

8-6-2010

# Using Benchmarking Methodology to Evaluate the Effectiveness of In-Home Parent-Child Interaction Therapy (PCIT)

Jessica R. Valente

Georgia State University, [jessicarvalente@gmail.com](mailto:jessicarvalente@gmail.com)

Follow this and additional works at: [http://digitalarchive.gsu.edu/iph\\_theses](http://digitalarchive.gsu.edu/iph_theses)

---

## Recommended Citation

Valente, Jessica R., "Using Benchmarking Methodology to Evaluate the Effectiveness of In-Home Parent-Child Interaction Therapy (PCIT)" (2010). *Public Health Theses*. Paper 135.

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

## **ABSTRACT**

JESSICA VALENTE

Using benchmarking methodology to evaluate the effectiveness of in-home Parent-Child Interaction Therapy (PCIT)

(Under the direction of Dr. Shannon Self-Brown, Faculty Member)

Benchmarking offers community practitioners more systematic judgments about research effectiveness when control groups are not feasible, while also providing a standard for program transportability from clinical to community settings. The purpose of the current study was to outline the necessary decisions, calculations, and strengths and limitations of applying benchmarking methodologies to a behavioral parent training (BPT) program, a field in which benchmarking remains relatively underutilized. The implementation of in-home Parent-Child Interaction Therapy (PCIT), an evidence-based practice shown to be successful in reducing child maltreatment and neglect, was evaluated as a case study of the application of benchmarking. Of those parents that completed in-home PCIT, a significant reduction was seen for pre-post ECBI scores. Six randomized controlled trials (RCTs) were established as benchmarks based on similarity in parent and child demographics as well as use of the ECBI as a primary measure. Effect sizes of each benchmark study were aggregated to create a single benchmark effect size for treatment and control groups, respectively. The effect size of the current study was found to be significantly superior to the control benchmark effect size but not significantly equivalent to the treatment benchmark effect size. Although the current study demonstrates the use of benchmarking in community research, the need for further guidelines is critical for researchers.

**INDEX WORDS:** benchmarking, effectiveness research, Parent-Child Interaction Therapy

USING BENCHMARKING METHODOLOGY TO EVALUATE THE EFFECTIVENESS OF

IN-HOME PARENT-CHILD INTERACTION THERAPY

by

JESSICA R. VALENTE

B.S., GEORGIA INSTITUTE OF TECHNOLOGY

A Thesis Submitted to the Graduate Faculty  
of Georgia State University in Partial Fulfillment  
of the Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA

2010

USING BENCHMARKING METHODOLOGY TO EVALUATE THE EFFECTIVENESS OF  
IN-HOME PARENT-CHILD INTERACTION THERAPY

by

JESSICA R. VALENTE

Approved:

Shannon Self-Brown, Ph.D.  
Committee Chair

Dan Whitaker, Ph.D.  
Committee Member

## ACKNOWLEDGEMENTS

I would like to first thank my thesis chairperson, Dr. Shannon Self-Brown, for not only her incredibly helpful advice throughout the thesis process, but also her dedication to her students as a mentor. The guidance and opportunities she has shown me both as a thesis student and graduate research assistant will be experiences that I will continue to be grateful for as I progress towards my future career goals. I would also like to express my gratitude to Dr. Daniel Whitaker, whose witty jokes made long meetings worthwhile, and the rest of faculty and staff at the Center for Healthy Development. I have been extremely lucky to work with such an extraordinary group of people. It will be difficult to find such a productive but fun environment like CHD. I must also extend a big thank-you to Rob Wild, whose statistical expertise (and patience with my inexperienced questions!) was truly a lifesaver.

This acknowledgements section would certainly not be complete without thanking the engineer in my life, Ryan DeMars, whose brains, patience, and understanding seemed to miraculously solve any problem I ran into along the course of graduate school. To my parents and sister, thank you for all of your support and tolerance with my lengthy life goals.

## AUTHOR'S STATEMENT

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

Jessica R. Valente\_\_\_\_\_

Signature of Author

## NOTICE TO BORROWERS

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

The author of this thesis is:

Jessica Renee Valente  
6640 Akers Mill Rd. #4338  
Atlanta, GA 30339

The Chair of the committee for this thesis is:

Shannon Self-Brown, Ph.D.  
Center for Healthy Development  
Georgia State University  
P.O. Box 3995  
Atlanta, GA 30302-3995

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

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

## CURRICULUM VITAE

**Jessica R. Valente**  
6640 Akers Mill Rd. #4338  
Atlanta, GA 30339  
jessicarvalente@gmail.com  
404-376-1424

### Education

**Georgia State University – Atlanta, GA**

Master of Public Health, 2010

**Georgia Institute of Technology– Atlanta, GA**

Bachelor of Science, Biology, Minor in Psychology 2008

### Professional Experience

**Georgia State University, Atlanta, GA**

Graduate Research Assistant

**Jan. – Aug. 2010**

**National SafeCare Training & Research Center, Atlanta, GA**

Curriculum Adaptations Practicum

**Jan. – Aug. 2010**

**Center for Leadership in Disability Fellowship Program**

**Jan. – Aug. 2009**

**St. Joseph's Mercy Care, Atlanta, GA**

Medical Referrals Volunteer

**Feb. – current**

### Presentations

Self-Brown, S., Massetti, G., Chen, J., Schulden, J., & **Valente, J.** (November, 2009). The Impact of the Washington, DC Sniper Attacks on Youth Mental Health. A poster presented at the 44th Annual Convention of the Association for Behavior and Cognitive Therapy, New York.

Self-Brown, S., Galanter, R., & **Valente, J.** (June, 2010). Evaluation of PCIT delivered in the Home Setting. Presentation at the American Professional Society on the Abuse of Children 18 Annual Colloquium, New Orleans.

**Valente, J.** (November 2010). Implementing Cognitive-Behavioral Therapy in the Community: Using Benchmarking to Validate the Effectiveness of PCIT. A poster presented at the 44<sup>th</sup> Annual Convention of the Association for Behavioral and Cognitive Therapy, San Francisco.

## TABLE OF CONTENTS

Acknowledgments.....	iii
Introduction.....	1
Public Health problem.....	1
Behavioral Parent Training programs as strategies.....	2
Home-based BPTs.....	2
Evaluation of program effectiveness in community settings .....	3
Benchmarking .....	4
Purpose of current study.....	6
Methods.....	7
Participants .....	7
Measures.....	8
Study Design and Statistical Analyses .....	8
Home-based PCIT .....	9
PCIT Training for Community Therapists .....	9
Benchmarking Decisions.....	10
Calculations.....	12
Results.....	14
Treatment Completion.....	14
ECBI scores pre-post change .....	15
Benchmarking Results.....	15
Discussion.....	17
Benchmarking Results.....	17
Limitations of Current Study and Future Directions with Benchmarking Methodology. ....	19
Conclusions.....	22
REFERENCES.....	23

## Introduction

### *Public Health problem*

According to the 4<sup>th</sup> National Incidence Study of Child Abuse and Neglect (NIS-4), child maltreatment, though still vastly unreported and underestimated, resulted in nearly 3 million child victims of abuse and neglect. Of these incident cases, approximately 835,000 children were abused and 2,251,600 children were neglected, with the risk of any form of child maltreatment affecting 1 in every 25 children (Sedlak, et al., 2010). The alarming prevalence of neglect and abuse of children, particularly those under the age of seven (U.S. Department of Health and Human Services, 2010), not only contributes to a growing public health problem, but child maltreatment also has lifelong negative consequences. Victims of child maltreatment are often at greater risk for pediatric hospitalization (Lanier, Jonson-Reid, Stahlschmidt, Drake, & Constantino, 2010), poor physical health outcomes as an adult (Irish, Kobayashi, & Delahanty, 2010), behavioral problems and learning disabilities (Thompson & Wyatt, 1999), risk-taking behaviors such as poor sexual decisions (Houck, Nugent, Lescano, Peters, & Brown, 2010), cigarette smoking (Topitzes, Mersky, & Reynolds, 2010), and substance abuse (Hussey, Chang, & Kotch, 2006), and numerous psychiatric disorders (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002). In an era where evidence-based programs provide the foundation for public health action, behavioral parent training programs (BPTs) have been suggested as an appropriate approach to the prevention of child maltreatment (Barth, 2009; Barth, et al., 2005; Dore & Lee, 1999; Whitaker, Lutzker, & Shelley, 2005).

### *Behavioral Parent Training programs as strategies*

Parents often struggle with their child's behaviors; in fact, many children who are referred to Child Protection Services have disproportionate rates of behavioral problems as well (Barth, 2009). BPTs, which can have dual purposes as agents of both prevention and intervention, usually focus on teaching parents the skills they need to manage their child's behavior. When effectively administered, BPTs can help parents avoid negative interactions such as physical discipline, as well as improve expectations of their children's behavior (Barth, 2009; Centers for Disease Control and Prevention, 2009; McMahon & Forehand, 2005). Although many of these programs are not directly intended to reduce neglect and abuse cases, the prevention of child maltreatment is a promising end result that can follow teaching at-risk parents positive interactions and empathy (Hakman, Chaffin, Funderburk, & Silovsky, 2009). However, of the 12 evidence-based BPTs currently available with promising research evidence or better according to the California Evidence Based Clearinghouse (<http://www.cebc4cw.org/>), only three have been formally shown to be effective in actually reducing child neglect and abuse: Parent-Child Interaction Therapy (PCIT) (Chaffin, et al., 2004), Triple P (Prinz, Sanders, Shapiro, Whitaker, & Lutzker, 2009), and SafeCare (Gershater-Molko, Lutzker, & Wesch, 2002). Yet, many are being implemented in child welfare systems to address abuse and neglect.

### *Home-based BPTs*

A large majority of BPTs were designed to be implemented in a clinic setting; thus, one of the shortcomings of even the most well-researched BPTs is the inability to reach populations most at-risk for substantiating child maltreatment (Gomby, Culross, & Behrman, 1999), who often have great difficulty completing clinic-based family intervention programs (Snell-Johns,

Mendez, & Smith, 2004). Although less common, BPTs like Project SafeCare are delivered entirely within the parent's home, bypassing the frequent obstacle of accessibility. These in-home programs offer advantages that typical clinic-based BPTs lack, such as eliminating the parent's need for childcare, reliable transportation, and adequate travel time during sessions (Schoenwald & Hoagwood, 2001). In this way, home visitation programs can be used to more often target parents who are of low socioeconomic status and education, and consequently are in most in danger of substantiating child abuse (Mersky, Berger, Reynolds, & Gromoske, 2009; Schumacher, Slep, & Heyman, 2001; Stith, et al., 2009).

#### *Evaluation of program effectiveness in community settings*

The transition towards home visitation programs is imminent, especially following the current recommendations by the CDC Task Force on Community Preventive Services for home visitation as the major treatment for at-risk parents (Briss, et al., 2000). Yet, the measures of evaluating program effectiveness considerably lag behind the urgency for the transition of BPT delivery from the clinic to the home (Newnham & Page, 2010). The increased variability in home visitation models makes effectiveness measures tremendously difficult (Sweet & Appelbaum, 2004), especially when programs are newly adapted for outside the clinical setting. Although randomized controlled trials (RCTs) offer the ability to rigorously control for extraneous confounding factors, these RCTs are usually difficult to implement in community-based settings where inclusion and exclusion criteria are less strictly imposed (Sibbald & Roland, 1998), and random assignment to treatment conditions can be very difficult (Soydan, 2008). Instead, community research participants are generally referred for services, and withholding this essential treatment from at-risk individuals can sometimes be considered unethical (Sibbald & Roland, 1998).

### *Benchmarking*

Without a control group for making comparisons of effectiveness, studies within the field of skills-based parent training are at a huge disadvantage when it comes to public health decision-making, especially if efficacy differs according to the community in which services are delivered (Gomby, et al., 1999). Benchmarking, which has its roots in medical laboratories, manufacturing, and engineering (Francis & Holloway, 2007), can offer a solution for making more systematic, evaluative research performance judgments, particularly when comparison groups are not feasible. This methodology, which provides a standard for program transferability from a clinical to community setting, can be especially beneficial for understanding program impact (Hunsley & Lee, 2007). Rather than compare community results to a control group, benchmarking utilizes the results of a meta-analytic compilation of RCTs in the literature to establish an aggregate effect size, using demographics and measures similar to the study in question (Weersing & Weisz, 2002). The general process of benchmarking involves four steps: (1) identify the problem, population, and treatment; (2) select or construct the “gold standard” benchmark, usually from clinical trials; (3) measure community outcomes that are similar to the benchmark and calculate the effect sizes; and (4) compare the effect sizes of the community and benchmark outcomes (Minami, Serlin, Wampold, Kircher, & Brown, 2006; Weersing, 2005). Since benchmarking measures effectiveness when control groups are missing, the ethics of community-based research are less of an obstacle, as researchers will no longer need to worry about denying services to certain groups of people or referrals.

Within the child maltreatment prevention and parent training fields, benchmarking remains relatively underutilized. A growing number of studies focused on mental health problems, such as adult depression (Minami, Wampold, Serlin, Hamilton, & Brown, 2008;

Minami, Wampold, Serlin, Kircher, & Brown, 2007) and Cognitive-Behavioral Therapy (CBT) with adolescent depression (Weersing, 2005; Weersing & Weisz, 2002), adult obsessive-compulsive disorder (Farrell, Schlup, & Boschen, 2010), bulimia (Tuschen-Caffier, Pook, & Frank, 2001), anxiety (Oei & Boschen, 2009), and social phobia (Lincoln, et al., 2003; McEvoy, 2007), have supported the use of benchmarking. However, Multisystemic Treatment (MST) is currently one of the only family-targeted interventions to undergo a benchmarking investigation as means of comparing community-based treatment outcomes of juvenile offenders to aggregate effect sizes from relevant RCTs. Here, researchers were able to demonstrate that MST delivered within the child's home, school, or other community setting produced similar results to MST RCTs and was superior to the community's treatment as usual services (Curtis, Ronan, Heiblum, & Crellin, 2009).

The scope and usability of such a powerful tool as benchmarking can help minimize many of the uncertainties associated with social services research. The applicability of benchmarking is a crucial addition to the parent-training discipline; nonetheless, much work must still be done to perfect the methodology for social sciences (Francis & Holloway, 2007), where the challenges of conducting RCTs are considerable. The current state of benchmarking literature extensively introduces the mathematics and formulas behind benchmarking calculations (Minami, et al., 2006; Minami, et al., 2008; Minami, et al., 2007), but offers much less straightforward approaches towards the decisions that must be made to establish comparable benchmarks. In order to increase the use of such novel methods within a field, guidelines should be created that outline the necessary decisions, appropriateness, and strengths and limitations of the benchmarking methods.

*Purpose of current study*

The current study outlines the transport of the normal clinic-based PCIT into the parents' home. PCIT, which is an evidence-based practice for reducing child externalizing disorders (Boggs, et al., 2004; S. M. Eyberg, et al., 2001; Hood & Eyberg, 2003), has shown positive effects for preventing child maltreatment recidivism when implemented within the clinic (Chaffin, et al., 2004). Only one previously published study by Ware and colleagues (2008) has examined the implementation of PCIT in the home setting. Although Ware (2008) was able to demonstrate that families who completed home-based PCIT treatment were less likely to have negative caregiver behavior, less frequent child behavior problems, and greater child compliance, the external validity of this particular study is compromised by the use of a single subject design with only three completing families (Ware, McNeil, Masse, & Stevens, 2008).

The current study is a secondary data analysis of a community family agency based in Durham, NC, the Family Exchange Center (FEC). The FEC primarily serves families with children over the age of two with behavioral problems as well as families who are involved with the Child Welfare System. Similar to Ware and colleagues' study in 2008, the FEC has been providing services within the home to its families for 16 years, integrating the PCIT program in 2005. Because a RCT study of this in-home PCIT implementation was not feasible, benchmarking methodology will be used as an alternative to study this program's effectiveness. Specifically, six RCTs were chosen as benchmarks based on similarity in parent and child demographics as well as use of the ECBI as a primary measure. In accordance with previously established benchmarking calculations (Hunsley & Lee, 2007; Minami, et al., 2006), the effect size for both subscales of the Eyberg Child Behavior Inventory (ECBI) will be calculated for the current study as well as each of the chosen benchmarks. The effect sizes for each benchmark

study will then be aggregated to create a single benchmark effect size for both the treatment and control groups. Finally, the effect size computed from the pre-post change for Family Exchange Center will be compared to the established effect sizes from existing research. It is hypothesized that the current study's effect size will exceed the aggregate benchmark effect size of the control groups and will be equal to the aggregate benchmark effect size of the treatment groups. By providing the home-based delivery of PCIT as an example scenario, this study will summarize the necessary decisions, calculations, and strengths and limitations of applying a benchmarking methodology to a study of cognitive-behavioral therapy in the community setting.

## Methods

### *Participants*

Clinical records of eighty-three families whose parent(s) had initiated PCIT services between January 2007 and January 2009 at FEC, a family support agency in a mid-size southeastern city, were reviewed. Family referral sources were diverse, including community health centers, hospitals, doctor's offices, school social workers, domestic violence programs, the juvenile justice system, mental health agencies, the Department of Social Services, friends, or by self-referral. Criteria for parent enrollment in PCIT services included: 1) children between the ages of two and ten years old, 2) regular parental contact with children, and 3) agree to services that required both parent and child participation. Of the parents enrolled in PCIT, 86.8% were females ( $n = 66$ ), and 13.2% were male ( $n = 10$ ), with an average female age of 30.08 (SD = 7.637) years old, and an average male age of 30.88 (SD = 4.912) years old. The family agency's clinic records also revealed an ethnically diverse group of parents, with 55% of parents reporting ethnic backgrounds as Latino/a ( $n = 46$ ), 37% African American ( $n = 31$ ), and 7% Caucasian ( $n=6$ ). Ten percent of the parents were court mandated for PCIT services. Other

sociodemographic variables such as parent education, socioeconomic status, or caregiver relationship were not provided in the clinic file.

### *Measures*

**Treatment Completion.** Parent-child dyad treatment completion was determined by reviewing family clinic records for PCIT completion.

**Eyberg Child Behavior Inventory (ECBI)** (S. M. Eyberg & Pincus, 1999; S. M. Eyberg & Ross, 1978). The ECBI is a 36-item parent-report measure for assessment of conduct behavioral problems in children ages 2 to 16 years old. The ECBI consists of two scales, Intensity and Problem. The Intensity subscale, rated on Likert scale ranging from 1 (“Never”) to 7 (“Always”), measures the frequency of child behaviors. The Problem subscale measures the parents’ perceptions of behaviors as problematic, using “Yes” or “No” responses (add ref). Raw scores of 131 or 15 for the Intensity and Problem subscales respectively, or T-scores of 60 or above for both subscales are considered clinically significant. The reliability and validity of the ECBI is also well-established (Boggs, Eyberg, & Reynolds, 1990; S. M. Eyberg & Pincus, 1999). This measure was completed by parent participants at both pre- and post-treatment.

### *Study Design and Statistical Analyses*

One-group pretest-posttest design and benchmarking methodology were used to evaluate the effectiveness of in-home PCIT services for this sample population. Using SPSS 17.0, a paired samples t-test was utilized to understand differences in ECBI Intensity and Problem subscales between pre- and post-treatment. The effect sizes from the pre and post ECBI scores were computed for the current study as well as selected benchmarks from the existing PCIT literature. Aggregate treatment and control effect sizes from the benchmarks were then

compared to the current study's aggregate effect size using the statistical software R version 2.11.0.

### *Home-based PCIT*

PCIT is a manualized, clinic-based intervention that consists of two stages: Child-Directed Interaction (CDI) and Parent-Directed Interaction (PDI). Parents are encouraged to develop high-quality relationships with their children through play therapy in the CDI segment, whereas parents learn specific and consistent disciplinary methods during the PDI segment (S. Eyberg & Robinson, 1983). Together, the two stages promote positive parenting skills and increase overall child compliance (Timmer, Zebell, Culver, & Urquiza, 2009). The home-based intervention delivered by the family agency adhered to the clinic-based PCIT manual, except when adaptations for delivery in the home were necessary. Rather than using the one-way mirror and bug-in-the-ear that is found in clinical PCIT, in-home PCIT requires the therapist to be in the same room as the parent and child during coaching sessions. Therefore, the child is informed of the therapist's presence, while the therapist relays coaching instructions to the parent via a hand-held radio. Furthermore, the parent is encouraged to interact with other children participating in home-based PCIT services. Home-based PCIT also permitted the therapist to make direct adaptations suitable for the variety of home settings.

### *PCIT Training for Community Therapists.*

The PCIT therapists at the family support agency were trained by two certified PCIT trainers over a period of 10 months in 2007 and 2008. Therapists completed workshop training that included didactics and role-play for the PCIT model. During the therapist and family's first ten sessions, post-workshops were monitored, either live or in-person, by PCIT trainers to ensure

competence and treatment integrity in service delivery. Since training was completed in 2008, the local trained PCIT supervisor at the agency conducts two fidelity monitoring sessions a year per therapist to ensure that they are remaining faithful to the PCIT model in service delivery with families.

### *Benchmarking Decisions*

**Methodology.** Without a previously established benchmark for the evaluation of PCIT effectiveness, numerous decisions were made to determine the most appropriate and reliable benchmark. Although many benchmarking methodologies existed, very few studies outline the steps necessary for determining and creating a novel benchmark. Previous benchmarking research has focused on three main methodologies involving the comparison of the community setting study to: 1) a single randomized controlled trial (RCT) (Hunsley & Lee, 2007; Weersing, 2005), 2) a comprehensive, previously-published, meta-analysis of aggregated effect sizes (Neill, 2003), or 3) individually aggregated effect sizes from more than one RCT (Curtis, et al., 2009; Minami, et al., 2007; Weersing & Weisz, 2002).

**Benchmarking Selection Criteria.** When deciding which methodology to use, various benchmark inclusion criteria are important. First, the study's sample demographics must be similar to the demographics of the possible selected RCTs. When community and efficacy studies are highly comparable, the selected benchmarks will have a better external validity, and thus, be more applicable to community-delivered services (Hunsley & Lee, 2007; Shadish, et al., 1997). Second, outcome variables must be measured using comparable measures. Although a wider range of measures provides a much better appreciation of treatment effectiveness, the selection of possible benchmark studies is limited to the measure most rigorously implemented

by the community agency. Benchmarks with similar measures are ultimately more useful for the understanding of transportability of efficacy trials to a community setting (Weersing, 2005).

Determination of Relevant RCTs. As previously mentioned, the PCIT field does not have an established benchmark to use as a desirable goal, or “gold standard”, for treatment comparisons. Thus, multiple literature searches in both PubMed and PsychInfo for previously published PCIT studies were conducted. The first search focused on PCIT meta-analyses. A recent meta-analysis of PCIT and Triple P was identified (Thomas & Zimmer-Gembeck, 2007). Although pertinent net effect sizes were given for PCIT parent reported negative child behaviors ( $d = -1.31$ ), this meta-analysis included over 13 PCIT RCTs, nonrandomized trials, and single cohort studies with measures such as the Eyberg Child Behavior Inventory (ECBI), Dyadic Parent-Child Interaction Coding System (DPICS), Child Behavior Checklist (CBCL), Behavior Assessment System for Children (BASC), Conner’s Teaching Rating Scales (CTRS), Diagnostic Statistical Manual-Oppositional Defiant Disorder (DSM-ODD), and Sutter-Eyberg Student Behavior Inventory (SESBI) (Thomas & Zimmer-Gembeck, 2007). The current study included both the ECBI and the DPICS; however, the DPICS was implemented without the blinded researchers used in clinic work. Thus, only the 8 RCTs that utilized the ECBI as the parent-report measure were included as possible benchmarks. From here, each of these RCTs was examined for demographic and pre-post study design similarities. Of the 8 RCTs from the Thomas & Zimmer-Gembeck (2007) meta-analysis, 2 RCTs were eliminated as benchmark prospects. One study was not available through the university library (McNeil, Capage, Bahl, & Blanc, 1999) and another was a one and two-year follow-up study to previously implemented PCIT services (Nixon, Sweeney, Erickson, & Touyz, 2004).

The second literature search focused on PCIT RCTs published after 2004, which was the inclusion criterion for the Thomas & Zimmer-Gembeck (2007) meta-analysis. Additional effectiveness studies were identified, but these studies were considered demographically dissimilar from the current study, as they included children with autism-spectrum disorder (Solomon, Ono, Timmer, & Goodlin-Jones, 2008), Chinese families (Leung, Tsang, Heung, & Yiu, 2009), and Puerto Rican families (Matos, Bauermeister, & Bernal, 2009), as well as some of these studies failed to meet RCT criteria. Consequently, these RCTs were not included as possible benchmarks. Thus, 6 total RCTs were established as benchmarks (see Table 2) based on both their use of the ECBI as a primary measure and similarity in parent and child demographics. As a result, the benchmarking guidelines that utilized aggregate effect sizes for more than one RCT was chosen as the most appropriate methodology for the current study.

### *Calculations*

In accordance with the benchmarking calculations established by Hunsley & Lee (2007), and as described in other benchmarking studies in the mental health field (Curtis, et al., 2009; Minami, et al., 2006; Minami, et al., 2007), the effect size (ES) for both subscales of the ECBI was calculated for the current study as well as each of the chosen 6 benchmarks. Because of the small sample sizes for the current study and the benchmarks, the ES calculation utilized included a sample size correction, as suggested by Minami et al. (2008). The following formulas, recommended by Minami et al. (2006), were used to calculate each individual ES and their corresponding variances:

$$d_{(1)} = \left(1 - \frac{3}{4n-5}\right) \frac{M_{post} - M_{pre}}{SD_{pre}} \quad (1)$$

$$\sigma_{d(i)}^2 = \frac{2(1-r)}{n} + \frac{d^2}{2n} \quad (2)$$

The  $r$ -value from the previous equation is used to determine the correlation between pre and posttreatment scores, and can be estimated with the following equation (Lipsey & Wilson, 2001):

$$r = \sqrt{\frac{d^2}{d^2 + 4}} \quad (3)$$

The individual effect sizes for both the ECBI Intensity and ECBI Problem subscales were aggregated to produce a single ECBI ES for the current study. To find the aggregate ES, the ratio of effect size to variance are summed for each benchmark. This value is then weighted by the summation of the inverse of the variances for each benchmark. This same aggregation was applied to the treatment and control groups as well, where the individual ESs of each ECBI subscale were combined into a single ES. The following equation was used for the aggregate ES:

$$d_{(B)} = \frac{\sum_i \frac{d_i}{\sigma_{d(i)}^2}}{\sum_i \frac{1}{\sigma_{d(i)}^2}} \quad (4)$$

To test the hypothesis that the current study's aggregate ES is clinically superior to the control benchmark's aggregate ES, a noncentral  $t$  distribution was employed with  $t$  at the 95<sup>th</sup> percentile, using the  $t_{(Control)\nu, \lambda:0.95}$  statistic with degrees of freedom as  $\nu = N - 1$ . Lambda, the noncentrality  $t$  parameter, can be found using the following formula:  $\lambda = \sqrt{N} (d_{B(Control)} + \Delta)$ , where  $\Delta = 0.2$ , the difference deemed small enough that the benchmark and current study can be considered clinically equivalent (Minami et al., 2006). With these parameters, an effect size

exceeding the calculated critical value of  $d_{CV(Control)}$  can be deemed more effective than the benchmark control:

$$d_{CV(Control)} = t_{(control)v, \lambda:95} / \sqrt{N} \quad (5)$$

To determine if the current study's effect size is similar to the treatment effect size, the same calculations and conditions would apply, with the exception of the critical value,  $d_{CV(treatment)} = t_{(treatment)v, \lambda:95} / \sqrt{N}$ , and  $\lambda = \sqrt{N} (d_{B(treatment)} - \Delta)$ . If the effect size from the current study is greater than the calculated critical value, the current study is considered clinically equivalent with the established treatment benchmark.

## Results

### *Treatment Completion*

Fifty-four parents completed in-home PCIT services, out of the original 83 parent-child dyads, resulting in an overall attrition rate of 34.9%. For those benchmarks that chose to report drop-out percentages, the attrition rate ranged from 18% to 34%, which is slightly lower than the current study's reported attrition rate of 34.9%. A significant number of the non-completers in the study were lost at follow-up and did not complete post-test data (73%). The demographic differences between those who completed in-home PCIT versus those who did not complete PCIT can be seen in Table 1. There were significant differences in ethnicity ( $p < .01$ ), language ( $p < .01$ ), and mandate for services ( $p < .05$ ) between completers and noncompleters, with completers being more likely to be Latino/a (64.8%), Spanish speaking (61.1%), and not mandated to receive services (92.6%) as compared to non-completers (37.9% Latino/a, 37.9%,

Spanish speaking, and 31% mandated for services, respectively) . Non-completers were not included in the benchmarking analyses.

### *ECBI scores pre-post change*

For both the ECBI Intensity and Problem subscales, significant decreases in parent-reported scores were noted following the receipt of in-home PCIT services. The pre-ECBI Intensity raw score of 134.28 was significantly higher than the post-ECBI Intensity raw score of 94.08 ( $t(52) = 7.563, p < 0.0001$ ). Similarly, for the ECBI Problem subscale, the mean pre raw score of 18.07 was also found to be significantly higher than the mean post raw score of 8.02 ( $t(53) = 6.915, p < 0.0001$ ).

### *Benchmarking Results*

Using the calculations for benchmarking as previously outlined and established by Minami et al. 2006, the current study's effect size (ES) for the Intensity and Problem subscales of the ECBI were calculated and then aggregated to give one overall ES for the in-home PCIT implementation. Because PCIT lacks a clearly defined benchmark, a literature search of previously published PCIT RCTs was conducted, revealing 6 RCTs that were included as benchmarks. These selected 6 RCTs included the ECBI measure as a major outcome variable and had similar sample characteristics as the current study. Table 2 allows comparison between the benchmark RCTs and the current study. The same effect size calculations were applied to the chosen benchmarks. However, effect sizes were also computed for both treatment and control groups of each benchmark and then aggregated across all the benchmarks to create an overall treatment benchmark effect size and control benchmark effect size.

In accordance with the non-central  $t$ -distribution and the non-centrality  $t$ - parameter, lambda, described earlier in the Measures section, the critical value of the effect size  $d_{CV}$  was determined using the following calculation:

$$d_{CV} = t_{\nu, \lambda, .95} / \sqrt{N}$$

The  $t$ -statistic used is dependent upon the condition for which the current study will be compared, either the treatment or control benchmark conditions. When comparing the current study ES to the treatment and control benchmarks, an effect size greater than the critical value for each condition would result in the current study being considered clinically equivalent with the treatment benchmark or clinically superior to the control benchmark, respectively.

The means, standard deviations, and sample sizes of the current study and the treatment and control groups of the 6 RCTs are shown in Table 3. Table 4 displays the benchmark group effect sizes, as well as the aggregate treatment and aggregate control effect sizes to which the current study's group effect size is compared against.

The current study resulted in an ES of 1.1333. When compared to the aggregated control benchmark effect size of 0.62418, a critical value ( $d_{CV(Control)} = 1.1140$ ,  $t(53) = 8.19$ ,  $\lambda = 6.056$ ,  $p < 0.0001$ ) was found. Because the current study (ES = 1.1333) exceeded the aggregate control benchmark critical value of 1.1140, the current study is considered clinically and significantly superior to the effect size found from the control conditions of the RCTs.

The aggregate treatment benchmark resulted in an effect size of 1.7262. After calculating the critical value ( $d_{CV(B)} = 1.9096$ ,  $t(53) = 14.033$ ,  $\lambda = 11.22$ ,  $p < 0.0001$ ), the current study's ES

of 1.1333 was significantly less than the benchmark critical value. Thus, the current study cannot be deemed clinically equivalent to the treatment conditions of the RCTs.

## Discussion

### *Benchmarking Results.*

This investigation demonstrates the application of benchmarking methodologies to understand the transportability of an evidence-supported clinical intervention to the community setting, using Parent Child Interaction Therapy (PCIT) as a case example. In the current study, participants were considered at-risk for child abuse and were referred to the family support agency for PCIT services. Benchmarking offered a useful approach to compare the effectiveness of the community-delivered PCIT to a standard benchmark established from RCTs examining the PCIT model. Benchmarking hypotheses were partially supported in this study.

As proposed, the in-home delivery of PCIT services resulted in more positive parent ratings of their child's behavior, as shown through the improved ECBI subscale scores. After using the benchmarking methodologies outlined by Minami (2007), the current study was shown to be superior in effectiveness to the selected RCT's control groups, as previously hypothesized. These data offer further validation to Ware's (2008) that PCIT delivered in the home is a promising approach for improving positive outcomes for families, even for those who are considered at-risk for maltreatment, as compared to families who do not receive any treatment. In this study, the benchmarking methodology aggregated the results from the various control conditions which included waitlist groups, where no treatment was received by families, and social validation (Nixon, 2001; Nixon, Sweeney, Erickson, & Touyz, 2003) for which families were not eligible for services because this group included nonproblem children. Therefore,

families assigned to these groups had more variability in their ECBI scores at baseline because the social validation groups from the Nixon studies included parents that reported no difficulty with their child's behavior and had children with ECBI scores in the normal range, as well as no previous diagnosis of ODD (Nixon, 2001; Nixon, et al., 2003). The inclusion of these groups into the control benchmarks could have provided a floor effect for ECBI scores and contributed to the lower control aggregate effect size.

Conversely, the implementation of in-home PCIT was not clinically equivalent to the RCT's aggregate treatment benchmark. Numerous factors intrinsic to the design of a RCT study can explain the discrepancy between the community and clinic-administered effectiveness trials. First, RCT studies are designed with strict inclusion and exclusion criteria. For example, all of the studies selected for this benchmarking investigation required a diagnosis of either Oppositional Defiant Disorder (Brestan, Eyberg, Boggs, & Algina, 1997; Nixon, 2001; Nixon, et al., 2003; Schuhmann, Foote, Eyberg, Boggs, & Algina, 1998) or other conduct behavior disorders (Hood & Eyberg, 2003). In the family agency's case, participants were enrolled in PCIT services if they had a child between the ages of 2 and 10, had regular contact with their child, and agreed to services that include parent and child participation. No prior diagnosis of a behavioral disorder was required; as a result, the current study's pre-ECBI scores were much lower than pre-ECBI scores reported by the RCTs, and thus there was less room for change among the current sample.

Second, the current sample was highly ethnically diverse, with minority participants representing 90% of the total parent sample. Although attempts were made to choose RCTs with participants as demographically similar to our study's samples, the relatively low number of available PCIT RCTs with families of diverse ethnic backgrounds made the selection of

appropriate benchmarks challenging. Furthermore, 61.1% of our sample received PCIT services delivered in Spanish, while all of the RCTs had predominantly English-speaking participants so cultural adaptations were not necessary. These sample-related differences could have attributed to a difference in effect size when compared to the aggregate treatment benchmarks.

*Limitations of Current Study and Future Directions with Benchmarking Methodology.*

Currently, there is very little guidance for the creation of benchmarks when this methodology is applied to a new field of study, and, thus, the current study results may be limited by the selection process for RCTs that were included. In this investigation, researchers followed the recommendations from Hunsley & Lee (2007) to create benchmarks from a collection of individual RCTs; however, another option would be to compare the group effect size from the current study to aggregate treatment and control benchmarks found from already-published meta-analyses (Neill, 2003). The only available PCIT meta-analysis (Thomas & Zimmer-Gembeck, 2007), however, included results from quasi-experimental studies, single case studies, and RCTs as well as measures outside the scope of the current study. When compared to the more inclusive PCIT meta-analysis, the FEC effect size ( $d = 1.1333$ ) was significantly lower than the net effect size (treatment effect size minus comparison effect size) reported ( $d = 1.31$ ), similar to the current benchmarking findings. Although this effect size provides a more representative standard for the effectiveness of general PCIT services, benchmarking against more selective RCTs was a more robust option. Curtis et al. (2009) reported similar issues when deciding whether or not to use a “best practice” meta-analysis to benchmark their study against. Rather than using their previous meta-analysis, the researchers chose RCTs ( $n = 3$ ) that represented the sample population, as the “best practice” study reported between-group effect sizes while their current study examined within-group effect sizes (Curtis, et al., 2009). As

benchmarking methodology becomes more standard across the behavioral sciences, further criteria and guidelines should be established to assist researchers in selecting the best benchmarking criteria for their particular project goals.

Another limitation in this study was the limited fidelity data available for the PCIT therapists in usual practice. In order to successfully utilize benchmarking in a community, therapist fidelity, training, and workload (Minami, et al., 2006) must be considered. The benchmark, or collection of benchmarks, serve as an identified “gold standard” for the overall effects of a particular intervention when delivered under ideal conditions; thus, the therapists delivering both the RCTs and the community interventions must properly follow program guidelines and procedures to ensure program quality (Carroll, et al., 2007; Hermann, et al., 2006; McLeod, Southam-Gerow, & Weisz, 2009). Discrepancies among therapist deliveries at different settings are difficult to resolve, but fidelity monitoring efforts should be made to ensure that effectiveness comparisons are appropriate. Benchmarking alone cannot account for these differences, so candidate benchmark studies must be chosen with care. Thus, it is imperative that, if benchmarking is to be appropriately utilized that community agencies be willing to monitor the fidelity of the providing interventionists.

An additional limitation of the current study was the small sample size of parents completing in-home PCIT services. The effect size calculations used included a correction for small sample sizes. However, the sample size is also a factor in the calculation of the non-centrality parameter,  $\lambda$ . Consequently, the sample size can significantly affect the computed critical values used for comparing the current study’s effect size to the treatment or control benchmarks, as small sample sizes ( $N < 100$ ) can produce higher critical values that must be exceeded to claim clinical equivalence (Minami, et al., 2007). Although Minami’s

recommendations suggest the application of benchmarking with samples greater than 100, community-based research samples often fail to meet such high participant numbers. Thus, benchmarking, which can be a very powerful effectiveness tool when applied correctly, must be interpreted cautiously when sample sizes are not sufficient. Future research should identify the lower limit threshold that is necessary to conduct benchmarking studies.

The current study was limited to one measure (ECBI) for benchmarking analyses, which may have impacted the results of this study. Dyadic Parent-Child Interaction Coding System (DPICS) data was collected, but this measure was not chosen as part of the benchmarking due our study's lack of a blinded research assistant administering this measure. In an ideal situation where a program evaluation could be planned a priori, implementing agencies could be required to send in random videotaped DPICS observations so that blind assessment ratings could be completed. Having a valid DPICS measure scored according to research criteria would vastly expand the research studies that could have been included in the aggregate benchmark scores.

Theoretically, the inclusion of a wider range of standardized measures into the usual care setting and benchmarking aggregates would provide for a stronger analysis of community implementation effectiveness. But practically, community providers must make strategic decisions about which tools to implement. These decisions are often dependent on the cost, expertise, and time required for such standardized measures. Benchmarking offers insight into the effectiveness of a community-based study using whichever measure is most relevant and easily implemented by the community providers. The use of measures commonly found in clinic-based PCIT would be beneficial to overall benchmarking results, but the reality of such implementations in the usual care or community interventions must be considered.

In addition to the recommendations for usual practice to assist in improving benchmarking studies, the clinic-based studies that the community setting must be compared to also need improvement. The required inclusion of attrition and descriptive data of the intervention and comparison sample groups of these studies would increase the external validity of clinic-based data to the community (Stewart & Chambless, 2009). This requirement would facilitate community researchers' decisions regarding the appropriate criteria for benchmark selection.

### Conclusions

The current research provides further support to the adaptation of PCIT for home-based delivery. Although the current study was not clinically equivalent to the treatment benchmark, it was clinically superior to the control benchmark, indicating that parents who received and completed in-home PCIT demonstrated significant improvement to parents who receive no services. Benchmarking methodology can be readily utilized by community practitioners to allow greater understanding of program effectiveness and transportability when control groups are not possible; however, additional research is necessary to further develop the decisions and criteria for the creation of field or measure-specific benchmarks.

## References

- Barth, R. P. (2009). Preventing child abuse and neglect with parent training: Evidence and opportunities. *Future of Children, 19*(2), 95-118.
- Barth, R. P., Landsverk, J., Chamberlain, P., Reid, J. B., Rolls, J. A., Hurlburt, M. S., et al. (2005). Parent-Training Programs in Child Welfare Services: Planning for a More Evidence-Based Approach to Serving Biological Parents. *Research on Social Work Practice, 15*(5), 353.
- Boggs, S. R., Eyberg, S. M., Edwards, D., Rayfield, A., Jacobs, J., Bagner, D., et al. (2004). Outcomes of parent-child interaction therapy: A comparison of dropouts and treatment completers one to three years after treatment. *Child & Family Behavior Therapy, 26*(4), 1-22.
- Boggs, S. R., Eyberg, S. M., & Reynolds, L. (1990). Concurrent validity of the Eyberg Child Behavior Inventory. *Journal of Clinical Child Psychology, 19*(19), 75-78.
- Brestan, E. V., Eyberg, S. M., Boggs, S. R., & Algina, J. (1997). Parent-Child Interaction Therapy: Parents' perceptions of untreated siblings. *Child & Family Behavior Therapy, 19*(3), 13-28.
- Briss, P. A., Zaza, S., Pappaioanou, M., Fielding, J., Wright-De Agüero, L., Truman, B. I., et al. (2000). Developing an evidence-based Guide to Community Preventive Services--methods. The Task Force on Community Preventive Services. *American Journal of Preventive Medicine, 18*(1 Suppl), 35-43.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 40.
- Centers for Disease Control and Prevention. (2009). *Parent training programs: Insight for practitioners*.
- Chaffin, M., Silovsky, J. F., Funderburk, B., Valle, L. A., Brestan, E. V., Balachova, T., et al. (2004). Parent-child interaction therapy with physically abusive parents: efficacy for reducing future abuse reports. *Journal of Consulting and Clinical Psychology, 72*(3), 500-510.
- Curtis, N. M., Ronan, K. R., Heiblum, N., & Crellin, K. (2009). Dissemination and effectiveness of Multisystemic Treatment in New Zealand: A benchmarking study. *Journal of Family Psychology, 23*(2), 119-129.
- Dore, M. M., & Lee, J. M. (1999). The Role of Parent Training with Abusive and Neglectful Parents. *Family Relations, 48*(3), 313-325.

- Eyberg, S., & Robinson, E. A. (1983). Conduct problem behavior: Standardization of a behavioral rating scale with adolescents. *Journal of Clinical Child Psychology, 12*, 347-354.
- Eyberg, S. M., Funderburk, B. W., Hembree-Kigin, T. L., McNeil, C. B., Querido, J. G., & Hood, K. K. (2001). Parent-child interaction therapy with behavior problem children: One and two year maintenance of treatment effects in the family. *Child and Family Behavior Therapy, 23*(4), 1-20.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory Sutter Student Behavior Inventory - Revised professional manual*. Odessa, FL: Psychological Assessment Resources.
- Eyberg, S. M., & Ross, A. W. (1978). Assessment of child behavior problems: The validation of a new inventory. *Journal of Clinical Child Psychology, 7*, 113-116.
- Farrell, L. J., Schlup, B., & Boschen, M. J. (2010). Cognitive-behavioral treatment of childhood obsessive-compulsive disorder in community-based clinical practice: Clinical significance and benchmarking against efficacy. *Behaviour Research and Therapy, 48*, 409-417.
- Francis, G., & Holloway, J. (2007). Themes from the literature on best-practice benchmarking. *International Journal of Management Reviews, 9*(3), 171-189.
- Gershater-Molko, R. M., Lutzker, J. R., & Wesch, D. (2002). Using Recidivism to Evaluate Project SafeCare: Teaching Bonding, Safety, and Health Care Skills to Parents. *Child Maltreatment, 7*(3), 277-285.
- Gomby, D. S., Culross, P. L., & Behrman, R. E. (1999). Home visiting: Recent program evaluations- Analysis and recommendations. *Future of Children, 9*(1), 4-26.
- Hakman, M., Chaffin, M., Funderburk, B., & Silovsky, J. (2009). Change trajectories for parent-child interaction sequences during parent-child interaction therapy for child physical abuse. *Child Abuse & Neglect, 33*, 461-470.
- Hermann, R. C., Mattke, S., Somekh, D., Silfverhielm, H., Goldner, E., Glover, G., et al. (2006). Quality indicators for international benchmarking of mental health care. *International Journal for Quality in Health Care, 18*(Supplement 1), 31-38.
- Hood, K. K., & Eyberg, S. M. (2003). Outcomes of Parent-Child Interaction Therapy: Mothers' reports of maintenance three to six years after treatment. *Journal of Clinical Child and Adolescent Psychology, 32*(3), 419-429.
- Houck, C. D., Nugent, N. R., Lescano, C. M., Peters, A., & Brown, L. K. (2010). Sexual abuse and sexual risk behavior: Beyond the impact of psychiatric problems. *Journal of Pediatric Psychology, 35*(5), 473-483.

- Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*(1), 21-33.
- Hussey, J. M., Chang, J. J., & Kotch, J. B. (2006). Child maltreatment in the United States: Prevalence, risk factors, and adolescent health consequences. *Pediatrics, 118*, 933-942.
- Irish, L., Kobayashi, I., & Delahanty, D. L. (2010). Long-term physical health consequences of childhood sexual abuse: A meta-analytic review. *Journal of Pediatric Psychology, 35*(5), 450-461.
- Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*.
- Lanier, P., Jonson-Reid, M., Stahlschmidt, M., Drake, B., & Constantino, J. (2010). Child maltreatment and pediatric health outcomes: A longitudinal study of low-income children. *Journal of Pediatric Psychology, 35*(5), 511-522.
- Leung, C., Tsang, S., Heung, K., & Yiu, I. (2009). Effectiveness of Parent-Child Interaction Therapy (PCIT) among Chinese families. *Research on Social Work Practice, 19*(3), 304-313.
- Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B., et al. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy, 41*, 1251-1269.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Matos, M., Bauermeister, J. J., & Bernal, G. (2009). Parent-Child Interaction Therapy for Puerto Rican preschool children with ADHD and behavior problems: A pilot efficacy study. *Family Process, 48*, 232-252.
- McEvoy, P. M. (2007). Effectiveness of cognitive behavioural group therapy for social phobia in a community clinic: A benchmarking study. *Behaviour Research and Therapy, 45*, 3030-3040.
- McLeod, B. D., Southam-Gerow, M. A., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *Social Psychology Review, 38*(4), 541-546.
- McMahon, R. J., & Forehand, R. L. (2005). *Helping the Noncompliant Child: Family-Based Treatment for Oppositional Behavior*. The Guilford Press.

- McNeil, C. B., Capage, L. C., Bahl, A., & Blanc, H. (1999). Importance of early intervention for disruptive behavior problems: Comparison of treatment and waitlist-control groups. *Early Education and Development, 10*, 445-454.
- Mersky, J. P., Berger, L. M., Reynolds, A. J., & Gromoske, A. N. (2009). Risk factors for child and adolescent maltreatment: A longitudinal investigation of a cohort of inner-city youth. *Child Maltreatment, 14*(1), 73-88.
- Minami, T., Serlin, R. C., Wampold, B. E., Kircher, J. C., & Brown, G. S. (2006). Using clinical trials to benchmark effects produced in clinical practice. *Quality & Quantity, 42*, 513-525.
- Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., & Brown, G. S. (2008). Benchmarking the effectiveness of psychotherapy treatment for Adult Depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology, 76*(1), 116-124.
- Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007). Benchmarks for psychotherapy efficacy in Adult Major Depression. *Journal of Consulting and Clinical Psychology, 75*(2), 232-243.
- Neill, J. T. (2003). Reviewing and benchmarking adventure therapy outcomes: Applications of meta-analysis. *Journal of Experiential Education, 25*(3), 316-321.
- Newnham, E. A., & Page, A. C. (2010). Bridging the gap between best evidence and best practice in mental health. *Clinical Psychology Review, 30*, 127-142.
- Nixon, R. D. V. (2001). Changes in hyperactivity and temperament in behaviourally disturbed preschoolers after Parent-Child Interaction Therapy (PCIT). *Behaviour Change, 18*(3), 168-176.
- Nixon, R. D. V., Sweeney, L., Erickson, D. B., & Touyz, S. W. (2003). Parent-Child Interaction Therapy: A comparison of standard and abbreviated treatments for Oppositional Defiant preschoolers. *Journal of Consulting and Clinical Psychology, 71*(2), 251-260.
- Nixon, R. D. V., Sweeney, L., Erickson, D. B., & Touyz, S. W. (2004). Parent-Child Interaction Therapy: One- and two-year follow-up of standard and abbreviated treatments for oppositional preschoolers. *Journal of Abnormal Child Psychology, 32*(3), 263-271.
- Oei, T. P. S., & Boschen, M. J. (2009). Clinical effectiveness of a cognitive behavioral group treatment program for anxiety disorders: A benchmarking study. *Journal of Anxiety Disorders, 23*, 950-957.
- Prinz, R. J., Sanders, M. R., Shapiro, C. J., Whitaker, D. J., & Lutzker, J. R. (2009). Population-based prevention of child maltreatment: The U.S. Triple P system population trial. *Prevention Science, 10*(1), 1-12.

- Schoenwald, S. K., & Hoagwood, K. (2001). Effectiveness, transportability, and dissemination of interventions: What matters when? *Psychiatric Services*, *52*(9), 1190-1197.
- Schuhmann, E. M., Foote, R. C., Eyberg, S. M., Boggs, S. R., & Algina, J. (1998). Efficacy of Parent-Child Interaction Therapy: Interim report of a randomized trial with short-term maintenance. *Journal of Clinical Child Psychology*, *27*(1), 34-45.
- Schumacher, J. A., Slep, A. M. S., & Heyman, R. E. (2001). Risk factors for child neglect. *Aggression and Violent Behavior*, *6*(2-3), 231-254.
- Sedlak, A. J., Mettenburg, J., Basena, M., Petta, I., McPherson, K., Greene, A., et al. (2010). *Fourth National Incidence Study of Child Abuse and Neglect (NIS-4): Report to Congress*.
- Shadish, W. R., Matt, G. E., Navarro, A. M., Siegle, G., Crits-Christoph, P., Hazelrigg, M. D., et al. (1997). Evidence that therapy works in clinically representative conditions. *Journal of Consulting and Clinical Psychology*, *65*(3), 355-365.
- Sibbald, B., & Roland, M. (1998). Understanding controlled trials: Why are randomised controlled trials important? *BMJ*, *316*, 201.
- Snell-Johns, J., Mendez, J. L., & Smith, B. H. (2004). Evidence-based solutions for overcoming access barriers, decreasing attrition, and promoting change with underserved families. *Journal of Family Psychology*, *18*(1), 19-35.
- Solomon, M., Ono, M., Timmer, S. G., & Goodlin-Jones, B. (2008). The effectiveness of Parent-Child Interaction Therapy for families of children on the Autism Spectrum. *J Autism Dev Disord*, *38*, 1767-1776.
- Soydan, H. (2008). Applying randomized controlled trials and systematic reviews in social work research. *Research on Social Work Practice*, *18*(4), 311-318.
- Stewart, R. E., & Chambless, D. L. (2009). Cognitive-Behavioral Therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology*, *77*(4), 595-606.
- Stith, S. M., Liu, T., Davies, C., Boykin, E. L., Alder, M. C., Harris, J. M., et al. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior*, *14*, 13-29.
- Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development*, *75*(5), 1435-1456.

- 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.
- Thompson, R. A., & Wyatt, J. M. (1999). Current research on child maltreatment: Implications for educators. *Educational Psychology Review*, 11(3), 173-201.
- Timmer, S. G., Zebell, N. M., Culver, M. A., & Urquiza, A. J. (2009). Efficacy of adjunct in-home coaching to improve outcomes in Parent-Child Interaction Therapy. *Research on Social Work Practice*, 20(1), 36-45.
- Topitzes, J., Mersky, J. P., & Reynolds, A. J. (2010). Child maltreatment and adult cigarette smoking: A long-term developmental model. *Journal of Pediatric Psychology*, 35(5), 484-498.
- Tuschen-Caffier, B., Pook, M., & Frank, M. (2001). Evaluation of manual-based cognitive-behavioral therapy for bulimia nervosa in a service setting. *Behaviour Research and Therapy*, 39(3), 299-308.
- U.S. Department of Health and Human Services. (2010). *Child Maltreatment 2008*.
- Ware, L. M., McNeil, C. B., Masse, J., & Stevens, S. (2008). Efficacy of in-home Parent Child Interaction Therapy. *Child & Family Behavior Therapy*, 30(2), 99-126.
- Weersing, V. R. (2005). Benchmarking the effectiveness of psychotherapy: Program evaluation as a component of evidence-based practice. *J. Am. Acad. Child Adolesc. Psychiatry*, 44(10), 1058-1062.
- Weersing, V. R., & Weisz, J. R. (2002). Community clinic treatment of depressed youth: Benchmarking usual care against CBT clinical trials. *Journal of Consulting and Clinical Psychology*, 70(2), 299-310.
- Whitaker, D. J., Lutzker, J. R., & Shelley, G. A. (2005). Child Maltreatment Prevention Priorities at the Centers for Disease Control and Prevention. [Journal; Peer Reviewed Journal]. *Child Maltreatment*, 10(3), 245-259.

Table 1.  
*Demographics of Parent and Child Completers and Non-completers*

Demographics	Completers		Non-completers		<i>t(d.f.)</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<b>Parent</b>					
Maternal age	29.96	7.003	30.17	9.485	N.S.
Paternal age	30.50	2.887	31.25	6.898	N.S.
Female n(%)	48 (88.9)		25 (86.2)		N.S.
Ethnicity					2.915(81)**
African American n(%)	14 (25.9)		17 (58.6)		
Latino/a n(%)	35 (64.8)		11 (37.9)		
Caucasian n(%)	5 (9.3)		1 (3.4)		
Spanish language n(%)	33 (61.1)		11 (37.9)		-2.044(81)*
Mandatory n(%)	4 (7.4)		9 (31.0)		2.50(37.7)*
No. of sessions	17.87	5.756	8.69	5.832	-6.90(81)**
<b>Child</b>					
Female age	4.69	2.056	4.58	1.782	N.S.
Male age	4.37	1.618	4.47	2.183	N.S.
Female n(%)	16 (29.6)		12 (41.4)		N.S.

*Note.* \* Significant at  $p < 0.05$ , \*\*Significant at  $p < 0.01$ .

Table 2.  
*Comparison of Demographics of Benchmark to Current Study*

Benchmark	Study design	Attrition (%)	Clinical Diagnosis	Child characteristics			Parent Characteristics	
				Female (%)	Mean age	Mean mother age	Mean father age	Minority (%)
<b>Current Study</b>	<b>T</b>	<b>34.9</b>	<b>None</b>	<b>29.6</b>	<b>4.46</b>	<b>30.0</b>	<b>30.5</b>	<b>90.7</b>
Eyberg et al., 1995	T,W	28	ODD	20	4.5	-	-	20
Brestan et al., 1997	T,W	-	ODD	17	4.53	T: 35 W: 29	T: 39 W: 33	T: 21 W: 40
Schuhmann et al., 1998	T,W	34	ODD, ADHD medicine	19	4.9	31.7	36.1	23
Nixon et al., 2001	T,W, SV	-	ODD, clinical ECBI	26	3.9	35	37.6	-
Hood & Eyberg, 2003	T,W	-	ODD or conduct disorder	30	4.7	36.04	-	17
Nixon et al., 2003	2T,W, SV	18	ODD, clinical ECBI	29	3.9	34.5	37.1	4

*Note.* T = treatment, W = waitlist, SV = social validation, ODD = oppositional defiant disorder, ADHD = Attention Deficit Hyperactivity Disorder, ECBI = Eyberg Child Behavior Inventory; Minority percentage includes all ethnicities except Caucasian.

Table 3.

*Treatment and Control Means, Standard Deviations, and Sample Sizes for Current Study and Benchmarks*

Study	N	ECBI Intensity			ECBI Problem		
		$M_{Pre}$	$M_{Post}$	SD	$M_{Pre}$	$M_{Post}$	SD
<b>Current study</b>	<b>54</b>	<b>134.3</b>	<b>94.08</b>	<b>33.65</b>	<b>18.07</b>	<b>8.019</b>	<b>9.099</b>
Treatment Benchmarks							
Eyberg 1995	10	159.5	117.5	16.6	20.7	6.6	4.8
Brestan 1997							
Mother	16	173	133	29.5	23	11	5.8
Father	9	169	137	24.1	22	14	3.3
Schuhmann 1998							
Mother	22	170.3	117.6	26.4	21.9	10.9	6.5
Father	12	159.6	126.8	25.2	20.5	10.2	5
Nixon 2001	17	166.58	125.24	18.93	-	-	-
Hood 2003	23	173.83	126.04	24.42	23.3	8.96	6.15
Nixon 2003							
Mother	17	166.59	125.24	18.93	-	-	-
Father	17	148.33	124.0	24.54	-	-	-
Control Benchmarks							
Eyberg 1995	6	170.7	177.2	40.3	23	21.5	10.3
Brestan 1997							
Mother	13	176	170	30.2	24	24	5.4
Father	7	181	185	41.2	25	24	10.2
Schuhmann 1998							
Mother	20	172.9	169.7	25.8	21.2	22.1	6.1
Father	10	167.7	160.9	36.5	24.6	17.4	5.7
Nixon 2001							
WL	17	173.82	148.35	22.7	-	-	-
SV	21	108.15	105.8	15.4	-	-	-
Hood 2003	27	172.37	118.85	22.3	21.2	11.35	6.41
Nixon 2003							
WL Mother	17	173.82	148.35	22.7	-	-	-
WL Father	17	147.47	134.13	26.0	-	-	-
SV Mother	21	108.81	105.8	15.3	-	-	-

*Note.* WL = Waitlist, SV = Social Validation; Nixon et al. (2001, 2003) did not measure ECBI Problem subscale.

Table 4.

*Group and Aggregate Effect Sizes for Current Study, Treatment Benchmarks, and Control Benchmarks*

	<u>Aggregate ES</u>		<u>Eyberg 1995</u>		<u>Brestan 1997</u>		<u>Schuhmann 1998</u>		<u>Nixon 2001</u>		<u>Hood 2003</u>		<u>Nixon 2003</u>	
	PCIT	WL	PCIT	WL	PCIT	WL	PCIT	WL	PCIT	WL/SV	PCIT	WL	PCIT	WL/SV
Current Study	<b>1.133</b>	-												
Group ES	<b>1.726</b>	<b>0.62418</b>	2.478	0.129	1.563	0.0928	1.664	0.325	2.080	0.5896	2.049	1.802	1.363	0.5696

*Note.* ES = Effect size; PCIT = Parent-Child Interaction Therapy; WL = Waitlist; SV = Social Validation; all reported benchmarks used ECBI, either the Intensity subscale, the Problem subscale, or both.