as acceptable, appropriate and feasible, and being willing to try it out" (p. 76). Actively applying a previously tested implementation model that is appropriate for the parameters of the proposed implementation seems like a good prescription for future preventative parenting programs. More specifically, developing a better sense for what is needed at helper and agency levels, while extending lines of communication into multiple levels of the agency, could maximize the likelihood of program success and continued use of the content. This includes all of the aforementioned considerations about agency readiness, local evidence gathering, and increasing program value in the eyes of its immediate beneficiaries.

Strengths and Limitations

The current evaluation of the BFT program has several strengths. First, it used an established implementation vocabulary and framework to guide hypothesis testing and implementation innovation in an out-of-home context. We have good evidence that well-established parenting interventions can be effective for displaced children, but more implementation-based work is needed to promote its penetrance to populations and agencies in need. Second, it employed novel delivery methods of PCIT content, including an abbreviated and preventative in-person format and a video-based online format. These innovations are still new but warrant attention in a resource-strapped child welfare system. Third, the multiple-outcome approach allowed for a better understanding of the kinds of measures that may be useful for future implementation efforts. In particular, the promising findings from strength-based measures (e.g., the PSICA) represent an exciting way for implementation researchers to operationalize and assess progress. Lastly, the current study established a very high interest and regard for this type of training amongst caregivers, and may be heartening for aspiring program designers.

Numerous limitations have been described throughout the discussion, but warrant some emphasis and general summary. The most obvious limitation was sample size, which hampered every analysis and meaningful inference attempted. Problems with sample size were multiply determined, including drop-out before the training, attrition during the training, and lack of engagement in the post-training activities. Conclusions about any of the analyses presented must be made with caution. Second, many of the questionnaires were not detailed enough to capture valuable information about the training experience. A mixed-methods approach may have clarified some of the puzzling findings from the questionnaires. Indeed, many participants offered their own reasons why they did not participate when the Barriers to Treatment

Participation Scale did not adequately capture their reasons in the 44-item list. A mixed-methods approach may have also clarified the meaning of some of the web-based metrics regarding video usage. Third, a unifying implementation framework was lacking. Using a previously tested framework would have likely improved multiple outcomes, most notably engagement and retention. Lastly, random assignment and multiple assessment points between training components would allow for stronger inferences about the effects of the training.

General Conclusions

This pilot study of a preventive parent training program study suggests that caregivers strongly approve of such a program, and many of them experienced benefits for their own families before they started hosting displaced children. There is good reason to believe that larger efforts of this nature will be successful and well-liked, though caution is warranted in making this conclusion. Any efforts to expand or replicate a program like BFT should thoroughly incorporate an implementation framework and additional implementation outcomes to maximize engagement. With planning, and some luck, future parenting programs that heed the current study's admonitions will provide a more robust buffer against the risks that displaced children and families face.

REFERENCES

Abidin, R. (2012). Parenting Stress Index 4th Edition-Professional Manual. *Lutz, FL: Par Inc*.

Acevedo-Polakovich, I. D., Niec, L. N., Barnett, M. L., Bell, K. M., Aguilar, G., Vilca, J., . . . Peer, S. O. (2014). Exploring the role of natural helpers in efforts to address disparities for children with conduct problems. *Children and Youth Services Review, 40*, 1-5.

Akin, B. A., Mariscal, S. E., Bass, L., McArthur, V. B., Bhattarai, J., & Bruns, K. (2014). Implementation of an evidence-based intervention to reduce long-term foster care: Practitioner perceptions of key challenges and supports. *Children and Youth Services Review, 46*, 285-293.

Barber, J. G., & Delfabbro, P. H. (2006). Psychosocial well-being and placement stability in foster care: Implications for policy and practice. *Promoting resilience in child welfare*, 157-174.

Barkley, R. A. (1987). What is the role of group parent training in the treatment of ADD children? *Journal of Children in Contemporary Society, 19*, 143-151.

Beveridge, R. M., Fowles, T. R., Masse, J. J., McGoron, L., Smith, M. A., Parrish, B. P., . . . Widdoes, N. (2015). State-wide dissemination and implementation of parent–child interaction therapy (PCIT): Application of theory. *Children and Youth Services Review, 48*, 38-48.

Boggs, S. R., Eyberg, S. M., Edwards, D. L., Rayfield, A., Jacobs, J., Bagner, D., & Hood, K. K. (2005). Outcomes of parent-child interaction therapy: A comparison of treatment completers and study dropouts one to three years later. *Child & Family Behavior Therapy, 26*, 1-22.

Brightman, R. P., Baker, B. L., Clark, D. B., & Ambrose, S. A. (1982). Effectiveness of alternative parent training formats. *Journal of Behavior Therapy and Experimental Psychiatry, 13*, 113-117.

Brown, J. D. (2008). Foster parents' perceptions of factors needed for successful foster placements. *Journal of Child and Family Studies, 17*, 538-554.

Burns, B. J., Phillips, S. D., Wagner, H. R., Barth, R. P., Kolko, D. J., Campbell, Y., & Landsverk, J. (2004). Mental Health Need and Access to Mental Health Services by Youths Involved With Child Welfare: A national survey. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*, 960-970.

Campbell, S. B. (1995). Behavior problems in preschool children: a review of recent research. *Child Psychology & Psychiatry & Allied Disciplines*.

Campbell, S. B., & Ewing, L. J. (1990). Follow-up of hard-to-manage preschoolers: Adjustment at age 9 and predictors of continuing symptoms. *Journal of Child Psychology and Psychiatry, 31*, 871-889.

Cavazzi, T., Guilfoyle, A., & Sims, M. (2010). A phenomenological study of foster caregivers' experiences of formal and informal support. *Illinois Child Welfare, 5*, 125-141.

Chaffin, M., Silovsky, J. F., Funderburk, B., Valle, L. A., Brestan, E. V., Balachova, T., . . . Bonner, B. L. (2004). Parent-child interaction therapy with physically abusive parents: efficacy for reducing future abuse reports. *Journal of Consulting and Clinical Psychology, 72*, 500.

Chaffin, M., Valle, L. A., Funderburk, B., Gurwitch, R., Silovsky, J., Bard, D., . . . Kees, M. (2009). A motivational intervention can improve retention in PCIT for low-motivation child welfare clients. *Child maltreatment, 14*, 356-368.

Chamberlain, P., & Mihalic, S. F. (1998). *Multidimensional treatment foster care*: Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado at Boulder.

Chamberlain, P., Price, J., Leve, L. D., Laurent, H., Landsverk, J. A., & Reid, J. B. (2008). Prevention of behavior problems for children in foster care: Outcomes and mediation effects. *Prevention Science, 9*, 17-27.

Chamberlain, P., Price, J. M., Reid, J. B., Landsverk, J., Fisher, P. A., & Stoolmiller, M. (2006). Who disrupts from placement in foster and kinship care? *Child Abuse & Neglect, 30*, 409-424.

Chambers, D. (2014). Guiding theory for dissemination and implementation research. In R. Beidas & P. Kendall (Eds.), *Dissemination and implementation of evidence-based practices in child and adolescent mental health*. New York, NY: Oxford University Press.

Clausen, J. M., Landsverk, J., Ganger, W., Chadwick, D., & Litrownik, A. (1998). Mental health problems of children in foster care. *Journal of Child and Family Studies, 7*, 283-296.

Cooley, M. E., Veldorale-Griffin, A., Petren, R. E., & Mullis, A. K. (2014). Parent–Child Interaction Therapy: A meta-analysis of child behavior outcomes and parent stress. *Journal of Family Social Work, 17*, 191-208.

Cooper, C. S., Peterson, N. L., & Meier, J. H. (1987). Variables associated with disrupted placement in a select sample of abused and neglected children. *Child Abuse & Neglect, 11*, 75-86.

Courrégé, S., White, J., Niec, L., Solomon, D., & Peer, S. (2015). *The problem with measuring only problems in clinical child research and practice: validation and proposed application of the child strengths inventory*. Paper presented at the PCIT International Convention, Pittsburgh, PA.

Daly, B. P., Buchanan, C., Dasch, K., Eichen, D., & Lenhart, C. (2012). Prevention of Conduct Disorder: Do We Have Reason to Be Hopeful? *The Oxford Handbook of Prevention in Counseling Psychology*, 268.

Dorsey, S., Farmer, E. M. Z., Barth, R. P., Greene, K. M., Reid, J., & Landsverk, J. (2008). Current status and evidence base of training for foster and treatment foster parents. *Children and Youth Services Review, 30*, 1403-1416.

Eyberg, S. (1988). Parent-child interaction therapy: Integration of traditional and behavioral concerns. *Child & Family Behavior Therapy, 10*, 33-46.

Eyberg, S., Nelson, M., & Boggs, S. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology, 37*, 215-237.

Eyberg, S., Nelson, M., Ginn, N., Bhuiyan, N., & Boggs, S. (2013). Dyadic Parent-Child Interaction Coding System (DPICS): Comprehensive Manual for Research and Training: Gainesville, FL: PCIT International.[Links].

Eyberg, S. M., Boggs, S. R., & Rodriguez, C. M. (1993). Relationships between maternal parenting stress and child disruptive behavior. *Child & Family Behavior Therapy, 14*, 1-9.

Festinger, T., & Baker, A. J. L. (2013). The quality of evaluations of foster parent training: An empirical review. *Children and Youth Services Review, 35*, 2147-2153.

Finn, C. A., & Sladeczek, I. E. (2001). Assessing the social validity of behavioral interventions: A review of treatment acceptability measures. *School Psychology Quarterly, 16*, 176.

Fisher, P. A., Burraston, B., & Pears, K. (2005). The early intervention foster care program: Permanent placement outcomes from a randomized trial. *Child maltreatment, 10*, 61-71.

Fisher, P. A., Gunnar, M. R., Chamberlain, P., & Reid, J. B. (2000). Preventive intervention for maltreated preschool children: Impact on children's behavior, neuroendocrine activity, and foster parent functioning. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*, 1356-1364.

Fisher, P. A., & Stoolmiller, M. (2008). Intervention effects on foster parent stress: Associations with child cortisol levels. *Development and psychopathology, 20*, 1003.

Foster, E. M., & Jones, D. (2006). Can a costly intervention be cost-effective?: An analysis of violence prevention. *Archives of General Psychiatry, 63*, 1284-1291.

Furtkamp, E., Giffort, D., & Schiers, W. (1982). In-class evaluation of behavior modification knowledge: Parallel tests for use in applied settings. *Journal of Behavior Therapy and Experimental Psychiatry, 13*, 131-134.

Garland, A. F., Hough, R. L., McCabe, K. M., Yeh, M., Wood, P. A., & Aarons, G. A. (2001). Prevalence of psychiatric disorders in youths across five sectors of care. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 409-418.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 1337-1345.

Hampson, R. B., & Tavormina, J. B. (1980). Feedback from the experts: A study of foster mothers. *Social Work, 25*, 108-113.

Harwood, M. D., & Eyberg, S. M. (2006). Child-directed interaction: Prediction of change in impaired mother–child functioning. *Journal of Abnormal Child Psychology, 34*, 323-335.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association, 82*, 1147-1149.

Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association, 81*, 991-999.

Hochstadt, N. J., Jaudes, P. K., Zimo, D. A., & Schachter, J. (1987). The medical and psychosocial needs of children entering foster care. *Child Abuse & Neglect, 11*, 53-62.

Horwitz, S. M., Chamberlain, P., Landsverk, J., & Mullican, C. (2010). Improving the mental health of children in child welfare through the implementation of evidence-based parenting interventions. *Administration and Policy in Mental Health and Mental Health Services Research, 37*, 27-39.

Jackson Foster, L. J., Phillips, C. M., Yabes, J., Breslau, J., O'Brien, K., Miller, E., & Pecora, P. J. (2015). Childhood behavioral disorders and trauma: Predictors of comorbid mental disorders among adult foster care alumni. *Traumatology, 21*, 119-127.

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

Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology, 36*, 567-589.

Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior Analysis, 13*, 259-273.

Kazdin, A. E. (2005). *Parent management training: Treatment for oppositional, aggressive, and antisocial behavior in children and adolescents*: Oxford University Press, New York, NY.

Kazdin, A. E., Holland, L., Crowley, M., & Breton, S. (1997). Barriers to Treatment Participation Scale: Evaluation and validation in the context of child outpatient treatment. *Child Psychology & Psychiatry & Allied Disciplines, 38*, 1051-1062.

Kelley, M. L., Heffer, R. W., Gresham, F. M., & Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology and Behavioral Assessment, 11*, 235-247.

Keppel, G., & Wickens, T. (2004). The linear model and its assumptions. In G. Keppel & T. Wickens (Eds.), *Design and analysis: A researcher's handbook*. Upper Saddle River, NJ: Pearson Education, Inc.

Lanier, P., Kohl, P. L., Benz, J., Swinger, D., Moussette, P., & Drake, B. (2011). Parent–child interaction therapy in a community setting: examining outcomes, attrition, and treatment setting. *Research on Social Work Practice, 21*, 689-698.

Leathers, S. J., Spielfogel, J. E., Gleeson, J. P., & Rolock, N. (2012). Behavior problems, foster home integration, and evidence-based behavioral interventions: What predicts adoption of foster children? *Children and Youth Services Review, 34*, 891-899.

Leve, L. D., Harold, G. T., Chamberlain, P., Landsverk, J. A., Fisher, P. A., & Vostanis, P. (2012). Practitioner review: Children in foster care–vulnerabilities and evidence-based interventions that promote resilience processes. *Journal of Child Psychology and Psychiatry, 53*, 1197-1211.

Linares, L. O., Montalto, D., Li, M., & Oza, V. S. (2006). A promising parenting intervention in foster care. *Journal of Consulting and Clinical Psychology, 74*, 32.

Lyon, A. R., & Budd, K. S. (2010). A community mental health implementation of parent–child interaction therapy (PCIT). *Journal of Child and Family Studies, 19*, 654-668.

MacGregor, T. E., Rodger, S., Cummings, A. L., & Leschied, A. W. (2006). The Needs of Foster Parents: A Qualitative Study of Motivation, Support, and Retention. *Qualitative Social Work: Research and Practice, 5*, 351-368.

Maughan, D. R., Christiansen, E., Jenson, W. R., Olympia, D., & Clark, E. (2005). Behavioral parent training as a treatment for externalizing behaviors and disruptive behavior disorders: A meta-analysis. *School Psychology Review, 34*, 267.

McLoughlin, C. S. (1985). Utility and efficacy of knowledge of behavioral principles as applied to children. *Psychological reports, 56*, 463-467.

McNeil, C., & Hembree-Kigin, T. (2010a). Teaching Child Directed Interaction. In C. McNeil & T. Hembree-Kigin (Eds.), *Parent-Child Interaction Therapy* (2nd ed.). New York, NY: Springer.

McNeil, C., & Hembree-Kigin, T. (2010b). Teaching Parent Directed Interaction. In C. McNeil & T. Hembree-Kigin (Eds.), *Parent-Child Interaction Therapy* (2nd ed.). New York, NY: Springer.

McNeil, C. B., Herschell, A. D., Gurwitch, R. H., & Clemens-Mowrer, L. (2005). Training foster parents in parent-child interaction therapy. *Education and Treatment of Children*, 182-196.

Mersky, J. P., Topitzes, J., Grant-Savela, S. D., Brondino, M. J., & McNeil, C. B. (2016). Adapting Parent–Child Interaction Therapy to Foster Care Outcomes From a Randomized Trial. *Research on Social Work Practice, 26*, 157-167.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychological Review, 100*, 674.

Newton, R. R., Litrownik, A. J., & Landsverk, J. A. (2000). Children and youth in foster care: Disentangling the relationship between problem behaviors and number of placements. *Child Abuse & Neglect, 24*, 1363-1374.

Niec, L. N., Acevedo-Polakovich, I. D., Abbenante-Honold, E., Christian, A. S., Barnett, M. L., Aguilar, G., & Peer, S. O. (2014). Working together to solve disparities: Latina/o parents' contributions to the adaptation of a preventive intervention for childhood conduct problems. *Psychological services, 11*, 410.

Niec, L. N., Peer, S. O., Courrégé, S., & Solomon, D. (2016). Psychosocial Strengths Inventory for Children and Adolescents: Preliminary psychometrics and potential applications. *Manuscript in preparation. Center for Children, Families, and Communities, Mt. Pleasant, MI.*

O'Dell, S. L., Tarler-Benlolo, L., & Flynn, J. M. (1979). An instrument to measure knowledge of behavioral principles as applied to children. *Journal of Behavior Therapy and Experimental Psychiatry, 10*, 29-34.

Olofsson, V., Skoog, T., & Tillfors, M. (2016). Implementing group based parenting programs: A narrative review. *Children and Youth Services Review, 69*, 67-81.

Owens, P. L., Hoagwood, K., Horwitz, S. M., Leaf, P. J., Poduska, J. M., Kellam, S. G., & Ialongo, N. S. (2002). Barriers to children's mental health services. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*, 731-738.

Pecora, P. J., Kessler, R. C., O'Brien, K., White, C. R., Williams, J., Hiripi, E., . . . Herrick, M. A. (2006). Educational and employment outcomes of adults formerly placed in foster care: Results from the Northwest Foster Care Alumni Study. *Children and Youth Services Review, 28*, 1459-1481.

Pecora, P. J., Kessler, R. C., Williams, J., O'Brien, K., Downs, A. C., English, D. J., . . . Holmes, K. (2005). Improving family foster care: findings from the Northwest Foster Care Alumni Study *Foster care alumni studies: stories from the past to shape the future* (pp. 64): Casey Family Programs.

Pecora, P. J., White, C. R., Jackson, L. J., & Wiggins, T. (2009). Mental health of current and former recipients of foster care: A review of recent studies in the USA. *Child & Family Social Work, 14*, 132-146.

Powell, N. R., Lochman, J. E., & Boxmeyer, C. L. (2007). The prevention of conduct problems. *International Review of Psychiatry, 19*, 597-605.

Price, J. M., Chamberlain, P., Landsverk, J., & Reid, J. (2009). KEEP foster-parent training intervention: Model description and effectiveness. *Child & Family Social Work, 14*, 233-242.

Price, J. M., Chamberlain, P., Landsverk, J., Reid, J. B., Leve, L. D., & Laurent, H. (2008). Effects of a foster parent training intervention on placement changes of children in foster care. *Child maltreatment, 13*, 64-75.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*, 65-76.

Romney, S., Israel, N., & Zlatevski, D. (2014). Exploration-stage implementation variation: Its effect on the cost-effectiveness of an evidence-based parenting program. *Zeitschrift für Psychologie, 222*, 37.

Rork, K. E., & McNeil, C. B. (2011). Evaluation of foster parent training programs: A critical review. *Child & Family Behavior Therapy, 33*, 139-170.

Self-Brown, S., Valente, J. R., Wild, R. C., Whitaker, D. J., Galanter, R., Dorsey, S., & Stanley, J. (2012). Utilizing benchmarking to study the effectiveness of parent–child interaction therapy implemented in a community setting. *Journal of Child and Family Studies, 21*, 1041-1049.

Shanley, J. R., & Niec, L. N. (2010). Coaching parents to change: The impact of in vivo feedback on parents' acquisition of skills. *Journal of Clinical Child & Adolescent Psychology, 39*, 282-287.

Shaw, D. S. (2013). Future directions for research on the development and prevention of early conduct problems. *Journal of Clinical Child & Adolescent Psychology, 42*, 418-428.

Solomon, D. T., Niec, L. N., & Schoonover, C. E. (2017). The impact of foster parent training on parenting skills and child disruptive behavior: a meta-analysis. *Child maltreatment, 22*, 3-13.

Spielfogel, J. E., Leathers, S. J., Christian, E., & McMeel, L. S. (2011). Parent management training, relationships with agency staff, and child mental health: Urban foster parents' perspectives. *Children and Youth Services Review, 33*, 2366-2374.

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical child and family psychology review, 13*, 254-274.

Tabak, R. G., Khoong, E. C., Chambers, D., & Brownson, R. C. (2013). Models in dissemination and implementation research: useful tools in public health services and systems research. *Frontiers in Public Health Services and Systems Research, 2*, 8.

Tarnowski, K. J., & Simonian, S. J. (1992). Assessing treatment acceptance: The abbreviated acceptability rating profile. *Journal of Behavior Therapy and Experimental Psychiatry, 23*, 101-106.

Thomas, R., & Zimmer-Gembeck, M. J. (2007). Behavioral outcomes of Parent-Child Interaction Therapy and Triple P-Positive Parenting Program: A review and meta-analysis. *Journal of Abnormal Child Psychology, 35*, 475-495.

Tiano, J. D., & McNeil, C. B. (2014). Mothers' and fathers' knowledge of behavioral principles as applied to children: Data from a normative sample. *Behavioral Development Bulletin, 19*, 25.

Timmer, S. G., Llrquiza, A. I., Herschell, A. D., McGrath, J. M., Zebell, N. M., Porter, A. L., & Vargas, E. C. (2006). Parent-child interaction therapy: application of an empirically supported treatment to maltreated children in foster care. *Child Welfare, 85*.

Timmer, S. G., Sedlar, G., & Urquiza, A. J. (2004). Challenging children in kin versus nonkin foster care: Perceived costs and benefits to caregivers. *Child maltreatment, 9*, 251-262.

Vanschoonlandt, F., Vanderfaeillie, J., Van Holen, F., De Maeyer, S., & Robberechts, M. (2013). Parenting stress and parenting behavior among foster mothers of foster children with externalizing problems. *Children and Youth Services Review, 35*, 1742-1750.

Ward, M. A., Theule, J., & Cheung, K. (2016). Parent–child interaction therapy for child disruptive behaviour disorders: A meta-analysis. *Child & Youth Care Forum*.

Zisser, A., & Eyberg, S. (2010). Parent-child interaction therapy and the treatment of disruptive behavior disorders. *Evidence-based psychotherapies for children and adolescents*, 179-193.