This is to certify that the
dissertation entitled

TRANSFORMING EVERYDAY PRACTICES
USING SCIENTIFIC EVIDENCE:
META-ANALYSIS OF A PARENT TRAINING PROGRAM

presented by

John R. Sougstad

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in

School Psychology

Major Professor's Signature

12/15/10

Date

MSU is an Affirmative Action/Equal Opportunity Employer

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

5/08 K:/Proj/Acc&Pres/CIRC/DateDue.indd

TRANSFORMING EVERYDAY PRACTICES USING SCIENTIFIC EVIDENCE: META-ANALYSIS OF A PARENT TRAINING PROGRAM

By

John R. Sougstad

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

School Psychology

2010

ABSTRACT

TRANSFORMING EVERYDAY PRACTICES USING SCIENTIFIC EVIDENCE: META-ANALYSIS OF A PARENT TRAINING PROGRAM

By

John R. Sougstad

Meta-analysis of existing research on the Incredible Years Parent Training Program (IYPTP) provided stronger evidence for stable, reliable, and valid estimates of benefits, than would be obtained by reviewing individual studies separately. Comparing outcomes from primary developer-based research studies with the benefits found in independent replication studies yielded stable and generalizable effects for significantly reducing child conduct problems (CCP) within a three-tiered intervention model. At Tier I, small benefits were found from studies using IYPTP as a universal program to prevent the development of CCP. At Tier II, small to moderate decreases in CCP were found for selective studies where parenting was known to be dysfunctional and/or CCP levels were not restricted to clinically-significant cases. At the Tier III indicated level, the most severe and clinically significant forms of CCP exhibited the greatest reductions with moderate to large effects found. Provisionally identified (Sougstad, Oka, Carlson & Tomac, 2008) significant differences between inventor-based and replication research studies at Tier III were spurious and attributable to a larger, double-dosage of IYPTP used within two primary studies. For the most severe CCP cases, the group-administered form of IYPTP that required a substantial investment of time and resources was found to produce significantly higher benefits (about one standard deviation) over that of the self-

administered form of IYPTP (about half-standard deviation) only when the dosage of the group sessions was doubled. The latter finding has not been published in a prior peer-reviewed study or reproduced by independent replication, although Webster-Stratton and Herman (2010) recently made a similar observation from review of prior data. The results of this research are consistent with several others showing that intervention effects are greater when initial child conduct problems are more severe. There was evidence suggesting that both primary and independent replication studies similarly reduced negative parenting while primary studies showed greater increases in positive parenting than did replications. Data across three levels of severity for CCP support the proposed alteration of norms for determining clinical-significance on the Eyberg Child Behavior Inventory (Colvin, Eyberg & Adams, 1999). Overall results from this meta-analysis provide ample evidence for the potency and robustness of IYPTP. The results of this study suggest that the feasibility of the group-administered form over that of the self-administered form within practice settings should be determined by the degree to which resources are available to ensure highly trained therapists and treatment integrity as well as substantial investments in a large dosage of sessions over a considerable length of time.

Copyright by
JOHN R. SOUGSTAD
2010

DEDICATION

My interest in parenting began very early in childhood. The high esteem I hold for parenting and its importance for our future was instilled in me by the many opportunities provided by my parents Daniel and Mary Sougstad. Their devotion to parenting me, my sister, my brother, and now their grandchildren is a living testament to their strong characters and commitment to a set of values that I aspire to emulate throughout my life's endeavors including my research. This dissertation has been built upon valuing lifelong learning and persistence taught to me by my parents.

This dissertation would not have been possible had it not been for the unwavering love, support, understanding and encouragement provided to me by my wife and soul-mate Barbara and our two sons, Ryan and Daniel. My pursuit of the doctoral degree has actually been a prolonged, selfish indulgence on my part requiring many sacrifices from others that I humbly remain so grateful for; especially those of Barb, Ryan and Danny. My years of step-parenting Michelle and Alicia and parenting our boys, has taught me great reverence and humility for how much parenting is a learning process; one we could all use ongoing training for.

> "While we try to teach our children all about life, our children teach us what life is all about" - Unknown

I thank each of you for your patient support of my lifelong learning pursuits and self-development. I dedicate this dissertation to our collective experiences and how much more enriched my life becomes while sharing our accomplishments together, like this one!

ACKNOWLEDGEMENTS

I am most appreciative of Dr. Evelyn Oka, my advisor, mentor and chair of my doctoral committees. Evelyn has been my greatest advocate, teacher and the most patient educator I have had the privilege to know. I am grateful for her benevolent guidance, helping me to believe in my self as a researcher, and to hold the highest expectations for my work. She has persistently encouraged my academic pursuits, despite my many shortcomings and multiple barriers potentially preventing program completion.

I am most appreciative of Dr. John Carlson. Without John I would not be accomplishing my academic goals. John provided the initial idea that ignited what became my dissertation. I most appreciate John's insightful reviews, his encouragement of my research and our shared pragmatism for investigating intervention parsimony.

I am most appreciative of Dr. Rena Harold. Rena has mentored not only my doctoral studies but also my many years of serving as a university instructor and guide to graduate social work students. Her warm encouragement, perceptive and practical advice provided me direction through many moments of despair and self-doubt.

I am most appreciative of Dr. John Mooradian. John has provided me with a most excellent personal example of academic accomplishment beyond a career in practice. I have found strength, determination and a needed sense of humor from John's guidance.

Finally, I would also like to recognize the tremendous benefits I continue to receive as a result of being a student of Dr. Jean Baker, member of my committee lost to breast cancer before dissertation completion. Her life example and research extolling the value of positive relationships that facilitate learning and development live on through my own personal and professional endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Rationale for the Study

In the world of psychological practice a dazzling array of choices promises to promote the wellbeing and mental health of children and their parents. How are interventions selected for use? Practitioner attitudes and beliefs as well as organizational characteristics rather than empirical evidence are the main predictors in decisions about whether to adopt psychosocial evidence-based interventions (EBI) in authentic community-based practice settings (Aarons, 2005; Aarons & Palinkas, 2007; Aarons & Sawitzky, 2006; Graham et al., 2006; Stahmer & Aarons, 2009; McNeil, 2006). Surveys of practitioners who recently completed an APA accredited doctoral program ranked their "belief" about whether an intervention would have positive effects on students as the most influential factor effecting their intervention adoption decisions, above that of knowing an intervention's empirical validity (Forman, Fagley, Steiner & Schneider, 2009). Despite shifts in policy mandating the use of EBIs, practitioners appear to view the research bases of interventions as irrelevant in their decision making (Hoagwood, Burns & Weisz, 2002). Audiences for much of the accumulated intervention research on young children with emotional and behavioral disorders has been limited to researchers, scientific grant-giving agencies, tenure review boards and graduate students, while practitioners have remained apart, absorbed in the daily demands of service provision (Kendziora, 2004).

Researchers and practitioners grapple daily with questions about how to best ameliorate mental health problems, but these efforts have typically remained separate and

disconnected (Chorpita, 2003; Kazdin, 2008; Kratochwill & Shernoff, 2003; LaGreca, Silverman & Lochman, 2009; Southam-Gerow, Ringeisen & Sherrill, 2006). Despite this gap between research and practice, legal mandates, third-party payers, professional organizations and agency policies are placing increased pressure on mental health practitioners to adopt and implement evidence-based interventions (EBI) (Aarons, 2005; Kazdin, 2008; Ollendick & Davis, 2004). Researchers and practitioners are members of distinctive cultures and communication across these borders of practice has been problematic. Researchers are commonly viewed as failing to generate 'usable knowledge' for practitioner decision-making and practitioners are viewed as relying more on experience than data (Graham et al, 2006; Rosenfield, 2000). The limited number of EBIs delivered in practice settings to date, has been most often viewed as the responsibility of practitioners, but researchers have a duty to diminish barriers toward EBI adoption in practice (Kratochwill & Shernoff, 2003).

Translation of research into practice rests on an assumption that practitioners will adopt interventions based on published research findings (McHugh & Barlow, 2010). Practitioners must weigh and draw conclusions about the relative merits of alternative interventions as more and more research gets produced (Mullen & Bacon, 2003; Watkins, 2009). Practitioners, however, lack available time, resources and systematic methods to review the large amounts of research generated; this is a daunting task to achieve over time (Glass, McGaw & Smith 1981; Ollendick & Davis, 2004; White and Kratochwill, 2005; Wolfe, 1986). Even single primary studies are tremendously complex to interpret given the inclusion of multiple indices of outcomes where some, but not others are found significant (Kazdin, 2008; Quintana & Minami, 2006). It is naïve to assume that

practitioners can assess, appraise and adopt new knowledge as it becomes available (McNeil, 2006). The traditional emphasis on knowledge creation within the academy has paid limited attention to methods for disseminating empirically-founded knowledge into daily practices (Fixsen, Blase, Naoom & Wallace, 2009).

No one intervention study presents perfect, indisputable findings easily translatable into practice decision-making. Objective decision-making in practice must consider systematic weighting of vast amounts of scientific evidence taking into account potential sources of error, as well as similarities and differences across studies (Beutler, 2009; Waas, 2005). Unfortunately, research in school psychology has hampered the practitioner's consumption of research by the prevailing use of inadequate statistical methodology leading to decision-making based more on 'fads and the bandwagon effect' rather than cumulative scientific knowledge (Kehl & Bray, 2007; Trachtman, 1981). Within the prevalent disconnected, one-shot model of intervention study publication process (Hoagwood & Johnson, 2003), practice-based decisions are highly susceptible to "cherry-picking" individual studies potentially based on preconceived notions without adequate syntheses of cumulative research (Beutler, 2009). Without better solutions that diminish the divide between research and practice those who would benefit most from evidence-based psychological interventions do not receive them (Hoagwood et al., 2002; Kazak, et al., 2010). Clients receiving psychological services have a right to receive interventions shown to be effective, supported by accumulated research, and demonstrated to be the best use of limited resources (Fixsen et al, 2009).

Adopting and researching evidence-based interventions (EBI's) within authentic community practice settings has not become a mainstream practice despite heightened

interest (Chafouleas & Riley-Tillman, 2005; Kazak et al., 2010; McHugh & Barlow, 2010) and thirty-plus years of abiding by a scientist-practitioner model of psychological training (Hughes, Kaufman & Miller, 2010). Deployment of EBIs with youth populations has been especially slow to evolve compared to that for adults (Silverman, Kurtines & Hoagwood, 2004). This is especially concerning when child and adolescent psychological interventions routinely provided in community practice settings without allegiance to a particular type of evidence-based intervention ("treatment as usual") have generally shown no benefit, and in some cases incur harm (Bickman, Lambert, Andrade & Penaloza, 2000; Weiss, Catron, Harris & Phung, 1999; Weisz, Huey and Weersing, 1998; Weisz, Jensen-Doss & Hawley 2006).

Meta-analyses of child and adolescent psychotherapy research studies have shown that the more studies resemble real-world practice settings, the less effective interventions are found to be (Weisz, et al., 2006; Weisz & Jensen, 2001). Almost all of the studies measuring benefits from evidence-based psychosocial interventions for youth have been conducted in or within substantial proximity to highly controlled university settings rather than authentic practice settings in communities using practitioner intervention delivery (Brestan & Eyberg, 1998; Hoagwood & Johnson, 2003, Steele, Elkin & Roberts, 2008; Weisz, Jenson-Doss & Hawley, 2005). Paradoxically, there is a wealth of research evidence showing youth psychological interventions not generally found in use within practice settings, do produce beneficial outcomes under ideal research conditions (Chambless & Ollendick, 2001; Kazdin, Bass, Ayers & Rodgers, 1990; Ollendick & King, 2004; Weisz, Weiss, Alicke & Klotz, 1987; Weisz, Weiss, Han, Granger & Morton, 1995).

Transforming Practice Decision-Making:
Meta-Analysis of an Evidence-Based Intervention

This study explores the practical benefits from public investments already made in a set of intervention studies, where the return from these types of research ventures has been quite poor within public service settings to date (Chorptita & Regan, 2009). The meta-analysis procedures followed for this investigation avoid the pitfalls of narrative reviews of research involving examination of studies one at a time. Meta-analysis is a well-established statistical method designed to synthesize accumulated research evidence from several like studies in a systematic, transparent manner (Borenstein, Hedges, Higgins, & Rothstein, 2009; Rosenthal & DiMatteo, 2001).

Meta-analytic aggregation of outcomes from successive implementations of an EBI has been regarded as a particularly appropriate method to evaluate program benefits by providing more stable, reliable, and valid estimates of its advantages than would be obtained by reviewing each study independently (Wolf, 1986). Practitioners will more likely connect research to adoption decisions in practice by selecting EBIs shown to produce potent benefits robustly across several replications of EBI implementation. How robust an intervention will be, for whom, and under what conditions may be discovered by studying data from repeated deployment of studies substantiating intervention effects (Weisz, et al., 2006). Synthesizing findings across research studies helps identify variables that moderate intervention effects, making it possible to hopefully translate similar benefits beyond controlled, university-based research settings (Dunst, Hamby & Trivette, 2004; Kazdin, 2008). Measured benefits from primary developer-based research may present reasonable and attainable benchmarks for subsequent implementations of the same intervention, offering empirical estimates for what a

practitioner might reasonably expect from adoption of an intervention (Hunsely and Lee, 2007; Southam-Gerow, Marder & Austin, 2008; Weersing & Weisz, 2002).

This meta-analysis includes roughly 30 years of research examining the Incredible Years Parent Training Program (IYPTP) which has been used to reduce young children's anti-social conduct behavior problems (Webster-Stratton, 2007). A preliminary meta-analysis (Sougstad, Oka, Carlson & Tomac, 2008) found studies by the primary developer and her associates using a clinic-based, group-administered form of IYPTP to yield roughly a one standard deviation benefit (g = -0.986; -1.211 to -0.760) in immediate reductions in child conduct problems. In contrast, independent replications of the same form of IYPTP realized small to moderate effects of a statistically-significant lower magnitude with 95% confidence (g = -0.383; -0.695 to -0.071). A third set of IYPTP studies conducted by the inventor's primary research team using a self-administered videotape form of IYPTP (that does not include clinical group intervention) yielded about a half standard deviation benefit (g = -0.531; -0.737 to -0.324). No statistically significant difference in the amount of small to moderate reduction in child conduct problems was found from the clinic-based, group-administered form of IYPTP compared to results from self-administered IYPTP when it was administered independent of the primary research team.

Since clinic-based, group-administered IYPTP is more costly to deliver and requires greater investments of time and resources than the self-administered form, further research is needed to discover what factors are associated with the greater magnitude of benefits found by the primary research team's implementation of the clinic-based, group-administered IYPTP but not by independent replications. Preliminary meta-

analysis suggests that there may be insufficient benefits found from independent replications of the clinic-based, group-administered IYPTP to warrant the more substantial investment needed to adopt this form of IYPTP in practice, rather than the self-administered form of IYPTP.

This study further examines the significant differences found between inventor-based research and that of independent replications of IYPTP by including additional studies published since preliminary data were collected over 2 years ago. In addition, this study identifies some factors most important to realizing maximum benefits (i.e. benchmark) from IYPTP in order to consider whether these could be feasibly incorporated into settings beyond the primary inventor-based research team. Preliminary meta-analysis did not consider other critical variables that might have moderated outcomes in addition to the type and setting of IYPTP delivery (e.g. severity of child problems and parenting at outset). Some of these factors are examined within this study. It is important to further investigate other plausible explanations for the disparity in results among forms and settings of IYPTP delivery to more confidently draw conclusions as well as to accurately discuss implications for future implementation of IYPTP. This study expands the scope of the initial meta-analysis by including more research studies and investigates additional independent and dependent variables.

Preliminary findings suggest it is difficult to achieve the same level of benefit from an EBI in independent replications that was found by inventor-based research. This study is especially timely since IYPTP was recently downgraded from being a 'well-established' intervention 10 years ago (Brestan & Eyberg, 1998) to a 'probably efficacious' intervention in an updated review of evidence-based psychosocial treatments

for youth exhibiting disruptive behaviors (Eyberg, et al., 2008). Due to a recording error in the direction of group differences for an independent replication of IYPTP in the earlier review (Spaccarelli, Cotler & Penman, 1992), IYPTP had been incorrectly classified as "well-established." Corrected data showed IYPTP no longer met the 'well established' criteria of having significant effects from the intervention demonstrated by at least 2 different investigators, at least one being conducted independent of the inventor. However, Eyberg and colleagues (2008) utilized an apparently arbitrary standard that at least 50% of the reliable and valid measures for child misbehavior within a study must be superior to the comparison condition to be supportive of the intervention. The current meta-analysis provides a more comprehensive criterion for evaluating intervention outcomes whereby rating scale and independent observations of child conduct behavior problems are aggregated to yield one standardized measurement of benefit for each study without excluding 50 % of the results. Meta-analysis also performs a weighting of the evidence relative to the size of each study to achieve a better estimate of actual program benefits. In addition, this study examines not only child outcomes as a measure of intervention benefit, but also examines data regarding changes in parenting. Changes in parenting are believed to be the mechanism of change for parent training programs, but research testing this empirically has been quite limited, especially with clinical samples (DeGarmo, Patterson & Forgatch, 2004).

Primary research offered by the program developer for the group-administered form of IYPTP offers tremendous hope for realizing large, potent benefits in reducing conduct problems in young children at a much greater level than prior meta-analyses had suggested for group administered parent training (Lundahl, Risser & Lovejoy, 2005;

Maughan, Christiansen, Jenson, Olympia, & Clark, 2005). This study examines the variety of outcomes obtained across multiple studies of IYPTP in order to form a greater understanding about what factors are associated with greater benefits.

If the significantly greater EBI benefits obtained by intervention developers are to be realized by authentic community-based practitioners, research must indentify variables that moderate benefits across multiple replications. This type of study is rarely produced within the current literature-base but is highly desirable to diminish the gap between research and practice (Kazdin, 2008). Knowing how moderators work across multiple conditions has important implications for potential EBI adoption by practitioners, and is essential to furthering research and theory directed toward alleviating undesirable consequences from psychological problems.

Practitioner opinions and expertise are crucial for deciding what evidence applies to individual situations. However, without weighing available empirical evaluations of interventions, practice decision-making more likely resembles the world of politics, which is often more about power than truth (McNeil, 2006). In the absence of considering applicable research findings, intervention adoption decisions among practitioners and program administrators may be unduly influenced by personal biases, interpersonal and organizational dynamics. Good intentions and years of professional experience do not ensure competent decision-making or favorable intervention outcomes (Watkins, 2009). Interventions may be adopted prematurely without consideration for how much benefit has been produced across differing conditions (Graham et al, 2006). There is also an alarming possibility practitioners will ignore empirical evidence that has found some interventions actually incur harm (Ang & Hughes, 2002; Arnold & Hughes,

1999; Dishion, McCord & Poulin, 1999; Dodge, Dishion & Lansford, 2006; Rhule, 2005).

CHAPTER 2

LITERATURE REVIEW

Early-Onset Conduct Problems and the Desirability of Parent Intervention

Students exhibiting antisocial conduct behavior problems present unique, difficult, often seemingly insurmountable challenges to educators, families and communities (Brophy, 1995; Fortin, 2003; Gerler, 2004; Glicken, 2004; Hughes, Crothers, Jimerson, 2008; Lahey, Miller, Gordon & Riley, 1999). Youth conduct behavior problems frequently disrupt family, school and community climates, perpetuate patterns of violence, embroil home-school discord, exacerbate parental and sibling conflict, and lead to school suspensions, expulsions, and dropping out of school (Frick, 2004; Lahey, Moffitt & Caspi, 2003; Lahey & Waldman, 2003; Olweus, Limber & Mihakic, 1999; Sprague & Walker, 2000; Walker, Ramsey & Gresham, 2004). Antisocial behavior in young children is a frequent precursor to chronic unemployment, uneducated/unskilled adults, substance abuse, higher rates of crime and victimization, domestic violence, child abuse & neglect, and costly legal prosecutions, litigation and incarceration (Fergusson, Horwood, & Ridder, 2005; Frick, 2004; Jaffee, Belsky, Harrington, Caspi & Moffitt, 2006; Jafee, Caspi, Moffitt & Taylor, 2004; Kokko & Pulkkinen, 2000; Moffitt, Caspi, Harrington, & Milne, 2002; Snyder and Stoolmiller, 2004).

Serious antisocial behavior among older children, adolescents and adults may be predicted from indicators evident in children as young as toddlers and preschoolers (Dodge, 2001; Eddy,Reid & Curry, 2002; Keenan, & Shaw, 2003; Lahey & Waldman, 2003; Patterson, 2002; Reid, 1993; Reid, Patterson & Snyder, 2002; Reid, Webster-

Stratton, & Baydar, 2004; Snyder & Stoolmiller, 2002; Tremblay, et al., 2004). Patterson and Yoerger (2002) contend that the ideal window for intervention to ameliorate youth antisocial behavior is during the toddler stage of development. However, these youth are frequently not identified until preschool where their behaviors reap havoc leading to repeated expulsions (Egger & Angold, 2006; Gilliam, 2005). The Office of Juvenile Justice and Delinquency Prevention's Study Group calculates children engaged in antisocial behavior by the ages of 7 to 12 years have a two to three times greater chance of developing chronic, serious antisocial behavior patterns in adulthood (Loeber, Farrington & Petechuk, 2003).

Two developmental pathways account for the development of antisocial conduct behavior patterns in youth (Bloomquist & Schnell, 2002; McCabe, Hough, Wood & Yeh, 2001; Nock, Kazdin, Hiripi, & Kessler, 2006; Patterson & Yoerger, 2002). The 'early-onset, life-persistent group', is known to exhibit antisocial behaviors beginning prior to entering elementary school. For this early-onset group there is an especially high association with familial variables related to both constitutional predispositions and psychosocial risk factors, especially ineffective parenting (Moffitt, 2005; Rutter, 2005). This group is also known to show more overt, severe and stable forms of antisocial behavior, higher rates of multiple psychiatric disorders, and to be more impervious to treatment efforts the later these are instituted (McCabe, et al. 2001; Moffitt, 2005; Patterson et al., 2002). The second, 'late-onset group', shows more moderate levels of antisocial behaviors, possesses greater social skills and shows a more favorable long-term prognosis than the early-onset. Ethnic minority status and exposure to deviant peers are more strongly related to the later-onset group (McCabe, et al, 2001).

In the absence of effective interventions younger children displaying antisocial conduct problem behaviors are known to become worse and increasingly resistant to change efforts (Reid, 1993; Shaw, Lacourse, & Nagin 2005; Watson, Fischer, Andreas & Smith, 2004). Children exhibiting antisocial behavior in middle childhood (between the ages of 7 to 12 years) are two to three times more likely to develop chronic, serious antisocial behavior patterns into adulthood (Loeber et al., 2003). Those youth exhibiting antisocial behavior patterns beyond age 8 years and completing the third grade typically display persistent symptoms of chronic disorders unlikely to be completely ameliorated (Patterson & Yoerger, 2002). Parents of early-onset youth are most in need of the earliest treatment to prevent a worsening of symptoms (Patterson et al., 2002).

Parenting plays a central role in the evolution and perpetuation of antisocial conduct problems among young children, as well as in their potential diminution (Berkowitz, 2003; Capaldi, DeGarmo, Patterson & Forgatch, 2002; Duncan, 1999; Finkelhor, Ormrod, Turner & Hamby 2005; Goldstein & Rider, 2006; Henggeler, Schoenwald, Borduin, Rowland & Cunningham, 2009; Kumpfer & Alvarado, 2003; Reid, et al., 2002; Reid et al., 2004). Family interactions account for over 60% of the variance in individual differences in youth deviancy (Patterson, et al., 2002). Many key risk factors for youth antisocial behavior involve parenting variables including attitudes and behaviors reinforcing violence, poor conflict resolution, neglect, abuse, and dysfunctional communication patterns (Jaffee, et al., 2004; McFarlane, Groff, O'Brien and Watson; 2003; Patterson, 2002). Even for the most severe form of antisocial conduct problems involving callous and unemotional traits (CU) higher levels of youth-perceived warmth and involvement from parents predicted decreases in CU traits and antisocial

behaviors. Similarly, parents perceived to display low warmth and involvement were associated with youth exhibiting higher CU traits. Lower levels of physical punishment used by parents were also associated with decreasing CU traits over time (Pardini, Lochman & Powell, 2007).

Accumulated research based on self-determination theory also convincingly points to parenting as a crucial factor leading children to not only comply with social standards of behavior but also to internalize and hold societal values intrinsically in the absence of external rewards for doing so (Grolnick, Deci & Ryan, 1997). Greater social competence, internalized regulation of behaviors and healthier adaptation is related to parents who provide an environment characterized by autonomy support, optimal structure and warm involvement. These parenting dimensions have been found to predict child outcomes including moral reasoning, behavior adjustment and self-regulation both at home and in school settings (Grolnick, et al., 1997). The converse of autonomy-supportive parenting is controlling parenting which is associated with children low in self-regulated behaviors, higher in acting out behaviors, lower in teacher rated competence, lower in academic achievement and grades (Grolnick & Apostoleris, 2002).

Parenting practices also appear to be a risk factor associated with the development of negative, hostile, abrasive and aggressive peer relationships (Anthonysamy & Zimmer-Gembeck, 2007; Grimes, Klein & Putallaz, 2004; Webster-Stratton & Lindsay, 1999). Negative, over-controlling, and critical parenting is related to more negative, abrasive and aggressive peer relationships, while a parenting style characterized by warmth, support, agreeableness and moderately directive tactics is associated with youth who have more positive and harmonious peer interactions (Grimes, et. al, 2004).

Attempts to change antisocial attitudes, beliefs and behaviors among youth across home, school and community settings will likely achieve limited success if changes are not reinforced within the family environment, especially with parents of younger children. There appears to be a general consensus regarding the types of parental conditions associated with pro-social behavior patterns in children including warm, responsive relations with a parental figure, a climate of caring in which this is modeled, reinforced and attributed to the child when relevant, and discipline conveys a commitment to pro-social values using the expression of adult feelings highlighting effects of behaviors on others, using clear rules and principles (Bronson, 2000).

Parent intervention is a vital and promising approach to reducing the prevalence and severity of conduct problems in young children at its earliest point of expression by "nipping it in the bud" (Webster-Stratton & Taylor, 2001). Recent meta-analysis of 71 study outcomes from treatment of antisocial behaviors found age to be a significant moderator of effect for the two prevalent treatments, whereby behavioral parent training was superior to that of cognitive-behavior therapy for the youngest children included in the study (McCart, Priester, Davies & Azen, 2006). Eyberg and colleagues (2008) most recent review of the extant empirical evidence for youth interventions for disruptive behavior led to the recommendation that "clinicians consider parent training as the first line approach for young children..." (p. 233).

Parent-focused intervention is the most extensively studied and supported form of treatment for youth conduct problems (Weisz, Hawley & Doss, 2004). A review of the literature regarding youth antisocial conduct problems led Bloomquist and Schnell (2002) to conclude parent and family skills training should be a mandatory intervention given

the clear role parent and family factors play in the escalation of these problems and the effectiveness of these interventions. Webster-Stratton (2007) and colleagues have been demonstrating for about 30 years that treating parents of children with early-onset conduct problems in groups using IYPTP is an effective, less costly and less time-consuming form of parenting intervention than treating each individual family separately.

The Incredible Years Parent Training Program

Webster-Stratton's Incredible Years Parent Training Program (IYPTP) is frequently recognized for having a substantial base of evidence as an intervention and prevention program for families and youth exhibiting conduct problems. IYPTP has been identified as one of only a few youth psychological interventions with an accumulation of scientific evidence sufficient to be a "probably efficacious" EBI according to the stringent "Chambless" criteria (Brestan & Eyberg, 1998; Eyberg, et al., 2008). The United States Office of Juvenile Justice and Delinquency Prevention (OJJDP) selected IYPTP as a "Model Blueprint Program" for early violence prevention (Mihalic, Irwin, Elliot, Fagan & Hansen, 2001; Webster-Stratton, C., Mihalic, S., Fagan, A., Arnold, D., Taylor, T., & Tingley, C., 2001). Within the National Registry of Evidence-Based Programs and Practices (NREPP) published by the Department of Health and Human Services, Substance Abuse and Mental Health Administration (SAMHSA, n.d.) the Incredible Years Program was reviewed, yielding a score of 4.0 on a scale from 0.0 to 4.0 for "Readiness for Dissemination." This highest rating possible was based on a review of the implementation materials, quality assurance characteristics, training and support available within this program. Most recently the "Futures Task Force on Family-School

Partnerships" (http://fsp.unl.edu/) conducted an extensive literature review of evidence-based family interventions naming the Incredible Years Training Series as one of fifteen family and family-school models of intervention showing a "promising evidence-base" (Sheridan & Kratochwill, 2007).

The Incredible Years focuses on intervening with children between the ages of 3 and 8 who are at-risk of developing or already exhibit conduct problems. The program has been further adapted to work with children between the ages of 2 to 12 years of age, as well as specific to frequently co-existing problems (e.g. ADHD). The goals of the program are twofold (Webster-Stratton, 2000; Webster-Stratton & Hancock, 1998; Webster-Stratton et al., 2001):

1.) To develop comprehensive treatment programs for young children with early onset conduct problems.

2.) Development of cost-effective, community-based universal prevention programs that all families and teachers of young children can use to promote social competence and to prevent children from developing conduct problems in the first place.

The Incredible Years Parent Training Program (IYPTP) uses manuals for the trainers (co-leaders) and parent participants, and is based on a collaborative group format of about 10 to 14 parents using video-taped modeling, group discussions, role-playing and rehearsal (Webster-Stratton, 1984). More recently Webster-Stratton and colleagues have developed Child Training and Teacher Training Programs and there is a Self-Administered form of IYPTP involving parents viewing the videotapes and completing

workbooks without a group experience (Webster-Stratton et al., 1988; Webster-Stratton, Reid, & Hammond, 2004).

The Incredible Years Parent Training Program includes many features that would suggest good transportability into various settings while maintaining fidelity of treatment (Webster-Stratton & Herman, 2010). IYPTP manuals are extensive and include scripts for role-plays and discussions. Videotape-based content increases the standardization of the treatment program, provides pro-social modeling (Kazdin, 2005) that potentially generalizes to situations outside the training. Videos showing less favorable parenting in some videos encourage parents to think realistically about barriers to desirable parenting and plan for ways to overcome these. Additionally, the program developers have made videotapes of practitioners modeling delivery of the treatment. Extensive training workshops and ongoing consultation are made available by the program inventors and other experts certified by the inventor. Practitioners can become certified in IYPTP by extensive review of their implementation by the inventor and her team.

The Incredible Years Parent Training Program (IYPTP) is one of the only parent training programs devoted to treating children's antisocial conduct problems that has an extensive array of research studies conducted by both the primary developer (Webster-Stratton et al., 2001; Webster-Stratton, 2007) and independent researchers spanning about 30 years, that has not had its effects summarized using meta-analysis. This study is timely relative to recent meta-analyses performed for Parent-Child Interaction Therapy and Triple-P Parent Training (de Graff et al., 2008; Thomas & Zimmer-Gembeck, 2007), allowing additional comparisons between these treatments that are all designed to address similar child behavior problems. Of importance to a main focus of the current study is

that neither of the recent meta-analyses of other parent intervention programs has contrasted studies from primary inventors with that of independent replications. This study introduces a novel comparison to the research literature base that may be potentially beneficial for future analyses of already existing data on other interventions which could influence practitioner adoption decisions. Kazdin (2008) contends that the integration of research science and practice is timely and more important to address than ever before when considering the magnitude of stakes involved in academic and clinical training, research, practice and health care in general.

Connecting Intervention Research Science with Practitioner Service

Several models have been proposed to diminish the research to practice gap. Chorpita and Daleiden (2009) advocate distillation of the most beneficial elements in common across controlled intervention research for particular disorders to instill best practices. Nutley, Walter and Davies (2009) identified three additional models for bringing research science to practice. The research-based practitioner model is the default, currently established model relying on practitioner responsibility to locate and digest research to integrate into practice decision-making. The embedded research model brings research into practice indirectly away from practitioner delivery by way of systems and standards (e.g. policies, frameworks, procedures). The organizational excellence model views the organizations where interventions are provided as the locus of experimentation, evaluation and practice development based on partnerships with universities to research local factors related to EBI implementation. These first three models work in a top-down fashion whereby controlled research findings are created by

researchers and findings are supposed to trickle down to research-informed local decision-making usually supported by workshops. The last model operates more from a bottom-up approach that has recently become more attractive (Fixsen et al., 2009). This approach acquires research findings using studies located within local practice settings thereby creating knowledge examining authentic community-based practice factors that may not be considered within most top-down, controlled university-based research.

In keeping with the latter, Kazak and colleagues (2010) assert that recent research-based implementation of EBI's into local practice settings will hopefully lead to "improvements in quality" by "normalization" and "demystification" of the scientific enterprise, that will "facilitate the constant reevaluation, refinement, and improvement of services" by creating "locally relevant evidence" (p. 94). It seems improbable, however, that research can ever consider all possible practice variables for every EBI (Kazdin , 2008). While customization and modification of evidence-based protocols to fit specific characteristics of clients makes sense intuitively, there is mixed empirical evidence supporting the superiority of customization over standardized interventions for youth (Castro, Barrera, & Steiker, 2010; Shirk & McMakin, 2008;). Regardless of whether future intervention research follows more traditional top-down or bottom-up approaches an essential need remains, that delivery of research findings be produced in usable forms that generalize beyond research into practitioner decision-making and behavior in service delivery (Chafouleas & Riley-Tillman, 2005; Graczyk, Domitrovich, Small & Zins, 2006; Kazdin, 2008; Rosenfield, 2000; Southam-Gerow et al., 2008). Researchers will produce information more easily translatable into practitioner knowledge when interventions are studied repeatedly by different researchers under various conditions, reported using

statistics about actual effects obtained in relation to variables across studies. Where adjustments and alterations of intervention replications are likely needed to bring evidence-based interventions into a variety of settings these differences can be studied systematically in relation to measured effects (Weisz, et al., 2006).

The most respected professional organizations representing psychologists serving within clinical and school settings requires significant benefits from an intervention independent of the inventor to be considered a "well-established" EBI (Brestan & Eyberg 1998, Eyberg,et al., 2008; Chambless & Holon, 1998; Kratochwill & Stoiber, 2002; Task Force on Evidence Based Interventions in School Psychology, 2003). Studying replication outcomes and associated factors provides a scientifically acceptable normative process by which facts are either falsified or established as reproducible phenomenon (Popper, 1954; Radder, 1996; Schmidt, 2009). Little attention has been paid to what roles replication plays in the development of knowledge, theory and practice in psychology (Schmidt, 2009). Alternative explanations to links between intervention characteristics and intended outcomes may be ruled out when research is replicated (Stoiber, 2002). Replications provide information regarding the probability of an intervention's future success while providing a means of controlling for variables that may limit and or enhance generalization of an intervention (Cancelli et al., 1989). Evidence for the effectiveness of an intervention will remain weak until an effect is found stable and generalizable across multiple study methods and samples (Wilson & Lipsey, 2001).

Combining results from multiple random controlled trial (RCT) studies of the same intervention provides the most accurate and reliable benchmarks of intervention

benefits (Hunsely and Lee, 2007). The number of studies and the range of conditions such as types of settings, clients, problems and therapists will enhance the level of confidence in the value of the intervention and the applicability of the benchmark in additional settings beyond research (Hunsely and Lee, 2007). For an EBI to have utility in practice, attention must be paid to the degree of robustness for an intervention across several deliveries accounting for the generality of effects across diverse recipients, intervention providers and settings, as well as the feasibility of implementing the EBI in authentic real-world settings in a cost effective manner (American Psychological Association, 2002; American Psychological Association's Presidential Task Force on Evidence-Based Practice, 2006; American Psychological Association Task Force on Evidence-Based Practice for Children and Adolescents, 2008). The generalizability and feasibility of implementing an evidence-based intervention and finding similar benefits beyond primary research is founded on a process of replicating interventions across multiple contexts while maintaining the integrity of core mechanisms of change for an EBI (Bhattacharyya, Reeves, & Zwarenstein, 2009; Fixsen et al., 2009).

Replications beyond primary, inventor-based research demonstrates the degree of generalizability for an intervention to the population of concern, as well as the robust potency of EBI across various contexts (Flay et al., 2005). While the magnitude of an effect from an EBI study (small, medium, or large) likely holds some significance, comparison of measured benefits relative to those previously obtained in the same research area or type of outcome provides the most useful information about the actual value of an EBI (Durlak, 2002).

Overcoming Statistical Barriers to Bringing Research into Practice

Increasing the number of researched implementations of EBI's within authentic practice settings may be a useful method for diminishing the research to practice gap. How research is reported no matter where interventions take place is a barrier overcome much more rapidly than deploying EBI's across multitudes of practice settings. The dominant use of null-hypothesis significance testing to demonstrate an intervention is likely effective has set a substantially low standard in terms of driving research questions and providing useful information (Beutler, 2009; Hinshaw & Park, 1999). It is difficult to translate findings of statistical significance into practical meaning (Kehle & Bray, 2006). Acceptance or rejection of the null hypothesis is largely contingent upon the statistical power of the study (largely determined by sample size) and fraught with potential error (Beutler, 2009; Kazdin, 2008; Schmidt, 2009; Volker, 2006). Some studies may yield statistically significant results because of their large sample sizes but have little practical meaning. Other studies may have practical significance but lack statistically significant results because of a lack of sufficient power due to a small sample size.

Rejecting a null hypothesis in favor of an unlikely, statistically significant result tells nothing about what the true value of an effect parameter is likely to be within a population. When we reject a null hypothesis that says the effects of two interventions (or an intervention and control group) are equal, we only know what the outcome is unlikely to be (e.g. that the intervention does not produce equal results). But we do not know how different their respective effects are within the population. In order to determine estimates of a parameter for an intervention within the population, effect sizes

must be calculated including confidence intervals (Kehle, Bray, Chafouleas & Kawano, 2007; Sanabria & Killeen, 2007).

Effect sizes provide a standardized unit of measure (*z*-scores) for interpreting the magnitude and direction of an intervention effect, and are useful for comparing results from study to study, where slightly different measures may have been used to measure the same category of outcome variable (Olejnik & Algina, 2000). Wilkinson & the American Psychological Association's Task Force on Statistical Inference (1999) recommended reporting effect sizes for primary outcomes whenever a *p*-value (probability) is reported, along with confidence intervals around observed effect sizes. These statistics facilitate conclusions from study findings that are more intuitive, understandable and translatable into practical meaning. Unfortunately the reporting of effect sizes in school psychology has been a rarity (Swaminathan & Rogers, 2007; Volker, 2006). Rather than referring to arbitrary interpretations of statistical significance decision-makers should be able to weigh the likely benefits from an intervention in relation to the level of resources needed to implement an intervention (Sanabria & Killeen, 2007). Meta-analysis of youth psychotherapy studies across the past four decades found most were underpowered due to small sample sizes (Weisz et al., 2005), supporting the notion that effects across multiple studies of an EBI are needed to adequately investigate the probable effects of youth mental health interventions within a population.

Effect sizes from studies in isolation provide limited information compared to what meta-analysis of multiple effect sizes and associated variables can tell us. For example, the rigorous coding system from the *Procedural and Coding Manual for the*

Identification of Evidence-Based Interventions (Task Force on Evidence Based Interventions in School Psychology, 2003) prescribes calculation of effect sizes along with tabular descriptions of study characteristics but falls short of using meta-analysis. Following these procedures Bates (2005) obtained a considerably wide range of effect sizes for various family-school interventions with preschool children that spanned from 0.16 (very small benefit, possibly none) to 1.9 (huge benefit exceeding most effect sizes for psychological interventions). Absent meta-analysis, Bates was left to conclude there were no emerging patterns among the 15 studies reviewed. In fact there may have been some patterns related to these studies. However, without meta-analytic methods to investigate this possibility narrative reviews are unlikely to uncover these, even when effect sizes are calculated for each study.

Prior Meta-Analyses and Multivariate Studies of Parent Training Interventions

Meta-analysis provides practitioners efficient access to a large body of evidence (Borenstein et al., 2009; Cooper & Hedges, 1994; Lipsey & Wilson, 2001a; Rosenthal, 1991; Wolf, 1986). Studies using meta-analysis have examined parent training in terms of broad categories (e.g., behavioral parent training) as well as specific to a particular parent training intervention.

Psychological treatment research is dominated by smaller studies with limited power statistically (Kazdin, 2008). A review of 383 psychological treatments for youth spanning from 1962 to 2002 found small sample sizes left studies underpowered and lacking in fair tests of impact (Weisz et al., 2006). Smaller less comprehensive studies must accumulate over time to provide a sketch of important patterns translatable to practice (Chorpita, 2003). Accumulating small scale trials of an intervention using meta-

analysis can lead to more effective practices in the absence of large, statistically powerful research studies. The belief that some interventions will have a more favorable impact when implemented by practitioners is most adequately articulated and disseminated when there is reliable data supporting the effectiveness of an intervention across rigorous trials (Bhattacharyya et al., 2009).

One of the first identified meta-analysis examining parent training effects was conducted roughly 20 years ago targeting one particular parent intervention (Cedar & Levant, 1990). Effects from twenty-six studies of the Parent Effectiveness Training Program (PET) spanning 15 years of implementation were aggregated. The author's conclusions purport a small effect size (0.33) favoring PET though methodological problems limit what conclusions may be drawn from this study. Effect sizes were not weighted for sample size hence most of the studies included within the meta-analysis that were quite small, biased the overall results. Equal weights for studies regardless of sample size allows smaller studies to exert more influence on the overall effect size than should be warranted proportionate to other studies containing larger sample sizes. Another concern was higher effect sizes were found among studies with elevated subject drop-out rates (e.g. results for subjects who dropped out and would most likely score lower on outcome measures were not included) which likely led to inflated outcomes. Even with these methodological flaws likely inflating results the author's optimistic conclusion that PET is effective seems contraindicated by mean outcomes reported immediately following PET that yielded statistically non-significant results (Table 1, p. 378). Had the authors calculated and reported confidence intervals around their small effect size for PET, it appears likely they would have found the range of possible effect

sizes included 0. They would have had to conclude there may be insufficient evidence to show PET actually yields a significant effect across their sample of studies.

The most intriguing finding reported by Cedar & Levant (1990) is that higher effect sizes were associated with studies implemented by researchers having an allegiance to the PET program. One plausible explanation for this observation not reported by the authors is that these results tapped into greater effects found among primary inventor-based research versus those who tried to independently replicate earlier results. Greater outcomes obtained by researchers most closely aligned with a particular approach is a common characteristic among *efficacy* research studies where there are usually more favorable characteristics including greater control over the research setting, recruited samples that are more homogeneous, more substantial training and supervision for treatment providers who may be considered specialists in implementing the particular intervention, and who also hold lower caseloads than is commonly found under conditions more typical of practice-settings (Chorpita. 2003; Weisz, Weis & Donenberg, 1992).

Serketich & Dumas (1996) conducted a meta-analysis examining 26 studies using behavioral parent training (BPT) to intervene with youth exhibiting antisocial behavior. Their analysis was restricted to studies displaying rigorous methodologies, from which general conclusions about the effect of BPT would be most likely made without undue influence of poor quality research studies. A moderately strong effect size of 0.86 was reported for BPT though effect size calculations were not weighted to account for differing sample sizes across studies. Evidence of how problematic this was is exemplified by the author's finding that sample size had a statistically significant

negative impact ($p < .05$) on the overall effect size obtained for this meta-analysis. Studies with larger samples tended to yield smaller effect sizes. Smaller studies exerted the same weight as larger studies and these smaller studies with higher effects exerted more influence on the overall effect size than would be warranted given the smaller proportion of the population they sampled.

Favorable to the focus of this study, Serketich & Dumas (1996) noted research studies using IYPTP (Webster-Stratton 1992; Webster-Stratton et al., 1988) were outstanding for exceptional methodological qualities even among their sample of rigorous studies. Samples were described extensively. Subject attrition rates, means and standard deviations for all variables at pre- and post-assessment, as well as F values for all group comparisons regardless of significance were also reported. These studies also included randomly assigned participants to treatment and control groups, utilized uniform observers to code child behaviors, and also collected a myriad of data on parental report, teacher report and parental adjustment. The accumulated research on this program likely holds important information not available among studies of lesser methodological quality.

Both the aforementioned studies (Cedar & Levant, 1990; Serketich & Dumas, 1996) did not systematically investigate within-group heterogeneity among effect sizes in their meta-analyses. This is highly problematic considering the relatively large standard deviations reported in each study. Inadequate homogeneity among groups of effect sizes warrants careful investigation to uncover what factors remain unaccounted for among wide dispersions of scores. It is extremely difficult to draw definitive conclusions about a set of studies that display substantially different results.

Lundahl and colleagues (2006) identified moderators of parent training effects by examining within-group heterogeneity. However, like the two prior meta-analyses of parent training, the effect size calculations used for this study were not weighted for sample size. Therefore overall findings are likely biased by smaller studies and should be interpreted cautiously and with reservations. Each study included in their meta-analysis was listed in a table along with their sample sizes. Hedges & Olkin (1985) found the most severe bias occurs among studies with sample sizes below 20, which constituted roughly half of the studies analyzed by Lundahl et al. (2006).

Maughan, et al (2005) provided a comprehensive and technically sophisticated meta-analysis of behavioral parent training (BPT). For BPT studies weighted effect sizes ranged from small to moderate. The between-subjects weighted effect size was .30, while the within-subjects weighted effect size was .68, and the single-subject design weighted effect size was .54, although single-subject effect sizes varied widely between two different methods of calculation. Researchers noted the higher effect size found for within-subjects studies is consistent with prior findings of Lipsey & Wilson (2001a), who concluded that one group pre/post designs potentially inflate effect sizes compared with the results of between-groups designs. This study provides the most empirically sound meta-analytic evidence indicating behavioral parent training is a successful intervention for reducing disruptive/externalizing behaviors in children. However, making use of this finding in practice remains difficult, generating a number of new questions needing to be addressed by further research.

Knowing that BPT significantly reduces disruptive-externalizing behaviors in children does not tell a practitioner which behavioral parent training intervention to

select. Are all BPT programs equal? Are some preferred in certain settings over others? One of the potential weaknesses of meta-analysis is making too broad a claim about a swath of interventions and assuming all the interventions within a class or brand name are similar (Beutler, 2009). More contextual information is needed to choose the most desirable intervention beyond knowing BPT is effective (Ingraham & Oka, 2006). There is a need to move beyond "what works" to "what effects can be expected using a given intervention in a given setting with a given student or family from a given background exhibiting a given type of concern at a given time in the student's development" (Sheridan, 2005; p. 519).

More recent meta-analyses have focused on specific parent training programs. Thomas & Zimmer-Gembeck (2007) conducted a meta-analysis comparing Parent-Child Interaction Therapy (Eyberg, Boggs & Algina, 1995) and the Triple P-Positive Parenting Program (Sanders, Markie-Dadds, Tully & Bor, 2000). Twenty-four studies were included in the final analysis which included reporting effects for various delivery formats for each of these treatments (e.g., standard individual, group, enhanced, self-directed and media forms of Triple P; standard, abbreviated and enhanced PCIT). Treatment effects depended on the outcome measure, pointing out how essential it is to dig deeper into data to understand more about a program's treatment effect beyond a single effect size representing a study for meta-analysis. Independent observations yielded smaller effects than parent reported effects. It was also found that mother and father reports generated different effects. Effects were also found to be different for the different delivery formats and the length of intervention. Inconsistencies in reports of

demographic variables made it difficult to form conclusions relative to issues like single versus two-parent homes, racial identity, and SES.

Subsequent meta-analysis of the Triple P-Positive Parenting Program (Sanders et al., 2000) was recently published (de Graff et al., 2008). The researchers focused on one intensive form of Triple P (Level 4) designed to be administered either individually or in groups to address severe behavioral difficulties in children. Effects on child behavior were measured by including only studies that utilized one particular measure of outcome, the Eyberg Child Behavior Inventory because the vast majority of studies utilized this measure. An overall large effect size (.88) was obtained for immediate treatment outcome and a larger (1.0) benefit was reported for long-term measurement of child behavior. However there was substantial heterogeneity among effects requiring further analysis. Outliers were removed from the analysis lowering an immediate effect size to 0.42, and at 6 months follow-up 0.49. Analysis of variables related to the size of outcomes found studies with a higher proportion of girls had larger long-term benefits, and those studies beginning with higher intensity scores on the ECBI in the clinical range had larger long-term effects on behavior problems than for those beginning with average scores in the nonclinical range on the ECBI.

Additional studies examining possible moderators of parent training have added to the aforementioned findings from meta-analysis using other methods of multivariate analysis. Nowak & Heinrichs (2008) studied the Triple P Program by using Hierarchical Linear Modeling to identify variables moderating outcomes for all levels of the intervention. Their findings replicate the prior results of Thomas and Zimmer-Gembeck (2007) indicating larger effects are found on parent report measures as compared with

observational measures. They did not replicate the aforementioned finding of de Graff and colleagues (2008) indicating severity of child conduct problems predicted higher benefit from changes in parenting. Most compelling is the finding of no significant differences among individual, group, and self-administered formats of Triple P which appears to stand in contrast to the preliminary-meta-analysis by Sougstad et al. (2008) that suggested significantly greater benefit for group-administered IYPTP versus self-administered in primary research but not in replications. Prior findings (deGraff et al., 2008; Lundahl et al., 2006) of no difference may be due to inadequate heterogeneity analysis. Perhaps greater effects from primary, inventor-based research over that of self-administered forms of parent training noted by Sougstad and colleagues (2008) in IYPTP studies may be evident among other parent training data that have not considered distinguishing between inventor-based and independent replications (deGraff et al., 2008; Lundahl et al., 2006).

Hartman and colleagues (2003) note about one-third of young children remain at a clinically significant level at follow-up assessments after IYPTP. They used Hierarchical Linear Modeling to examine what factors affected treatment outcomes for primary studies of IYPTP. Differing levels of maternal depression, negative family stress and family socio-economic status did not predict outcomes from treatment at one month and one year follow-up assessments. Boys with inattention problems along with conduct problems benefitted at least as well if not better than those with conduct problems without inattention problems.

Beauchaine, Webster-Stratton & Reid (2005) utilized Growth Curve Analysis to identify mediators, moderators and predictors of treatment response from measures at

baseline, immediately following intervention and 1-year outcomes from the delivery of various combinations of Incredible Years Interventions including Parent, Teacher and Child Training components delivered to 514 children ages 3 to 8.5 studied across 6 randomized trials by the primary researcher and her associates. While Parent Training (IYPTP) was not a distinct and separate treatment, some of the growth curve analysis considered those treatments that included (in combination with teacher and/or child training) and did not include parent training.

Better 1-year outcomes for interventions including parent training (IYPTP) versus those that did not, were found for mother-reported youth externalizing problem behaviors. These were evident when mothers reported lower marital satisfaction and when mothers scored above a median score of 8 on the Beck Depression Inventory (more symptomatic). The same trend was noted for children of fathers with a history of substance abuse. Children scoring below the median T score of 56 on the Child Behavior Checklist Anxious/Depressed subscale (less of these problems) fared better when provided intervention that included IYPTP, whereas those children scoring above the median score responded equally regardless of treatment combinations. Children of mothers without a partner responded more favorably when IYPTP was included than when it was not. Children with a semiskilled or unskilled parent responded more favorably when IYPTP was included than when it was not. Across mother-report and observations for the seven significant moderating variables identified, interventions with parent training (IYPTP) were never less effective than interventions without IYPTP. Parent training using IYPTP produced the most consistent benefits across all moderating variables. These results led investigators to conclude IYPTP should be the "front-line"

intervention for young children with conduct problems with other components added on when warranted (e.g. Child training may be more beneficial for those children showing higher rates of depression/anxiety in addition to conduct problems; Teacher training provides added benefits for children with attention problems in addition to conduct problems).

Literature Review Summary

Cohen (1998) estimates each at-risk youth who is prevented from following an antisocial - criminal path into adulthood saves our society approximately 1.7 million dollars. Similarly, Muntz, Hutchings, and Edwards (2004) estimated the long-term economic costs associated with unresolved conduct problems is estimated to exceed $1 million per individual over a lifetime. Youth exhibiting early-onset antisocial conduct problems are most likely to develop a life-persistent pattern of dysfunction with grave consequences (Dodge, 2001; Keenan, & Shaw, 2003; Lahey & Waldman, 2003; Loeber et al., 2003; Patterson, 2002; Reid et al., 2002; Reid, 1993; Snyder & Stoolmiller, 2002; Tremblay et al., 2004).

One of the most well researched areas of youth psychological intervention to date is treating antisocial conduct behavior problems using parent training (Olendick, 2005; Weisz et al., 2004). The preponderance of reviewed intervention research addressing young children exhibiting these serious behavioral problems indicates parent training should be the first-line approach (Eyberg et al., 2008; Webster-Stratton & Taylor, 2001). However, roughly 90% of services delivered to address youth antisocial behavior within authentic community settings lack an evidence base to support them (Kumpfer &

Alvarado, 2003; Satcher, 2001). Science cannot yet tell us whether EBI's provide an advantage over treatment-as-usual in practice (Kazak et al., 2010).

Prevailing methods of collecting and reporting intervention research has not generally provided information leading to practitioner knowledge and implementation of EBI (Beutler, 2009; Hoagwood & Johnson, 2003; Kehle et al., 2007; Trachtman, 1981). Research findings must produce usable forms that generalize beyond research into practitioner decision-making and service delivery (Chafouleas & Riley-Tillman, 2005; Graczyk, et al., 2006; Kazdin, 2008; Kratochwill & Shernoff, 2003; Rosenfield, 2000; Southam-Gerow, et al., 2008). Meta-analysis has recently been found beneficial for evaluating cumulative evidence for parent training programs (deGraaf, et al., 2008; Thomas & Zimmer-Gembeck, 2007) but none of these studies have examined differences between primary inventor-based research and independent replications of the intervention. Those EBI's most worthy of being transported into practice settings should already show potent, robust and reproducible benefits of a similar magnitude in replications beyond primary inventor's research. Otherwise there may be little to no reason to believe an intervention holds promise when a practitioner would choose to invest limited resources to replicate that EBI within their own setting. Inventor-based primary studies provide reasonable and attainable benchmarks with which to compare replication studies (Southam-Gerow, et al., 2008; Weersing & Weisz, 2002).

This current research study updates and more fully develops a preliminary meta-analysis examining extant research using the Incredible Years Parent Training Program (IYPTP) when used to diminish early-onset antisocial conduct behavior problems (Sougstad et al., 2010). Over 30 years of IYPTP accumulated research studies support its

efficacy in two primary forms, clinic-based/group-administered and self-administered. IYPTP contains many properties suggesting it is an attractive, transportable intervention from ideally-controlled research into authentic community practice settings with desirable integrity (Kazdin, 2005; SAMHSA, n.d.; Sheridan, 2005; Webster-Stratton, 2007).

Preliminary meta-analysis (Sougstad et al., 2008) suggested that independent replications of the clinic-based group-administered IYPTP yielded small effects that were less than half of the large effects reported by the inventor-based research team. Effects from group-administered replications were also not significantly different from effects produced by the self-administered form (Sougstad et al., 2008). Since group-administered IYPTP requires substantially more time and resources than the self-administered form, practitioners considering adoption of IYPTP would benefit from knowing whether factors most associated with the large benefits reported by the inventor-based research studies can be replicated within authentic practice settings.

This dissertation study extends the preliminary meta-analysis by adding studies meeting inclusion criteria that have been published since the meta-analysis was done 2 and one-half years ago. The current research systematically examines variables within studies that may relate to the replicability of IYPTP beyond the primary inventor-based researcher and her team. Study variables accounting for benefits over time, comorbidity, characteristics intervention recipients, intervention providers and outcome measures are examined systematically. This study aspired to illustrate the beneficial role that meta-analysis may play toward understanding what variables matter most to reduce the gap

between research science and practitioner service using existing bodies of research about an EBI such and IYPTP.

Research Questions and Hypotheses

Research Question 1: Are significant differences between reductions in child conduct problems for primary and replication research studies using the Incredible Years Parent Training Program (IYPTP) previously found in the preliminary meta-analysis (Sougstad et al., 2008) retained with the addition of a broader range of studies produced over the last 2.5 years?

Hypothesis 1a: It was hypothesized for studies measuring effects from IYPTP that the grand mean effect size representative of a group of primary studies conducted by inventor Dr. Carolyn Webster-Stratton and her colleagues at the University of Washington would show significantly greater reductions in child conduct behavior problems than the grand mean effect size representative of replication studies, at a 0.05 level of statistical significance. Less favorable effects have been found for child and adolescent interventions separate from primary research demonstrating efficacy (Weiss, et al., 1999; Weisz, Donenberg, Han, & Weis, 1995; Weisz et al., 1998; Weisz & Jensen, 2001; and Weisz et al., 2006). Preliminary results of IYPTP meta-analysis (Sougstad et al., 2008) continued to support these findings. In accordance with the Task Force on Evidence Based Interventions in School Psychology (2003) and the review of intervention studies reducing youth disruptive behavior disorder symptoms provided by Brestan & Eyberg (1998), independent replications are defined by the same intervention, with the same target problem, independent of the program author group and institution. This

dichotomous distinction was made with caution given Chorpita's (2003) observation that studies exist along a continuum from those most closely tied to a primary researcher who developed the intervention (Efficacy Studies) to being completely independent (see Table 1). To make relevant comparisons addressing this issue, this dichotomy served a practical purpose of providing a starting point for this analysis.

Table 1: Chorpita (2003) research study continuum:

Research Type:	**Distinguishing Features:**	**Potential Conclusions:**
Type I: *Efficacy*	• highly controlled; subjects carefully selected yielding homogeneous samples • highly trained treatment agents with explicit allegiance to investigator/approach used • intensive supervision by national expert typical	• "A particular form of parent training is efficacious for oppositional youth"
Type II: *Transportability*	• Laboratory conditions without exclusionary selection criteria • Wider range of client conditions closer to practice settings • Maximum treatment agent and supervision under ideal conditions	• "A particular form of parent training is a promising approach for real world cases of oppositional youth."

Table 1 continued:

Type III: Dissemination	• Usage of system/agency employees • Supervision provided by research investigator under typical practice setting conditions	• "A particular form of parent training is a promising approach for school-based treatment agents serving oppositional youth."
Type IV: System Evaluation	• System of intervention evaluated by completely independent investigator having no influence on treatment delivered	• A particular parent training approach leads to positive outcomes for oppositional youth, entirely on it's own.

Hypothesis 1b: It was hypothesized that the grand mean effect size for reductions in child conduct problems would be significantly greater for parent rating scale outcome measures than for direct observation of parent-child behavior, at a 0.05 level of statistical significance. It has been generally found that parent rating scale results showing parental perceptions of child behavior show more benefit from intervention than independent observations of child behaviors. This was most recently noted within the meta-analysis of Parent-Child Interaction Therapy and Triple-P Parent Training studies (Thomas & Zimmer-Genbeck, 2007).

Hypothesis 1c: It was hypothesized that for studies measuring effects from IYPTP that the grand mean effect size representative of a group of primary studies conducted by inventor Dr. Carolyn Webster-Stratton and her colleagues at the University of

Washington would show significantly greater reductions in negative parenting than the grand mean effect size representative of replication studies, at a 0.05 level of statistical significance.

Hypothesis 1d: It was hypothesized that for studies measuring effects from IYPTP that the grand mean effect size representative of a group of primary studies conducted by inventor Dr. Carolyn Webster-Stratton and her colleagues at the University of Washington would show significantly greater improvements in positive parenting than the grand mean effect size representative of replication studies, at a 0.05 level of statistical significance.

Research Question 2: What variables moderate greater benefit from IYPTP for primary and replication research as separate groups? Are they similar or different? Moderators are variables that are present at baseline and differentiate under what conditions and for whom an intervention is effective. (Beauchaine et al., 2005; Shadish & Sweeney, 1991).

Hypothesis 2a: The grand mean effect size for a group of studies treating the most severe child conduct problems using IYPTP will be significantly different from the grand mean effect size for a group of studies treating the least severe child conduct problems using IYPTP, at a 0.05 level of statistical significance. Numerous studies have found that the greater the severity of a problem at the onset of treatment, the greater benefit (Beauchaine, et al., 2005; deGraff et al., 2008; Spirito et al., 2009), although there have been exceptions to this trend (Nowak & Heinrichs, 2008; Weisz et al., 2006).

Hypothesis 2b: The grand mean effect size for child conduct problems will be significantly better for each group of studies, according to the amount of training

therapists delivering IYPTP received, at a statistically significant level of 0.05. Lochman and colleagues (2009) looked at training in a unique and highly understudied manner by considering training on a relative scale of 3 conditions involved in transporting an EBI into a school setting. Benefits were significantly affected by whether therapists had a level of no training, basic training or training plus ongoing feedback. More benefits in terms of reducing behavioral problems in youth were also recently associated with adherence to treatment protocol as facilitated by ongoing supervision (Schoenwald, Sheidow & Chapman, 2009). For this meta-analysis it is anticipated studies would be grouped according to Lochman's distinctions of no training, basic training and training plus ongoing supervision.

Hypothesis 2c: The grand mean effect size for a group of studies reporting use of methods to ensure treatment fidelity will be significantly greater at the 0.05 level of significance, than the grand mean effect size for a group of studies not reporting use of methods to insure treatment fidelity.

CHAPTER 3

METHOD

Literature Search Procedure

In order to locate a representative set of research studies for potential inclusion in this meta-analysis a multiple-step strategy including computerized searches of several databases, manual inspections of reference lists from prior meta-analyses of parent training interventions, and reviews of websites listing published research by authors who have published in this area of study was used (Durlak, 2003: Eyberg et al., 2008; Lucas & Cutspec, 2005). A comprehensive electronic search of the professional literature used *PsycInfo, ERIC, MEDline, GoogleScholar, and Web of Science* using the terms "Incredible Years." Additionally abstracts for these citations were searched for only those studies specifically implementing the Incredible Years Parent Training Program.

Inclusion Criteria for this Study

To be included in this meta-analysis, studies met the following criteria:

1.) The Incredible Years Parent Training Program in its BASIC form (Webster-Stratton, 1997) is reported to be a distinct and primary behavior intervention within the study, whereby treatment effects are directly measured.

2.) The studied intervention must address child conduct behavior problems under the umbrella term of disruptive behavior disorders (McMahon, Wells & Kotler, 2006; Wilson, Lipsey & Derzon, 2003). This would include oppositional, rule-breaking, aggressive, antisocial behavior problems shown by both empirical factor analysis and clinical DSM taxonomic classification systems to form the category of Externalizing

Behavior Problems (Achenbach, Dumenci & Rescorla, 2003). Clinical diagnoses of Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) are common clinical manifestations of these behaviors when they are exhibited at severe levels (Quay, 1999). Attention Deficit Hyperactivity Disorder (ADHD) is also conceptualized as a third type of externalizing, disruptive behavioral disorder that frequently coexists with ODD/CD (Burns & Walsh, 2002) though has been regarded as separate and distinct disorder relative to ODD/CD (Hinshaw, 1987). Roughly half of youth with ADHD do not develop ODD or CD (Jensen, Martin & Cantwell, 1997). Studies solely focused on intervention for children with ADHD were not included in the preliminary meta-analysis (Sougstad et al., 2008) but these are included for this study. Recent research by Chronis and colleagues (2004) suggests it is often difficult to separate out ADHD from other disruptive behavior disorders, especially in younger children that the IYPTP focuses on.

3.) Outcomes are provided using valid and reliable standardized measures of child conduct problems.

4.) The preliminary meta-analysis (Sougstad et al., 2008) was limited to only studies using a *between-subjects group design*. This design was defined as an experimental design in which a no-treatment control group and one or more treatment groups were employed. This type of design is the most highly regarded in terms of definitive demonstrations of treatment effects. To be more inclusive than the aforementioned preliminary meta-analysis w*ithin-subjects* studies were included for this more current study, requiring the use of additional statistical formulas.

5.) Studies must provide mean scores, standard deviations and sample sizes for purposes of initially calculating *Cohen's d* effect size (1988).

Definition of Dependent Variables and Coding of Selected Studies

This meta-analysis focuses on child outcomes as the principal dependent variable and parent outcomes as secondary dependent variables. This study is primarily interested in treatment outcomes specific to a problem often warranting intervention in practice, that being child conduct behavior problems. These types of problems are the most frequent reason for referral to psychological treatment for children (Bloomquist & Schnell, 2002; Kamphaus & Frick, 2002). Treatments that reduce child conduct behavior problems must be implemented to meet these needs. A focus on child conduct behavior outcomes measured by adequate psychometric methods provides the most direct method of establishing an intervention is effective at helping children (Durlak, 2002). Changes in positive and negative parenting are also included within this study. Measured changes in parenting associated with parent training intervention is believed to be the mechanism of change responsible for reductions in child conduct problems (Kazdin, 2008).

Child conduct problems were measured by parent rating scale results, and third-party direct observations of child behavior reported within the studies sampled. Parent rating scales quantified parent perceptions about children's conduct problems and included measures like the Eyberg Child Behavior Inventory (ECBI) including the Intensity and Problems subscales (Robinson, Eyberg & Ross, 1980), the Strengths & Difficulties Questionnaire (SDQ) including the Hyperactivity, Conduct Problems, Antisocial subscales and Total Difficulties scale (Goodman, 1997), and the Achenbach

Child Behavior Checklist (CBCL) including measures of Externalizing behavior problems (Achenbach & Edelbrock, 1991). Independent observations of child conduct behavior problems were measured systematically by coding brief samples of parent-child interactions using the Interpersonal Behavior Construct Scale (Kogan & Gordan, 1975) or the Dyadic Parent-Child Interaction Coding System (DPICS; original and revised) that included subscales for Total Child Deviance and Noncompliance Ratio (Robinson & Eyberg, 1981; Webster-Stratton, 1989). Utilizing results from parent rating scales and direct observations of child behaviors avoids an undesirable reliance on a single category of informants (Hinshaw & Park, 1999). This method is also consistent with best practices for the assessment of child conduct problems, relying on more than one source and more than one method for assessments (Kamphaus & Frick, 2002). The most recent primary study included within this meta-analysis (Webster-Stratton, Reid, & Hammond, 2004) reported composite scores for each dependent variable (i.e., child conduct problems, negative and positive parenting) but did not report the individual measures comprising each of these composites. While they reported that composite scores were based on 49 separate summary scores comprised of approximately two-thirds independent observations and one-third reports there was no way from a review of the study to code single measured outcomes. Since this meta-analysis was interested in the same three composite measures this lack of specific scale scores was not detrimental.

Interview with parent, found in only a few studies, is the one type of outcome measure for child conduct problems that was excluded from this meta-analysis. The interview methods also differed across studies (e.g., daily telephone to pre- and post-interviews) making interpretations of these results across studies challenging.

The negative and positive parenting outcome variables were measured by studies included within this meta-analysis by parent self-report rating scales and independent observations of parenting behaviors. Parent self-report rating scales were all taken from the Parenting Practices Inventory (PPI) which included subscale scores for a variety of constructs including harsh discipline, praise and physical punishment (Conduct Problems Prevention Research Group, 1996). Observations of parenting behaviors in relation to the child were measured systematically using the Dyadic Parent-Child Interaction Coding System (DPICS; original and revised) that included subscales for Total Praise and Total Criticism (Robinson & Eyberg, 1981; Webster-Stratton, 1989).

Calculation of Effect Size Procedures

Many studies in this meta-analysis reported more than one measure of a dependent variable of interest (i.e., parent reports of child behavior and third-party direct observation of child behavior). Therefore, when there are scores reported for multiple subgroups (i.e. mothers and fathers separately), child behavior outcomes (e.g. Eyberg Total Problem and Intensity scores) and parenting changes (e.g. observations and self reports of parenting) separate effect sizes were calculated for each of these. This method allows for the fullest empirical examination of possible relationships between different methods of operationalizing multiple measures during later analysis (Lipsey & Wilson, 2001b). However, to avoid problems of dependence, only one effect size can represent each study in meta-analysis of study groups (Lipsey & Wilson, 2001a). Therefore, all effect sizes were initially combined, and averaged to yield one mean effect size for each dependent variable studied by this meta-analysis (child conduct problems; negative and positive parenting). Larger negative effect sizes are more desirable for the principal

dependent variable of reducing children's antisocial conduct problem behaviors and for the secondary dependent variable of reducing negative parenting. Positive effects are desirable for increasing the secondary dependent variable of positive parenting.

All studies included in this meta-analysis reported pre-and post- mean scores, standard deviations and sample sizes for at least an intervention group (*within subjects*) or more ideally for both intervention and control groups (*between-groups*). *Single-subject* case studies were not included within this meta-analysis given the ongoing controversy regarding appropriate model(s) for conducting meta-analysis with these studies. A recent prior meta-analysis for BPT yielded wildly different effect size results for single-case studies, depending on the method used (Maughan et al., 2005).

Between-group effect sizes calculated for this study using pre- and post- mean scores for both intervention and control groups is considered the standard of accuracy because these most closely approximate current conceptualizations of desired effect sizes (Carlson & Schmidt, 1999; Hedges & Olkin, 1985). Frequently meta-analyses compare only the differences between post-means for intervention and control groups but this method ignores possible difference between the two groups on pre-intervention measures, potentially inflating or obscuring differences. This would be most detrimental on those studies where intervention and control groups are not equivalent on the variable of interest before the intervention takes place, which is more likely to happen in research studies outside of highly controlled university settings. The approach used for this meta-analysis using differences between pre- and post- measures for both intervention and control groups (when these are available) allows for more control for pre-existing differences between these groups and allows for estimation of intervention effects even

when intervention and control groups may differ on the variable of interest prior to intervention (Morris, 2008; Morris & DeShon, 2002; Ortego & Botella, 2010).

Lipsey and Wilson (1993) found among 45 meta-analyses examining studies using one group pre/post designs without a comparison or control group yielded effect sizes 61% larger than those studies using a control or comparison group. They concluded as did Maughan and colleagues (2005) that one group pre/post designs potentially inflate mean effect sizes and should therefore be separated when reporting results within meta-analyses.

Several steps were followed to calculate effect sizes depending on the type of study (i.e. *between-groups* versus *within-groups*). For both types of studies raw mean differences were calculated by hand for all variables of interest within each study. For *within-group* studies unweighted effect sizes were calculated by subtracting the post-intervention mean by the pre-intervention mean, divided by the pre-intervention standard deviation which produces a standardized mean change (Becker, 1988). For *between-groups* studies each intervention group (I) and control group (C), post-intervention means were subtracted from the pre-intervention means for each group. For each dependent measure (parent rating scale, direct observation) this quantifies the average amount of change gained by each group (intervention and control) during the time from pre- to post-intervention of the study (Becker, 1988). These unweighted mean difference scores for each control group ($C_2 - C_1$) were then subtracted from the unweighted mean difference for each intervention group ($I_2 - I_1$) to arrive at an unweighted mean difference in change between the intervention and control group (as represented in the numerator of Formula 1).

$$d = \frac{[(I2 - I1) - (C2 - C1)]}{SIC} \quad (1)$$

The mean difference score between each intervention and control group was then divided by the pooled standard deviation from the intervention and control groups pre-intervention, as represented in the denominator of Formula 1 (Carlson & Schmidt, 1999; Morris & DeShon, 2002). This standardizes (normalizes) difference between gains made by control and intervention groups into one common metric called an effect size (*d*) often referred to in the literature as *Cohen's d* (1963). The result of these calculations for groups receiving intervention indicates by how many standard units those receiving IYPTP have improved over time relative to their own baseline score; and for those in *between-groups* studies, relative to the improvement found by the comparison control group.

For intervention and control groups Becker (1988) advises using the pre-intervention standard deviations, in order to most closely approximate the true variance within the population prior to any potential impact incurred by conducting the experiment. Post-intervention standard deviations are not used because these may have been affected by activities of the experiment, thus altering the variance away from the population value. Individual differences among participants in a study may interact with the intervention methods employed, resulting in different rates of change among participants (Glass, McGaw & Smith, 1981). For example, exposure to dependent variable measures may exert influence on some participants more than others. Another possibility is that some participants may receive more or less treatment because of their attendance, attention, or other personal factors affecting their participation (Carlson & Schmidt, 1999).

For *between-group* studies the aforementioned calculation of mean difference scores between intervention and control groups as well as the pooling of pre-intervention standard deviations accounting for sample sizes of each group was completed by use of an Excel-based program available for download from the Center for Evaluation and Monitoring at Durham University in the United Kingdom (www.cemcentre.org). The resulting effect sizes (*d*) and their confidence intervals were obtained using this program.

Effect sizes are known to be upwardly biased for small sample sizes, particularly under 20 (Hedges, 1981). The aforementioned pooled standard deviation for intervention and control groups (represented as S_{IC} in formula 1) was calculated to adjust for sample size using Hedges & Olkin's (1985) formula (2), where SD = the standard deviation for each of the intervention and control groups, n_i = sample size for intervention group, n_c = sample size for control group, and N = total sample.

$$SIC = \sqrt{\left((n_i - 1)SDI^2 + (n_c - 1)SDC^2\right)/(N - 2)} \qquad (2)$$

An additional correction for sample size is recommended and was applied to each effect size, as defined in formula 3 where N_i=sample size for intervention group and N_c=sample size for control group (Hedges & Olkin, 1985). The resulting effect size is commonly referred to as Hedges' *g* which produces the most conservative effect size value for studies with smaller sample sizes. Hedges & Olkin (1985) report *d* tends to *g* in probability as *N* becomes larger. Hence *d* and *g* become essentially the same estimator for larger samples. Because many studies in social and behavioral sciences including those reviewed in prior meta-analyses of parent training contain some studies with smaller sample sizes, these steps were regarded as crucial for obtaining the most unbiased

estimate of program effect size. All formulas hereafter refer to individual effect sizes as "*g*."

$$d \times 1 - \frac{3}{4(Ni + Nc) - 9} = g \tag{3}$$

Because each effect size is an estimate of a population value it is important to calculate a confidence interval for each effect size, as well as for the overall mean effect size for a group of studies (Hedges, 1994; Lipsey & Wilson, 2001a). Confidence intervals give the range within which the true mean effect size likely occurs for the group of studies according to a pre-defined level of probability acceptable to the researcher. For example, in this study the conventional 95% confidence level is utilized, whereby the resulting interval describes the range of scores within which the true effect size occurs every 95 times out of 100 that a population of studies is sampled theoretically. To calculate a confidence interval, the standard error of the effect size is required. The standard error for each effect size was calculated using formula 4.

$$SE_g = \sqrt{\frac{ni + nc}{ni + nc} + \frac{gic^2}{2(ni + nc)}} \tag{4}$$

To calculate confidence intervals for each effect size the standard error estimate (ES_{ic}) is multiplied by a critical *z*-value representing the desired confidence level for the study (e.g. Z = 1.96 for $\alpha = .05$); the product is then added to the mean effect size to establish the upper limit of the confidence interval; and the product is also subtracted from the mean effect size to establish the lower limit of the confidence interval (Formula 5).

$$\bar{g}_U = \bar{g} + z_{(1-\alpha)}(SE_{\bar{g}})$$
$$\bar{g}_L = \bar{g} - z_{(1-\alpha)}(SE_{\bar{g}}) \quad (5)$$

Obtained effect sizes (Hedges *g*) and their confidence intervals were utilized for meta-analysis using a central database file contained in a statistical software package called *Comprehensive Meta-Analysis, Version 2.0* (Borenstein, et al., n.d.). Meta-analysis was conducted using this software package. Cooper & Hedges (1994) and Lipsey & Wilson (2001a) recommend utilizing software packages specifically designed for performing the multiple calculations involved in meta-analysis.

Before effect sizes from individual studies (Hedges *g*) can be combined to reveal an overall grand mean effect size ($\bar{g}$) that is representative of a group of studies, each study's effect size must be weighted by their reciprocal of variance (Hedges & Olkin, 1985). The precision of each study's effect size estimate determines the degree to which it contributes to an overall grand mean effect size for a group of studies. The less error (variance) associated with each study's effect size, the closer it approximates the grand mean for a group of studies. Increases in sample size also improve the likelihood an effect size is coming closer to approximating the grand mean effect size for a group of studies. Effect sizes with smaller variances provide more reliable estimates of the true grand mean of a group of studies and thus provide more weight in the calculation of an overall grand mean for a group of studies (Borenstein et al., 2009; Hedges, 1994).

Meta-Analysis Assumptions and Procedures

The effect size for each study estimates the true value of a grand mean effect size for a group of studies. Meta-analysts must choose between fixed-effect or random-

effects models to conduct a meta-analysis amidst subtle differences between the two models, requiring considerable subjective judgment (Cooper & Hedges, 1994). The fixed-effect model assumes the only source of uncertainty is the within-study (sampling estimation) error. The random-effects model includes the aforementioned source of uncertainty in addition to variance between studies. The study-to-study variance (tau-squared) is not assumed to be the same for all subgroups within the random-effects model, and is therefore not pooled across subgroups but for each subgroup separately.

The variance, standard error and confidence intervals for a summary effect size will always be larger under a random-effects model than under the fixed-effect model (Borenstein, et al., 2009). The weights assigned to studies are more balanced within a random-effects model than those assigned under a fixed-effect model. Under the random-effects model, more balanced study weights lead to larger studies being assigned less relative weight and smaller studies more relative weight. Moving from fixed-effect to random-effects models leads to extreme studies having less influence when they are larger and gaining influence if they are smaller.

Because this meta-analysis is quite narrow in terms of focusing on one particular intervention program and includes a fairly small number of studies, it is tempting to utilize a fixed-effect model that would assume studies originate from the same population and that differences among the effect sizes for a group of studies are only due to sampling error (e.g. IYPTP studies all approximate one, unknown population parameter for IYPTP intervention). If this meta-analysis were about a medication and all studies were performed the same way, by the same researchers, and drew from the same population then a fixed-effect model would be supportable. Under these conditions this would be

most accurately termed a "common-effect meta-analysis" (Borenstein et al., 2009). Under these conditions meta-analysis seeks to identify a grand mean effect size for a group of studies that is an observed score to estimate an unknown true effect size within a population.

However, the studies included in this meta-analysis are not so homogeneous thus a fixed-effect model must be ruled out. The decision to employ either a random-effects model or a fixed-effect model should be based on an understanding about whether or not all studies share a common effect size, and should not be solely reliant on a statistical test for heterogeneity that often suffers from low power, which is definitely anticipated within this meta-analysis given the small number of studies focused on one particular intervention (Borenstein et al., 2009).

Therefore given the hypotheses for this study and initial findings of a preliminary meta-analysis of IYPTP (Sougstad et al., 2008) suggesting effects on child conduct problems from IYPTP may be diverse rather than homogeneous, the random-effects model will be employed. Borenstein and colleagues (2009) indicate when sampling from the professional literature, a random-effects model is typically most appropriate. The random-effects model estimates the mean of a distribution of effect sizes. This would seem especially appropriate for this study because we already know researchers operating independently are unlikely to produce functionally equivalent studies. In fact it is this variation we are interested in examining. Borenstein and colleagues (2009) further advocate use of a Mixed-Effects Model whereby random-effects are calculated within a subgroup of studies and then a fixed-effect model is used to calculate across these subgroups to determine their degree of similarity or difference.

The variability among study effect sizes within a group was calculated using a *Q*-statistic. (Formula 8) which is distributed as a chi-square, with *k*-1 degrees of freedom and *k* represents the total number of studies contributing to a grand mean effect size for the group and w_i is the individual weight for each effect size (g_i) (Borenstein, 2009; Hedges, 1994).

$$Q = \sum_{j-1}^{k} wi(gi - \bar{g})^2 \qquad (8)$$

A level of significance was pre-determined as equal to or below 0.05 for the test of variability (*Q*-statistic). A *Q*-statistic value having a probability level exceeding 0.05 was not regarded as significant, hence the variability of effect sizes within the group would be regarded as adequately homogeneous. When a group of effect sizes are relatively homogeneous an overall grand mean effect size for this group is more easily interpreted as representing this group of studies. If on the other hand a *Q*-statistic value has a probability level equal to or below 0.05, this indicates a significant amount of heterogeneity among effect sizes within the group suggesting interpretation of an overall grand mean effect size for this group would be too difficult to make. Too much heterogeneity suggests there are other variables impacting effect size results among some studies that are not adequately accounted for by this grouping. In this latter case, the meta-analysts job then becomes one of investigating possible sources of difference among effect sizes for this group of studies to uncover sources of too much heterogeneity (Borenstein et al., 2009; Lipsey & Wilson, 2001a; Cooper & Hedges, 1994). The hypotheses for this study served as a means of testing multiple variables identified *a priori* for purposes of uncovering unique relationships within the accumulated data.

When the meta-analyst finds comparable characteristics behave or operate similarly across various studies of the same intervention, this provides accumulated evidence about what really matters relative to producing the most desirable results from an intervention (Dunst et al., 2004).

CHAPTER 4

RESULTS

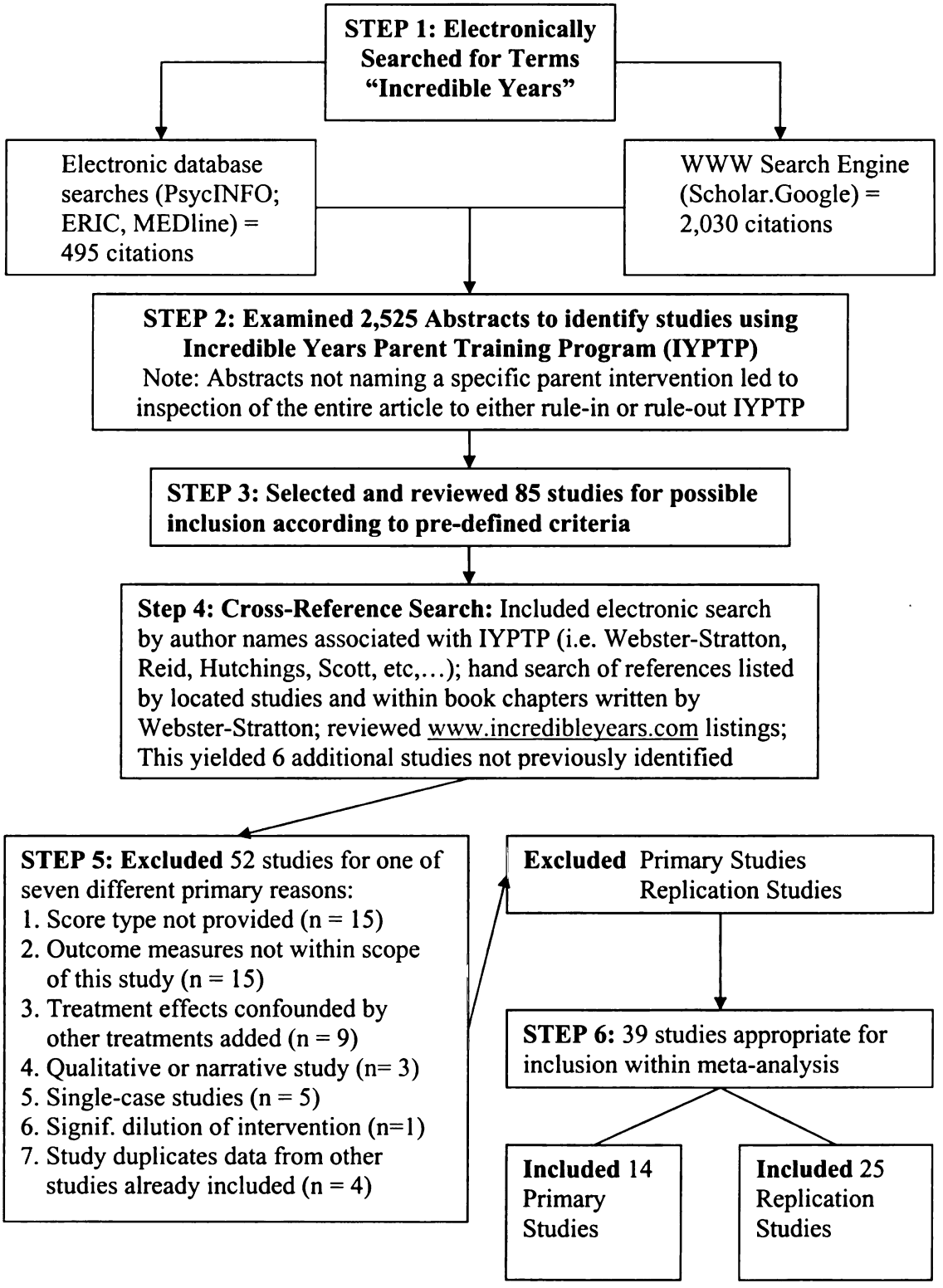
Findings from Systematic Selection of Research Studies

To locate a representative set of research studies for potential inclusion in this meta-analysis, I utilized multiple strategies including computerized searches of several databases, manual inspections of reference lists as well as reviewing websites listing research by authors who have published in this area of study (Durlak, 2003; Lucas & Cutspec, 2005). Figure 1 provides the steps followed for this search. A comprehensive electronic search of the professional literature used *PsycInfo, ERIC, MEDline* and *GoogleScholar* using the term, "Incredible Years". These combined methods yielded 2,525 citations (495 from *PsycInfo, ERIC, MEDline* and 2,030 from *GoogleScholar*). Abstracts were reviewed electronically to locate studies for possible inclusion in this meta-analysis. Where there was not a specific parent training program mentioned in the abstract, the article was reviewed to determine if IYPTP was the intervention studied. This procedure yielded 85 studies for further inspection. Cross-referencing was also employed by hand searching lists of references in articles and book chapters and also checking the Incredible Years website (www.incredibleyears.com) where there are published and unpublished studies available. The latter procedure resulted in 6 more studies. A total of 91 studies were reviewed to determine whether they met criteria for inclusion in this meta-analysis. Inspection of the 91 studies led to exclusion of 52 studies (57 %) and the inclusion of 39 studies (43%).

Studies were excluded from this meta-analysis for seven primary reasons. Fifteen studies (16%) did not provide the needed scores for this meta-analysis (e.g. means,

standard deviations and sample sizes). Another fifteen studies (16%) provided outcome measures not within the scope of this meta-analysis (e.g. parent self-efficacy and child autonomy). Nine of the studies (10%) were excluded because IYPTP effects were confounded by the inclusion of additional interventions (e.g. individual child intervention, summer programs, or school interventions). Five studies (5%) were excluded because they were single-case study designs. Four studies (4%) were excluded because they reported on data already included within a prior, included study. Three studies (3%) were excluded because they were qualitative studies. One study (1%) was excluded because the treatment was significantly diluted (only provided two sessions of IYPTP). This summary of reasons for exclusion provides some confidence that the sampling for this meta-analysis did not exclude any one type of study at a significant magnitude.

Figure 1: Literature search process for locating and selecting studies:



Coding of Study Variables

Each study accepted for this meta-analysis was reviewed using a coding form for purposes of gathering data relevant to *a priori* hypotheses. This coding form may be reviewed within Appendix A.

Examination of Possible Bias in Results for Between-Group Studies

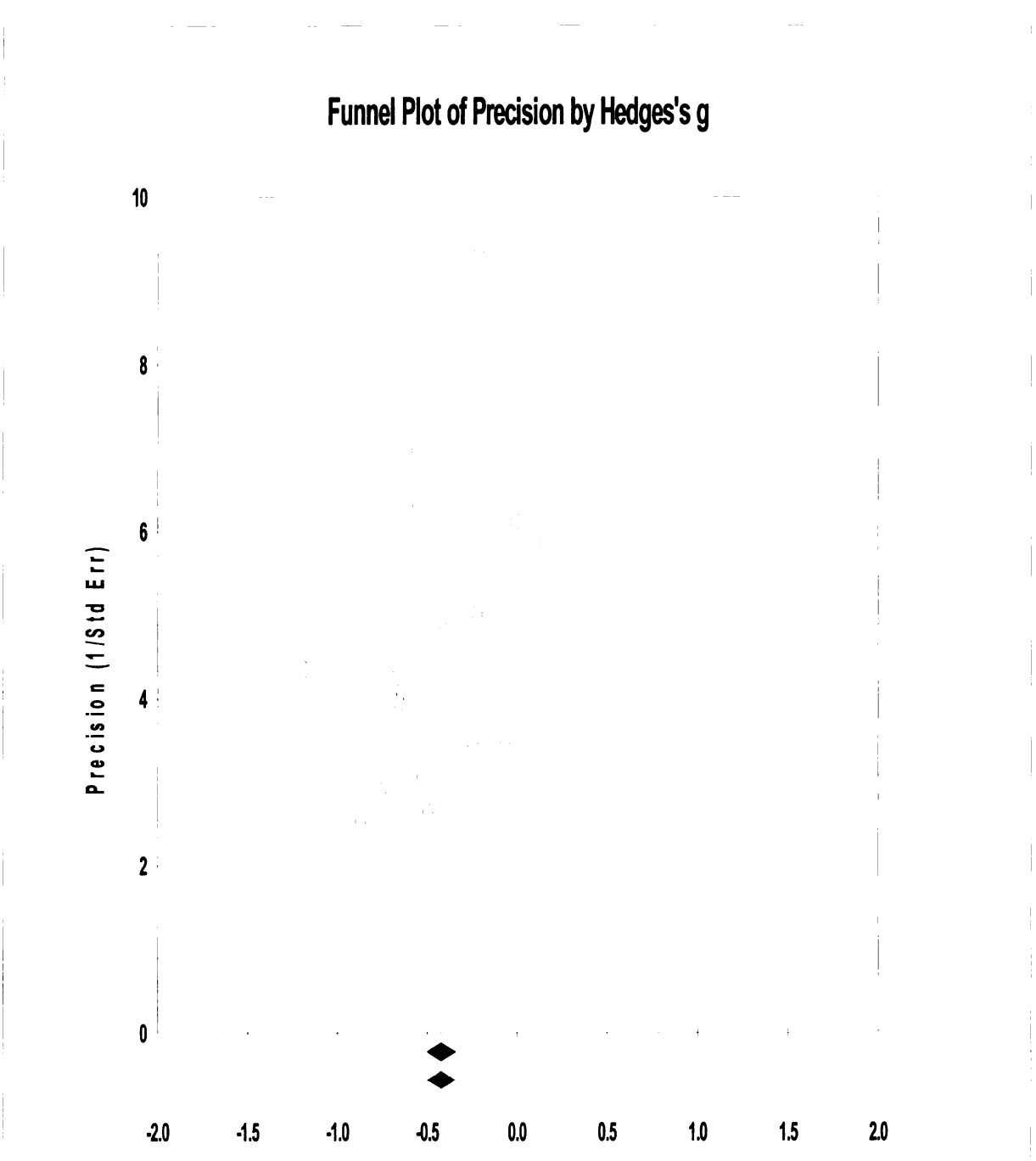
Studies that report larger effect sizes are more likely to be submitted and accepted for publication than studies producing small to moderate effects (Kratochwill & Shernoff, 2004). This trend results in a less than random selection process whereby studies included in a meta-analysis likely overestimate the true effect of the intervention studied. It is important to assess the extent to which results of a meta-analysis may have publication bias and subsequently consider the potential impact on conclusions drawn as a result of this consideration. Before conducting this analysis it is important to recognize that effect sizes generated from between-groups and within-group studies are generated quite differently, thus requiring not only separate meta-analyses but also separate tests of publication bias. Two methods were employed to evaluate possible publication bias for both types of studies.

First, a funnel plot was constructed as a graphic means of examining the distribution of effect sizes found in each study relative to their standard error. When publication bias is absent, the plot of studies should be distributed symmetrically about the combined effect size (depicted as a perpendicular line) such that the plot resembles a proportioned funnel shape. If there is a higher concentration of studies plotted on the side of the combined effect size representing favorable change, this would reflect the presence

of some publication bias. When bias is present smaller studies reporting statistically significant findings are published and included in the meta-analysis, while there is a presumption that other smaller studies that did not yield statistically significant results were not published and are therefore not included within the study. Sampling is supposed to be random and include the population of all studies, not just those reporting larger effect sizes. For this study desirable effects are negative, reflecting reductions in child conduct behavior problems.

Examination of the funnel plot for between-group studies (Figure 2) found most included studies fell within a moderate size range as depicted by the clustering in the middle of the graph. At the very bottom of the graph there is more representation of favorable effects among some smaller studies perhaps suggesting some undue bias of publication effect. However the impact of these studies appears negligible in relation to the weight of the symmetrical bulk of larger studies falling in the middle of the graph as well as a larger study falling toward a less desirable effect.

Figure 2: Funnel plot for between-groups studies:



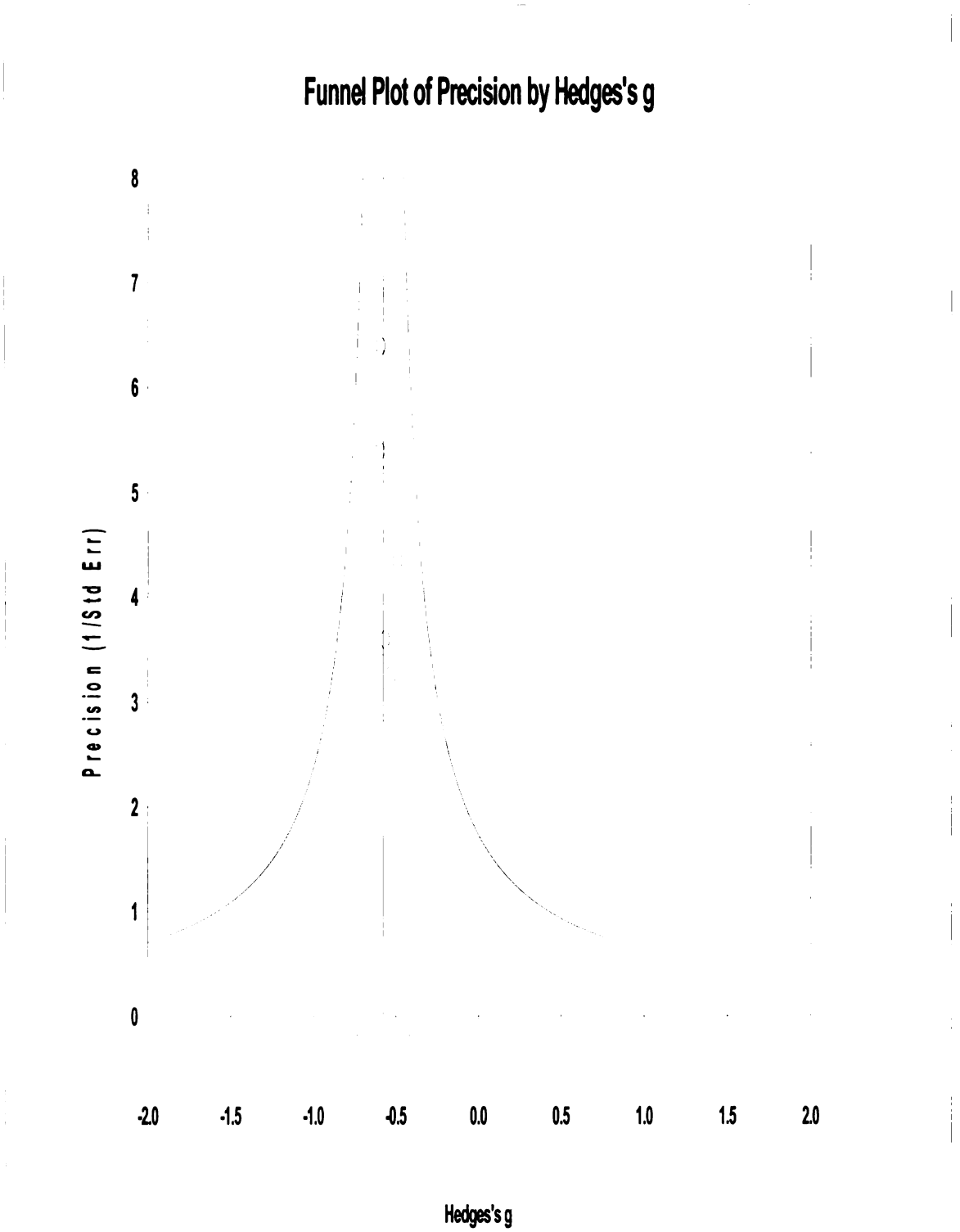
To further consider possible publication bias a second method of analysis was employed. Rather than speculating about the potential impact of missing, unpublished studies remaining in a "file drawer," Rosenthal (1979) recommends a method for calculating the number of missing studies that would lead to nullifying the found effect in the analysis. If this number is quite small there is more reason for concern given their potential to change the overall outcome of the study, making conclusions based on the current results more tenuous. If on the other hand a large number of studies would be required to nullify the effect, then conclusions using the studies already obtained may be made more confidently even though there may be some slight inflation incurred by using only published studies.

The between-group studies included within this meta-analysis yielded a Z-value of – 10.24 and corresponding p-value of 0.0000. The fail-safe N calculation resulted in a figure of 657. This means we would need to locate 657 'null' studies in order for the combined 2-tailed p-value to exceed 0.050 significance. It seems unlikely there are that many studies finding no effects. The fail-safe N supports the interpretation of the funnel plot, suggesting that publication bias is not a problem in this current study. It should also be noted that the traditional fail-safe N algorithm uses probability levels (p-values) for each study, and then combines those values. The more accepted method that was used for this research, calculates effect sizes for each study, combine these, and then compute the p-value for the combined effect.

Examination of the funnel plot for within-group studies (Figure 3) included in this meta-analysis shows six of the seven studies produced very similar benefit for reductions in conduct problems, at a moderate level. The outlier score showing a more significant,

desirable result is actually an unpublished study which would contraindicate the hypothesis of there being publication bias.

Figure 3: Funnel plot for within-groups studies



The within-group studies used in this meta-analysis included two unpublished studies, thus two of the four replication studies were unpublished (Clondalkin, n.d. & Rogers, 2007). The seven within-group studies in this meta-analysis yielded a Z-value of -6.81915 and corresponding p-value of 0.0000. The fail-safe N calculation resulted in a figure of 78. This means we would need to locate 78 'null' studies in order for the combined 2-tailed p-value to exceed 0.050 significance. It seems unlikely there are that many studies finding no effects. The fail-safe N supports the interpretation of the funnel plot, suggesting publication bias in this study is not a significant problem. It should also be noted here that the traditional fail-safe N algorithm uses probability levels (p-values) for each study, and then combines those values. Whereas the more accepted method that was used for this study is to calculate effect sizes for each study, combine these, and then compute the p-value for the combined effect.

Primary Hypotheses Test Findings

For this meta-analysis statistical tests were conducted separately for studies using *between-group* comparisons and studies using *within-group* comparisons, since the effect sizes were calculated differently for each type of study and would not be comparable. Lipsey and Wilson (1993) found that effect sizes for studies using within-group comparisons yielded significantly larger benefits than those using between-group comparisons where control groups were used. This led them to conclude that one group pre/post within-group studies may inflate true effects from an intervention compared with those studies using a control group. Maughan and colleagues (2005) meta-analysis of behavioral parent training intervention found similar results, further supporting the parceling of studies according to design.

Research Question 1: Are significant differences between reductions in child conduct problems for primary and replication research studies using the Incredible Years Parent Training Program (IYPTP) previously found in the preliminary meta-analysis (Sougstad et al., 2008) retained with the addition of a broader range of studies produced over the last 2.5 years?

Hypothesis 1a: It was hypothesized for studies measuring effects from IYPTP that the grand mean effect size representative of a group of primary studies conducted by inventor Dr. Carolyn Webster-Stratton and her colleagues at the University of Washington would show significantly greater reductions in child conduct behavior problems than the grand mean effect size representative of replication studies, at a 0.05 level of statistical significance.

Results for Hypothesis Test 1a: Given the principal hypothesis for this study that primary and replication studies would yield significantly different effect sizes, a test of homogeneity was calculated for each group (primary and replication studies) separately to ensure that the grand mean effect size depicted adequately represents each group of effect sizes. The Q-statistic tests whether the observed variance among effect sizes within a group are larger than expected from sampling error. The effect size for each study estimates the true value of a grand mean effect size for a group of studies. To report a grand mean effect size as representative of a group of studies, the effect sizes for studies within that group should display an acceptable level of homogeneity. Statistically-significant heterogeneity contraindicating the reporting of a grand mean effect size for a group of studies will be indicated by a Q-Statistic below the pre-defined cut-off of a p-value of 0.050.

Table 2 provides combined effect sizes (Hedges *g*) and associated statistics for each within-groups study of IYPTP included in this meta-analysis. Studies were grouped according to whether they were a primary or a replication study. A test of homogeneity was conducted for each group of primary and replication within-group studies using a *Q*-statistic. For the primary within-group studies of IYPTP the *Q*-value of 4.613 with 2 degrees of freedom yielded a non-significant p-value of 0.100. Similarly the *Q*-value of 2.172 with 5 degrees of freedom yielded a non-significant p-value of 0.825 for the replication studies of IYPTP. Since the grand means representing each group of studies appears to represent a fairly homogeneous set of findings, these may be compared statistically to determine their degree of similarity.

Borenstein and colleagues (2009) advocate use of a Mixed-Effects Model for comparing subgroup means, which uses the Random-Effects Model within subgroups and a Fixed-Effect Model across subgroups. Using Mixed-Effects Analysis to compare the grand mean effect size for the primary and replication studies yielded a *Q*-value of 0.261 with 1 degree of freedom and a non-significant p-value of 0.609.

Additionally, for illustration purposes the confidence interval for the Hedges's *g* effect size -0.727 (-1.115 to -0.338 with 95% confidence) for primary studies overlaps with the confidence interval for the replication studies, Hedges's *g* effect size -0.616 (-0.787 to -0.445 with 95% confidence). This significant amount of overlap of about one-half standard deviation further demonstrates the non-significant p-value showing no statistically-significant difference in reductions for child conduct problems between primary and replication within-groups studies of IYPTP. The grand mean effect size of -0.634 (-0.791 to -0.478 with 95% confidence) indicates a moderate benefit accrued from

IYPTP, with child conduct problems reduced by about two-thirds of a standard deviation for within-group studies.

Table 2: Within-groups IYPTP studies, immediate reductions in child conduct problems

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
W-S & Shoecraft 2009	-0.488	0.230	0.053	-0.940	-0.037	-2.118	0.034
W-S, 1982b	-0.526	0.307	0.094	-1.128	0.076	-1.713	0.087
W-S, 1994	-1.026	0.160	0.026	-1.340	-0.712	-6.401	0.000
Primary,Random Model:	***-0.727***	***0.198***	***0.039***	***-1.115***	***-0.338***	***-3.667***	***0.000***
Replication Studies:							
Axberg, 2007	-0.596	0.186	0.034	-0.960	-0.232	-3.210	0.001
Clondalkin, 2004	-0.984	0.359	0.129	-1.689	-0.280	-2.740	0.006
Fergusson et al., 2009	-0.677	0.140	0.020	-0.952	-0.402	-4.829	0.000
Manby, 2005	-0.564	0.279	0.078	-1.111	-0.017	-2.022	0.043
McIntyre, 2008	-0.328	0.382	0.146	-1.077	0.420	-0.860	0.390
Rogers, 2007	-0.502	0.204	0.041	-0.901	-0.103	-2.464	0.014
Replication Random Model:	***-0.616***	***0.087***	***0.008***	***-0.787***	***-0.445***	***-7.061***	***0.000***
Overall Primary & Replication. Random Model:	***-0.634***	***0.080***	***0.006***	***-0.791***	***-0.478***	***-7.940***	***0.000***

Table 3 provides a listing of effect sizes (Hedges's *g*) and associated statistics for each between-groups study included within this meta-analysis, along with the summary grand mean effect size for each group of studies (primary versus replication). Given the principal hypothesis for this study, that primary and replication studies would yield significantly different effect sizes a test of homogeneity was calculated for each group separately to ensure that the grand mean effect size depicted in Table 3 adequately

represents each group of effect sizes. The Q-value for primary studies of 42.940 with 8 degrees of freedom yielded a significant p-value of 0.000. This finding indicates this group of primary studies has too much variance to use a grand mean effect size to represent these studies. Similarly the Q-value for replication studies of 28.487 with 14 degrees of freedom yielded a significant p-value of 0.012. This finding also indicates too much variance to use a grand mean effect size to summarize these studies. Given these results for between-groups studies of IYPTP, further meta-analysis is required prior to concluding whether there is a significant difference between the benefits of primary and replication studies for IYPTP.

Table 3: Between-Groups Studies for IYPTP Immediate Reduction in Child Conduct Problems

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
Goss, Fogg, Web-St. 2003	0.104	0.171	0.029	- 0.231	0.439	0.611	0.541
Kim, Cain, Web-St. 2008	- 0.228	0.374	0.140	- 0.962	0.505	- 0.610	0.542
Webster-Stratton 1982a	- 0.528	0.327	0.107	- 1.169	0.114	- 1.611	0.107
Webster-Stratton 1984	- 0.871	0.396	0.157	- 1.647	- 0.094	- 2.198	0.028
Webster-Stratton 1998	- 0.210	0.107	0.012	- 0.420	0.001	- 1.950	0.051
Webster-Stratton 2001	- 0.160	0.132	0.017	- 0.418	0.099	- 1.212	0.226
Webster-Stratton et al 2004	- 1.009	0.193	0.037	- 1.388	- 0.631	- 5.227	0.000
Webster-Stratton et al 1988	- 0.827	0.206	0.042	- 1.230	- 0.423	- 4.018	0.000
Webster-Stratton et al 1997	- 1.189	0.230	0.053	- 1.640	- 0.738	- 5.168	0.000
Random Model:	***-0.524***	***0.155***	***0.024***	***-0.827***	***-0.221***	***-3.389***	***0.001***
Replication Studies:							
Brotman, Gouley et al 2005	0.357	0.207	0.043	- 0.049	0.763	1.723	0.085
Brotman, Klein et al 2003	-0.701	0.351	0.124	- 1.389	- 0.012	- 1.993	0.046
Bywater et al In Press	- 0.066	0.292	0.085	- 0.637	0.506	- 0.225	0.822
Gardner, Burton, Klimes 2006	-0.652	0.251	0.063	- 1.143	- 0.161	- 2.602	0.009
Hutchings, Bywater et al 2007	- 0.656	0.183	0.034	- 1.016	- 0.297	- 3.576	0.000
Jones, Daley et al 2007	- 0.713	0.234	0.055	- 1.171	- 0.255	- 3.050	0.002
Larsson, Fossum et al 2009	- 0.558	0.184	0.034	- 0.918	- 0.198	- 3.038	0.002
LeTarte et al 2010	- 0.459	0.367	0.135	- 1.179	0.261	- 1.250	0.211
McIntyre 2008	-0.246	0.290	0.084	- 0.813	0.322	- 0.848	0.396
Patterson, Barlow et al 2002	-0.222	0.200	0.040	- 0.613	0.170	- 1.109	0.268
Scott, O'Connor et al 2006	- 0.010	0.164	0.027	- 0.331	0.311	- 0.061	0.952
Scott, Spender et al 2010	- 0.238	0.232	0.054	- 0.692	0.216	- 1.027	0.305
Scott, Sylva, et al 2010	- 0.417	0.202	0.041	- 0.813	- 0.020	- 2.059	0.039
Taylor et al 1998	- 0.775	0.342	0.117	- 1.446	- 0.104	- 2.263	0.024
Random Model:	***- 0.362***	***0.089***	***0.008***	***- 0.537***	***- 0.188***	***- 4.075***	***0.000***

Primary and replication studies of IYPTP can only be compared fairly when studies using different types and usages for IYPTP are matched. IYPTP between-groups studies contain one of two IYPTP forms (group-administered and self-administered) whereas the within-groups studies only included group-administered IYPTP. The group-administered IYPTP in its Basic form includes roughly 10-12 sessions with a group of parents conducted by co-leaders (Webster-Stratton, 1998). The self-administered form of IYPTP has parents view the same video vignettes that are used in the group form, but to facilitate integration of content each parent completes self-administered workbooks without any group intervention or regular contact with a therapist or group leader (Webster-Stratton, 1988).

There are studies included within this meta-analysis that examined the effects of the self-administered form of IYPTP. Table 4 includes the 3 primary studies of the self-administered form of IYPTP. No replication studies met inclusion criteria for this study. Kratochwill, Elliot, Loitz, Sladeczek & Carlson (2003) reported on the use of the self-administered form of IYPTP but their child outcome variable was an omnibus measure that included both externalizing and internalizing problems. The current meta-analysis is restricted to only those child outcome measures having to do with externalizing conduct problems. The only other two studies located employing the self-administered form of IYPTP, were single-case studies (Ogg & Carlson, 2009, Walcott, Carlson & Beamon, 2009). Case studies were excluded from this meta-analysis. Examination of the references listed within each of these studies found no additional citations for studies using the individually-administered IYPTP, not already considered for this meta-analysis. For the primary studies using the self-administered form of IYPTP the Q-value of 0.568

with 2 degrees of freedom and a p-value of 0.753 showed good homogeneity among these studies results. The resulting effect size (Hedges' *g*) for the primary, individually-administered IYPTP is -0.531 (-0.737 to -0.324, with 95% confidence). This would indicate that the average child benefiting from self-administered parent training improved their behavior at a level superior to roughly 70 % of those in control groups who did not receive this intervention.

Table 4: Between-groups, Self-Administered IYPTP (primary studies only available)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
W-S, 1990	-0.531	0.201	0.040	-0.925	-0.137	-2.642	0.008
W-S, 1992	-0.604	0.157	0.025	-0.912	-0.296	-3.841	0.000
W-S et al, 1988	-0.412	0.201	0.040	-0.805	-0.018	-2.052	0.040
Primary, Random Model:	***-0.531***	***0.105***	***0.011***	***-0.737***	***-0.324***	***-5.037***	***0.000***

The absence of independent replication of between-groups studies of the self-administered form of IYPTP prevents any comparison with an effect size representing primary studies of self-administered IYPTP. Finding no published independent between-group comparisons of the self-administered form of IYPTP was surprising given the significant benefits reported by the primary studies, and the likelihood that the self-administered form may be more easily replicated without needing therapist intervention.

In addition to recognizing two forms of IYPTP (self-administered and group-administered) Webster-Stratton (1998b) also distinguishes between two primary uses of IYPTP. There is a *clinical treatment* of families who have a child exhibiting significant conduct problems and a second form used as *community prevention* to improve parenting and child social functioning (e.g. Head Start families). Both forms of IYPTP include four primary components:

1.) Interactive play and involvement

2.) Reinforcement techniques like praise and rewards

3.) Limit setting

4.) Discipline (e.g. nonviolent time-out, ignoring, logical and natural consequences)

A close inspection of all between-groups studies using the group-administered form of IYPTP included in this meta-analysis (Table 3) led to a determination that the group-administered form of IYPTP is actually used for not just two purposes as was suggested by Webster-Stratton in 1998, but has actually been studied for three purposes. Studies of IYPTP fit within a public-health model inclusive of three-tiers of intervention. The prior, preliminary meta-analysis of IYPTP (Sougstad et al., 2008) only distinguished between the 2 original uses but Figure 4 illustrates the three-tiered usage of IYPTP.

First, studies examine IYPTP as a form of Tier One, primary, universal-prevention (e.g. within Head Start and other preschool environments) to potentially "inoculate" youth and families against future conduct problem development. Second, studies examine IYPTP as a form of Tier Two selective-interventions provided to youth and families showing risk factors known to be associated with the development of conduct problems. Third, studies examine IYPTP as a Tier Three indicated-intervention to address cases where clinically-significant conduct symptoms are present requiring the most intensive intervention.

Figure 4: Three-Tiered Usage of Incredible Years Parent Training Program:

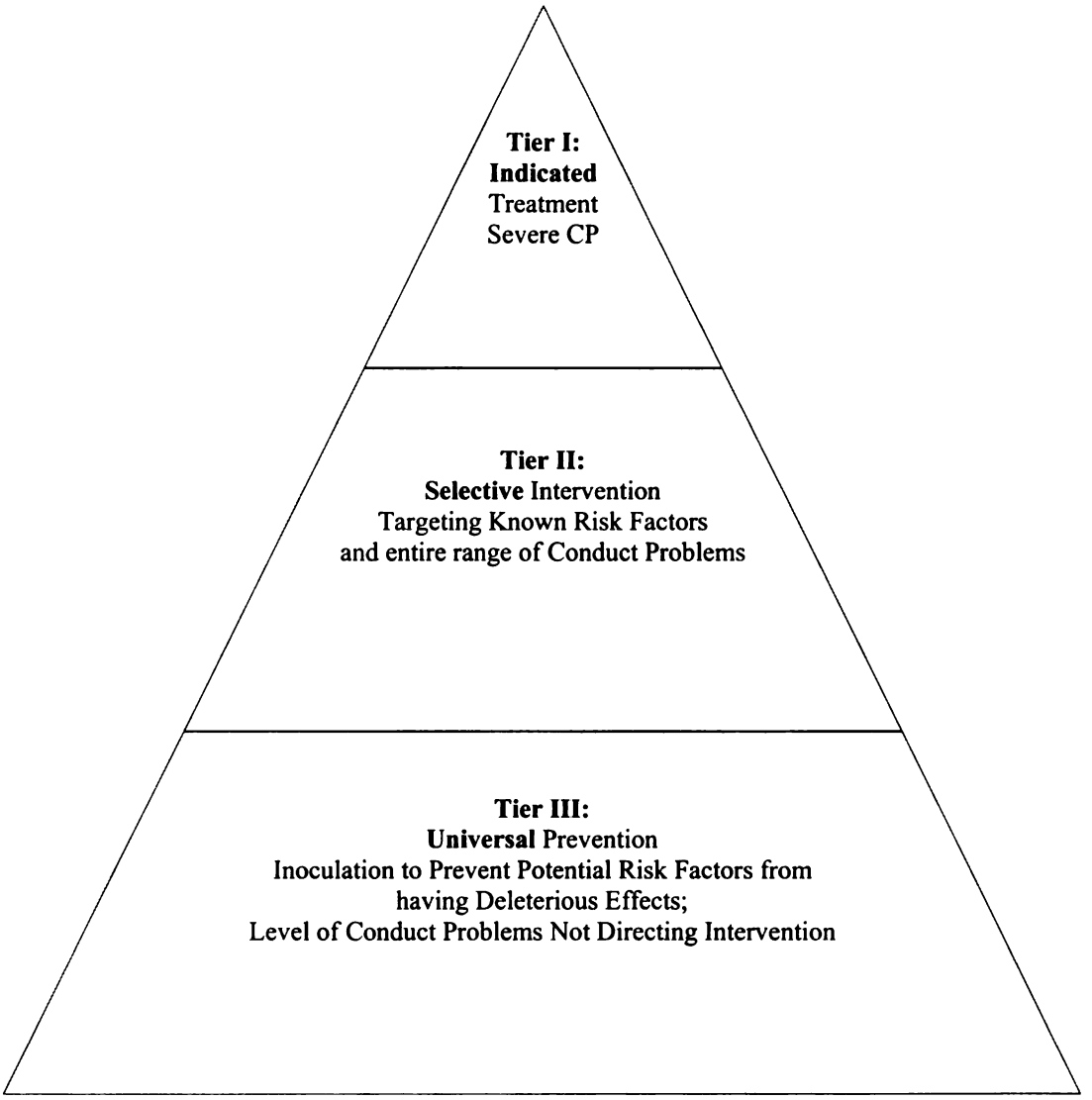


Table 5 summarizes the primary and replication studies where IYPTP was used as a Tier One intervention. These studies were grouped together because the selection of intervention recipients for these studies was not determined by a risk factor uniquely associated with the development of conduct problems, or according to a score on a

clinical measure of conduct problems. For example, Webster-Stratton, Reid & Hammond (2001) studied delivery of IYPTP with 272 mothers having a child attending Head Start. Yet Webster-Stratton & Hammond (1998b) found a sample of 394 Head Start families in the Northwest region of the United States where their studies were conducted, consisted of only 35% possessing at least three or more risk factors (e.g., single parenthood, poverty, depression, life stress, psychiatric illness, parent history of drug abuse, child abuse and spouse abuse), and between 40-45% having mothers who display high rates of harsh or physically negative parenting. Since it is estimated that the majority of the subjects in these two Head Start studies did not display risk factors at the aforementioned magnitudes, and IYPTP was provided at the Head Start Center level (not selected toward any group of at-risk students or families identified by a number of risk factors or conduct symptoms within the Head Start population) it was concluded this type of study is actually a universal prevention study. Similarly, Goss and colleagues (2003), including Dr. Webster-Stratton provided IYPTP to parents of children at day care centers where there was again poverty and other risk factors evident, but many of these are risk factors for not only conduct problems but also many other forms of child and family dysfunction.

Scott, O'Connor and Futh (2006) provided IYPTP across multiple settings identified for their impoverished circumstances in the United Kingdom (U.K.), similar to the Head Start studies in the U.S.A. The U.K. study deployed a gating procedure for making sure that the most high-risk students/families received intervention before those in less need. However, once these higher risk subjects were ensured of immediate intervention they were mixed into groups with lower risk subjects. While this step ensured an ethically supportable expediency to treatment delivery for those most in need,

the intervention effects were measured by overall changes by groups that were not defined by risk factors unique to conduct problems in children. It is reasonable to assume primary prevention studies included some higher risk students/families but in all of these studies the unit of analysis was change over time from treatment (versus control groups) at a level not unique to any particular characteristic other than circumstances of poverty entitling them to a preschool education program. Included along with these primary prevention studies was research by Kim, Cain and Webster-Stratton (2008) and McIntyre (2008b), who delivered IYPTP as demonstration projects with populations not known to be at high risk for conduct problems, but nevertheless may benefit from parent training (i.e. Korean mothers; Parents of children with a developmental disability such as Autism or Mental Retardation).

The primary studies' Q-Statistic of 2.541 with 3 degrees of freedom yielded a non-significant p-value of 0.468 indicating a roughly homogeneous set of effect sizes for this group. Similarly the Q-Statistic of 0.502 with 1 degree freedom yielded a non-significant p-value of 0.479 indicative of an absence of excessive heterogeneity. These results suggest it is reasonable to statistically compare the grand mean effect size representing each group, according to the predicted hypothesis of there being a significant benefit from primary over replication studies.

Comparison of the grand mean effect sizes (Hedges's g) for primary and replication studies using IYPTP as a tier one prevention program using Mixed-Effects Analysis yielded a Q-statistic of 0.191 with 1 degree of freedom and a non-significant p-value of 0.662. Therefore there is no statistically-significant difference between the benefits reported by these two groups of studies.

The obtained Hedge's *g* effect size of -0.122 suggests a very small benefit in the reduction of conduct problems for IYPTP used for primary prevention. For this group of studies, the unknown, true effect size lies between -0.250 and 0.006, ninety-five times out of one-hundred (95% confidence interval). Considering the Tier One, universal-primary prevention group of studies includes subjects where conduct problems were not identified as being highly problematic, this small effect size is not at all surprising.

Table 5: IYPTP Tier One, Primary-Universal Prevention (between-groups studies)

	Hedges's *g*	**Stnd. Error**	**Variance**	**Lower limit**	**Upper limit**	**z-value**	**p-value**
Primary Studies:							
Goss, Fogg, W-S, 2003	0.104	0.171	0.029	-0.231	0.439	0.611	0.541
Kim, Cain, W-S, 2008	-0.228	0.374	0.140	-0.962	0.505	-0.610	0.542
W-S, 1998	-0.210	0.107	0.012	-0.420	0.001	-1.950	0.051
W-S, 2001	-0.160	0.132	0.017	-0.418	0.099	-1.212	0.226
Primary, Random Model:	***-0.137***	***0.073***	***0.005***	***-0.281***	***0.007***	***-1.885***	***0.062***
Replication Studies:							
McIntyre, 2008	-0.246	0.290	0.084	-0.813	0.322	-0.848	0.396
Scott et al., 2006	-0.010	0.164	0.027	-0.331	0.311	-0.061	0.952
Replications, Random Model:	***-0.067***	***0.142***	***0.020***	***-0.346***	***0.212***	***-0.470***	***0.638***
Combined Primary & Replications, Random Model	***-0.122***	***0.065***	***0.004***	***-0.250***	***0.006***	***-1.873***	***0.061***

Table 6 contains those studies examining the use IYPTP as Tier Two selective intervention targeting groups specific to a particular risk factor(s). Inclusion in the Tier Three indicated intervention group required either a minimum cut-off score for clinically-significant conduct problem severity or clinic referral to treat high levels of conduct problems which the Tier Two group of studies does not contain. Tier Two studies are also differentiated from the aforementioned Tier One prevention studies because IYPTP is being used to specifically target groups where parenting and/or child functioning are

known to be at least somewhat problematic, suggesting an at-risk for conduct problems status.

The studies by Brotman and colleagues (2003 and 2005) studied IYPTP effects on children having a formal record of antisocial behavior in their immediate family history (e.g. adjudicated sibling, etc,...). Bywater and colleagues (In press) studied IYPTP effects on children served by foster care parents, where they cite about four times the rate of conduct disorder is found among this population than in the general population of the United Kingdom. Nilsen (2007) also studied the effects of IYPTP on foster parents. LeTarte and colleagues (2010) studied the effects of IYPTP in parents known to be neglectful and were already being served within the child welfare system in Canada. Patterson and colleagues (2002) studied IYPTP delivered to parents of children known to be above the 50^{th} percentile on a behavior inventory measuring conduct problems. Because this cut-off score is well below the clinically-significant range of conduct problems (falls within normal limits) this study was regarded as a form of selective intervention where children/families were selected because of an at-risk status rather than a clinical level of severity separate from most in the general population. There were no primary studies regarded as falling within a Tier Two level of selective intervention.

Table 6: IYPTP Tier Two, Selective Intervention Targeting At-Risk Youth/Families (between-groups studies)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies: NONE							
Replication Studies:							
Brotman et al., 2005	0.357	0.207	0.043	-0.049	0.763	1.723	0.085
Brotman et al., 2003	-0.701	0.351	0.124	-1.389	-0.012	-1.993	0.046
Bywater et al., In Press	-0.066	0.292	0.085	-0.637	0.506	-0.225	0.822
Letarte et al., 2010	-0.459	0.367	0.135	-1.179	0.261	-1.250	0.211
Nilsen, 2007	-0.495	0.380	0.145	-1.240	0.250	-1.302	0.193
Patterson et al., 2002	-0.222	0.200	0.040	-0.613	0.170	-1.109	0.268
Replications, Random Model:	***-0.195***	***0.165***	***0.027***	***-0.519***	***0.129***	***-1.179***	***0.239***

Examination of Table 6 found the Brotman et al. (2005) study yielded a Hedges's g effect size that is a substantial outlier to all other studies in this group (and that of other groups as well). The only measure comprising this effect size was based on a tool developed by the researchers called Observed Peer Play in Unfamiliar Settings (OPPUS) which is quite different from the other parent rating scale and independent observation measures included within other studies examined for this analysis. Because this measure accounted for 22% of the group mean effect size this study was removed.

Table 7: IYPTP Tier Two, Selective Intervention Targeting At-Risk Youth/Families with Outlier Removed (between-groups studies)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies: NONE							
Replication Studies:							
Brotman et al., 2003	-0.701	0.351	0.124	-1.389	-0.012	-1.993	0.046
Bywater et al., In Press	-0.066	0.292	0.085	-0.637	0.506	-0.225	0.822
Letarte et al., 2010	-0.459	0.367	0.135	-1.179	0.261	-1.250	0.211
Nilsen, 2007	-0.495	0.380	0.145	-1.240	0.250	-1.302	0.193
Patterson et al., 2002	-0.222	0.200	0.040	-0.613	0.170	-1.109	0.268
Replications, Random Model:	***-0.318***	***0.130***	***0.017***	***-0.573***	***-0.063***	***-2.445***	***0.014***

Table 7 more accurately represents the effects for this group of Tier Two selective intervention studies with the previously identified outlier removed. The Q-value of 2.531 with 4 degrees of freedom yielded a non-significant p-value of 0.639. This non-significant p-value at the 0.05 level indicates an acceptable level of homogeneity to use a grand mean effect size to represent this group of studies. The Tier Two selective intervention studies using IYPTP yielded a Hedges's g effect size of -0.318 with a 95% confidence interval between -0.573 to -0.063. On average this represents about a one-third standard deviation benefit from IYPTP Tier Two. The average child benefiting from IYPTP as a Tier Two intervention improved their behavior at a level superior to roughly 66 % of those in control groups who did not receive this intervention.

Table 8 shows Tier Three indicated intervention between-group studies in which IYPTP was used to treat clinically-significant child conduct problem symptoms. These studies involved intervening with parents of children demonstrated to show a clinically-significant magnitude of conduct problem symptoms based on standardized measures (e.g. roughly above the 90th percentile on the Intensity scale of the Eyberg Child

Behavior Inventory and/or a minimum number of problems such as 10 reported on the Problems scale of this same instrument).

Table 8: IYPTP Tier Three, Indicated Intervention Treating Clinically-Significant Child Conduct Problems (between-groups studies)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
Webster-Stratton 1982a	- 0.528	0.327	0.107	- 1.169	0.114	- 1.611	0.107
Webster-Stratton 1984	- 0.871	0.396	0.157	- 1.647	- 0.094	- 2.198	0.028
Webster-Stratton et al 2004	- 1.009	0.193	0.037	- 1.388	- 0.631	- 5.227	0.000
Webster-Stratton et al 1988	- 0.827	0.206	0.042	- 1.230	- 0.423	- 4.018	0.000
Webster-Stratton et al 1997	- 1.189	0.230	0.053	- 1.640	- 0.738	- 5.168	0.000
Primary, Random Model:	***-0.937***	***0.109***	***0.012***	***-1.149***	***-0.724***	***-8.634***	***0.000***
Replication Studies:							
Gardner, Burton, Klimes 2006	-0.652	0.251	0.063	- 1.143	- 0.161	- 2.602	0.009
Hutchings, Bywater et al 2007	- 0.656	0.183	0.034	- 1.016	- 0.297	- 3.576	0.000
Larsson, Fossum et al 2009	- 0.558	0.184	0.034	- 0.918	- 0.198	- 3.038	0.002
Scott, Spender et al 2010	- 0.238	0.232	0.054	- 0.692	0.216	- 1.027	0.305
Scott, Sylva, et al 2010	- 0.417	0.202	0.041	- 0.813	- 0.020	- 2.059	0.039
Taylor et al 1998	- 0.775	0.342	0.117	- 1.446	- 0.104	- 2.263	0.024
Replications Random Model:	***-0.533***	***0.089***	***0.008***	***-0.707***	***-0.359***	***-6.005***	***0.000***
Primary&Replications Combined, Random Model:	***-0.695***	***0.069***	***0.005***	***-0.830***	***-0.560***	***-10.115***	***0.000***

Tests of homogeneity were performed on the Tier Three studies primary and replication groups separately. The primary studies Q-value of 3.134 with 4 degrees of freedom yielded a non-significant p-value of 0.536. The replication studies Q-value of 3.149 with 5 degrees of freedom yielded a non-significant p-value of 0.677. These results suggest an acceptable level of homogeneity to report one grand mean effect size representing each group of studies.

However, visual inspection of the primary studies found a wide variation for obtained effect sizes spanning over one standard deviation (-1.189) to about one-half of a standard deviation (-0.537) thus making an interpretation of one overall effect size representative of this group quite difficult given this wide level of dispersion. Borenstein and colleagues (2009) note that a non-significant p-value may be indicative of low power (p. 113). Examination of statistics for this group indicates within-study variance was non-significant with a p-value of 0.711. The low number of studies combined with a low number of subjects in these studies seems to be the most likely explanation for the non-significant finding for heterogeneity. These observations warrant further examination of the primary Tier Three studies, since it is likely these do not comprise one group of homogeneous studies that would be adequately represented by one grand mean effect size.

Close inspection of the primary Tier Three studies found dosage (number of sessions) of IYPTP varied considerably. Differences in the dosage of IYPTP treatment was not hypothesized as a tested variable for this meta-analysis *a priori*. However given the wide variation in dosages across primary studies, this was tested first prior to a statistical comparison between primary and replication Tier Three studies, and before any *a priori* hypotheses are tested. Variations in IYPTP dosage clearly need to be accounted for prior to considering if other hypothesized variables differentially impact on IYPTP study effect sizes.

Test for Dosage Effect for IYPTP

Webster-Stratton (1982a) studied an early form of IYPTP employing only 4 weekly sessions of two-hours, and this produced the lowest Hedges *g* effect size for

between-groups primary studies. Since this dosage level is less than half the established dose for IYPTP this study was removed from further analysis. It was further noted that two of the other between-groups primary studies (Webster-Stratton & Hammond, 1997; Webster-Stratton, et al., 2004) reported using 22-24 two-hour, weekly parent group sessions. These two studies were found to have the largest Hedges *g* effect sizes and their relative weights account for 60% of the mean effect size for the primary between-group studies after the 1982a study was removed.

Within the "Handbook of Parent Training" (Schaefer & Briesmeister, 1998) Webster-Stratton and Hancock (1998b) reported IYPTP in its original, BASIC form consisted of 12 weeks using 10 videotapes with more than 250 vignettes. By the third edition of this handbook (Briesmeister & Schaefer, 2007) Webster-Stratton's chapter (2007) again reported the original BASIC program of 12 sessions was developed and found effective across several studies for young children diagnosed with Oppositional Defiant Disorder. However, at this point ten years after the first book chapter summarizing IYPTP, Dr. Webster-Stratton was recommending a combination of IYPTP BASIC and ADVANCED programs for this same population of children that takes between 20 to 24 weeks to complete.

Neither the Webster-Stratton and Hammond (1997), or the Webster-Stratton, Reid and Hammond (2004) studies included in this meta-analysis described the addition of the ADVANCED content to comprise their 22-24 weeks of intervention. Data from the Webster-Stratton (1994) that tested the addition of the ADVANCED program was not included in this meta-analysis because it was regarded as an addition to the BASIC program. The BASIC program was employed within all studies contained in this meta-

analysis. Because the two studies using 22-24 sessions of IYPTP indicate usage of only the BASIC program and not the ADVANCED program, they were retained.

To test whether dosage accounts for a statistically-significant greater benefit for IYPTP Tier Three studies, a comparison was made between the effects from the two studies using 22 - 24 two-hour sessions, with that of other studies (primary and replication) which all reported using 9 - 16 two-hour sessions. The range of sessions among studies within this meta-analysis suggests (with the exception of the two studies using larger dosages) studied group-administered IYPTP generally includes a dosage range of 12 sessions either minus 3 or plus 4. Table 9 shows study dosage comparison.

Table 9: Contrast IYPTP Tier Three Dosages (22-24 versus 9-16 two-hour sessions)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Studies using 22-24 Two-Hour Sessions:							
W-S et al 2004	- 1.009	0.193	0.037	- 1.388	- 0.631	- 5.227	0.000
W-S et al 1997	- 1.189	0.230	0.053	- 1.640	- 0.738	- 5.168	0.000
22-24Session Random Model:	***-1.084***	***0.148***	***0.022***	***-1.374***	***-0.794***	***-7.326***	***0.000***
Studies using 9-16 Two-Hour Sessions:							
Gardner, Burton, Klimes 2006	-0.652	0.251	0.063	- 1.143	- 0.161	- 2.602	0.009
Hutchings, Bywater et al 2007	- 0.656	0.183	0.034	- 1.016	- 0.297	- 3.576	0.000
Larsson, Fossum et al 2009	- 0.558	0.184	0.034	- 0.918	- 0.198	- 3.038	0.002
Scott, Spender et al 2010	- 0.238	0.232	0.054	- 0.692	0.216	- 1.027	0.305
Scott, Sylva, et al 2010	- 0.417	0.202	0.041	- 0.813	- 0.020	- 2.059	0.039
Taylor et al 1998	- 0.775	0.342	0.117	- 1.446	- 0.104	- 2.263	0.024
Webster-Stratton 1984	- 0.871	0.396	0.157	- 1.647	- 0.094	- 2.198	0.028
Webster-Stratton et al 1988	- 0.827	0.206	0.042	- 1.230	- 0.423	- 4.018	0.000
9-16 Session Random Model:	***-0.591***	***0.080***	***0.006***	***-0.747***	***-0.435***	***-7.403***	***0.000***

For the 22-24 sessions group a test of homogeneity yielded a *Q*-Statistic of 0.359 with 1 degree of freedom and a non-significant o-value of 0.549. For the 9-16 sessions group the test of homogeneity yielded a *Q*-statistic of 5.384 with 7 degrees of freedom and a non-significant p-value of 0.613. The relative homogeneity of each group warranted a comparison of the grand mean effect size representing each group of studies. Mixed-effects analysis yielded a *Q*-Statistic of 8.593 with 1 degree of freedom and a significant p-value of 0.003. This finding indicates a statistically-significant effect for a

higher dosage of 22-24 two-hour sessions of IYPTP over that of more commonly encountered dosages of 9-16 two-hour sessions within a Tier Three usage.

The dosage of 22-24 sessions of IYPTP Tier Three yielded a Hedges's *g* effect size of -1.084 (-1.374 to -0.794 with 95% confidence). This result suggests about a one-standard deviation reduction in child conduct problems. The average child benefiting from 22-24 sessions of IYPTP as a Tier Three intervention improved their behavior at a level superior to roughly 84 % of those in control groups who did not receive this intervention.

The dosage of 9-16 sessions of IYPTP Tier Three yielded a Hedges's *g* effect size of -0.591 (-0.747 to -0.435 with 95% confidence). This result suggests about two-thirds a standard deviation of benefit. The average child benefitting from 9-16 sessions of IYPTP as Tier Three intervention improved their behavior at a level superior to roughly 73% of those in the control groups who did not receive this intervention. This latter finding is a result of combining both primary and replication studies that all used dosages between 9-16 sessions of IYPTP as a Tier Three intervention. In order to equitably compare primary to independent replications of the Tier Three use of IYPTP the two primary studies that used a significantly greater dosage (22-24 sessions) were removed from further comparison yielding Table 10.

Table 10: IYPTP between-groups studies using 9-16 two-hour sessions as a Tier Three Intervention to Treat Clinically-Significant Child Conduct Problems

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
Webster-Stratton 1984	- 0.871	0.396	0.157	- 1.647	- 0.094	- 2.198	0.028
Webster-Stratton et al 1988	- 0.827	0.206	0.042	- 1.230	- 0.423	- 4.018	0.000
Primary, Random Model:	***-0.836***	***0.183***	***0.033***	***-1.194***	***-0.478***	***-4.579***	***0.000***
Replication Studies:							
Gardner, Burton, Klimes 2006	-0.652	0.251	0.063	- 1.143	- 0.161	- 2.602	0.009
Hutchings, Bywater et al 2007	- 0.656	0.183	0.034	- 1.016	- 0.297	- 3.576	0.000
Larsson, Fossum et al 2009	- 0.558	0.184	0.034	- 0.918	- 0.198	- 3.038	0.002
Scott, Spender et al 2010	- 0.238	0.232	0.054	- 0.692	0.216	- 1.027	0.305
Scott, Sylva, et al 2010	- 0.417	0.202	0.041	- 0.813	- 0.020	- 2.059	0.039
Taylor et al 1998	- 0.775	0.342	0.117	- 1.446	- 0.104	- 2.263	0.024
Replications, Random Model:	***-0.533***	***0.089***	***0.008***	***-0.707***	***-0.359***	***-6.005***	***0.000***
Combined Primary & Replications Random	***-0.591***	***0.080***	***0.006***	***-0.747***	***-0.435***	***-7.403***	***0.000***

Table 10 illustrates the most equitable comparison of primary versus replication studies of Tier Three usage of IYPTP to treat significant child conduct symptoms, where dosages are roughly similar. The test of homogeneity for primary studies yielded a *Q*-Statistic of 0.010 with 1 degree of freedom and a non-significant p-value of 0.921. The test of homogeneity for replication studies yielded a *Q*-statistic of 3.149 with 5 degrees of freedom and a non-significant p-value of 0.677. The within-group variance *Q*-Statistic of 3.159 with 6 degrees of freedom yielded a non-significant p-value of 0.789. These

findings suggest an absence of significant heterogeneity within each group making it appropriate to compare the grand mean effect sizes (Hedges's *g*) representing each group. The Mixed-Effects Analysis yielded a *Q*-Statistic of 2.225 with 1 degree of freedom and a resulting non-significant p-value of 0.136. Hence there is no statistically significant difference between the benefits of primary and replication studies for the Tier Three usage of IYPTP to treat significant symptoms of child conduct problems using between 9 and 16 two-hour sessions.

The Grand Mean Effect Size (Hedges's *g*) for Tier Three usage of 9 to 16 two-hour sessions of IYPTP to treat significant child conduct problems across both primary and replication studies is -0.591 (-0.747 to -0.435 with 95% confidence). This suggests about two-thirds of a standard deviation benefit from Tier Three IYPTP (9-16 sessions). The average child benefiting from 9-16 sessions of IYPTP as a Tier Three intervention improved their behavior at a level superior to roughly 73 % of those in control groups who did not receive this intervention.

Hypothesis 1b: It was hypothesized that the grand mean effect size for reductions in child conduct problems would be significantly greater for parent rating scale outcome measures than for direct observation of parent-child behavior, at a 0.05 level of statistical significance. It has been generally found that parent rating scale results showing parental perceptions of child behavior show more benefit from intervention than independent observations of child behaviors. This was most recently noted within the meta-analysis of Parent-Child Interaction Therapy and Triple-P Parent Training studies (Thomas & Zimmer-Genbeck, 2007).

Results for Hypothesis Test 1b: Table 11 provides a summary of study effect sizes (Hedges's *g*) according to only parent-rating scale results, contrasted for primary and replication studies. The test of homogeneity for primary studies yielded a *Q*-Statistic of 0.567 with 1 degree of freedom and a non-significant p-value of 0.451. The test of homogeneity for replication studies yielded a *Q*-statistic of 2.791 with 4 degrees of freedom and a non-significant p-value of 0.593. The comparison of these two adequately homogeneous groups using the Mixed Effects Model yielded a *Q*-Statistic of 3.570 with 1 degree of freedom and a non-significant p-value of 0.059. This non-significant p-value and the considerable overlap between the confidence intervals for primary and replication grand mean effect sizes for each group suggests adequate homogeneity to interpret one overall, grand mean effect size as representative of this entire group of studies. The Hedges's *g* grand mean effect size of -0.726 (-0.894 to -0.558 with 95% confidence) is an appropriate measure to compare with the results listed in Table 11 showing effect size outcomes for independent observations of child conduct problems.

Table 11: Parent rating scale results (no independent observations) for IYPTP between-groups studies using 9-16 sessions (contrasting primary vs. replications)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
Webster-Stratton 1984	- 1.316	0.413	0.171	-2.126	-0.507	-3.186	0.001
Webster-Stratton et al 1988	- 0.968	0.209	0.044	-1.377	-0.559	-4.653	0.000
Primary Random Effects Model:	***-1.039***	***0.186***	***0.035***	***-1.404***	***-0.673***	***-5.574***	***0.000***
Replication Studies:							
Gardner, Burton, Klimes 2006	-0.721	0.253	0.064	-1.217	-0.225	-2.846	0.004
Hutchings, Bywater et al 2007	-0.835	0.185	0.034	-1.197	-0.473	-4.521	0.000
Larsson, Fossum et al 2009	-0.558	0.184	0.034	-0.918	-0.198	-3.038	0.002
Scott, Sylva, et al 2010	-0.417	0.202	0.041	-0.813	-0.20	-2.059	0.039
Taylor et al 1998	-0.775	0.342	0.117	-1.446	-0.104	-2.263	0.024
Replications Random Effects Model:	***-0.642***	***0.096***	***0.009***	***-0.831***	***-0.453***	***-6.659***	***0.000***
Total Primary & Replications Random Effects:	***-0.726***	***0.086***	***0.007***	***-0.894***	***-0.558***	***-8.476***	***0.000***

Table 12 displays effect sizes (Hedges's *g*) and associated statistics for only those studies reporting independent observations of child conduct behavior in relation to parents. The test of homogeneity for primary studies yielded a *Q*-Statistic of 0.067 with 1 degree of freedom and a non-significant p-value of 0.796. The test of homogeneity for replication studies yielded a *Q*-statistic of 0.501 with 1 degree of freedom and a non-significant p-value of 0.479. The comparison of these two adequately homogeneous groups using the Mixed Effects Model yielded a *Q*-Statistic of 0.358 with 1 degree of freedom and a non-significant p-value of 0.549. The latter result supports using the grand mean effect size for this entire group of studies, to compare with the grand mean effect size for the studies reporting only parent rating scale results listed in Table 11.

Table 12: Independent observations results (no Parent rating scale) for IYPTP between-groups studies using 9-16 sessions (contrasting primary vs. replications)

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
Webster-Stratton 1984	-0.425	0.378	0.143	-1.166	0.317	-1.123	0.261
Webster-Stratton et al 1988	-0.535	0.199	0.040	-0.926	-0.145	-2.687	0.007
Primary Random Effects Model:	***-0.511***	***0.176***	***0.031***	***-0.857***	***-0.166***	***-2.901***	***0.004***
Replication Studies:							
Gardner, Burton, Klimes 2006	-0.514	0.245	0.060	-0.994	-0.034	-2.099	0.036
Hutchings, Bywater et al 2007	-0.298	0.181	0.033	-0.653	0.056	-1.649	0.099
Replications Random Effects Model:	***-0.374***	***0.146***	***0.021***	***-0.660***	***-0.089***	***-2.573***	***0.010***
Total Primary & Replications Random Effects:	***-0.430***	***0.112***	***0.013***	***-0.650***	***-0.210***	***-3.831***	***0.000***

The grand mean effect size of -0.430 (-0.650 to -0.210 with 95% confidence) for independent observations outcomes is not significantly different at the pre-set level of 0.05 than the grand mean effect size of -0.726 (-0.894 to -0.558 with 95% confidence) representing parent rating scale outcomes. These confidence intervals overlap slightly though this non-significant outcome may also be attributable to the small sample of studies representing these outcomes.

Hypothesis 1c: It was hypothesized that for studies measuring effects from IYPTP that the grand mean effect size representative of a group of primary studies conducted by inventor Dr. Carolyn Webster-Stratton and her colleagues at the University of Washington would show significantly greater reductions in negative parenting than the grand mean effect size representative of replication studies, at a 0.05 level of statistical significance.

Results for Hypothesis Test 1c: Table 13 illustrates measured changes in negative parenting across primary and replication studies. The primary studies' Q-Statistic of 0.155 with 1 degree of freedom yielded a non-significant p-value of 0.694. Similarly the replication studies Q-Statistic of 4.453 with 2 degrees of freedom yielded a non-significant p-value of 0.108. These results suggest acceptable levels of homogeneity within each group, where the Q-statistic of 4.607 with 3 degrees of freedom yielded a non-significant p-value of 0.203. A Mixed Effects Analysis yielded a Q-Statistic of 1.516 with 1 degree of freedom and a non-significant p-value of 0.218. The latter finding supports interpretation of a grand mean effect size for both primary and replication studies of -0.491 (-0.714 to -0.269 with 95 % confidence) for reductions in negative parenting. This represents about one-half a standard deviation reduction in negative parenting. The average parent benefiting from IYPTP as a Tier Three intervention reduced their negative parenting behaviors at a level superior to roughly 70 % of those in control groups who did not receive this intervention.

Table 13: Reductions in Negative Parenting for IYPTP between-groups studies

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
W-S et al., 1988	-0.660	0.202	0.041	-1.066	-0.265	-3.270	0.001
W-S et al., 1997	-0.545	0.214	0.046	-0.964	-0.126	-2.549	0.011
Primary RandomEffect Model:	***-0.606***	***0.147***	***0.022***	***-0.893***	***-0.318***	***-4.128***	***0.000***
Replication Studies:							
Gardner et al., 2006	-0.692	0.247	0.061	-1.177	-0.207	-2.797	0.005
Hutchings et al., 2007	-0.060	0.171	0.029	-0.395	0.276	-0.349	0.727
Scott et al., 2010	-0.314	0.210	0.044	-0.726	0.098	-1.492	0.136
Replications RandomEffect Model:	***-0.321***	***0.179***	***0.032***	***-0.671***	***0.030***	***-1.794***	***0.073***
Combined Primary & Replications RandomEffect:	***-0.491***	***0.113***	***0.013***	***-0.714***	***-0.269***	***-4.329***	***0.000***

Hypothesis 1d: It was hypothesized that for studies measuring effects from IYPTP that the grand mean effect size representative of a group of primary studies conducted by inventor Dr. Carolyn Webster-Stratton and her colleagues at the University of Washington would show significantly greater improvements in positive parenting than the grand mean effect size representative of replication studies, at a 0.05 level of statistical significance.

Results for Hypothesis Test 1d:

Table 14 summarizes the Tier Three primary and replication studies reporting data amenable to this meta-analysis, about changes in positive parenting via self-report of parents and/or independent observations. It was noticed that the reporting of data on changes in parenting behaviors appeared less consistent even for the same researcher (e.g. Webster-Stratton) where variables were reported separately in one study and aggregated

into larger categories in others as well as being reported using different types of scores. This likely makes meta-analysis of these data a bit more challenging to ensure equitable comparisons across studies.

Table 14: Increases in Positive Parenting for IYPTP between-groups studies

	Hedges's *g*	Stnd. Error	Variance	Lower limit	Upper limit	z-value	p-value
Primary Studies:							
W-S et al., 1984	2.375	0.496	0.246	1.403	3.347	4.790	0.000
W-S et al., 1988	1.194	0.213	0.046	0.776	1.613	5.594	0.000
W-S et al., 1997	0.874	0.220	0.048	0.443	1.306	3.971	0.000
Primary RandomEffect Model:	***1.331***	***0.318***	***0.101***	***0.707***	***1.955***	***4.182***	***0.000***
Replication Studies:							
Gardner et al., 2006	0.464	0.242	0.059	-0.011	0.940	1.916	0.055
Hutchings et al., 2007	0.448	0.171	0.029	0.112	0.783	2.615	0.009
Scott et al., 2010	0.285	0.211	0.045	-0.129	0.699	1.349	0.177
Replications RandomEffect Model:	***0.402***	***0.117***	***0.014***	***0.174***	***0.631***	***3.448***	***0.001***
Combined Primary & Replications RandomEffect:	***0.512***	***0.110***	***0.012***	***0.297***	***0.727***	***4.676***	***0.000***

The replication studies reporting changes in positive parenting yielded a *Q*-statistic of 0.444 with 2 degrees of freedom and a non-significant p-value of 0.801. This suggests a fairly homogeneous set of data for this group. The primary studies reporting changes in positive parenting yielded a *Q*-statistic of 7.717 with 2 degrees of freedom and a significant p-value of 0.021, suggesting an unusual amount of heterogeneity. Visual inspection of the data found one study (Webster-Stratton, 1984) produced an overall effect size (Hedges's *g*) of 2.375 which is over double that of the other two primary studies and 4 to 5 times greater than the replication study effect sizes. Table 15 shows

changes in positive parenting effects minus the outlier study. With this study removed the *Q*-statistic of 1.089 with 1 degree of freedom yields a non-significant p-value of 0.297. With both primary and replication groups showing adequate homogeneity, a Mixed Effects Analysis was performed comparing the mean effect sizes from each group. This yielded a *Q*-statistic of 10.343 with 1 degree of freedom and a significant p-value of 0.001. The obtained primary studies Hedges's *g* effect size of 1.039 (0.725 to 1.352 with 95% confidence) is significantly greater than replication studies the Hedges's *g* effect size of 0.402 (0.174 to 0.631). This greater degree of improvement in positive parenting for both primary and replication studies was primarily based on independent observations of parenting behaviors.

Table 15: Increases in Positive Parenting for IYPTP between-groups studies (Outlier removed)

	Hedges's *g*	**Stnd. Error**	**Variance**	**Lower limit**	**Upper limit**	**z-value**	**p-value**
Primary Studies:							
W-S et al., 1988	1.194	0.213	0.046	0.776	1.613	5.594	0.000
W-S et al., 1997	0.874	0.220	0.048	0.443	1.306	3.971	0.000
Primary RandomEffect Model:	***1.039***	***0.160***	***0.026***	***0.725***	***1.352***	***6.494***	***0.000***
Replication Studies:							
Gardner et al., 2006	0.464	0.242	0.059	-0.011	0.940	1.916	0.055
Hutchings et al., 2007	0.448	0.171	0.029	0.112	0.783	2.615	0.009
Scott et al., 2010	0.285	0.211	0.045	-0.129	0.699	1.349	0.177
Replications RandomEffect Model:	***0.402***	***0.117***	***0.014***	***0.174***	***0.631***	***3.448***	***0.001***
Combined Primary & Replications RandomEffect:	***0.623***	***0.094***	***0.009***	***0.438***	***0.808***	***6.612***	***0.000***

Secondary Hypotheses Testing Findings

The second stage of meta-analysis was intended to further examine what factors moderate an anticipated greater benefit from primary studies over that of independent replication studies for IYPTP. However, this principal hypothesis was not supported. While there were no significant differences found for child conduct problem outcomes between primary and replication studies, several planned hypotheses tests offer additional information regarding moderators of IYPTP benefits across both study types, and are discussed below. Moderators are variables that are present at baseline and differentiate under what conditions and for whom an intervention is effective. (Beauchaine et al., 2005; Shadish & Sweeney, 1991).

Research Question 2: What variables moderate greater benefit from IYPTP for primary and replication research as separate groups? Are they similar or different? Moderators are variables that are present at baseline and differentiate under what conditions and for whom an intervention is effective. (Beauchaine et al., 2005; Shadish & Sweeney, 1991).

Hypothesis 2a: The grand mean effect size for a group of studies treating the most severe child conduct problems using IYPTP will be significantly different from the grand mean effect size for a group of studies treating the least severe child conduct problems using IYPTP, at a 0.05 level of statistical significance. Numerous studies have found that the greater the severity of a problem at the onset of treatment, the greater benefit (Beauchaine, et al., 2005; deGraff et al., 2008; Spirito et al., 2009), although there have been exceptions to this trend (Nowak & Heinrichs, 2008; Weisz et al., 2006).

Results for Hypothesis Test 2a: Fortunately the vast majority of the Tier Three intervention studies used the Eyberg Child Behavior Inventory (ECBI) as a measure of

child conduct problems, allowing for comparisons of child conduct severity between studies. The ECBI Intensity Scale measures the frequency of child disruptive behavior as rated on a scale from 1 (never) to 7 (always). The ECBI Problem Scale measures how problematic a child's behavior is perceived based on yes/no answers to statements about whether a behavior is present (Butler, Brestan & Eyberg, 2008). The norms for the ECBI indicate Intensity scores at or above 127 have been regarded as clinically-significant while Problem scores at or above 11 have been regarded as clinically-significant (Eyberg & Ross, 1978). These cut-off scores have been regarded as being at the 93rd percentile within the norm sample. However a more recent re-standardization of ECBI norms has suggested these cut-off scores should now be 132 for the Intensity Scale and 15 for the Problems Scale (Colvin, Eyberg & Adams, 1999). These have not been published, however, in a peer-reviewed journal or validated independently as was the earlier norms.

Five out of the six replication Tier Three studies reported ECBI scores. All five cited ECBI Intensity scores, and 4 out of 5 reported ECBI Problem scores. Of the Tier Three primary studies three out of four studies also reported ECBI scores. All three cited ECBI Intensity scores and two out of three reported ECBI Problem scores. These rates of score reports suggest reasonably good representative data across the primary and replication study samples for comparisons to be made.

Mean scores were reported separately for experimental and control samples as well as frequently reported separately for mothers and fathers. The mean ECBI Intensity, Problem Scale scores and their standard deviations for each study, and then for groups of studies were calculated by hand. Each reported mean and standard deviation within a study was multiplied by the size of the sample on which this was based upon

(experimental or control). These products were then added together for all scores reported for a scale (Intensity or Problem) in each study. The total of all these products were then divided by the total number of respondents who provided these scores (i.e. all sample group members). This method accounted for different sample sizes from which means and standard deviations were reported. In this way one mean score from a smaller sample would not have equal influence as a larger sample on the overall mean for a scale. Results from these calculations are listed in Table 16.

For the Tier Three independent replication studies the ECBI Intensity mean score was 141.4 and the ECBI Problem mean score was 18.5. The one study (Scott et al., 2010) not reporting a Problem score was also noted to report a Intensity mean score of 117.6 that was significantly below the mean for the other three studies (149), hence the latter likely best represents the mean for Tier Three studies. For Tier Three primary studies the ECBI mean Intensity score was 153.85 and the Problem mean score was 19.94. Combining these results yielded a mean ECBI Intensity Scale score of 151.4 and an ECBI Problem mean score of 19.22 for primary and replication Tier Three studies. For the Tier Two independent replication studies (there were no primary studies at this tier) the mean ECBI Intensity Scale score was 119.97 and the ECBI Problem mean score was 12.39 based on three studies. For the Tier One primary studies (no ECBI scores listed for replications) the mean ECBI Intensity Scale score was 90.83 and the ECBI Problem Scale score was 9.9.

Table 16: Severity of Initial Conduct Problems by Intervention Tier

	ECBI Intensity Score	ECBI Problem Score
TIER III (22-24 sess) ES*g*= -1.084	Primary Studies = 159.15 (*n* of individuals = 305) Replication Studies = None	Primary Studies = not reported (*n* of individuals = 305) Replication Studies = None
TIER III (9-16 sess) ES*g* = -0.591 Total *n* of individuals = 587	Primary Studies = 151.2 (26.4) (*n* of individuals = 202) Replication Studies = 149 (29.06) (*n* of individuals = 385)	Primary Studies = 19.4 (6.7) (*n* of individuals = 202) Replication Studies = 18.5 (7.03) (*n* of individuals = 385)
TIER II ES*g* = -0.318 Total *n* of individuals = 197	Primary Studies = None Replication Studies= 119.7 (31.79) (*n* of individuals = 197)	Primary Studies = None Replication Studies = 12.39 (8.25) (*n* of individuals = 197)
TIER I ES*g*=-0.122 Total *n* of individuals = 435	Primary Studies = 90.83 (26.35) (*n* of individuals = 435) Replication Studies=none reported	Primary Studies = 9.9 (8.0) (*n* of individuals = 435) Replication Studies=none reported

Note: Mean Scores reported (standard deviations in parentheses)

Several findings are notable relative to the data presented within Table 15. First, for ECBI Intensity and Problem scores are each significantly different in magnitude between the three tiered levels according to One-Way ANOVA (Fisher F statistic of 564.544 with 2 degrees of freedom and 191.614 with 2 degrees of freedom respectively) that yielded statistically significance (p = 0.000). This data supports the initial hypothesis that more severe conduct problems initially, would be associated with larger effect sizes

(listed in first column to the left in table 16). However, the severity of Intensity and Problem scores are essentially the same for Tier Three studies using either 9-16 sessions or 22-24 sessions, where the latter is shown to have a much larger effect size. This suggests that beyond initial severity of conduct problems other variables also have influence upon overall effects from the intervention. It is likely that the increased dosage has an additional, favorable impact on the outcome's effect size, although there may be additional variables not yet identified that affect this larger outcome. It is important to note that the separation of the IYPTP studies according to usage (three tiers) is validated by the data contained within Table 16. Finally, the data in Table 16 lends further support to the more recent re-standardized norms for the ECBI that suggest using a cut-off for the Problems score of 15 (rather than the former 11) to distinguish clinical significance (Colvin, Eyberg & Adams, 1999).

One type of study not included within Table 16 is the primary studies using the self-administered form of IYPTP. All three of these studies reported ECBI scores for the Intensity scale and two of the three reported scores on the Problems Scale. The overall mean ECBI Intensity score was 152.74 which is essentially the same as Tier Three studies. Similarly, the mean ECBI Problems score was 19.70 which are also comparable to the Tier Three studies. When comparing the obtained effect size (Hedges's *g*) of -0.531 (-0.737 to -0.324 with 95% confidence) for the Self-Administered IYPTP with that of the Tier Three (9-16 sessions) obtained effect size (Hedges's *g*) of -0.591 (-0.747 to -0.435 with 95% confidence) there is no statistical evidence of an improved effect from the more labor intensive and time-consuming group-administered form of IYPTP over that of the self-administered form. This lack of difference is not due to a differing

level of severity treated by each type of IYPTP. Additionally, comparison of the group-administered IYPTP at an increased dosage of 22-24 sessions with a Hedges's *g* effect size of -1.084 (-1.374 to -0.794 with 95% confidence) with that of the Self-Administered outcome (-0.531; -0.737 to -0.324 with 95% confidence) results in an unlikely, statistically-significant difference from the 22-24 session group-administered form over that of the self-administered form. The lack of overlap between these two different sets of confidence intervals suggests there is a significantly greater benefit from intervening with the most severe forms of child conduct problems when the group-administered form is at twice the typical dosage for the BASC IYPTP (22 to 24 sessions). Certainly the low number of primary studies and no replication studies for the 22-24 sessions suggests these findings are preliminary.

Hypothesis 2b: The grand mean effect size for child conduct problems will be significantly better for each group of studies, according to the amount of training therapists delivering IYPTP received, at a statistically significant level of 0.05. Lochman and colleagues (2009) looked at training in a unique and highly understudied manner by considering training on a relative scale of 3 conditions involved in transporting an EBI into a school setting. Benefits were significantly affected by whether therapists had a level of no training, basic training or training plus ongoing feedback. More benefits in terms of reducing behavioral problems in youth were also recently associated with adherence to treatment protocol as facilitated by ongoing supervision (Schoenwald, Sheidow & Chapman, 2009). For this meta-analysis it is anticipated studies would be grouped according to Lochman's distinctions of no training, basic training and training plus ongoing supervision.

Results for Hypothesis Test 2b: (see "Results for Hypothesis Test 2c")

Hypothesis 2c: The grand mean effect size for a group of studies reporting use of methods to ensure treatment fidelity will be significantly greater at the 0.05 level of significance, than the grand mean effect size for a group of studies not reporting use of methods to insure treatment fidelity.

Results for Hypothesis Test 2c: It was hypothesized that studies using some measure of treatment fidelity and those studies using more highly trained therapists (group leaders) would have greater benefits. While collecting data from the articles included within this meta-analysis it became evident that methods for insuring treatment integrity and the level of therapist training were inextricably tied together. The Tier Three primary studies (with one exception) listed several methods for insuring treatment integrity. These are listed below:

1.) Therapists co-led their first group with a supervisor.

2.) Sessions followed content directed from a treatment manual.

3.) Therapists kept detailed notes for each treatment session, including weekly protocol checklists for standards to be covered within each session (e.g. agenda, number of vignettes, role-plays to be completed and themes to be discussed). These were monitored weekly to ensure delivery of all components.

4.) Sessions were videotaped for review and feedback during weekly supervision meetings; and these were randomly selected for integrity checks which were always reported to be very high.

Examination of the Tier Three independent replications studies found that all of these contained similar methods of insuring treatment fidelity but at a lesser degree of

comprehensiveness. All of the replication studies made mention of ensuring that group leaders had multiple prior experiences with delivering the program. There were also some studies mentioning the use of the IYPTP manual where it is probably reasonable to assume all did this at some level. Most of the replication studies did not make mention of reviewing videotaped sessions, using checklists, or independent, random checks of integrity. There was consistent mention of weekly supervision meetings to ensure treatment integrity and ongoing training of therapists.

Both primary and independent replication studies using IYPTP as a Tier Three intervention reported high levels of therapist (group leader) training. All studies reported therapists had received specialized training in IYPTP. Four out of the six independent replications reported therapists had received 3 – 4 day trainings from either Dr. Webster-Stratton or another certified trainer endorsed by Dr. Webster-Stratton. Two of the six replication studies also indicated therapists were certified by Dr. Webster-Stratton.

CHAPTER 5

DISCUSSION

Common criticisms of meta-analysis include the assertion that combining different studies is like comparing apples to oranges (Borenstein et al., 2009) and that meta-analytic aggregation of studies yields claims about a swath of interventions too broad to provide meaningful conclusions (Beutler, 2009). This meta-analysis avoided these potential criticisms by including several features. First, this study focused on studies for one particular intervention. Second, the principal hypothesis for this study was evaluated using one type of outcome (child conduct problems) measured with only two methods (rating scales and independent observations) where many studies used the same or similar standardized measures. Third, this study combined IYPTP studies separately according to type of research study (between-groups versus within-groups) since effect sizes were calculated using different forms of data (pre-post changes for groups in isolation and compared with a control group). Fourth, studies were grouped by the particular use of IYPTP described (i.e. prevention, selective and indicated). These separations were further validated by divergent levels of conduct problem severity within each tier.

Another potential criticism of meta-analysis is that it ignores important data. This study used methods to acknowledge and avoid this potential problem. There was a careful, thorough and systematic search for data pertaining to the focus of this study. The nature of studies not included in this meta-analysis was tabulated to inform readers about areas of potential inadequate representation. Examination of the reasons studies were excluded could inform future research. Methods to evaluate for potential publication bias

and missing studies deployed in this meta-analysis suggested insufficient reason to suspect these were a significant problem for this meta-analysis. Additionally, the literature search found and included two pieces of unpublished research. While this could potentially cause a problem in terms of the lack of peer-review and quality assurance for these studies, these did not present a unique influence on the findings for this study. It should also be noted that relying on only narrative reviews of the professional literature suffers from the same potential publication bias as meta-analysis, but it is easier to ignore this with the former, and can at least be considered and evaluated within the latter (Borenstein et al., 2009).

Narrative reviews of studies are susceptible to weighting evidence non-systematically, thereby potentially ignoring valid data and/or validating problematic data. Interpretations of results for multiple variables, reported by a mixture of statistics (e.g., effect sizes, p-values, F-statistics), within and across studies, makes it difficult to form any consolidated conclusions for a set of studies (Olejnik & Algina, 2000). This is especially true when studies produce a mixture of results within and across similar research domains. Interestingly, the most recent primary study included in this meta-analysis (Webster-Stratton et al., 2004) did not cite each of the large number of outcome variables used separately, but did report composite scores comprised of the same two types of measures reported in this meta-analysis (e.g. ratings and observations) for the same three dependent variables of interest in this meta-analysis (i.e. child conduct problems, negative and positive parenting). This meta-analysis weighted studies to account for, and minimize variances within and between studies so that conclusions could be based on homogeneity of data produced by systematic and replicable procedures. This

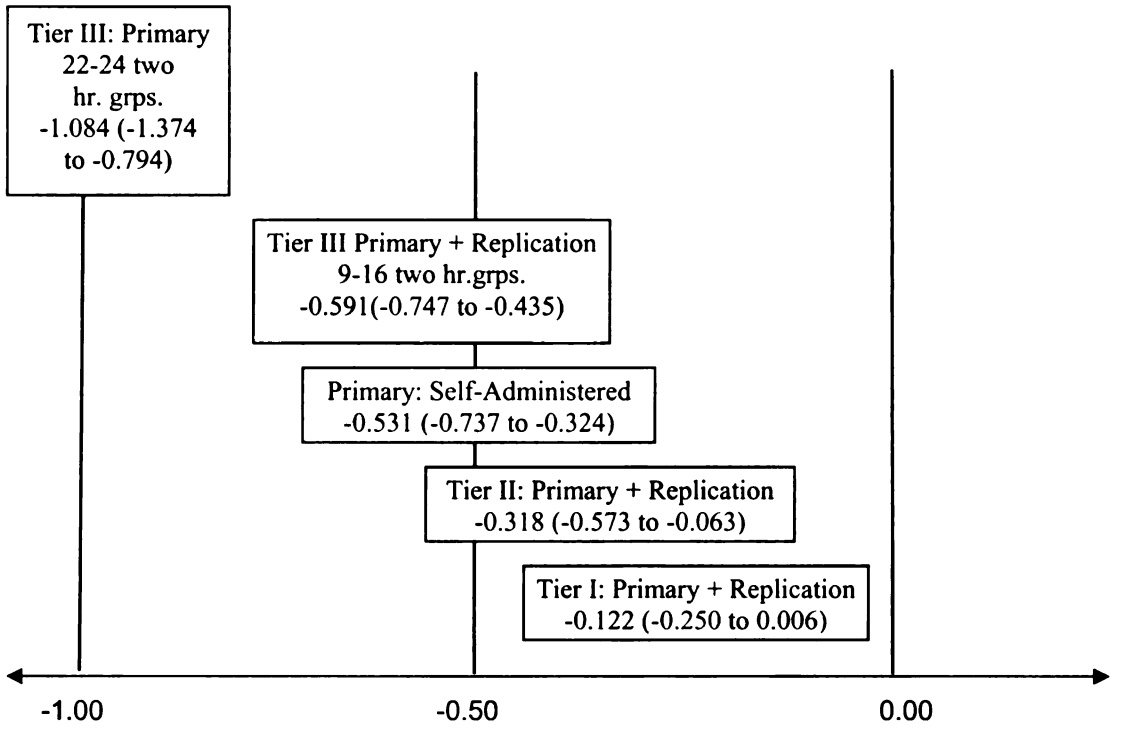
study avoided limitations of earlier meta-analyses of parent training that did not account for sample size and did not analyze significant sources of heterogeneity (Cedar & Levant, 1990; Lundahl et al., 2006; Serketich & Dumas, 1996). This study also adds to earlier meta-analytic findings touting the benefits of behavioral parent training as a general category of intervention (Maughan et al., 2005), by more focused evaluation of a specific program that could be transported into practice. This study also provides a systematic means of comparing meta-analytic results for IYPTP with the meta-analytic results from two recent studies examining other parent training programs intended to address child conduct problems (deGraaf et al., 2008; Thomas & Zimmer-Genbeck, 2007).

The principal hypothesis of this meta-analysis study was that primary research effect sizes (Hedges's *g*) would be greater than those for independent replications of IYPTP at a pre-set 0.05 level of statistical significance. Preliminary, significantly greater reductions of child conduct problems found by primary inventor-based research studies over that of independent replication studies (Sougstad et al., 2008), were not found within this more comprehensive meta-analysis. For the with-in groups studies of IYPTP there was no statistically-significant difference found between primary and replication study reductions in child conduct problems. For these within-groups studies the grand mean effect size (Hedges's *g*) of -0.634 (-0.791 to -0.478) represents a benefit of about two-thirds of a standard deviation. Examination of the within-groups studies suggested diverse uses of IYPTP (prevention, selected and indicated uses), but not enough studies for any particular use to support further grouping and statistical analysis. This makes final interpretations of findings for the within-groups studies difficult. Because there was a larger sampling of between-groups studies (includes control group) and this type of

study is regarded as the gold standard for demonstrating effects from an intervention, the remainder of this investigation concentrated on the between-groups data.

Figure 5 provides a summary for results from between-groups primary and replication studies of IYPTP across all uses and types examined within this meta-analysis. For between-groups Tier One Universal Prevention studies of IYPTP there were no statistically-significant differences between primary and independent replication effect sizes yielding a grand mean effect size (Hedges's *g*) of -0.122 (-0.250 to 0.006 with 95% confidence). This is a quite small effect, but not surprising considering subjects may not have shown any conduct problems to begin with. For between-groups Tier Two selective intervention studies there were no primary studies that met inclusion criteria for this meta-analysis. The between-groups replication studies using IYPTP as a Tier Two selective intervention yielded a grand mean effect size (Hedges's *g*) of -0.318 (-0.573 to -0.063 with 95% confidence). This represents about one-third a standard deviation of benefit. Tier Two studies focused on groups specific to a particular risk factor(s) (e.g. parenting known to be neglectful, youth placed in foster care, conduct problems including those below a clinically-significant level, etc,...).

Figure 5: Summary Table for IYPTP Between-Groups Studies Effects (Hedges's *g*)



Use of IYPTP as a Tier Three indicated intervention to treat the highest levels of clinically-significant child conduct problems yielded similar benefits for primary and replication studies when 9-16 two-hour sessions were used. These studies provided a grand mean effect size (Hedges's *g*) of -0.591 (-0.747 to -0.435 with 95% confidence) representing about two-thirds of a standard deviation reduction in child conduct problems. The primary studies using 22-24 two-hour sessions of IYPTP as a Tier Three indicated intervention yielded a statistically-significant higher Hedges's *g* mean effect size of -1.084 (-1.374 to -0.794 with 95% confidence) representing about 1 standard deviation of reduction in child conduct problems.

No replication studies of the self-administered form of IYPTP met inclusion criteria for this study. The primary studies of self-administered IYPTP yielded a grand

mean effect size (Hedges's *g*) of -0.531 (-0.737 to -0.324 with 95% confidence) which represents about one-half of a standard deviation of benefit.

The overall effect size for IYPTP Tier Three intervention is generally comparable to the measured benefit found from meta-analysis of the Triple-P Parent Training Program Level 4 that is used to address the highest levels of conduct problems (de Graff, et al., 2008), though the increased dosage studies of IYPTP (22-24 two-hour sessions) grand mean effect size (Hedges's g) appears to be significantly greater than the benefits reported for Triple-P Level 4. The absence of significantly different benefits from the self-administered form of IYPTP with that of the BASIC group-administered form of IYPTP at the 9-16 two-hour dosage level is not consistent with a similar comparison made by deGraff and colleagues (2008) for Triple-P. They found significantly greater benefits for a self-directed form of Triple-P over that of group administered. However Nowak and Heinrichs (2008) for Triple-P study did not find significant differences between the self-directed and group-administered Triple-P interventions.

While the number of studies is small, those IYPTP studies that used double the BASIC IYPTP program dosage (22-24 two-hour sessions) yielded twice the measured benefit from this more intensive dosage. Webster-Stratton and Herman (2010) recently mentioned a similar finding embedded within their recent article. They recently found significantly greater benefit from 20 weeks of two-hour IYPTP over that of 10 weeks IYPTP for families of children dually diagnosed with Oppositional Defiant Disorder and Attention Deficit Hyperactivity Disorder. They also cited a similar finding by Henggeler, Schoenwald, Liao, Letournea and Edwards (2002), that 20 sessions or more was

associated with significantly higher benefits treating a similar, older population of youth with conduct problems receiving Multi-Systemic Therapy.

Summary

Overall, there were no statistically-significant differences in reductions for child conduct problems between primary and replication research studies found in this meta-analysis when comparisons were made fairly, based on the usage and dosage of IYPTP across a three-tiered public health model of prevention and intervention. An initially encountered difference between primary and replication studies in the preliminary meta-analysis study (Sougstad et al., 2008) and also encountered in this current study for Tier Three IYPTP were spurious, due to the effect of an intervening higher dosage used in two primary studies. Once these two studies were removed there was no significant difference between effect sizes for primary and replication studies of IYPTP using similar dosages, for the same levels of child conduct problems.

The severity of conduct problems was found to be higher for each tier of intervention at a statistically significant level. Greater intensity and frequency of conduct problems coincided with a higher tier use of IYPTP, and a greater overall mean effect size from IYPTP intervention. This finding coincides with the findings of several other studies (Beauchaine, et al., 2005; deGraff et al., 2008; Spirito et al., 2009) and contradicts others reports on this issue (Nowak & Heinrichs, 2008; Thomas and Zimmer-Gembeck, 2007; Weisz et al., 2006).

Additional review of primary and replication Tier Three studies found these were more similar rather than different in terms of the level of training held by therapists

(group leaders) delivering IYPTP. The studies were also quite similar in terms of a pervasive use of the IYPTP Manual, prescribed videotapes, role-plays and coverage of content for each session, as well as use of weekly supervision meetings sometimes involving review of videotaped sessions. The latter form of supervision was most prevalent among primary studies as was the use of random checks of therapist completed checklists and submitted videos of sessions to verify treatment fidelity. The incidence and discrimination of differences in these forms of training and fidelity for each study were so small and difficult to detect differentially across this small sample of studies that statistical analysis was not appropriate. Finding these elements generally present across both primary and replication research studies seems to further help explain no significant differences between these two types of studies. The presence of these features for the primary and replication studies further reinforces the initial premise for this study, that IYPTP contains many features that make it a more likely candidate for effective replication and transportation beyond primary researchers. The finding of significant benefits for studies containing treatment agents highly trained and regularly supervised for the integrity of deployment is consistent with the recent findings of Lochman and colleagues (2009) as well as Schoenwald and colleagues (2009).

Dr. Webster-Stratton, her colleagues, and other researchers have invested a great deal of thought and resources to ensure successful transportability of IYPTP with good fidelity and integrity beyond the primary research team. Perhaps the distinction made for this study, between primary and independent replication of research studies according to authorship oversimplifies a complex set of issues involved with transporting, disseminating and evaluating an evidence-based intervention. Future examination of this

differentiation has important ramifications for current systems of evaluating the evidence-base for interventions. Currently, the highest levels of evidence are reserved for "independent" verification of benefits. Yet when primary researchers and colleagues are involved with training, ongoing supervision of EBI deployment and reviews of treatment integrity for certification in the IYPTP does this refute independence? Using Chorpita's typology (Table 1) this meta-analysis found that beyond the primary inventor's efficacy studies (Type I) there is a range of studies showing significant and robust benefits from IYPTP as a transported (Type II) and disseminated (Type III) EBI, while efforts to ensure the highest levels of training and fidelity of intervention deployment may "muddy the waters" for determining whether there is absolute independence. Data across three levels of severity for child conduct problems collected within this meta-analysis were supportive of more recent, proposed alteration of norms for determining clinical-significance on the Eyberg Child Behavior Inventory (Colvin, Eyberg & Adams, 1999).

The absence of no independent replications of between-group studies of the self-administered form of IYPTP prevented any comparison with an effect size representing primary studies of self-administered IYPTP. The finding of no between-groups or within-group studies using the self-administered form of IYPTP is quite surprising given the significant benefits reported by primary research, and the more plausible likelihood that the self-administered form would be more easily replicable without needing to ensure therapist integrity for group intervention.

For the most severe child conduct cases the group-administered form of IYPTP requiring a substantial investment of time and resources was only found to produce significantly higher benefits (about one standard deviation) over that of the self-

administered form of IYPTP (about half-standard deviation) when the dosage of two-hour group sessions was doubled to between 22 to 24. The latter finding has not been reproduced by independent replication studies but was recently reported by Webster-Stratton and Herman (2010) for youth diagnosed with Oppositional Defiant Disorder and Attention Deficit Hyperactivity Disorder.

There was some preliminary evidence collected for this meta-analysis demonstrating an improvement in parenting behaviors with IYPTP. Primary and replication studies showed statistically significant, moderate to large changes in reductions of negative parenting and increases in positive parenting. The latter finding was significantly greater for primary over that of replication studies. Because parenting measures varied greatly within different studies it was not possible to test to what degree initial measures of positive and negative parenting relate to found benefits from IYPTP.

Limitations

A potential limitation of this meta-analysis is that it was conducted on a fairly small number of studies. However, it is now possible to systematically compare these results to meta-analyses of other parent training programs (de Graff et al., 2008) and to conduct future systematic research that may also be compared to these findings. These cross-checks across similar research over time can produce meaningful conclusions for practice and future research. While a small number of studies included in this meta-analysis may be a limitation, this did allow for a careful analysis of dispersion across studies that might otherwise be obscured by larger data sets (e.g. larger data sets become more homogeneous by virtue of larger sampling). Small sample size is also not unique to

this meta-analysis. Weisz and colleagues (2005) noted small sample sizes within their meta-analysis of youth psychotherapy research studies from the past four decades and this has been noted for the studies across the larger field of psychological intervention (Kazdin, 2008).

Effect sizes calculated for studies included within this meta-analysis tell how much benefit was accrued from IYPTP intervention, avoiding the pitfalls of limitations inherent with null-hypothesis testing (Beutler, 2009; Hinshaw & Park, 1999; Kazdin, 2008; Kehle & Bray, 2006; Kehle et al., 2007; Olejnik & Algina, 2000; Sanabria & Killeen, 2007; Schmidt, 2009; Swaminathan & Rogers, 2007; Volker 2006). The meta-analytic procedures conducted for this study exceeded the rigorous standards of calculating effect sizes in isolation of one and other prescribed by the *Procedural and Coding Manual for the Identification of Evidence-Based Interventions* (Task Force on Evidence Based Interventions in School Psychology, 2003) by further synthesizing various effect sizes relative to their levels of common homogeneity across multiple primary and replication studies of IYPTP. This methodology sought to maintain the integrity of the core mechanism of change for IYPTP while systematically considering what variables relate to outcome differences for whom and under what conditions (Bhattacharyya et al., 2009; Fixsen et al., 2009; Ingraham & Oka, 2006; Sheridan 2005). While the magnitude of an effect from an EBI study (small, medium or large) holds some significance (Cohen, 1988), comparison of measured benefits relative to those previously obtained in the same research area or type of outcome as was done for this study, provides useful information about the actual value of an EBI (Durlak, 2002).

The bulk of meta-analysis was performed on between-groups studies of IYPTP while the within-groups studies lacked adequate numbers to perform additional analyses. The effect sizes used in the between-groups meta-analysis (Hedges' *g*) were calculated using the most rigorous methodology available, whereby differences between pre- and post- means were divided by the pooled pre-intervention standard deviations (Becker, 1988) and were adjusted for small sample sizes (Carlson & Schmidt, 1999; Hedges & Olkin, 1985). These methods produced the most conservative effect size value for studies with smaller sample sizes. Nevertheless, separation of studies across different forms and uses of IYPTP produced comparisons between small groups of studies. Considering the small number of studies and the small number of subjects within a group of studies (summarized in Table 16) one should use caution in interpreting the results from these comparisons. One way to do this is by interpretation of effect sizes within the context of their confidence intervals as was done for this study. Finding no significant differences between confidence intervals for grand mean effect sizes representing two groups of studies (primary and replication) may be at least partially an artifact of smaller sample sizes, warranting further confirmation as more studies may be added to address this primary research question. One other potential limitation of this study is that effect sizes representing each study were calculated by combining parent rating scale scores with independent observations of behavior for both child and parent dependent variables while excluding interview data that was only available within some studies. Further analysis of this excluded data would clarify to what degree this restriction of data may have affected results from this study.

Careful review of the between-groups studies resulted in conducting meta-analysis for three tiers of IYPTP intervention. Separating between-groups studies into three tiers of intervention supported equitable comparisons between primary and replication studies of IYPTP where unique benefits would be anticipated for differing uses (e.g. universal prevention, selective and indicated intervention). Webster-Stratton and Herman (2010) refer to IYPTP distinguish between universal prevention, selective and indicated interventions somewhat differently than was determined for this meta-analysis. For example, they consider IYPYP in Head Start Centers as a form of selective intervention whereas this meta-analysis categorized these as universal prevention since poverty is a risk factor for many child and family dysfunctions not just conduct problems. The distinction between three levels of tiered service delivery lies at the heart of the National Association of School Psychologists "Blueprint for Training and Practice, Third Edition" (2006). Evidence-based interventions should be viewed according to the purpose(s) they serve across tiers. This meta-analytic summary of a roughly thirty year history of IYPTP intervention studies supports deployment at all three tiers of intervention, where significant benefits were found in ameliorating child conduct problems and improving parenting behaviors.

Future Directions for Research

Several recommendations for future research are offered based on the findings of this meta-analysis. First, the absence of replication between-groups and within-groups studies of the self-administered form of IYPTP points to a notable gap within the research literature on IYPTP needing to be filled by researchers independent of the primary

researchers. There would seem to be a good likelihood that replication studies of the self-administered form of IYPTP would find significant benefits in reducing child conduct problems and that this form may be quite amenable to reproductions with good fidelity in settings beyond the primary researchers. Surprisingly the self-administered form of IYPTP has not been evaluated with sub-clinical child conduct problem populations where this may be found beneficial pointing to another area of needed research.

Second, some child conduct problem cases may be more amenable and adequately treated by the self-administered form of IYPTP than others, thus research investigating how to make this determination may be beneficial (e.g. What benchmark can be used to increase the level of intervention from a tier two to a tier three?). Case study research as well as group-level research could add helpful information in this regard. Research studies driven by a tiered model of intervention described within this research study may offer clearer methods of measuring response to intervention, and subsequent delivery of greater resources based on demonstrated needs.

Third, Webster-Stratton and colleagues have more recently developed two additional intervention components for the Incredible Years (IY) brand, adding child and teacher training to IYPTP. Subsequent meta-analyses may be beneficial in collecting primary and replication studies that include these additional components once there are more replication studies available. Since these additional treatment components represent a large amount of additional investment of time and resources a similar analysis to the one performed for this study will be beneficial in ascertaining what degree of benefit is found from these additional investments, and whether benchmarks set by the primary studies can be replicated.

Fourth, Dr. Webster-Stratton initially argued and has subsequently provided empirical support along with many others for the effective treatment of young children's conduct problems using a parenting group format as an economical alternative to having therapists intervene with each family separately. Additional replication studies directly comparing the benefits of the group-administered IYPTP to that of other parent training interventions implemented with families separately would provide additional data regarding to what degree benefits may be similar or different from group- versus individual-family treatment of child conduct problems. These findings would also hold significant implications for deployment of limited resources based upon predicted benefits from each type of intervention.

Fifth, IYPTP in its original, BASIC form consists of about 12-14 weekly, two-hour sessions (Webster-Stratton, 1996; Webster-Stratton, 2001; Webster-Stratton & Reid, 2010; Weisz, 2004). This meta-analysis found studies using 9-16 two-hour group sessions produced significant, immediate, moderate reductions in child conduct symptoms whereas 22-24 two-hour sessions of group-administered IYPTP accrued a large effect, roughly doubling the immediate benefit. These greater benefits from increased dosage still need to be verified by independent replication beyond primary researchers. Additional research is needed to demonstrate whether the significant benefit for 22-24 session dosage remains greater than that of the 9-16 session dosage for long term outcomes beyond immediate effects. A recent systematic review of child psychotherapy research studies found roughly fifty percent reported differential effects from varying dosage level, suggesting the presence of an active treatment ingredient (Jensen, Weersing, Hoagwood & Goldman, 2005). Further independent replication

research studies of IYPTP at an increased dosage would provide additional verification of the active treatment agent for IYPTP and its greater benefits with the most severe forms of child conduct problems and their parents at increased dosages.

Sixth, more research is needed to identify the mechanism(s) of change within an intervention such as IYPTP (Kazdin, 2008). Jensen and colleagues (2005) asserted studies generally lack adequate controls to account for nonspecific therapeutic factors of positive expectancies, therapeutic alliance and attention which likely play a role in intervention effectiveness. Future IYPTP primary and replication research studies that measure these variables would provide additional information regarding possible IYPTP moderators of change. Eames and colleagues (2009) recently developed the "Leader Observation Tool" (TOL) to study implementer's process skills using IYPTP. They used this research tool during the collection of data for one of the studies included in this meta-analysis (Hutchings et al., 2007). They found larger changes in positive parenting measured by independent observations were related to group leader's process skills. Among leader behaviors including listening, empathy, physical encouragement, and negative behaviors the category of "positive behavior" was found to be significantly related to differences in parent-reported parenting styles, while empathy skills showed a similar trend approaching statistical significance. Of particular relevance to the aforementioned discussion about needing to locate mechanisms of change, Eames and colleagues (2009) found changes in positive parenting predicted improvements in child conduct behaviors. This demonstrates changes in parenting from IYPTP do serve as a mechanism of change for improvements in child conduct problems. Further research delving into IYPTP group-leader process skills using the TOL or other similar measures

would add significant and important information toward additional successful replications of IYPTP by practitioners within authentic community settings. Additional measures of more nonspecific factors such as the favorability of therapist-parent relationship would also add to our knowledge about what works best and for whom. Webster-Stratton and Herman (2010) place a premium value on not only adequate fidelity of treatment content, but also on the meaningful interpersonal processes achieved within the training milieu. The latter qualities go beyond checklists for coverage of prescribed content and delve into the therapist's adaptation of the intervention to meet the unique needs of various treatment recipients. At this time, these qualities are only judged by those certified as mentors in the IYPTP, using reviews of videotaped session delivery. What qualities are evaluated at this level of clinical implementation and treatment adaptation should be further identified and studied using controlled research to further define what matters most in treatment deployment beyond the number of role-plays and video vignettes presented.

Seventh, the number of studies that report on observed changes in positive and negative parenting were quite small for the group included within this meta-analysis pointing to a need for these measures to be included within future research on parent training addressing child conduct problems. Only a handful of studies have already demonstrated changes in parenting mediate changes in child conduct problems (Beauchaine et al., 2005; DeGarmo et al., 2004; Eames et al., 2009). This meta-analysis found significant, moderate effects for reductions in negative parenting and increases in positive parenting consistent with these findings. Additionally, increases in positive parenting were significantly greater for primary versus replication studies, although the

small number of studies this finding was based on preclude any meaningful conclusions. Eames and colleagues (2009) finding of a significant correlation between positive group-leader behavior and that of improvements in positive parenting that are associated with changes in child conduct problems strongly points to the importance of further research examining the role of positive behaviors and their benefits toward altering undesirable child behaviors. Kazdin (2005) noted IYPTP goes beyond the typical behavioral, operant conditioning methods of many other parent training programs, and similarly the manual for IYPTP places a high value on working to establish positive relationships between parents and children before proceeding with more behaviorally-oriented methods of child management (Webster-Stratton, 2001). It is hypothesized for future research, that the magnitude of change in positive parenting plays a significant role in achieving higher benefits from IYPTP. Unfortunately for the small number of studies included in this meta-analysis there was not enough data to support research methods such as meta-regression to test this hypothesis.

Eighth, recognizing studies in psychology are often small in size would seem to warrant calculation of effect sizes using methods such as those contained within this meta-analysis that account for smaller sample sizes (Hedges' *g*). If the sampling of studies found for this research is any indication of the larger reporting of effect sizes within the profession, one should maintain caution and reservation regarding effect sizes that do not fully account for sample size, especially where samples are lower than twenty. Additionally, where effect sizes were reported in studies it was rare to also find reporting of confidence intervals for these effect sizes. In the absence of the latter, reporting of effect sizes as absolute values is just as problematic as reporting of significant p-values.

To properly interpret effect sizes and to compare them, confidence intervals are needed to estimate the accuracy of an effect size estimate relative to the larger group this statistic is believed to represent.

Ninth, meta-analyses are most productive when theory is integrated with empirical findings across studies such as was done for this study. Future meta-analyses must go beyond removal of outliers to obtain adequately homogeneous sets of studies supportable by current theories. The most stable, reliable and valid estimates of intervention potency and robustness can be demonstrated best when evidence contained within multiple studies of an intervention are weighted and systematically examined using meta-analysis rather than using vote counting and narrative reviews to determine an evidence base for an intervention.

Tenth, an emphasis on evidence-based intervention requires careful attention to many complex issues, not the least of which is answering, "What is an evidence-based intervention?" Many professional organizations within psychology and allied professions have organized thoughtful sets of criteria by which interventions are bestowed the "evidence-based" badge (Task Force on Evidence Based Interventions in School Psychology, 2003). These procedures often distinguish between higher and lower levels of evidence such as "promising" and "well established" (Eyberg et al., 2008). However, many of the methods used for evaluating research studies of an EBI amount to "vote counting" (Borenstein et al., 2009) whereby the credibility for an EBI is determined by the accumulated presence of several statistically significant results. This is highly problematic from a statistical perspective given the undeserving equal weight each p-value is afforded using this methodology. Null hypothesis tests reporting p-values tell

nothing about the true effect from the intervention (Hinshaw & Park, 1999; Kazdin 2008; Kehle & Bray 2006; Sanabria & Killeen, 2007). Borenstein and colleagues (2009) acknowledge the intuitive appeal, yet unsupportable notion of adding up the number of p-values (i.e. vote counting) to provide more accurate decisions than those based on isolated significance test results. However, obtained p-values are largely a function of study power. Non-significant p-values do not necessarily mean an effect is absent, and significant p-values may represent a small effect with little practical significance. Without the systematic weighing of evidence for an intervention as was done by this study, so called evidence-based practice decision-making may likely resemble the world of politics (e.g. vote counting) or popularity contests (e.g. cherry picking), which may be more about factors other than truth (McNeil, 2006). The latter undesirable practices make the profession more vulnerable to decision-making based more on "fads and the bandwagon effect' rather than cumulative scientific knowledge (Kehle & Bray, 2007; Trachtman, 1981). The methodology used to downgrade IYPTP from a "well-established" to a 'probably efficacious' intervention (Eyberg et al., 2008) essentially amounts to vote-counting of significant p-values and relied on a criterion of finding only fifty percent of reliable and valid outcome measures needing to show superiority for intervention versus control groups. However, this meta-analysis weighted findings for both IYPTP primary and replication research studies meeting inclusion criteria for this study using all parent rating scale and independent observation data reported. This procedure yielded strong and compelling reasons to assert IYPTP has a "well established" evidence-base across both primary and independent replication studies. For combined parent rating scale and independent observation outcomes (Table 10) and for

each of these outcomes separately (Tables 11 and 12 respectively) there are statistically significant benefits found for both primary and replication studies amounting to small to moderate to large effects depending on use. Differential outcomes relative to varying dosages as was found from this meta-analysis suggests future reviews of interventions need to account for dosage levels when ascribing a level of evidence base.

Eleventh, this meta-analysis relied on quantitative studies of IYPTP. These findings would be supplemented by additional qualitative research. For example, combining quantitative and qualitative research methods to study IYPTP implementation within a "real world community prevention practice" identified several barriers practitioners encountered relative to treatment fidelity (Stern, Alaggia, Watson & Morton, 2008). The researchers found a group of community based practitioners implemented IYPTP with a high degree of adherence to the treatment manual with three exceptions. Practitioners deviated from prescribed treatment in their use of role plays, the dosages of videotaped modeling and their use of "buddy calls." These findings have important implications relative to insuring treatment integrity and replication of IYPTP. Another qualitative study of a different parent training program implementation highlighted important, unintended consequences from the intervention not identified by quantitative methods (Mockford & Barlow, 2004). These researchers noted many mothers participating in the intervention experienced difficulties gaining the support of their partners in using new methods taught in the training, and that this led to increased discrepancies in parenting between partners, and increased parent conflicts. This finding has important implications to consider relative to a variable that may need to be accounted for relative to measuring benefits from an intervention such as IYPTP.

Twelfth, related to the aforementioned statistical support for IYPTP across primary and replication research studies, what methods are most appropriate to determine the level of evidentiary support for an EBI should be reviewed. Alternatives such as the methodology used for this meta-analysis might be considered for those interventions amassing multiple replication studies. While this method is much more time consuming, practice-based decision-making about EBI adoption should be made from "usable knowledge" (Rosenfield, 2000) and recipients deserve decisions based on the best evidence available (Fixsen et al., 2009). To make research usable by practitioners requires more studies like this one, that take a "Consumers Report" approach to weighing accumulated studies of an intervention using systematic, reproducible and transparent methods. Where there has been a recent proliferation of internet web sites purportedly describing levels of evidence-base for interventions, the professions of psychology and other allied mental health providers would be most accurately guided by research when this is accumulated and summarized regularly using accepted and transparent methods of meta-analysis. There is a frontier of research still to be done in order to make sense of the substantial number of research studies accumulating within a vast and ever-expanding research literature. More research is not necessarily better or more helpful unless we can make sense of, and integrate new information with what has already been studied.

Thirteenth, statistics differentiating IYPTP Tier Two studies from Tier Three studies revealed statistically significant differences for ECBI scores between these two groups. These findings were consistent with more recent, suggested re-standardization norms for the ECBI. Colvin and colleagues (1999) have suggested using a cut-off score of 15 and above to define clinical significance on the Problems Scale rather than the

former norm of 11 and above. The findings for this meta-analysis provide further evidence for this suggested change.

Implications for Practice

This meta-analysis study offers several important implications for practice. First, practitioner's selection of an intervention should rely on more than just whether it has produced a level of statistical significance unlikely produced by chance. Practitioners should make informed decisions to adopt an intervention based on answers to several important questions, including but not limited to these:

1.) What is the anticipated level of benefit for a particular use (i.e. universal prevention, selected or indicated intervention) and for whom?

2.) Have these benefits been replicated by others outside of the primary researchers, and do they achieve similar benefits?

3.) What level of problem severity is to be addressed?

4.) What level of dosage is necessary to achieve the desired benefits?

5.) What level of resources will be needed to implement with adequate fidelity?

Second, specific to this meta-analysis, IYPTP is found to be an intervention with a substantial evidence base well beyond the primary researchers that warrants adoption by practitioners in authentic community-based practice settings to address young children either at-risk or currently exhibiting child conduct problems. It has been successfully transported and disseminated by replication research studies outside the realm of primary research, across a three-tiered model of prevention and intervention. Several of these replications have used practitioners in authentic community settings to deliver the

intervention versus highly-controlled university research contexts. When use and dosage of IYPTP are accounted for, IYPTP produces similar, significant levels of benefit in reducing child conduct problems for both primary and replication research studies. The findings from this study suggest inventor-based benchmark levels of benefit can be achieved beyond the primary researcher when there is a high degree of therapist training and qualifications along with weekly supervision specific to the delivery of IYPTP. These findings suggest practitioners maintaining fidelity of treatment using IYPTP can expect similar, significant benefits in reductions for child conduct problems, diminution of negative parenting and increases in positive parenting. For clinically-significant child conduct problem cases practitioners should find the average benefit from BASIC IYPTP at a dosage of 9–16, two-hour sessions to be about a half-standard deviation reduction in child conduct problems while an increased dosage of between 22–24 two-hour sessions may lead to an average benefit of one standard deviation benefit in child conduct problems. These benefits would be according to combined parent rating scale and independent observations of child behaviors.

Third, findings from this meta-analysis suggest practitioners implementing IYPTP should likely find average benefits of about one half-standard deviation in reductions of negative parenting. Replication studies suggest about the same level of benefit in increased positive parenting while primary research suggests much greater, one standard deviation changes may be possible. Eames and colleagues (2009) recent research suggests positive and empathic group-leader behaviors during IYPTP delivery serves a particular benefit for increasing positive parenting behaviors and subsequent reductions in child conduct problems. Practitioner deliveries of group-administered IYPTP are

encouraged to emphasize positive interactions with parents that include empathy for their experiences of parenting a child with significant conduct problems. Considering these findings practitioners implementing IYPTP are strongly advised to not lose sight for how important it is to devote initial and ongoing efforts toward the development of strategies for encouraging positive relationships between group leaders and participants, as well as between parents and their children prior to engaging in implementation of more aversive strategies for child behavior management.

Fourth, the more time-consuming and high demand of resources needed for the group-administered form of IYPTP over that of the more economical self-administered form only appears warranted for the most severe cases of child conduct problems. The feasibility of investing resources in practice necessary for the group-administered form of IYPTP over that of the self-administered form for treating clinically-significant child conduct problems should be determined by the degree to which resources are available to ensure therapists are highly trained and supervised by qualified mentors of IYPTP, and treatment integrity is ensured by high levels of weekly expert supervision. The self-administered form of IYPTP may serve as an initial gateway procedure for determining a need for the consumption of much greater resources to deliver the group-administered form. In other words, initial treatment may be provided by use of the self-administered form for those families amenable to this level of intervention. Those not responding adequately would then be afforded the group-administered form. Foster and Roberts (2007) found about one-third of parents receiving a videotape parent training (not IYPTP) to handle clinic referred disobedient preschool children needed therapist assistance for initial compliance training. Sougstad and colleagues' (2008) meta-analytic comparison

of primary studies using self-administered IYPTP with that of self-administered IYPTP along with therapist consultation did not find significantly better outcomes from the additional provision of therapist consultation to the self-administered program. While the self-administered form of IYPTP appears to be an attractive intervention it also remains limited. Benefits are about half that of an intensive 22-24 week, two-hour deployment of the group-administered form and there is a large number of cases that do not adequately benefit from the self-administered form even with therapist consultation.

Appendix A

IYPTP Meta-Analysis Dissertation (Sougstad, 2010)

Title of Study: _____

Publication Name: _____

Authors Names: _____ Pub Date: _____

1.) ☐ Incredible Years Parent Training Interven. & ☐ Child Conduct Problems Outcome

2.) ☐ Primary Inventor-Based OR ☐ Independent Replication

3.) ☐ Random Assignment (exp vs. control grp) ☐ Other: _____

4.) ☐ Measures of treatment fidelity included (Type: _____

5.) ☐ Paid Participants OR ☐ Unpaid Participants

6.) ☐ Recruited Participants OR ☐ Referred for Treatment

7.) Parents: ☐ unskilled OR ☐ semi-skilled OR ☐ skilled

8.) ☐ 2-parent family OR ☐ Single-parent family

9.) ☐ Conduct Probs with ADD/Inattention, ☐ Conduct Probs with ADHD, Combined

☐ ADHD Only ☐ Conduct Problems Only

10.) ☐ United States, ☐ Canada, ☐ Denmark, ☐ Norway, ☐ Sweden, ☐ Ireland,

☐ Wales, ☐ England, ☐ South America: Venezuela, other: _____

☐ S. Korea, ☐ _____ ☐ _____ ☐ _____

11.) Therapist Characteristics:

Therapist Training Background:

☐ Non-professional (para-pro, etc,...), ☐ Graduate stds. ☐ Profess. (MA, MSW, PhD)

IYPTP level of Therapist Training:

☐ No formal training reported,

☐ Some training not from credentialed IY staff

☐ 3-day training from credentialed IY staff

☐ 3-day training from credentialed staff PLUS additional supervision

☐ Experts (certified in IYPTP, affiliated with IY and/or U. of Wash; Par Clinic)

12.) Mother Report: ☐, Father Report: ☐, Mother & Father Combined Rpt.: ☐

13.) Equivalence of control and experimental groups measured: _____

14.) Initial mean conduct problems score based on rating scale score: _____

_____

15.) Initial mean conduct problems score based on indep. observer: _____

_____

16.) Initial mean Negative parenting score (self-report rating): _____

_____

17.) Initial mean Negative parenting score (indep. obs.): _____

_____

18.) Initial mean Positive parenting score (self-report rating): _____

_____

19.) Initial mean Positive parenting score (indep. obs.): _____

_____

References

* = Primary Studies

= Independent Replication Studies

Aarons, G. A. (2005). Measuring provider attitudes toward evidence based practice: Organizational context and individual differences. *Child and Adolescent Psychiatric Clinics of North America, 14*, 255–271.

Aarons, G. A., & Palinkas, L. A. (2007). Implementation of evidence-based practice in child welfare: Service Provider Perspectives. *Administration and Policy in Mental Health and Mental Health Services Research, 34*, 411-419

Aarons, G. A., & Sawitzky, A. (2006). Organizational culture and climate and mental health provider attitudes toward evidence based practice. *Psychological Services, 3*, 61–72.

Achenbach, T. M., & Edelbrock, C. S. (1991). *Manual for the child behavior checklist and revised child behavior profile.* Burlington, VT: University Associates in Psychiatry.

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology, 32,* 328-340.

American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist, 57*, 1052-1059.

American Psychological Association's Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271-285.

American Psychological Association Task Force on Evidence-Based Practice for Children and Adolescents. (2008). *Disseminating evidence-based practice for children and adolescents: A systems approach to enhancing care.* Washington, DC: American Psychological Association.

Ang, R. P., & Hughes, J. N. (2002). Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review, 31*, 164-185.

Anthonysamy, A., & Zimmer-Gembeck, M. J. (2007). Peer status and behaviors of maltreated children and their classmates in the early years of school. *Child Abuse & Neglect, 31,* 971-991.

Arnold, M. E., & Hughes, J. N. (1999). First do no harm: Adverse effects of grouping deviant youth for social skills training. *Journal of School Psychology, 37*, 99–115.

#Axberg, U., Hansson, K., & Broberg, A. G. (2007). Evaluation of the Incredible Years Series – An open study of its effects when first introduced in Sweden. *Nordic Journal of Psychiatry, 61*, 143-151.

Bates, S. L. (2005). Evidence-based family-school interventions with preschool children. *School Psychology Quarterly, 20*, 352-370.

Beauchaine, T. P., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems. A latent growth curve analysis. *Journal of Consulting and Clinical Psychology, 73*, 371–388.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257-278.

Berkowitz, S. J. (2003). Children exposed to community violence: The rationale for early intervention. *Clinical Child and Family Psychology Review, 6*, 293-302.

Beutler, L. (2009). Making science matter in clinical practice: Redefining psychotherapy. *Clinical Psychology: Science and Practice, 16*, 301-317.

Bhattacharyya, O., Reeves, S., & Zwarenstein, M. (2009). What is implementation research? Rationale, concepts, and practices. *Research on Social Work Practice, 19*, 491-502.

Bickman, L., Lambert, E. W., Andrade, A. R., & Penaloza, R. (2000). The Fort Bragg continuum of care for children and adolescents: Mental health outcomes over 5 years. *Journal of Consulting and Clinical Psychology, 68*, 710-716.

Bloomquist, M. L. & Schnell, S. V. (2002). *Helping children with aggression and conduct problems, best practices foriIntervention*. New York: The Guilford Press.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (n.d.). *Comprehensive Meta-Analysis (Version 2.0)* [Computer software]. Englewood, N.J.: Authors.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, United Kingdom: John Wiley and Sons Ltd. Publications.

*Brotman, L. M., Gouley, K. K., Chesir-Teran, D., Dennis, T., Klein, R. G., & Shrout, P. (2005). Prevention for preschoolers at high risk for conduct problems: Immediate outcomes on parenting practices and child social competence. *Journal of Clinical Child and Adolescent Psychology, 34*, 724-734.

*Brotman, L. M., Klein, R. G., Kamboukos, D., Brown, E. J., Coard, S. I., & Sosinsky, L. S. (2003). Preventive intervention for urban, low-income preschoolers at familial risk for conduct problems: A randomized pilot study. *Journal of Child and Adolescent Psychology, 3*, 246-257

Brestan, E. V., & Eyberg, S. M. (1998). Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of Clinical Child Psychology, 27,* 180-189.

Bronson, M.B. (2000). *Self-regulation in early childhood, nature and nurture.* New York: Guilford.

Brophy, J. (1995). *Elementary teachers perceptions of and reported strategies for coping with twelve types of problem students.* East Lansing, MI: Institute for Research on Teaching, Michigan State University. ED 389 390. Retrieved from the World Wide Web October 30, 2007: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/44/bf.pdf

Burns, G. L., & Walsh, J. A. (2002). The influence of ADHD-hyperactivity/impulsivity symptoms on the development of oppositional defiant disorder symptoms in a 2-year longitudinal study. *Journal of Abnormal Child Psychology, 30*, 245-256.

*Bywater, T., Hutchings, J., Linck, P., Whitaker, C. J., Daley, D., Yeo, S., & Edwards, R. T. (In Press). Incredible Years Parent Training support for foster carers in Wales: A multi-centre feasibility study. Retrieved from http://incredibleyears.com/Library/paper.asp?nMode=1&nLibraryID=614

Cancelli, A. A., Lange, S. M., Lopez, E., Kerwin, C., Matalon, T., Perlman, L., . . . Hanson, J. (1989). A study of guidelines for the conduct and analysis of simultaneous clinical replications. *Professional School Psychology, 4*, 295-305.

Capaldi, D., DeGarmo, D., Patterson, G. R., & Forgatch, M. (2002). Contextual risk across the early life span and association with antisocial behavior. In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.) *Antisocial Behavior in Children and Adolescents: A Developmental Analysis and Model for Intervention.* (pp. 45-64) Washington, D.C.: American Psychological Association.

Carlson, K.D., & Schmidt, F.L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology, 84*, 851-862.

Castro, F. G., Barrera, M., & Steiker, L. K. H. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology,* (Vol. 6, pp. 213-239).

Cedar, B., & Levant, R. F. (1990). A meta-analysis of the effects of parent effectiveness training. *The American Journal of Family Therapy, 18*, 373-384.

Chafouleas, S. M., & Riley-Tillman, T. C. (2005). Accepting the gap: Introduction to the special issue of bridging research and practice. *Psychology in the Schools 42*, 455-458.

Chambless, D. L., & Hollon, S. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology. 66*, 7-18.

Chambless, D. L. & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and Evidence. *Annual Review of Psychology, 52,* 685-716.

Chorpita, B. F. (2003). The frontier of evidence-based practice. In A. E. Kazdin and J. R. Weisz (Eds.) *Evidence-based psychotherapies for children and adolescents* (pp. 42-59). New York: Oxford.

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77,* 566-579.

Chorptita, B. F., & Regan, J. (2009). Dissemination of effective mental health treatment procedures: Maximizing the return on a significant investment. *Behaviour Research and Therapy,* xxx, 1-4.

Chronis, A. M., Chacko, A., Fabiano, G. A., Wymbs, B. T., & Pelham, W. E. (2004). Enhancements to the behavioral parent training paradigm for families of children with ADHD: Review and future directions. *Clinical Child and Family Psychology Review,7*, 1-27.

*Clondalkin Parnership (2006). *Incredible Years Basic Parenting Programme: The Clondalkin Partnership Study in Ireland.* Retrieved from: http://www.incredibleyears.com/Library/paper.asp?nMode=1&nLibraryID=594

Cohen, J. (1988). *Statistical power for the behavioral sciences.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Cohen, M. A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology, 14*, 5-33.

Colvin, A., Eyberg, S. M., & Adams, C. D. (1999). *Restandardization of the Eyberg Child Behavior Inventory.* [On-line]. Available: http://pcit.phhp.ufl.edu/Measures/ECBI%20Restandardization.pdf Last accessed: 11/06/10.

Conduct Problems Prevention Research Group (1996). *Parenting Practices Inventory.* Unpublished manuscript. University of Washington: Seattle.

Cooper, H., & Hedges, L. V. (1994). Potentials and limitations of research synthesis. In H. Cooper & L. V. Hedges, (Eds.). *The handbook of research synthesis.* (pp. 521-528). New York: Russell Sage Foundation.

DeGarmo D. S., Patterson G. R., & Forgatch M. S. (2004). How do outcomes in a specified parent training intervention maintain or wane over time? *Prevention Science,* 5, 73-89.

deGraaf, I., Speetjens, P., Smit, F., de Wolff, M., & Tavecchio, L. (2008). Effectiveness of the Triple-P Positive Parenting Program on behavioral problems in children. *Behavior Modification, 32,* 714-735.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54,* 755-764.

Dodge, K. A. (2001). The science of youth violence prevention; Progressing from developmental epidemiology to efficacy to effectiveness to public policy. *American Journal of Preventive Medicine, 20,* 63-70.

Dodge, K. A., Dishion, T. J., & Lansford, J. E. (2006). *Deviant peer influences in programs for youth: problems and solutions.* New York: Guilford.

Duncan, R. D. (1999). Peer and sibling aggression: An investigation of intra- and extra-familial bullying. *Journal of Interpersonal Violence, 14,* 871-886.

Dunst, C. J., Hamby, D. W., & Trivette, C. M. (2004). Guidelines for calculating effect sizes for practice-based research syntheses. *Centerscope, 3,* 1 – 10.

Durlak, J. A. (2002). Evaluating evidence-based interventions in school psychology. *School Psychology Quarterly, 17,* 475-482.

Durlak, J. A. (2003). Appendix B: General guidelines for conducting a literature review. In *Procedural and Coding Manual for the Identification of Evidence-Based Interventions.* Retrieved July 1, 2007 from Task Force on Evidence-Based Interventions in School Psychology Web site: http://www.sp-ebi.org/documents/_workingfiles/EBImanual1.pdf

Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Jones, K., Hughes, J. C., & Bywater, T. (2009). Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training. *Child: care, health and development, 35,* 603-612.

Eddy, J. M., Reid, J. B., & Curry, V. (2002). The etiology of youth antisocial behavior, delinquency, and violence and a public health approach to prevention. In M. R. Shinn, H. M. Walker & G. Stoner (Eds.) *Interventions for Academic and Behavior Problems II: Preventive and Remedial Approaches.* (pp. 27-52) Bethesda, MD: National Association of School Psychologists.

Egger, H.L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology and epidemiology. *Journal of Child Psychology and Psychiatry, 47*, 313-337.

Eyberg, S. M., Boggs, S. R., & Algina, J. (1995). Parent child interaction therapy: A psychosocial model for the treatment of young children with conduct problems behavior and their families. *Psychopharmacology Bulletin, 31*, 83-91.

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

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.

Fergusson, D. M., Horwood, L. J., & Ridder, E. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology & Psychiatry, 46*, 837-849.

*Fergusson, D., Stanley, S. L., & Horwood, J. (2009). Preliminary data on the efficacy of the Incredible Years Basic Parent Programme in New Zealand. *Australian and New Zealand Journal of Psychiatry, 43,* 76-79.

Finkelhor, D., Ormrod, R., Turner, H., & Hamby, S. L. (2005). The victimization of children and youth: a comprehensive, national survey. *Child Maltreatment, 10*, 5-25.

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice, 19*, 531-540.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, xxx,* 1-25

Forman, S. G., Fagley, N. S., Steiner, D. D., & Schneider, K. (2009). Teaching evidence-based interventions: Perceptions of influences on use in professional practice in school psychology. *Training and Education in Professional Psychology, 3*, 226-232.

Fortin, L. (2003). Students' antisocial and aggressive behavior: Development and prediction. *Journal of Educational Administration, 41*, 669-688.

Foster, B. T. & Roberts, M. W. (2007). Training with videotapes: Recognizing limitations. *Child & Family Therapy, 29*, 21-35.

Frick, P. J. (2004). Developmental pathways to conduct disorder: implications for serving youth who show severe aggressive and antisocial behavior. *Psychology in the Schools. 41*, 823-834.

*Gardner, F., Burton, J., & Klimes, I. (2006). Randomised controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: Outcomes and mechanisms for change. *Journal of Child Psychology and Psychiatry, 47*, 1123-1132.

Gerler, E. R. (2004). *Handbook of school violence.* New York: Haworth Press.

Gilliam, W. S. (2005). *Prekindergartners Left Behind: Expulsion Rates in State Prekindergarten Programs.* New York: Foundation for Child Development.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-Analysis in social research.* Newbury Park, CA: Sage Publications.

Glicken, M. D. (2004). *Violent young children.* Boston, MA: Pearson Education, Inc.

Goldstein, S., & Rider, R. (2006). Resilience and the disruptive disorders of childhood. In S. Goldstein & R.B. Brooks (Eds.). *Handbook of resilience in children.* (pp. 203-222). New York: Springer.

Goodman R (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586.

#Gross, D., Fogg, L. Webster-Stratton, C., Garvey, C., Julion, W., & Grady, J. (2003). Parent training with families of toddlers in day care in low-income urban communities. *Journal of Consulting and Clinical Psychology*, 71, 261-278.

#Kim, E., Cain, K. C., & Webster-Stratton, C. (2008). The preliminary effect of a parenting program for Korean American mothers: A randomized controlled experimental study. *International Journal of Nursing Studies, 45,*1261-1273.

Graczyk, P. A., Domitrovich, C. E., Small, M., & Zins, J. E. (2006). Serving all children: An implementation model framework. *School Psychology Review, 35,* 266-274.

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W. & Robinson, N. (2006). Lost in knowledge translation: Time for a map? *Journal of Continuing Education in the Health Professions, 26*, 13-24.

Grimes, C. L., Klein, T. P. & Putallaz, M. (2004). Parents' relationships with their parents and peers: Influences on children's social development. In J.B. Kupersmidt & K.A. Dodge (Eds.) *Children's peer relations.* (pp. 141-158). Washington, DC: American Psychological Association.

Grolnick, W. S., & Apostoleris , N. H. (2002). What makes parents controlling?. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 161-181). Rochester, NY: University Of Rochester Press.

Grolnick, W. S., Deci, E. L., & Ryan, R. M. (1997). Internalization within the family: The self-determination theory perspective. In J. E. Grusec & L. Kuczynski (Eds.),

Parenting and children's internalization of values: A handbook of contemporary theory (pp. 135-161). New York: Wiley.

Hartman, R. R., Stage, S. A., & Webster-Stratton, C. (2003). A growth curve analysis of parent training outcomes: Examining the influence of child risk factors (inattention, impulsivity, and hyperactivity problems), parental and family risk factors. *Journal of Child Psychology and Psychiatry, 44*, 388-398.

Hedges, L.V. (1994). Statistical considerations. In H. Cooper & L.V. Hedges, (Eds.). *The handbook of research synthesis.* (pp. 29-38). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Toronto, Ontario, Canada: Academic Press.

Henggeler, S. W., Schoenwald, S. K., Borduin, C. M., Rowland, M. D., & Cunningham, P. B. (2009). *Multisystemic treatment of antisocial behavior in children and adolescents* (2nd ed.). New York: Guilford Press.

Hinshaw, S. P. (1987). On the distinction between attention deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Bulletin, 101,* 443-463.

Hinshaw, S.P., & Park, T. (1999). Research problems and issues: Toward a more definitive science of disruptive behavior disorders. In H.C. Quay & A.E. Hogan (Eds.) *Handbook of disruptive behavior disorders.* (pp. 593-620) New York: Kluwer Academic/Plenum Publishers.

Hoagwood, K., Burns, B.J., & Weisz, J.R. (2002). A profitable conjunction: From science to service in children's mental health. In B.J. Burns & K. Hoagwood (Eds.) *Community treatment for youth, evidence-based interventions for severe emotional and behavioral disorders.* (pp. 327-338) New York: Oxford University Press.

Hoagwood & Johnson (2003). School psychology: a public health framework I. From evidence-based practices to evidence-based policies. *Journal of School Psychology. 41*, 3-21.

Hughes, T.L., Crothers, L.M., & Jimerson, S.R. (2008). *Identifying, assessing and treating conduct disorder at school.* New York: Springer.

Hughes, T. L., Kaufman, J., & Miller, J. A. (2010). Is everything old new again: School psychology training past, present, and future. In T. L. Hughes & Kaufman, J. (Eds.) *Handbook of education, training and supervision of school psychologists in school and community.*

Hunsely, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies and beyond. *Professional Psychology: Research and Practice, 38,* 21-33.

*Hutchings, J., Gardner, F., Bywater, T., Daley, D., Whitaker, C., & Jones, K. (2007). Parenting intervention in sure start services for children at risk of developing conduct disorder: Pragmatic randomised controlled trial. *BMJ: British Medical Journal, 334* (7595)

Ingraham, C. L., & Oka, E. R. (2006). Multicultural issues in evidence-based interventions. In B. K. Nastasi (Ed.) *Multicultural issues in school psychology*. New York: Haworth Press.

Jaffee, S. R., Belsky, J., Harrington, H., Caspi, A., & Moffitt, T. E. (2006). When parents have a history of conduct disorder: How is the caregiving environment affected? *Journal of Abnormal Psychology, 115*, 309-319.

Jaffee, S. R., Caspi, A., Moffitt, T. E., & Taylor, A. (2004). Physical maltreatment victim to antisocial child: Evidence of an environmentally mediated process. *Journal of Abnormal Psychology, 113*, 44-55.

Jensen, P. S., Martin, D., & Cantwell, D. P. (1997). Comorbidity in ADHD: Implications for research, practice, and DSM-V. *Journal of the American Academy of Child and Adolescent Psychiatry, 36,* 1065-1079.

Jensen, P. S., Weersing, R., Hoagwood, K. E., Goldman, E. (2005). What is the evidence for evidence-based treatments? A hard look at our soft underbelly. *Mental Health Services Research, 7,* 53 – 74.

*Jones, K., Daley, D., Hutchings, J., Bywater, T., & Eames, C. (2007). Efficacy of the Incredible Years Basic parent training programme as an early intervention for children with conduct problems and ADHD. *Child: care, health and development, 33*, 749-756.

Kamphaus, R. W., & Frick, P. J. (2002). *Clinical Assessment of child and adolescent personality and behavior* (2nd ed.). Boston, MA: Allyn and Bacon.

Kazak, A. E., Hoagwood, K., Weisz, J. R., Hood, K., Kratochwill, T. R., Vargas, L. A., & Banez, G. A. (2010). A meta-systems approach to evidence-based practice for children and adolscents. *American Psychologist, 65*, 85-97.

Kazdin, A. E. (2005). *Parent management training; treatment for oppositional, aggressive, and antisocial behavior.* New York: Oxford University Press.

Kazdin, A. E. (2008). Evidence-based treatment and practice, New Opportunities to bridge clinical research and practice, enhance knowledge base, and improve patient care. *American Psychologist, 63*, 146-159.

Kazdin. A. E., Bass, D., Ayers, W. A., & Rogers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology, 58,* 729-740.

Keenan, D., & Shaw, D. S. (2003). Starting at the beginning: Exploring the etiology of antisocial behavior in the first years of life. Causes of conduct disorder and juvenile delinquency. In B. B. Lahey, T. E. Moffitt, & A. Caspi (Eds.) *Causes of conduct disorder and juvenile delinquency.* (pp. 153-181). New York: Guilford Press.

Kehle, T. J., & Bray, M. A. (2007). On such a full sea and we still are not afloat: An introduction to statistical reform in school psychology. *Psychology in the Schools, 44*, 415.

Kehle, T. J., Bray, M. A., Chafouleas, S. M., & Kawano, T. (2007). Lack of statistical significance. *Psychology in the Schools, 44*, 417-422.

Kendziora, K. T. (2004). Early intervention for emotional and behavioral disorders. In R. B. Rutherford, M. M. Quinn & S. R. Mathur (Eds.) *Handbook of research in emotional and behavioral disorders.* (pp. 327-351) New York: Guilford.

Kogan, K. L. & Gordon, B. M. (1975). A mother-instruction program: Documenting change in mother-child interactions. *Child Psychiatry and Human Development, 5*, 189-200.

Kokko, K., & Pulkkinen, Lea. (2000). Aggression in childhood and long-term unemployment in adulthood: A cycle of maladaptation and some protective factors. *Developmental Psychology, 36,* 463-472.

Kratochwill, T. R., Elliot, S. N., Loitz, P. A., Sladeczek, I. & Carlson, J. S. (2003). Conjoint consultation using self-administered manual and videotape parent-teacher training: Effects on children's behavioral difficulties. *School Psychology Quarterly, 18*, 269-302.

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17,* 341-389.

Kratochwill, T. R. & Shernoff, E. S. (2003). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Quarterly, 18*, 389-408.

Kratochwill, T. R. & Shernoff, E. S. (2004). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Review, 33*, 34-48.

Kumpfer, K. L. & Alvarado, R. (2003). Family-strengthening approaches for the prevention of youth problem behaviors. *American Psychologist, 58*(6/7), 457-465.

La Greca, A. M., Silverman, W. K., & Lochman, J. E. (2009). Moving beyond efficacy and effectiveness in child and adolescent intervention research. *Journal of Consulting and Clinical Psychology, 77*, 373-382.

Lahey, B., Miller, T. I., Gordon, R. A., & Riley, A. W. (1999). Developmental epidemiology of the disruptive behavior disorders. In H. C. Quay & A. E. Hogan (Eds.) *Handbook of disruptive behavior disorders.* (pp. 23-48) New York: Kluwer Academic/Plenum Publishers.

Lahey, B. B. Moffitt, T. E., & Caspi, A. (Eds.). (2003). *Causes of conduct disorder and juvenile delinquency.* New York: Guilford.

Lahey, B., & Waldman, I. D. (2003). A developmental propensity model of the origins of conduct problems during childhood and adolescence. In B. B. Lahey, T. E. Moffitt & A. Caspi (Eds.) *Causes of conduct disorder and juvenile delinquency.* (pp. 76-117) New York: Guilford.

*Larsson, B., Fossum, S., Clifford, G., Drugli, M. B., Handegard, D. H., & Morch, W. T. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children. *European Child and Adolescent Psychiatry, 18,* 42-52.

*Letarte, M., Normandeau, S., & Allard, J. (2010). Effectiveness of a parent training program "Incredible Years" in a child protection service. *Child Abuse & Neglect, 34,* 253-261.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment; Confirmation from meta-analysis. *American Psychologist, 48,* 1181-1209.

Lipsey, M. W., & Wilson, D. B. (2001a). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Lipsey, M. W., & Wilson, D. B. (2001b). The way in which studies have "personality" and why it is important to meta-analysis. *Evaluation & the Health Professions, 24,* 236-254.

Lochman, J. E., Boxmeyer, C., Powell, N., Qu, L., Wells, K., & Windle, M. (2009). Dissemination of the Coping Power Program: Importance of intensity of counselor training. *Journal of Consulting and Clinical Psychology, 77,* 397–409.

Loeber, R., & Farrington, D. P. (2000). *Child delinquents: development, intervention, and service needs.* Thousand Oaks, CA: Sage Publications, Inc.

Loeber, R., Farrington, D. P., & Petechuk, D. (2003). *Child Delinquency: Intervention and Prevention.* Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Lucas, S. M., & Cutspec, P. A. (2005). The role and process of literature searching in the preparation of a research synthesis. *Centerscope: Evidence-Based Approaches to Early Childhood Development, 4,* 1-26.

Lundahl, B., Risser, H. J., & Lovejoy, M. C. (2005). A meta-analysis of parent training: Moderators and follow-up effects. *Clinical Psychology Review, 26*, 86-104.

*Manby, M. (2005). Evaluation of the impact of the Webster-Stratton parent-child videotape series on participants in a midlands town in 2001-2002. *Children & Society, 19*, 316-328.

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*, 267-286.

McCabe, K. M., Hough, R., Wood, P. A., & Yeh, M. (2001). Childhood and adolescent onset conduct disorder: A test of the developmental taxonomy. *Journal of Abnormal Child Psychology, 29*, 305-316.

McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth" A meta-analysis. *Journal of Abnormal Child Psychology, 34*, 527-543.

*McIntyre, L. L. (2008a). Adapting Webster-Stratton's Incredible Years parent training for children with developmental delay: Findings from a treatment group only study. *Journal of Intellectual Disability Research, 52*, 1176-1192.

*McIntyre, L. L. (2008b). Parent training for young children with developmental disabilities: Randomized controlled trial. *American Journal of Mental Retardation, 113*, 356-368.

McFarlane, J. M., Groff, J. Y., O'Brien, J. A., & Watson, K. (2003). Behaviors of children who are exposed and not exposed to intimate partner violence: An analysis of 330 Black, White and Hispanic children. *Pediatrics, 112*, 202-207.

McHugh, R.K., & Barlow, D.H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65,* 73-84.

McMahon, R. J., Wells, K. C., & Kotler, J. S. (2006). Conduct problems. In E. J. Mash & R. A. Barkley (Eds.) *Treatment of childhood disorders* (3rd ed.). (pp. 137-268) New York: The Guilford Press.

McNeil, T. (2006). Evidence-based practice in an age of relativism: Toward a model for practice. *Social Work, 51*, 147-156.

Michael, K. D., Huelsman, T. J., & Crowley, S. L. (2005). Interventions for child and adolescent depression: Do professional therapists produce better results? *Journal of child and Family Studies, 14*, 223-236.

Mihalic, S., Irwin, K., Elliot, D., Fagan, A., & Hansen, D. (July, 2001). Blueprints for violence prevention. *Juvenile Justice Bulletin.* Washington, DC: United States Department of Justice, Office of Juvenile Justice and Delinquency Prevention. Retrieved from the World Wide Web on July 1, 2007 from: http://www.ncjrs.gov/pdffiles1/ojjdp/187079.pdf

Mockford, C., & Barlow, J. (2004). Parenting programmes: Some unintended consequences. *Primary Health Care Research & Development, 5,* 219 – 227.

Moffitt, T.E. (2005). The new look of behavioral genetics in developmental psychopathology: Gene-environment interplay in antisocial behaviors. *Psychological Bulletin,131,* 533-554.

Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology, 14,* 179–207.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent group designs. *Psychological Methods, 7,* 105-125.

Mullen, E. J., & Bacon, W. F. (2003). Practitioner adoption and implementation of evidence-based effective treatments and issues of quality control. In A. Rosen & E. K. Proctor (Eds.), *Developing practice guidelines for social work intervention: Issues, methods, and a research agenda.* New York City: Columbia University Press.

Muntz, R., Hutchings, J., & Edwards, R. T. (2004). Economic evaluation of treatments for children with severe behavioural problems. *Journal of Mental Health Policy and Economics, 7,* 177–189.

National Association of School Psychologists (2006). *Blueprint for training and practice* (3rd ed.). Silver Springs, MD: Author.

*Nilsen, W. (2007). Fostering futures: A preventive intervention program for school-age children in foster care. *Clinical Child Psychology and Psychiatry, 12,* 45-63.

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. *Psychological Medicine, 36,* 699–710.

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple-P Positive Parenting Program using Hierarchical Linear Modeling: Effectiveness and Moderating variables. *Clinical Child and Family Psychology Review, 11,* 114-144.

Nutley, S., Walter, I., & Davies, H. T. O. (2009). Promoting evidence-based practice: Models and mechanisms from cross-sector review. *Research on Social Work Practice, 19*, 552-559.

Ogg, J. A., & Carlson, J. S. (2009). The self-administered Incredible Years Parent Training Program: Perceived effectiveness, acceptability, and integrity with children exhibiting symptoms of Attention-Deficit/Hyperactivity Disorder. *Journal of Evidence-Based Practices for Schools, 10*, 143-166.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241-286.

Ollendick, T. H. (2005). Evidence-based parent and family interventions in school psychology: A commentary. *School Psychology Quarterly, 20*, 512-517.

Ollendick, T. H., & Davis, T. E., III. (2004). Empirically supported treatments for children and adolescents: Where to from here? *Clinical Psychology: Science and Practice, 11*, 289–294.

Ollendick, T. H., & King, N. J. (2000). Empirically supported treatments for children and adolescents. In P. C. Kendall (Ed.), *Child and adolescent therapy: Cognitive behavioral procedures* (2nd ed.). (pp. 386–425). New York: Guilford Press.

Olweus, D., Limber, S., & Mihakic, S. (1999). Bullying prevention program. In D.S. Elliott (Ed.) *Blueprints for violence prevention: Book Nine.* Boulder, CO: Center for the Study and Prevention of Violence.

Ortego, C., & Botella, J. (2010). The hard but necessary task of gathering order-one effect-size indices in meta-analysis. *Psicologica, 31*, 291-315.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8*, 157-159.

Pardinini, D. A., Lochman, J. E., & Powell, N. (2007). The development of callous-unemotional traits and antisocial behavior in children: Are there shared and/or unique predictors? *Journal of Clinical Child and Adolescent Psychology, 36*, 319-333.

Patterson, G. R. (2002). The early development of coercive family process. In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.) *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention.* (pp. 25-44) Washington, D.C.: American Psychological Association.

Patterson, G. R., & Yoerger, K. (2002). A developmental model for early- and late- onset delinquency. . In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.) *Antisocial behavior*

in children and adolescents: A developmental analysis and model for intervention. (pp. 147-172) Washington, D.C.: American Psychological Association.

*Patterson, J., Barlow, J., Mockford, C., Klimes, I., Pyper, C., & Stewart-Brown, S. (2002). Improving mental health through parenting programmes: Block randomised controlled trial. *Archives of Disease in Childhood*, 87:472–477

Popper, K. R. (1954). Degree of confirmation. *The British Journal for the Philosophy of Science, 18*, 143-149.

Quay, H. C. (1999). Classification of the disruptive behavior disorders. In H. C. Quay & A. E. Hogan (Eds.) *Handbook of Disruptive Behavior Disorders.* (pp. 3-21) New York: Kluwer Academic/Plenum Publishers.

Quintana, S. M., & Minami, T. (2006). Guidelines for meta-analyses of counseling psychology research. *The Counseling Psychologist, 34*, 839-877.

Radder, H. (2003). *The philosophy of scientific experimentation.* Pittsburg, PA: University of Pittsburg Press.

Reid, J. B. (1993). Prevention of conduct disorder before and after school entry: Relating interventions to development findings. *Journal of Development and Psychopathology, 5,* 243-262.

Reid, J. B., Patterson, G. R., & Snyder, J. (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention.* Washington, D.C.: American Psychological Association.

Reid, M. J., Webster-Stratton, C., & Baydar, N. (2004). Halting the development of conduct problems in head start children: The effects of parent training. *Journal of Clinical Child and Adolescent Psychology, 33*, 279-291.

Rhule, D. M. (2005). Take care to do no harm: Harmful interventions for youth problem behavior. *Professional Psychology: Research and Practice, 36*, 618-625.

Robinson, E. A., & Eyberg, S. M. (1981). The dyadic parent-child interaction coding system: Standardization and validation. *Journal of Consulting and Clinical Psychology, 49*, 245-250.

Robinson, E. A., Eyberg, S. M., & Ross, A. W. (1980). The standardization of an inventory of child conduct problem behaviors. *Journal of Clinical Child psychology, 9*, 22-28.

*Rogers, R. D. (2007). *Evaluating the effectiveness of the Incredible Years Parenting Education and Support Program* (Unpublished doctoral dissertation). James Madison University, Harrisonburg, VA.

Rosenfield, S. (2000). Crafting usable knowledge. *American Psychologist,* 1347-1355

Rosenthal, R. (1979). The "file-drawer problem" and tolerance for null results. *Psychological Bulletin*, 638-641.

Rosenthal, R. (1991). *Meta-Analytic procedures for social research.* Newbury Park, CA: Sage.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59-82.

Rutter, M. (2005). Environmentally mediated risks for psychopathology: Research strategies and findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 3-18.

Sanabria, F., & Killeen, P. R. (2007). Better statistics for better decisions: Rejecting null-hypotheses statistical tests in favor of replication statistics. *Psychology in the Schools, 44*, 471-481.

Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple-P Positive Parenting Program: A comparison of enhanced, standard, and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology, 68*, 624-640.

Satcher, D. (2001). *Youth violence: A report of the Surgeon General.* Washington, D.C.: U.S. Public Health Service, U.S. Department of Health and Human Services.

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology, 13*, 90-100.

Schoenwald, S. J., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology, 77*, 410-421.

*Scott, S., O'Connor, T., & Futh, A. (2006). What makes parenting programmes work in disadvantaged areas? The PALS trial. Downloaded from: http://ssiacymru.org.uk/media/doc/3/i/What_makes_parenting_progs_effective.pdf

*Scott, S., Spender, Q., Doolan, M., Jacobs, B. & Aspland, H. (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal, 323,* 1-6.

*Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Crook, C., & Landau, S. (2010). Randomised controlled trial of parent groups for child antisocial behavior targeting multiples risk factors: The SPOKES project. *Journal of Child Psychology and Psychiatry, 51*, 48-57.

Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy, 27*, 171-186.

Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology, 59*, 883-893.

Shaw, D., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry,45*, 1–12.

Sheridan, S. (2005). Commentary on evidence-based parent and family interventions: Will what we know now influence what we do in the future? *School Psychology Quarterly, 20*, 518-524.

Sheridan, S., & Kratochwill, T. (2007). *Conjoint behavioral consultation: promoting family-school connections and interventions* (2nd ed.) New York: Springer – Verlag.

Shirk, S. & McMakin, D. (2008). Client, therapist, and treatment characteristics in EBT's for children and adolescents. In *Handbook of evidence-based therapies for children and adolescents* (pp. 471-484). New York: Springer.

Silverman, W. K., Kurtines, W. M., & Hoagwood, K. (2004). Research progress on effectiveness, transportability, and dissemination of empirically supported treatments: Integrating theory and research. *Clinical Psychology: Science and Practice, 11,* 295–299.

Snyder, J., & Stoolmiller, M. (2002). Reinforcement and coercive mechanisms in the development of antisocial behavior: Peer relationships. In J. Reid, G. Patterson & J. Snyder (Eds.) *Antisocial Behavior in Children and Adolescents: A Developmental Analysis and Model for Intervention.* (pp. 65-100). Washington, DC: American Psychological Association.

Sougstad, J. R., Oka, E. R., Carlson, J. S., & Tomac, R. (2008, August). *Meta-Analysis of the Incredible Years Parent Training Intervention Program: Synthesis of primary and replication research across the frontier from efficacy to effectiveness.* Poster Presentation to the 116th Annual Convention of the American Psychological Association; Boston, MA.

Southam-Gerow, M. A., Marder, A. M., & Austin, A. A. (2008). Transportability and dissemination of evidence-based manualized treatments in clinical settings. In R. G. Steele, T. D. Elkin & M. C. Roberts (Eds.). *Handbook of evidence-based therapies for children and adolescents* (pp. 447-469). New York: Springer

Southam-Gerow, M. A., Ringeisen, H. L., & Sherrill, J. T. (2006). Integrating interventions and services research: Progress and prospects. *Clinical Psychology: Science and Practice, 13,* 1-8.

Spaccarelli, S., Cotler, S., & Penman, D. (1992). Problem-solving skills training as a supplement to behavioral parent training. *Cognitive Therapy and Research, 16*, 1-18.

Spirito, A., Abebe, K. Z., Iyengar, S., Brent, D., Vitiello, B., Clarke, G.,. . . Keller, M. (2009). Sources of site differences in the efficacy of a multisite clinical trial: The treatment of SSRI-resistant depression in adolescents. *Journal of Consulting and Clinical Psychology, 77,* 439–450.

Sprague, J., & Walker, H. (2000). Early identification and interventions for youth with antisocial and violent behavior. *Exceptional Children, 66*, 367-379.

Stahmer, A. C., & Aarons, G. A. (2009). Attitudes toward adoption of evidence-based practices: a comparison of Autism early intervention teachers and children's mental health providers. *Psychological Services, 6*, 223-234.

Steele, R. G. , Elkin T. D., & Roberts M. C. (Eds.). (2008). *Handbook of evidence-based therapies for children and adolescents*. New York: Springer.

Stern, S. B., Alaggia, R., Watson, K., & Morton, T. R. (2008). Implementing an evidence-based parenting program with adherence in the real world of community practice. *Research on Social Work, 18,* 543-554.

Stoiber, K. C. (2002). Revisiting efforts on constructing a knowledge base of evidence-based intervention within school psychology. *School Psychology Quarterly, 17*, 533-546.

Swaminathan, H., & Rogers, H. J. (2007). Statistical reform in school psychology research: A synthesis. *Psychology in the Schools, 44*, 543-549.

Task Force on Evidence Based Interventions in School Psychology. (2003). *Procedural and coding manual for the identification of evidence-based interventions.* Retrieved July 1, 2007 from Task Force on Evidence-Based Interventions in School Psychology Web site: http://www.spebi.org/documents/_workingfiles/EBImanual1.pdf

Taylor, T. K., Schmidt, F., Pepler, D., & Hodgins, C. (1998). A comparison of eclectic treatment with Webster-Stratton's parents and children series in a children's mental health center: A randomized controlled trial. *Behavior Therapy, 29*, 221-240.

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.

Trachtman, G. M. (1981). On such a full sea. *School Psychology Review, 10,* 138-181.

Tremblay, R. E., Nagin, D. S., Seguin, J. R., Zoccolillo, M, Zelazo, Boivin, M., Perusse, D., & Japel, C. (2003). Physical Aggression During Early Childhood: Trajectories and Predictors. *Pediatrics, 114*, 43-50.

van de Wiel, N., Matthys, W., Cohen-Kettenis, P. C., & van Engeland, H. (2002). Effective treatments of school-aged conduct disordered children: recommendations for changing clinical and research practices. *European Child & Adolescent Psychiatry, 11*, 79-84.

Volker, M. A. (2006). Reporting effect size estimates in school psychology research. *Psychology in Schools, 43*, 653-672.

Waas, G. A. (2002). Identifying evidence-based interventions in school psychology: Building a bridge or jousting with windmills? *School Psychology Quarterly, 17*, 508-517.

Walcott, C. M., Carlson, J. S., & Beamon, H. L. (2009). Effectiveness of a self-administered training program for parents of children with ADHD. *School Psychology Forum: Research in Practice, 3*, 43-61.

Walker, H., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices, Second Edition.* Belmont, CA: Wadsworth.

Watkins, M. (2009). Errors in diagnostic decision making and clinical judgment. In T. B. Gutkin & C. R. Reynolds (Eds.) *Handbook of School Psychology* (4th ed.). Hoboken, NJ: Wiley & Sons.

Watson, M. W., Fischer, K. W., Andreas, J. B., & Smith, K. W. (2004). Pathways to aggression in children and adolescents. *Harvard Educational Review, 74, 404-430.*

#Webster-Stratton, C. (1982a). The long-term effects of a videotape modeling parent training program: Comparison of immediate and 1 year follow-up results. *Behavior Therapy 13*, 702-714.

#Webster-Stratton, C. (1982b). Teaching mothers through videotape modeling to change their children's behaviors. *Journal of Pediatric Psychology* 7, 279-294.

#Webster-Stratton, C. (1984). A randomized trial of two parent training programs for families with conduct-disordered children. *Journal of Consulting and Clinical Psychology 52,* 666-678.

Webster-Stratton, C. (1989). *Dyadic parent-child interaction coding system – revised.* Unpublished manuscript. University of Washington: Seattle.

#Webster-Stratton, C. (1990a). Enhancing the effectiveness of self-administered videotape parent training for families with conduct-problem children. *Journal of Abnormal Child Psychology 18*, 479-492.

#Webster-Stratton, C. (1992). Individually administered videotape parent training: "Who benefits?" *Cognitive Therapy and Research, 16*, 31-51.

Webster-Stratton, C. (1994). Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology, 62*, 583-593.

Webster-Stratton, C. (1996). Early intervention with videotape modeling: Programs for families of children with oppositional defiant disorder or conduct disorder. In M. J. Guralnick (Ed.). *The Effectiveness of early intervention: second generation research. (pp. 435-473).* Baltimore, MD: Paul Brookes Publishing.

#Webster-Stratton, C. (1998). Preventing conduct problems in Head Start children: Strengthening parent competencies. *Journal of Consulting and Clinical Psychology 66*, 715-730.

Webster-Stratton, C. (2000). The incredible years training series. *Juvenile Justice Bulletin.* Washington, DC: United States Department of Justice, Office of Juvenile Justice and Delinquency Programs.

Webster-Stratton, C. (2001). *The parents and children series, BASIC program leader's guide.* Seattle, WA: Author.

Webster-Stratton, C. (2007). Tailoring the incredible years parent programs according to children's developmental needs and family risk factors. In J. M. Briemeister & C. E. Schaefer (Eds.). *Handbook of parent training, helping parents prevent and solve problem behaviors* (3rd ed.). (pp. 305-344). Hoboken, NJ: Wiley & Sons.

#Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology* 65, 93-109.

Webster-Stratton, C., & Hancock, L. (1998). Parent training of young children with conduct problems: Content, methods, and therapeutic processes. In C. E. Schaefer & J. M. Briesmeister (Eds.). *Handbook of parent training,* New York, NY: John Wiley & Sons. pp. 98 – 152.

Webster-Stratton, C., & Herman, K. C. (2010). Disseminating Incredible Years series early-intervention programs: Integrating and sustaining services between school and home. *Psychology in the Schools, 47*, 36-54.

#Webster-Stratton, C., Kolpacoff, M., and Hollinsworth, T. (1988). Self-administered videotape therapy for families with conduct-problem children: Comparison with two cost-effective treatments and a control group. *Journal of Consulting and Clinical Psychology, 56*, 558-566.

Webster-Stratton, C., & Lindsay, D.W. (1999). Social competence and conduct problems in young children: Issues in assessment. *Journal of Clinical Child Psychology, 28*, 25-43.

Webster-Stratton, C., Mihalic, S., Fagan, A., Arnold, D., Taylor, T., & Tingley, C. (2001). *The Incredible Years: Parent, Teacher And Child Training Series:*

Blueprints for violence prevention, book eleven. blueprints for violence prevention series. Boulder, CO: Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado.

Webster-Stratton, C., & Reid, M. J. (2010). The Incredible Years Parents, Teachers, and Children Training Series, A multifaceted treatment approach for young children with conduct problems. In J. Weisz & A. Kazdin (Eds.). *Evidence-based psychotherapies for children and adolescents* (2nd ed.). (pp. 224-240). New York: Guilford Publishing.

#Webster-Stratton, C., Reid, M. J. and Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in head start. *Journal of Clinical Child Psychology, 30*, 283-302.

#Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33*, 105-124.

#Webster-Stratton, C. & Shoecraft, K. (2009, unpublished). Delivering of Incredible Years Parenting Program. *Washington State Child Welfare Report.* Retrieved from: http://incredibleyears.com/Library/items/washington-state-child-welfare-report_6-25-09%20.pdf

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0 to 8 years). *Prevention Science, 2*, 165-192.

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,* 299–310.

Weiss, B., Catron, T., Harris, V. & Phung, T. M. (1999). The effectiveness of traditional child psychotherapy. Journal of Consulting and Clinical Psychology, 67, 82-94.

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples.* Cambridge, UK: Cambridge University Press.

Weisz, J. R., Donenberg, G. R. Han, S. S., & Weis, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 688-701.

Weisz, J. R., Hawley, K. M., & Doss, A. J. (2004). Empirically tested psychotherapies for youth internalizing and externalizing problems and disorders. *Child and Adolescent Psychiatric Clinics of North America, 13*, 729-815.

Weisz, J. R., Huey, S. M., & Weersing, V. R. (1998). Psychotherapy outcome research with children and adolescents: The state of the art. In T. H. Ollendick & R. J. Prinz (Eds.), *Advances in clinical child psychology* (Vol. 20). (pp. 49–92). New York: Plenum Press.

Weisz, J. R., & Jensen, A. L. (2001). Child and adolescent psychotherapy in research and practice contexts: Review of the evidence and suggestions for improving the field. *European Child & Adolescent Psychiatry, 10*, 12-18.

Weisz, J. R. Jensen-Doss, A. & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist, 61*, 671-689.

Weisz, J. R. Jensen-Doss, A. & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology, 56,* 337-363.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542-549.

Weisz, J. R., Weis, B., & Donenberg, G. R. (1992). The lab versus the clinic; Effects of child and adolescent psychotherapy. *American Psychologist, 47*, 1578-1585.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin, 117*, 450–468.

White, J. L. & Kratochwill, T. R. (2005). Practice guidelines in school psychology: Issues and directions for evidence-based interventions in practice and training. *Journal of School Psychology, 43*, 99-115.

Wilkinson, L. & the American Psychological Associatiation Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 549-604.

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*, 413-429.

Wilson, S. J., Lipsey, M. W. & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136-149.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Newbury Park, CA: Sage.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 03220 8674