



# Effectiveness of the Incredible Years parent training to modify disruptive and prosocial child behavior: A meta-analytic review<sup>☆</sup>



Ankie T.A. Menting<sup>a,\*</sup>, Bram Orobio de Castro<sup>a</sup>, Walter Matthys<sup>b</sup>

<sup>a</sup> Developmental Psychology, Utrecht University, Utrecht, The Netherlands

<sup>b</sup> Department of Child and Adolescent Studies, Utrecht University, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands

## HIGHLIGHTS

- We examined effectiveness of the IYPT as treatment and preventive intervention.
- Results of 50 studies showed that the IYPT is effective regarding child behavior.
- Effects with regard to distinct outcomes and distinct informants were found.
- Initial severity of child behavior was the strongest predictor of effects.
- The IYPT meets criteria for a well-established intervention.

## ARTICLE INFO

### Article history:

Received 27 February 2012

Revised 26 April 2013

Accepted 16 July 2013

Available online 22 July 2013

### Keywords:

Meta-analysis

Parent training

Incredible Years

Disruptive behavior

Prosocial

## ABSTRACT

The present meta-analytic review examined effectiveness of the Incredible Years parent training (IYPT) regarding disruptive and prosocial child behavior, and aimed to explain variability in intervention outcomes. Fifty studies, in which an intervention group receiving the IYPT was compared to a comparison group immediately after intervention, were included in the analyses. Results showed that the IYPT is an effective intervention. Positive effects for distinct outcomes and distinct informants were found, including a mean effect size of  $d = .27$  concerning disruptive child behavior across informants. For parental report, treatment studies were associated with larger effects ( $d = .50$ ) than indicated ( $d = .20$ ) and selective ( $d = .13$ ) prevention studies. Furthermore, initial severity of child behavior revealed to be the strongest predictor of intervention effects, with larger effects for studies including more severe cases. Findings indicate that the IYPT is successful in improving child behavior in a diverse range of families, and that the parent program may be considered well-established.

© 2013 Elsevier Ltd. All rights reserved.

## Contents

1.	Introduction . . . . .	902
1.1.	Intervention characteristics . . . . .	903
1.2.	Child characteristics . . . . .	903
1.3.	Family characteristics . . . . .	903
1.4.	Methodological features . . . . .	903
1.5.	The current study . . . . .	904
2.	Method . . . . .	904
2.1.	Inclusion criteria . . . . .	904
2.2.	Literature search . . . . .	904
2.3.	Coding of study characteristics . . . . .	904
2.3.1.	Differential effectiveness: study context and standard IYPT . . . . .	905
2.3.2.	Moderators . . . . .	905

<sup>☆</sup> The original research project is funded by the Dutch Ministry of Security & Justice, Directorate-General for Youth and Sanctions, Judicial Youth Policy Department.

\* Corresponding author at: Developmental Psychology, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands. Tel.: +31 30 2531971; fax: +31 30 2534718.  
E-mail address: A.T.A.Menting@uu.nl (A.T.A. Menting).

2.4.	Coding of effect size statistics . . . . .	907
2.5.	Procedure and statistical analysis . . . . .	907
3.	Results . . . . .	907
3.1.	Study characteristics . . . . .	907
3.2.	Relations between study characteristics . . . . .	907
3.3.	Analyses . . . . .	908
3.3.1.	Overall effect sizes . . . . .	908
3.3.2.	Differential effectiveness: study context and standard IYPT . . . . .	908
3.3.3.	Moderators of effect size for parental report . . . . .	908
4.	Discussion . . . . .	909
4.1.	Overall effect sizes . . . . .	909
4.2.	Differential effectiveness . . . . .	909
4.3.	Moderators of effect size . . . . .	910
4.4.	Limitations . . . . .	910
4.5.	Future research . . . . .	910
4.6.	Implications . . . . .	911
	References . . . . .	911

## 1. Introduction

Behavioral parent training (BPT) has been proven to be the most effective intervention method for pre-school and school-aged youth with antisocial behavior problems (McCart, Priestler, Davies, & Azen, 2006). Although it is clear that training parents to use behavioral techniques can affect their children's behavior, it is less clear which factors influence the effectiveness of BPT programs (Hinshaw, 2002). Even with successful intervention programs, studies reveal substantial variability in outcomes. Identification of those who respond differently to intervention programs is necessary, because this points to interesting groups and possibilities to optimize interventions (Hinshaw, 2002). Also intervention effects may be associated with non-participant related factors, such as intervention formats and methodological features. Examination of these features may also yield possibilities for optimization of intervention programs in general, but also for specific programs. Examining to what extent a specific BPT program is effective in specific populations, with certain formats or within certain contexts, may guide better targeting of intervention programs and help to optimize the specific intervention program. Examination of a specific BPT program allows for more specific inspection of factors that might influence its effects. This improved understanding of a specific BPT program might guide agencies' or practitioners' choices to adopt a specific BPT program. Furthermore, knowledge considering factors that influence effectiveness of a specific BPT program might also be relevant in view of other BPT programs.

In examining specific BPT programs, it seems particularly useful to focus on the Incredible Years parent training (IYPT; Webster-Stratton, 2001), which is considered a "blueprint" for violence prevention (e.g., Mihalic & Irwin, 2003). The IYPT has a group based format, distinguishing the IYPT from most other BPT programs. During the IYPT, parents of young children view videotapes depicting parent models interacting with their children in various situations. In collaboration with two group leaders, who use an empowering approach, parents discuss these video vignettes and put learned techniques into practice through role-plays. In addition, home assignments are used to encourage parents to practice parenting skills at home. In the BASIC IYPT the subjects play skills, praise and rewards, limit setting and handling misbehavior are discussed, while the supplementary ADVANCE component goes further into interpersonal issues such as communication and problem solving (Webster-Stratton, 2002). Together with the Baby and Toddler Series, School Readiness Program, Incredible Years Child Programs and Incredible Years Teacher Programs these programs form the comprehensive set of Incredible Years curricula.

The IYPT has seen widespread growth and is used and studied internationally as a treatment for children with severe conduct problems, as well as a preventive intervention. The IYPT has proven to be an effective

intervention for disruptive child behavior according to numerous effect studies (e.g., Scott, Spender, Doolan, Jacobs, & Aspland, 2001; Webster-Stratton, 1984; Webster-Stratton & Hammond, 1997) and reviews (e.g., Bauer & Webster-Stratton, 2006; Brestan & Eyberg, 1998). However, other studies revealed less convincing evidence of effectiveness (e.g., Helfenbaum-Kun & Ortiz, 2007; Webster-Stratton, 1998). In addition, the initial classification of the IYPT as a well-established treatment in a review-article by Brestan and Eyberg (1998) was later recanted by the authors: it was believed to be based on erroneous recording of the direction of group differences. According to a later review, the IYPT should be considered rather as probably efficacious, because of a lack of supportive replications by independent researchers, in addition to a lack of studies comparing the IYPT to other treatments (Eyberg, Nelson, & Boggs, 2008). However, it is unclear why independent studies by, for example, Drugli and Larsson (2006), Gardner, Burton, and Klimes (2006), and Scott et al. (2001) were not included in the review by Eyberg et al. (2008).

Furthermore, because of the inclusion criterion in the aforementioned reviews – only studies regarding youth with significant levels of disruptive behavior were included – a number of IYPT studies regarding youth with lower levels of disruptive behavior, particularly preventive studies, have not been included. Also new studies appeared since previous reviews were published and conclusions about the effectiveness of the IYPT differ among studies and reviews. Therefore, more clarity is desirable about the extent to which this specific parent training program works, both as a treatment program and a preventive program, and both "standard IYPT" and variants of the IYPT program. The large variation in outcomes of IYPT effect studies also raises the question 'what determines the effectiveness of the IYPT?'

The aims of the present study are to provide a meta-analytic test of the effectiveness of the IYPT in improving child behavior and to examine potential moderators of effectiveness. In this study, we use *effectiveness* as a generic term encompassing both efficacy and effectiveness studies. As discussed by others (e.g., La Greca, Silverman, & Lochman, 2009), efficacy and effectiveness may be considered along a continuum rather than as sharply distinct categories, and this continuum will likely be most informative for both practice and research. By examining moderators across this continuum, we hope to gain insight into factors critical when designing and studying effective parenting interventions.

Dissimilarities among studies of parenting interventions might partly explain differential conclusions about effectiveness between studies. Reviews of parent training literature and empirical studies of the IYPT and comparable interventions suggest a number of factors that may influence child outcomes of the IYPT. In the next paragraphs, we discuss suggested factors related to intervention characteristics, child characteristics, family characteristics, and methodological features within studies. Additionally, characteristics that describe the general study context may be relevant,

because they might be proxies for substantive or methodological variables (Lipsey & Wilson, 2001). For example, meta-analytical evidence suggests considerably larger effect sizes for studies conducted by developers-as-evaluators than for independent trials. This may stem from both systematic bias due to problematic decisions during the study and better implementation quality in developer-led studies (Petrosino & Soydan, 2005). Therefore, descriptive characteristics such as the involvement of the IYPT's developer Dr. Webster-Stratton might influence child outcomes of the IYPT.

### 1.1. Intervention characteristics

The IYPT is applied in various settings. The originally clinic-based treatment program has also been used as a prevention program for various high-risk populations, such as Head Start children (e.g., Webster-Stratton, 1998), pre-school children with adjudicated siblings (Brotman et al., 2005), and children placed in foster care (Linares, Montalto, Li, & Oza, 2006). Furthermore, the training has been offered as a stand-alone intervention as well as in packages with other elements, including complementary curricula from the Incredible Years Training Series. Moreover, precursors of the IYPT (e.g., Webster-Stratton, 1982, 1984) and deliberately stripped formats, such as individual parent training (Kratochwill, Elliott, Loitz, Sladeczek, & Carlson, 2003; Webster-Stratton, Kolpacoff, & Hollinsworth, 1988; Webster-Stratton, 1992), are variants of the well-known IYPT, for which effectiveness has also been reported. These features of the intervention protocol and delivery of sessions may be relevant regarding effectiveness. Hence, it might be relevant to examine *study context*, *training components*, *number of IYPT sessions offered*, *total number of sessions*, and *number of sessions attended*.

Study context concerns whether the intervention should be classified as a treatment (intervention for families who sought help), indicated prevention (intervention aimed at children who are identified as having minimal, but detectable signs or symptoms), selective prevention (intervention targeted at children who are considered at high-risk due to biological, psychological or contextual factors), or as a universal prevention (intervention targeted at children who have not been identified on the basis of individual risk; Mrazek & Haggerty, 1994). A common thought is that prevention trials reveal smaller effects than treatment studies. In prevention studies, there may be less room for improvement, because of lower initial severity of child behavior and inclusion of a high number of participants who do not actually need intervention: "false positives" (Bennett, Lipman, Racine, & Offord, 1998). Yet, type of preventive intervention was not related to effect size in a study by McMahon, Holly, Harrington, Roberts, and Green (2008) for the prevention of conduct disorder.

The distinction between prevention and treatment studies may be related to initial problem severity of participating children. Probably, treatment studies generally include participants with more severe problems than prevention studies. However, study context and initial problem severity are not identical. It is quite possible that a selective prevention trial in a high-risk environment includes participants with problems as severe as participants in treatment studies elsewhere. Therefore, we distinguish separate variables concerning the study context (intervention characteristic) and the initial problem severity (participant characteristic; see below).

Regarding training components, Foster, Olchowski, and Webster-Stratton (2007) concluded that stacking intervention components of the Incredible Years program is cost-effective, based on per-child treatment costs and child behavior outcomes. However, other studies revealed that a package of interventions does not always provide better outcomes (Beauchaine, Webster-Stratton, & Reid, 2005; Kaminski, Valle, Filene, & Boyle, 2008; Thomas & Zimmer-Gembeck, 2007).

Also for training content received by parents, the number of sessions offered as well as the number of sessions attended may be relevant. Evidence regarding treatment dosage is, in general, inconclusive. Wilson

and Lipsey (2001) mentioned that dose is associated with effect size variance, but noted that treatment intensity and duration are probably confounded with other study features. However, Serketic and Dumas (1996) found no relation between treatment length and child outcomes for BPT. Additionally, for the IYPT, training engagement and outcome are positively associated in a dose-response manner (Baydar, Reid, & Webster-Stratton, 2003). However, this was not supported regarding children's externalizing behavior (Beauchaine et al., 2005).

### 1.2. Child characteristics

Characteristics of the target child, such as *age*, *gender*, *initial severity of child behavior*, the extent to which the child's behavior is considered a *problem*, and *clinical symptom levels*, might also moderate training effectiveness. Child's age and gender may not be predictive of treatment effect for parent training programs in general and the IYPT (Beauchaine et al., 2005; Lundahl, Risser, & Lovejoy, 2006). Yet another IYPT study revealed relatively better intervention outcomes for young children and for boys, in which the latter moderator effect was primarily because boys in the control group fared worse than girls (Gardner, Hutchings, Bywater, & Whitaker, 2010). Regarding initial severity of child behavior, some narrative reviews suggest that more severe behavior predicts reduced responsiveness to parent training (Assemany & McIntosh, 2002; Kazdin, 1997; Nixon, 2002). However, a meta-analytic review by Lundahl et al. (2006) showed, with respect to BPT, that larger effect sizes were observed in studies including clinical symptom levels at pre-treatment than in studies which included non-clinical or mixed symptom levels. Moreover, an IYPT study suggests that children with more and less difficulties have an equal chance of responding well to the program (Gardner et al., 2010).

### 1.3. Family characteristics

The main family characteristics that may be associated with intervention outcomes are *single parenthood*, *ethnic minority* status, mother's level of *education*, and *at-risk* populations. Numerous studies suggest that single-parent households profit less from parent training (e.g., Assemany & McIntosh, 2002; Kazdin, 1997; Nixon, 2002; Reyno & McGrath, 2006). However, other studies (Serketic & Dumas, 1996), including IYPT studies (Beauchaine et al., 2005; Fossum, Mørch, Handegard, Drugli, & Larsson, 2009; Gardner et al., 2010), do not support this assumption. Several studies suggest that ethnic minority membership is related to poor treatment response (e.g., Assemany & McIntosh, 2002; Nixon, 2002). Likewise, parent training is expected to be less beneficial for low-educated, young mothers (e.g., Assemany & McIntosh, 2002). However, replication of this assumption failed in some IYPT studies (Beauchaine et al., 2005; Fossum et al., 2009; Gardner et al., 2010). To summarize, families which can be considered at-risk, because of various parental, contextual or family characteristics, may benefit less from parent training interventions. However, findings regarding at-risk samples are inconclusive, and might be different for IYPT studies.

### 1.4. Methodological features

Methodological features may be important in explaining differential effectiveness (Wilson & Lipsey, 2001). In IYPT studies, a range of child outcomes has been examined and different methods have been used across different studies to measure similar behaviors. Main methodological moderators may be related to assessment (e.g., targeted behavior, used informants, and the use of specific questionnaires such as the Eyberg Child Behavior Inventory [ECBI]), the type of *assignment*, and the *nature of the comparison group*.

First, regarding assessment, different informants may yield dissimilar results. For example, direct observation may yield different conclusions than parental report. Scott (2001) stated that direct observation

shows much smaller effects than parental questionnaires in several intervention studies. However, observed behavior was reported to be more sensitive to intervention effects than parent ratings of behavior in some IYPT studies (e.g., Brotman et al., 2008; Gardner et al., 2006; Posthumus, Raaijmakers, Maassen, Van Engeland, & Matthys, 2012). Also within informant dissimilarity may arise from differences in assessment: Scott (2001) revealed variation in effect sizes according to the outcome measured and the instrument used in a trial of the IYPT. Semi-structured interviews and user satisfaction showed more improvement than child symptoms and screening instruments. Furthermore, general screening instruments might be less sensitive to change than specific scales, such as the ECBI. In addition, Kaminski et al. (2008) found effect size for child outcomes to be related with outcome category. Measurements of externalizing behavior revealed larger effect sizes than measurements of social skills and prosocial behavior.

Second, assignment may affect differences between intervention and comparison groups. Random assignment should result in statistically equal groups and should control for all variables other than intervention that may interfere with causal conclusions about intervention effects. Hence, randomized controlled trials are expected to yield findings closer to the true effects than findings from studies using quasi-experimental or case-control designs. It has been suggested that these trials produce a minimum estimate of effect sizes, and may therefore yield smaller effect sizes than non-randomized studies (McCall & Green, 2004).

Third, differences in the nature of the comparison group may be related to differences in effect sizes: comparison with a non-treatment group might lead to larger effect sizes than comparison with a group that received alternative treatment (Wilson & Lipsey, 2001). However, Kaminski et al. (2008) found similar effects in studies of parent training interventions with various types of comparison groups.

### 1.5. The current study

The current study has two aims. The first aim is to examine the overall effectiveness of the IYPT with respect to child behavior, which includes both disruptive behavior and prosocial behavior, as well as behavior as observed by parents, teachers and observers. Given the wide dissemination of the IYPT and the growing number of IYPT studies, it would be useful to summarize the effects of this program quantitatively and separately from other parent training programs. Although meta-analyses regarding parent trainings in general or specific parenting programs like Parent-Child Interaction Therapy (PCIT) and Triple P – Positive Parenting Program (Triple P) exist (De Graaf, Speetjens, Smit, De Wolff, & Tavecchio, 2008; Nowak & Heinrichs, 2008; Thomas & Zimmer-Gembeck, 2007), there is no meta-analytic review to our knowledge specifically aimed at the IYPT.

The second aim is to examine variability in intervention outcomes and to determine whether intervention characteristics, child characteristics, family characteristics, and methodological features may explain differences in children's outcomes. Differential effectiveness regarding study context and standard IYPT versus variants of IYPT is examined first, because these are fundamentally distinct types of intervention, which may possibly not be meaningfully aggregated. Examination of the comparative influence of various study characteristics may yield critical information regarding the IYPT and, in that respect, may also contribute to knowledge in a field wishing to answer the question what works, when, and for whom, in parenting interventions (Brestan & Eyberg, 1998).

## 2. Method

### 2.1. Inclusion criteria

Studies had to meet the following inclusion criteria: (a) effects of the IYPT, as a stand-alone intervention or in a package with other

components or interventions, were examined immediately after intervention; (b) the effectiveness was examined by comparing an intervention group to a comparison group; (c) the study reported at least one quantitative measure of disruptive or prosocial child behavior, which was measured equally among participants; and (d) sufficient empirical data was reported to enable the calculation of standardized mean difference effect sizes or standardized mean difference effect sizes were reported in text. If several manuscripts regarding the same data were available, manuscripts with primary analyses and published manuscripts were preferred. There were no further restrictions on the types of research designs or comparison groups to be included in the meta-analytic review.

### 2.2. Literature search

A systematic and comprehensive search for studies was conducted for the period 1980 until April 2010. In total, 231 citations were checked for eligibility, after removing duplicates. Of the 231 citations surveyed: 86 citations were deemed irrelevant based on the title and abstract, 105 citations were deemed irrelevant based on the full text, 1 citation could not be retrieved, and 39 citations fulfilled inclusion criteria. Regarding citations deemed irrelevant, 132 citations (69.1%) were rejected primarily because they did not fulfill Criterion a; in 112 citations effectiveness was not examined, 13 citations did not encompass the IYPT, and in seven citations only long-term effectiveness was examined. Furthermore, 19 citations (9.9%) were primarily rejected because no comparison group was used (Criterion b), 12 citations (6.3%) because no quantitative measure of disruptive or prosocial child behavior was reported (Criterion c), and one citation (0.5%) because of insufficient data to calculate effect sizes (Criterion d). In addition, 27 citations (14.1%) were secondary analyses of relevant studies and did not provide any additional information.

Several approaches were used to identify the 39 relevant manuscripts. First, a large set of studies was retrieved by searches, using the search term “incredible years”, in the online data bases PsychINFO, Scopus, Web of Science, PubMed, Cochrane library, and ERIC. These searches, in which 95 studies were checked for eligibility, resulted in 16 manuscripts fulfilling inclusion criteria. Second, the Incredible Years library ([http://www.incredibleyears.com/Library/show\\_all.asp](http://www.incredibleyears.com/Library/show_all.asp)) was searched for eligible studies. This search, in which 173 studies were checked for eligibility, resulted in an additional 17 manuscripts. Third, bibliographies of retrieved and related studies were examined for eligible studies, resulting in five additional manuscripts. Fourth, a personal request for unpublished material and/or additional data sent by electronic mail to several researchers resulted in one additional manuscript.

Regarding this selection of 39 manuscripts, three manuscripts (Drugli & Larsson, 2006; Drugli, Larsson, & Clifford, 2007; Larsson et al., 2009) pertained to the same study, but each contained information about different relevant aspects. Therefore, these three manuscripts were combined and dealt with as one manuscript in our data. Furthermore, in 11 manuscripts two relevant studies could be distinguished, and in one manuscript three relevant studies. That is, manuscripts could contain multiple studies, because more than one IYPT condition could be compared to more than one comparison condition; each comparison of an IYPT condition to a comparison condition was considered a study. Therefore, 50 studies were included in the meta-analysis. See Table 1 for a list of included studies and study details.

### 2.3. Coding of study characteristics

Study characteristics of eligible studies were coded by six coders using a detailed coding schedule. Twenty-seven studies (54.0%) were coded by two independent coders. Pearson's *r* correlation coefficients were computed for continuous variables, and Cohen's *kappas* were calculated for categorical variables. Inter-rater reliability was sufficient. For continuous variables, all coefficients exceeded .78, with an average of

**Table 1**  
Articles included in the analyses.

Study	n <sup>a</sup>	Study context <sup>b</sup>	Standard IYPT	Initial severity/Assignment/ Comparison group <sup>c</sup>	Effect sizes				
					Disruptive	Prosocial	Parents	Teacher	Observation
August, Realmuto, Hektner, and Bloomquist (2001)	110, 84	Indicated	Yes	–/B/N	–0.02	–0.03	–0.09	0.04	–
Barrera et al. (2002)	123, 125	Indicated	Yes	–/R/N	0.14	–	0.07	0.12	0.34
Brotman et al. (2003)	16, 12	Selective	Yes	–/R/N	0.65	–	0.65	–	–
Brotman et al. (2005)	40, 38	Selective	Yes	51.0/R/N	–0.35	0.53	–	–	0.09
Connolly, Sharry, and Fitzpatrick (2001) <sup>d</sup>	27, 18	Treatment	Yes	65.6/N/W	0.15	–	0.15	–	–
Connolly et al. (2001) <sup>d</sup>	58, 18	Treatment	Yes	67.5/N/W	0.17	–	0.17	–	–
Cummings and Wittenberg (2008)	18, 17	Treatment	Yes	65.0/R/T	0.23	–	0.52	–	–0.34
Drugli and Larsson (2006) <sup>d,e</sup>	45, 28	Treatment	Yes	67.4/R/W	0.48	0.39	0.59	0.07	–
Drugli and Larsson (2006) <sup>d,e</sup>	52, 28	Treatment	Yes	67.3/R/W	0.39	0.55	0.44	0.47	–
Gardner et al. (2006)	39, 32	Treatment	Yes	66.2/R/W	0.59	–	0.49	–	0.78
Gross, Fogg, and Tucker (1995)	11, 6	Indicated	No	56.3/R/N	0.18	–	0.64	–	–0.74
Gross et al. (2003) <sup>d</sup>	52, 54	Selective	Yes	50.4/B/W	0.07	–	–0.11	0.20	0.29
Gross et al. (2003) <sup>d</sup>	54, 54	Selective	Yes	50.9/B/W	0.10	–	–0.04	0.11	0.37
Helfenbaum-Kun and Ortiz (2007)	7, 9	Selective	No	46.5/R/N	0.39	–	0.39	–	–
Hutchings et al. (2007)	104, 49	Indicated	Yes	63.6/B/W	0.48	–	0.55	–	0.26
Kim et al. (2008)	20, 9	Selective	Yes	43.9/R/N	0.30	0.45	0.35	–	–
Kratochwill et al. (2003) <sup>d</sup>	34, 11	Indicated	No	–/R/N	–0.05	0.19	0.05	0.01	–
Kratochwill et al. (2003) <sup>d</sup>	34, 14	Indicated	No	–/R/T	–0.42	–0.46	–0.83	–0.04	–
Lavigne et al. (2008) <sup>d</sup>	37, 31	Treatment	Yes	66.6/B/P	0.32	–	0.32	–	–
Lavigne et al. (2008) <sup>d</sup>	49, 31	Treatment	Yes	66.6/B/P	–0.06	–	–0.06	–	–
Letarte and Normandeau (2008)	35, 27	Treatment	Yes	55.6/–/T	0.41	–	0.41	–	–
Letarte et al. (2010)	26, 9	Selective	Yes	54.0/N/W	0.44	–	0.44	–	–
Linares et al. (2006)	68, 36	Selective	Yes	51.7/R/N	0.14	–	0.18	0.05	–
Lindsay et al. (2008) <sup>d</sup>	204, 356	Not classified	Yes	–/N/T	0.23	0.33	0.28	–	–
Lindsay et al. (2008) <sup>d</sup>	204, 501	Not classified	Yes	–/N/T	0.09	0.19	0.14	–	–
McIntyre (2008)	21, 23	Selective	Yes	62.6/R/W	0.26	–	0.26	–	–
Nilsen (2007)	11, 7	Selective	Yes	–/N/N	0.57	–	0.57	–	–
Patterson et al. (2002)	46, 50	Indicated	No	58.0/B/N	0.25	0.25	0.25	–	–
Raaijmakers et al., unpublished manuscript <sup>f</sup>	72, 72	Indicated	Yes	58.3/N/N	0.24	0.16	0.25	–0.19	0.40
Reid et al. (2007)	97, 89	Indicated	Yes	54.4/B/N	0.28	0.20	0.27	0.15	0.41
Scott et al. (2001)	90, 51	Treatment	Yes	73.5/B/W	0.67	–	0.53	–	–
Scott et al. (2010)	58, 51	Indicated	Yes	55.4/R/N	0.40	–	0.42	0.20	–
Spaccarelli et al. (1992) <sup>d</sup>	16, 16	Not classified	No	58.0/R/W	0.28	–	0.28	–	–
Spaccarelli et al. (1992) <sup>d</sup>	21, 16	Not classified	No	59.6/R/W	1.01	–	1.01	–	–
Taylor et al. (1998) <sup>d</sup>	15, 17	Treatment	Yes	61.6/B/W	0.43	0.14	0.70	0.12	–
Taylor et al. (1998) <sup>d</sup>	38, 32	Treatment	Yes	64.0/B/T	0.20	–0.28	0.51	–0.31	–
Webster-Stratton (1982)	16, 19	Not classified	No	55.5/R/W	0.55	–	0.45	–	0.62
Webster-Stratton (1984) <sup>d</sup>	13, 11	Treatment	No	64.9/R/W	0.83	0.04	0.92	–	0.56
Webster-Stratton (1984) <sup>d</sup>	13, 11	Treatment	No	67.8/R/T	–0.19	0.02	–0.23	–	–0.26
Webster-Stratton (1992)	57, 39	Treatment	No	67.2/R/W	0.48	0.40	0.57	0.34	0.31
Webster-Stratton (1998)	264, 130	Selective	Yes	55.2/B/N	0.31	0.17	0.12	0.06	0.44
Webster-Stratton and Hammond (1997) <sup>d</sup>	26, 22	Treatment	Yes	69.5/R/W	0.71	0.24	1.23	–0.47	0.36
Webster-Stratton and Hammond (1997) <sup>d</sup>	22, 22	Treatment	Yes	68.8/R/W	0.77	0.57	1.04	0.35	0.65
Webster-Stratton et al. (1988) <sup>d</sup>	27, 27	Treatment	Yes	67.4/R/W	0.71	0.28	0.74	0.57	0.64
Webster-Stratton et al. (1988) <sup>d</sup>	27, 27	Treatment	No	67.0/R/W	0.49	0.42	0.61	0.30	0.32
Webster-Stratton et al. (2001)	191, 81	Selective	Yes	48.2/B/N	0.27	0.46	0.05	0.28	0.43
Webster-Stratton et al. (2004) <sup>d</sup>	31, 26	Treatment	Yes	–/R/W	0.47	–0.09	0.24	0.15	0.44
Webster-Stratton et al. (2004) <sup>d</sup>	24, 26	Treatment	Yes	–/R/W	0.28	0.06	0.63	–0.07	0.26
Webster-Stratton et al. (2004) <sup>d</sup>	25, 26	Treatment	Yes	–/R/W	0.37	0.33	0.31	0.13	0.44
Williford and Shelton (2008)	50, 33	Indicated	No	–/N/N	0.55	–	0.38	0.72	–

Note. Dashes indicate that no data was reported. If a manuscript provided multiple studies, a listing for each study was included.

<sup>a</sup> The first value refers to the intervention group at post-intervention, the second value refers to the control group at post-intervention (uncorrected for double/triple sample inclusion).

<sup>b</sup> Selective = selective prevention, Indicated = indicated prevention.

<sup>c</sup> The first space provides information on initial severity of child behavior (T-score), the second space provides information on assignment (N = non-random assignment; B = random after blocking or matching; R = random assignment), and the third space provides information about the nature of the comparison group (N = receives nothing; W = waiting list; P = attention placebo; T = alternative treatment).

<sup>d</sup> Manuscript provided multiple studies.

<sup>e</sup> Including Drugli et al. (2007) and Larsson et al. (2009).

<sup>f</sup> See also Posthumus et al. (2012).

.96. For categorical variables, all kappas exceeded .73, with an average of .92. Disagreements among coders were resolved through consulting the studies and discussion between the first two authors.

### 2.3.1. Differential effectiveness: study context and standard IYPT

To allow examination of differential effectiveness of distinct types of intervention, we coded the study context and whether the intervention may be considered standard IYPT or a variant of IYPT. For *study context*, studies were coded, based on the classification of Mrazek and Haggerty (1994), as: treatment, selective prevention or indicated prevention. If studies could not be classified in any single category, studies remained

“not classified”. Studies were considered *standard IYPT* if at least eleven IYPT sessions were offered in a group format. Studies in which fewer than eleven sessions were offered or where sessions were delivered individually were considered variants of IYPT.

### 2.3.2. Moderators

**2.3.2.1. Descriptive characteristics.** We coded five characteristics, which broadly describe the study: Webster-Stratton, publication type, publication year, number of children, and number of families. In addition, we coded one characteristic related to the effect sizes: intention-to-treat.

Stem-and-Leaf Plots for Disruptive Behavior, Prosocial Behavior, Parental Report, Teacher Report, and Observation

Disruptive behavior		Prosocial behavior		Parental report		Teacher report		Observation	
Stem	Leaf	Stem	Leaf	Stem	Leaf	Stem	Leaf	Stem	Leaf
				-.8	3				
				-.7				-.7	4
				-.6				-.6	
				-.5				-.5	
-.4	2	-.4	6	-.4		-.4	7	-.4	
-.3	5	-.3		-.3		-.3	1	-.3	4
-.2		-.2	7	-.2	3	-.2		-.2	6
-.1	9	-.1	29	-.1	0	-.1	9	-.1	
-.0	156	-.0	145	-.0	368	-.0	46	-.0	
.0	79	.0	356889	.0	446	.0	04467	.0	9
.1	0335689	.1	448	.1	23568	.1	112244	.1	
.2	3335578889	.2	2289	.2	4445678	.2	007	.2	569
.3	11789	.3	146	.3	11478	.3	035	.3	01356
.4	013467789	.4	256	.4	124458	.4	7	.4	003344
.5	4578			.5	1224678	.5	7	.5	5
.6	57			.6	13459	.6		.6	145
.7	017			.7	3	.7	2	.7	8
.8	2			.8					
.9				.9	2				
1.0	1			1.0	13				
				1.1					
				1.2	3				

Fig. 1. Stem-and-leaf plots for disruptive behavior, prosocial behavior, parental report, teacher report, and observation.

Whether Dr. Webster-Stratton was involved as an author or co-author was coded to reflect whether the developer of the IYPT was involved. *Publication type* was coded to reflect whether the study was published as a journal article. *Publication year*, *number of children* and *number of families* at the start of the study were coded directly. *Intention-to-treat* reflected whether the effect sizes were based on intention-to-treat analyses.

**2.3.2.2. Intervention characteristics.** We coded four intervention characteristics: training components, number of IYPT sessions offered, total number of sessions offered to parents, and number of sessions attended. To obtain the variable *training components* three variables were coded: (1) whether the IYPT sessions were delivered in group format, (2) whether other components of the Incredible Years (IY) program were added to the IYPT, and (3) whether parents also received intervention elements other than the IYPT. These three variables were combined into five categories: (1) individual IYPT, (2) IYPT, (3) IYPT + other IY component(s), (4) IYPT + other component(s) (not IY), and (5) IYPT + other IY component(s) + other component(s) (not IY). The *number of IYPT sessions offered* was categorized as few (less than 11 sessions), normal (11 to 13 sessions) or many (more than 13 sessions). *Total number of sessions* was coded as a continuous variable, which included the number of IYPT sessions as well as other sessions delivered to parents (e.g., home visits). *Number of sessions attended* was also based on the total number of sessions, if possible.

**2.3.2.3. Child characteristics.** We coded five characteristics of the target child: age, gender, initial severity of child behavior (by coding the pre-treatment ECBI intensity score and Child Behavior Checklist [CBCL] externalizing score), the extent to which behavior was considered a problem, and clinical symptom levels. Mean *age* of children was coded directly or estimated by averaging the minimum and the maximum age reported. Percentage of boys was coded as a measure of *gender*. To be able to combine mean pre-treatment ECBI intensity scores and CBCL externalizing scores, where both reflect intensity of children's problem behavior, *T-scores* were coded or calculated based on ECBI (Eyberg & Pincus, 1999) or CBCL (Achenbach & Rescorla, 2001) norms. If both CBCL externalizing and ECBI intensity scores were available, ECBI intensity scores were preferred and included in the resulting

variable *initial severity of child behavior*. This was because the ECBI is a specific scale commonly used as an outcome measure for interventions targeting disruptive child behavior. Pre-treatment ECBI problem scores reflected the extent to which children's behaviors were considered a *problem* by their parents. As a measure of *clinical symptom levels*, the percentage ECBI scores indicating clinical range was coded.

**2.3.2.4. Family characteristics.** We coded four family characteristics: single parenthood, ethnic minority, education, and at-risk. Percentage of *single parenthood* and percentage of *ethnic minority* populations in the study were coded directly. If this latter percentage was not reported, it was estimated by adding percentages of participants from minority groups (other than Caucasian origin). Percentage of low educated mothers was coded as a measure of *education*. If maternal education was not specified, parental education was used as an estimate. Whether a sample could be considered *at-risk* was coded based on reported parental, contextual or family characteristics. Sample selection purely on grounds of behavioral problems did not count as at-risk.

**2.3.2.5. Methodological features.** We coded three methodological features. For the information used to assess behavior, we coded whether the ECBI was used as a measure of child behavior. For the comparison group, we coded: *assignment* and *nature of the comparison group*. Type of assignment was categorized as non-random assignment, random after blocking or matching, and random assignment. Nature of the comparison group was categorized as receives nothing, waiting list, attention placebo, and alternative treatment.

In addition, to quantify the quality of studies, we estimated overall study rigor using a 7-point scale covering six indicators of methodological strength, similar to Lundahl, Tollefson, and Lovejoy (2008). For group assignment, studies earned 2 points if groups were randomly assigned, 1 point if the groups were statistically equivalent despite using a nonrandom assignment, and no points if groups were not statistically equivalent or information about group assignment was missing. Studies earned 1 point for each of the following characteristics: (1) two or more measurement approaches, (2) well-known or standardized outcome measure, (3) use of the treatment manual reported,

(4) treatment fidelity was judged as adequate, and (5) means and standard deviations were reported.

#### 2.4. Coding of effect size statistics

We used the unbiased standardized mean difference effect size, or Cohen's *d*, as the measure of effect size (see [Lipsey & Wilson, 2001](#)). Effect sizes were calculated for all outcome measures regarding disruptive or prosocial child behavior. That is, multiple effect sizes per study could be calculated. When multiple effect sizes concerning a single outcome were available, these effect sizes were averaged into a single mean effect size for the study. However, effect sizes concerning different outcomes (e.g., parental report and teacher report) were calculated separately and used in separate analyses. Decreases in disruptive behavior and increases in prosocial behavior in the intervention group, relative to the comparison group, resulted in positive effect sizes, whereas increases in disruptive behavior and decreases in prosocial behavior resulted in negative effect sizes.

When manuscripts provided more than one relevant study, multiple comparisons with one sample were possible. For instance, in a study with two relevant IYPT conditions, both IYPT conditions were compared to the same comparison group. To avoid double/triple counting of participants and samples contributing too much to the effect size mean, the number of participants in each sample was divided by the number of occasions that this sample was included in the meta-analysis for standard errors and inverse variance weights.

#### 2.5. Procedure and statistical analysis

Two meta-analytic questions were studied. First, we tested whether the overall effect sizes were significantly larger than zero for all outcome constructs. To avoid manifold use of studies' effect sizes within one meta-analysis, we distinguished five separate outcome constructs, which were analyzed in separate analyses and based on the mean effect size of possibly several effect sizes within a study. Two outcome constructs, disruptive behavior and prosocial behavior, reflected targeted child behavior. Three outcome constructs reflected commonly used informants: parental report, teacher report, and observation. The robustness of effects was tested by calculating fail-safe numbers. For all outcome constructs, overall effect sizes for studies which did not compare the IYPT to alternative treatment, but to a waiting list condition or no treatment, were reported alongside the main analysis. These separate analyses were conducted because these alternative treatments may have been effective too; exclusion of these studies resulted in a more uniform comparison, in which the IYPT was compared to comparison groups in which no intervention effects were to be expected.

Second, the distribution of heterogeneous sets of studies was analyzed. Whether variability in effect sizes beyond subject-level sampling error can be explained by moderator variables was tested.

### 3. Results

#### 3.1. Study characteristics

Fifty studies with 4745 participants (2472 for intervention groups and 2273 for comparison groups) were included. Twenty-two studies (44%) were identified as treatment studies, 12 studies (24%) as selective prevention, 11 studies (22%) as indicated prevention, and 5 studies (10%) could not be classified. Thirty-seven studies (74%) were considered standard IYPT.

In 17 studies (34%) the IYPT's developer, Dr. Webster-Stratton, was an author or co-author. Most studies (92%) were published journal articles, and most of them (72%) appeared after the year 2000. Thirty percent of studies contained original sample sizes of more than 100 children and families. In 32% of the studies intention-to-treat analyses were conducted.

In most studies (60%) the IYPT was delivered in a group format, without adding other components. In four studies (8%) an individual format was used. In nine studies (18%) only IY components were added to the IYPT, while in two studies (4%) other components were also added to the IY components. Because of the small set of studies that evaluated both additional IY components and additional other components, the addition of other components was ignored in further analyses, i.e. both categories were treated as IYPT + other IY components. In five studies (10%) only other components were added to the IYPT. In 19 studies (38%) 11 to 13 IYPT sessions were offered, while fewer sessions were offered in 11 studies (22%) and more sessions were offered in 20 studies (40%). The total number of sessions offered ranged from 4 to 60. The mean number of sessions attended by parents ranged from 4 to 28.

The mean age of the child sample ranged from 3 to 9.2 years. The percentage of boys ranged from 38.9 to 90.5%, most studies (95.6%) included more boys than girls. Mean *T*-scores regarding children's problem behavior ranged from 43.9 to 73.5, while mean pre-treatment ECBI problem scores ranged from 7.2 to 22. Nine out of the 19 studies reporting the percentage of children with ECBI scores indicating clinical range, reported clinical symptom level for all participating children.

The percentage of single parenthood ranged from 0% to 79.1% ( $M = 38.6$ ). The percentage of participants from minority groups ranged from 0.8% to 100% ( $M = 44.7$ ). The percentage of low educated mothers ranged from 0% to 60.6% in the 18 studies that reported maternal education. Eighteen studies (36%) concerned at-risk samples.

To assess behavior, 36 studies (72%) used the ECBI. Assignment was random in 28 studies (56%), random after blocking or matching in 13 studies (26%), and non-random in eight studies (16%). The intervention condition was compared to a waiting list condition in 24 studies (48%). The comparison group received nothing in 17 studies (34%) and received an alternative treatment in seven studies (14%). The overall study rigor ranged from 2 to 7 on the 7-point scale ( $M = 6.0$ ;  $SD = 1.2$ ), with 22 studies rated as the most rigorous study designs.

#### 3.2. Relations between study characteristics

Study characteristics were not distributed independently over studies. The descriptive characteristics number of children and number of families were highly interrelated. Therefore, only number of children was included in further analyses.

Study context was strongly positively related to the child characteristics problem, clinical symptom level, and initial severity of child behavior (which also were highly interdependent). Furthermore, study context was strongly negatively related to the family characteristics ethnic minority and at-risk (which were also highly interdependent). Given several variables were strongly related to study context and because of the theoretical relevance of this distinction between prevention and treatment studies, we chose to examine effect sizes along this distinction, and to exclude study context from moderator analyses. To control for these confounding variables in further moderator analyses, we first included initial severity of child behavior (which was also negatively related to minority and at-risk) in moderator analyses, and excluded problem and clinical symptom level. For family characteristics, we included ethnic minority, and excluded at-risk because of theoretical considerations.

The intervention characteristic total number of sessions was highly positively related to number of sessions attended. Because of theoretical relevance, we included number of sessions attended and excluded total number of sessions.

The methodological feature 'assignment' was related to the child characteristic clinical symptom levels; lower clinical symptom levels were to be found in studies that used random assignment after blocking or matching. Given we controlled for initial severity of child behavior, which was highly related to clinical symptom levels, inclusion of assignment yielded no apparent problems.

### 3.3. Analyses

#### 3.3.1. Overall effect sizes

Stem-and-leaf plots of weighted effect sizes for disruptive behavior, prosocial behavior, parental report, teacher report, and observation are displayed in Fig. 1. The distributions of effect sizes evidenced no extreme deviations from the normal distribution or outliers.

The overall weighted effect size (across informants) of the set of 50 studies concerning disruptive behavior was .27 (95% CI = .21–.34,  $p < .001$ ). Effect sizes ranged from  $-.42$  to  $1.01$ . Calculation of the fail-safe number revealed that 1351 additional studies with non-significant or adverse results have to exist in order to reduce the overall effect size for disruptive behavior to a statistically non-significant overall effect, with a conventional significance level of .05. Exclusion of studies in which the IYPT was compared to alternative treatment revealed an overall weighted effect size of .31 (95% CI = .24–.38,  $p < .001$ ; fail-safe number = 1142) for the set of 43 studies concerning disruptive behavior.

The overall weighted effect size of the set of 26 studies concerning prosocial behavior was .23 (95% CI = .15–.31,  $p < .001$ ). Effect sizes ranged from  $-.46$  to  $.57$ . Calculation of the fail-safe number revealed that 300 additional studies with non-significant or adverse results have to exist in order to reduce the overall effect size for prosocial behavior to a statistically non-significant effect. Exclusion of studies in which the IYPT was compared to alternative treatment revealed an overall weighted effect size of .25 (95% CI = .16–.34,  $p < .001$ ; fail-safe number = 195) for the set of 21 studies concerning prosocial behavior.

The overall weighted effect size of the set of 49 studies concerning parental report was .30 (95% CI = .22–.39,  $p < .001$ ). Effect sizes ranged from  $-.83$  to  $1.24$ . Calculation of the fail-safe number revealed that 1207 additional studies with non-significant or adverse results have to exist in order to reduce the overall effect size for parents to a statistically non-significant overall effect. Exclusion of studies in which the IYPT was compared to alternative treatment revealed an overall weighted effect size of .28 (95% CI = .21–.35,  $p < .001$ ; fail-safe number = 912) for the set of 42 studies concerning parental report.

The overall weighted effect size of the set of 25 studies concerning teacher report was .13 (95% CI = .05–.22,  $p = .001$ ). Effect sizes ranged from  $-.47$  to  $.72$ . Calculation of the fail-safe number revealed that 71 additional studies with non-significant or adverse results have to exist in order to reduce the overall effect size for teacher report to a statistically non-significant overall effect. Exclusion of studies in which the IYPT was compared to alternative treatment revealed an overall weighted effect size of .15 (95% CI = .06–.23,  $p = .001$ ; fail-safe number = 78) for the set of 23 studies concerning teacher report.

The overall weighted effect size of the set of 23 studies concerning observation was .37 (95% CI = .29–.46,  $p < .001$ ). Effect sizes ranged from  $-.74$  to  $.78$ . Calculation of the fail-safe number revealed that 576 additional studies with non-significant or adverse results have to exist in order to reduce the overall effect size for observation to a statistically non-significant overall effect. Exclusion of studies in which the IYPT was compared to alternative treatment revealed an overall weighted effect size of .39 (95% CI = .30–.48,  $p < .001$ ; fail-safe number = 564) for the set of 21 studies concerning observation.

Significant heterogeneity in effect sizes was only found for parental report ( $Q(48) = 70.68, p = 0.02$ ), but not for the other four outcome constructs. Therefore, moderator analyses were only pursued for parental report. Overall study rigor was not significantly related to any of the five overall effect sizes.

#### 3.3.2. Differential effectiveness: study context and standard IYPT

Regarding parental report, we calculated overall effect sizes for treatment and prevention studies (i.e., study context) and for studies considered standard IYPT or variants of IYPT (see Table 2). Effect sizes differed significantly for study context. Study context explained

**Table 2**  
Differential effectiveness.

Study context	Standard IYPT		IYPT variant		All studies	
	<i>k</i>	<i>d</i>	<i>k</i>	<i>d</i>	<i>k</i>	<i>d</i>
Selective prevention	10	.13	1	.39	11	.13
Indicated prevention	6	.21	5	.10	11	.20
Treatment	18	.50	4	.54	22	.50
Not classified	2	.21	3	.58	5	.25
All studies	36	.25	13	.35	49	.30

significant variability ( $Q_B(2) = 19.65; p < .001$ ) and accounted for sufficient excess variability ( $Q_W(41) = 45.76; p = .28$ ). However, analog to ANOVA analyses revealed no differences in effect sizes between standard IYPT and variants of IYPT.

#### 3.3.3. Moderators of effect size for parental report

Weighted regression analyses were conducted to test potential moderators regarding parental report. First, all potential moderators were

**Table 3**  
Potential predictors entered separately and together with initial severity of child behavior.

Potential moderator	Entered separately			Initial severity of child behavior entered first		
	<i>p</i>	<i>B</i>	$\beta$	<i>p</i>	<i>B</i>	$\beta$
<i>Descriptive characteristics</i>						
Webster-Stratton	.97	-.00	-.00	.99	-.00	-.00
Publication type	.55	.05	.07	.98	-.00	-.00
Publication year	.18	-.01	-.16	.53	-.00	-.09
Number of children	.17	-.00	-.16	.07 <sup>†</sup>	-.00	-.29
Intention-to-treat	.69	-.03	-.05	.36	-.08	-.14
<i>Intervention characteristics</i>						
Training components						
Individual IYPT <sup>a</sup>	.80	-.04	-.03	.65	.08	.07
IYPT + other IY component(s) <sup>a</sup>	.03*	-.15	-.26	.68	.04	.07
IYPT + other components(s) (not IY) <sup>a</sup>	.31	.13	.12	.15	.21	.22
Number of IYPT sessions offered						
Few <sup>b</sup>	.77	.03	.04	.71	.04	.06
Many <sup>b</sup>	.83	-.02	-.03	.28	.10	.19
Number of sessions attended	<.001***	.03	.55	.04**	.03	.37
<i>Child characteristics</i>						
Age	.65	-.01	-.05	.25	.04	.17
Gender	.005**	.01	.35	.16	.01	.27
Initial severity of child behavior	<.001***	.02	.62	na		
<i>Family characteristics</i>						
Single parenthood	.08 <sup>†</sup>	-.00	-.22	.35	-.00	-.15
Ethnic minority	.11	-.00	-.21	.93	.00	.02
Education	.32	.00	.22	.76	-.00	-.09
<i>Methodological features</i>						
ECBI	.02*	.15	.27	.09 <sup>†</sup>	.26	.29
Assignment						
Non-random <sup>c</sup>	.09 <sup>†</sup>	-.14	-.23	.11	-.24	-.26
Random after blocking or matching <sup>c</sup>	.008**	-.20	-.36	.01*	-.22	-.40
Nature of the comparison group						
Waiting list <sup>d</sup>	<.001***	.31	.52	.61	.06	.11
Alternative treatment <sup>d</sup>	.67	.03	.05	.85	.03	.03

Note: na = not applicable.

<sup>a</sup> IYPT is the reference group.

<sup>b</sup> Normal is the reference group.

<sup>c</sup> Random is the reference group.

<sup>d</sup> Receives nothing is the reference group.

<sup>†</sup>  $p < .10$ .

\*  $p < .05$ .

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .



entered separately. As shown in Table 3, seven potential moderators were significant predictors of effect size. Two intervention characteristics (training components and number of sessions attended), two child characteristics (gender and initial severity of child behavior), and three methodological features (ECBI, assignment, and nature of the comparison group) were predictors of effect size, when entered separately. None of the descriptive characteristics or family characteristics produced significant regression coefficients, although a trend was found for the family characteristic single parenthood.

Second, the child characteristic initial severity of child behavior was entered in the weighted regression analysis together with potential moderators, to control for confounding variables and to study relative contribution of variables. As shown in Table 3, the intervention characteristic number of sessions attended and the methodological feature assignment remained significant predictors of effect sizes. Furthermore, trends were found for the descriptive characteristic number of children and the methodological feature ECBI.

Third, predictors that remained significant in the second step were entered simultaneously in a fixed effects weighted regression analysis. Hence, the final model, as shown in Table 4, consisted of the intervention characteristic number of sessions attended, the child characteristic initial severity of child behavior, and the methodological feature assignment. These predictors explained 68% of variability in effect sizes, in which initial severity of child behavior explained most variance in effect sizes.

#### 4. Discussion

The results from our meta-analyses show that the IYPT is an effective intervention regarding child behavior as measured immediately after intervention. Positive effects were found for disruptive behavior ( $d = .27$ ) and prosocial behavior ( $d = .23$ ). Mean effect sizes based on observations ( $d = .37$ ) were larger than mean effect sizes based on parents' ( $d = .30$ ) or teachers' judgments ( $d = .13$ ). Furthermore, parent-rated effect sizes differed between treatment studies ( $d = .50$ ) and prevention studies ( $d = .13$  for selective prevention;  $d = .20$  for indicated prevention). Intervention characteristics, child characteristics, and methodological features explained variability in parent-rated intervention outcomes of the IYPT. Pre-treatment intensity of children's problem behavior proved to be the strongest predictor of the IYPT's intervention effects on parental report, with larger effects for studies which included more severe cases.

##### 4.1. Overall effect sizes

Analyses of 50 IYPT studies revealed that the IYPT is effective in diminishing disruptive behavior and increasing prosocial behavior, according to parents, teachers, and observers. Moreover, large fail-safe numbers suggest that these results are robust against the possibility of

missing studies. The effectiveness of the IYPT is in line with the effectiveness of BPT in general (e.g., *McCart et al., 2006*), and is also in line with the presumed effectiveness of the IYPT in specific as put forward in previous narrative reviews (e.g., *Bauer & Webster-Stratton, 2006; Brestan & Eyberg, 1998*). Moreover, the parent-rated effect size for IYPT treatment studies ( $d = .50$ ) is higher than the parent-rated effect size for BPT treatment studies in general ( $d = .38$ ) reported by *McCart et al. (2006)*. Hence, our results put numerical weight behind the presumed effectiveness of this specific BPT program.

This meta-analytic review shows the IYPT to meet criteria of a well-established intervention. The effectiveness of the IYPT is underscored by studies comparing the IYPT with alternative treatments (e.g., *Cummings & Wittenberg, 2008; Taylor, Schmidt, Pepler, & Hodgins, 1998; Webster-Stratton, 1984*), independent replications (e.g., *Drugli & Larsson, 2006; Gardner et al., 2006; Scott et al., 2001*), a treatment manual (*Webster-Stratton, 2001*), and specification of participant characteristics in individual studies. Since these are criteria for well-established treatments (*Chambless & Ollendick, 2001*), the IYPT seems to currently meet these criteria.

The effects for disruptive behavior and prosocial behavior were highly similar. This result is somewhat contrary to results of *Kaminski et al. (2008)*, who found larger effect sizes for externalizing behaviors than for social skills and social competence. However, it is unclear to what extent operationalizations of this study's prosocial behavior and *Kaminski et al.'s (2008)* social skills and social competence overlap. Conceptually, social competence seems like an umbrella term, and likewise social skills incorporate other skills besides prosocial skills. However, this study's operationalization of prosocial behavior may be broader than intended, and may include other behaviors besides helping behavior. Therefore, more conceptual clarity regarding prosocial behavior and more focused measures of prosocial behavior are needed in future research (*Eisenberg & Fabes, 1998*).

The effects for teacher report were smaller than for parental report and observation. Discrepancies between parents' and teacher's ratings are well-known, and may be caused by genuine contextual differences and more similarity in criteria as used by different teachers than in criteria as used by different parents (*Scott, 2001*), or insufficient generalization of the intervention effect from home to school settings. However, similarity to effects based on observation, which is considered the 'gold standard', supports effectiveness as rated by parents.

##### 4.2. Differential effectiveness

For parent-rated outcomes, analyses revealed larger effect sizes in treatment studies when compared to prevention studies, but no differences between standard IYPT and variants of IYPT. Treatment studies differ from prevention studies in several respects, all of which may contribute to the established difference in effectiveness. Most likely, differences in initial levels of problem behavior contribute highly to this difference. For prevention studies, the relatively low initial levels of problem behavior may leave less room for change during the intervention period (e.g., *Kaminski et al., 2008*). Interdependence between treatment studies and initial levels of problem behavior, and initial level of problem behavior as the strongest predictor of parent-rated effectiveness are in line with this assumption. However, there may be differences between treatment and prevention studies, which were not investigated in this meta-analytic review. For example, setting and motivational issues may also play a role. Larger effect sizes were found for studies conducted in clinical settings than studies in non-clinical settings (*McCart et al., 2006*). Furthermore, parents in treatment studies are probably more in need of help than parents in prevention studies, because they typically seek help themselves (*Mrazek & Haggerty, 1994*). As a consequence, these parents may be more motivated to accept help and modify their own behaviors. Since motivation, or intention to attend sessions, may also be related to attendance (*Sheeran,*

**Table 4**  
Significant univariate predictors, when entered simultaneously.

Predictor	<i>p</i>	<i>B</i>	$\beta$
Number of sessions attended	.11	.03	.36
Initial severity of child behavior	.01*	.02	.47
Assignment			
Non-random <sup>a</sup>	.13	-.29	-.28
Random after blocking or matching <sup>a</sup>	.14	-.17	-.28
Model		$Q(4) = 27.10, p < .001$	
Residual		$Q(19) = 12.89, p = .84$	
R-square		.68	

<sup>a</sup> Random is the reference group.

\*  $p < .05$ .

Aubrey, & Kellett, 2007), higher attendance and treatment dosage are expected characteristics in highly motivated parents.

Interestingly, no differences were found between standard IYPT and variants of IYPT. It is still unclear why variants of IYPT and standard IYPT yielded similar effect sizes. Possibly, variants of IYPT are as effective as standard IYPT, because they are both based on the same principles. However, classification based on other aspects of the IYPT might reveal meaningful differences in effectiveness.

#### 4.3. Moderators of effect size

Our results shed light on which characteristics influence children's outcomes, as reported by their parents, immediately after delivery of the IYPT to their parents. Initial severity of problem behavior was revealed to be the strongest predictor of the IYPT's intervention effects. Although the IYPT is most effective in severe cases, there might be reasons to offer the IYPT to populations with less severe behavior problems.

Moderator analyses also revealed some less powerful predictors. Two intervention characteristics were related to intervention effects when entered separately. First, used training components were related to effect; the addition of other IY components was related to smaller effect sizes compared to the IYPT without other components. However, when initial severity of behavior was taken in account, training components were no longer predictive. Therefore, this association seems to be caused by a relatively low number of studies with severe samples in which other IY components are added to the IYPT. Indeed, for example, classroom-based components are likely to be added in prevention studies, which will include children with relatively few behavior problems. Second, the number of sessions attended by parents was positively related to intervention effects according to parents, even when initial severity of behavior was taken into account. Although we could obviously not examine causal relations, it seems likely that higher treatment dosages lead to more responses. However, in theory, it is also possible that this relation is affected by parental perceptions. Parents who experience improvements are most likely to continue attending and their perceptions of improvements might be influenced by their attendance and efforts (avoidance of cognitive dissonance).

Three methodological features were related to intervention effects when entered separately. First, studies in which the ECBI was used revealed relatively large improvements in child behavior as reported by parents. When initial severity of child behavior was taken into account, a trend remained. As suggested by Scott (2001), the ECBI may yield relatively large effect sizes because the ECBI is a specific scale concentrating on behaviors likely to change during treatment. Second, the type of assignment was related to intervention outcomes. Random assignment after blocking or matching resulted in smaller effect sizes compared with random assignment without blocking or matching. Probably, this association is caused by relatively large selective prevention studies within the set of studies with random assignment after blocking or matching. These studies had relatively small effect sizes that were weighted considerably due to their large sample sizes. Furthermore, a trend for smaller effects in non-randomized studies was found, when type of assignment was entered separately. Third, for the nature of the comparison group, comparison to participants on a waiting list yielded larger effect sizes than comparison to participants who received no intervention. The association was not found when initial severity of behavior was taken into account. This is probably caused by relatively severe problems in studies using a waiting list, because waiting lists are preferred over non-treatment in samples that are in need for help.

Possibly unexpected, none of the descriptive or family characteristics proved to be a predictor of effectiveness. The descriptive characteristics, particularly Dr. Webster-Stratton's involvement as an author, being unrelated to effect sizes might be surprising. Inflated effect sizes in developers' evaluation studies have been suggested (e.g., Eisner, 2009), because of developers' involvement and interests. A possible

explanation for our finding might be that treatment fidelity is generally high. Larger effects reported by developers evaluating their own programs might occur because they are able to achieve high fidelity (Petrosino & Soydan, 2005). Although differences in treatment fidelity may exist between studies (see below), the developer's firm recommendation to have group leaders accredited for conducting studies and to let them receive consultation is suggestive of high fidelity in IYPT studies. For example, authorized three day IY workshops, a group leader certification/accreditation process, a detailed treatment manual, and checklists may ensure a high level of treatment fidelity (Webster-Stratton, 2004). Current findings do not suggest inflated effect sizes for a developer-as-evaluator implying that developers' studies should not be disregarded routinely because of presumed systematic bias. Instead, the extent of systematic bias should be examined for individual programs.

Also family characteristics, such as ethnic minority status and single parenthood, not being related to intervention effects might be surprising in light of the numerous studies suggesting that high-risk families may benefit less from parenting interventions (e.g., Assemany & McIntosh, 2002; Kazdin, 1997; Nixon, 2002; Reyno & McGrath, 2006). Family characteristics may be unrelated to the outcomes of the IYPT for several reasons. First, the IYPT might have a capacity to be tailored to specific characteristics and needs of families, in spite of being a manualized group training. For example, group leaders can achieve flexible applications of the manual and help parents learn to use the parenting principles to achieve their own determined goals (Webster-Stratton, 2009). When the IYPT is tailored in this way differences between families might become less important. Second, delivery using a group format might have made differences between families less pronounced. Group discussions during parenting groups might normalize experiences of parents and might lead to the discovery that they have many issues in common (Webster-Stratton, 2009). Third, this finding might also be an artifact of coding. Manuscripts do not uniformly report about family characteristics like SES, parental depression, and maternal education. As a consequence, we could not include SES and parental depression and chose to reflect risk status in a broader way. Furthermore, a small number of studies could be inserted in analyses regarding, for example, maternal education. Broad measures and missing values might worsen the possibility to detect moderating effects.

#### 4.4. Limitations

An important limitation of this meta-analytic review pertains to training quality. Authorized three-day IY workshops and a group leader certification/accreditation process are offered to maximize the performance of the group leaders (Incredible Years, 2011). However, it is unclear to what extent training quality differs between individual studies. Unfortunately, it is often difficult to trace information about trainers' individual characteristics and treatment fidelity in individual studies, and to quantify this information for meta-analyses.

For some variables the extent of missing data was considerable. Therefore, we could not include all potentially interesting moderators in moderator analyses and low statistical power may have hindered the detection of moderator effects regarding other variables. This absence of desired information in manuscripts is common in meta-analytic reviews and has therefore also been commented on by other authors (e.g., Kaminski et al., 2008; McCart et al., 2006).

#### 4.5. Future research

Similar to other BPT programs, a premise of the IYPT is that children's functioning is influenced by parenting behaviors, and that modifying parenting behaviors will result in long-lasting changes for children and their parents. Although this meta-analytic review addresses an important part of this chain, other parts were not examined and left open for further research. First, we did not examine long-term

effects of the IYPT. Examination of follow-up studies may elucidate durability of effects, but may also be hampered by an absence of follow-up data collected in comparison groups. Since a waiting list is the most common method across IYPT studies to constitute a comparison group long-term between-group comparisons would be unfeasible in most cases. Nonetheless, examination of long-term effects by within-group comparisons is also recommended. Understanding the long-term effectiveness of the IYPT is important to optimize maintenance of child behavior change, with the view to alter children's pathways to antisocial and delinquent behaviors amongst other things.

Second, although the IYPT's effects on child behavior are probably mediated by modifications in parenting behaviors, parenting behaviors have not been examined in this meta-analytic review. We did not examine whether parenting behaviors improved and whether modifications in child behavior are influenced by modifications in parenting behaviors. However, individual mediation studies regarding the IYPT suggest that parenting domains such as critical, harsh parenting, inconsistent discipline, verbal criticism, positive parenting, responsive parenting, and stimulating parenting (Beauchaine et al., 2005; Brotman et al., 2009; Fossum et al., 2009; Gardner et al., 2006, 2010; Letarte, Normandeau, & Allard, 2010; Posthumus et al., 2012), mediate outcomes. Furthermore, ample evidence for modifications in parenting behaviors can be found in individual IYPT studies. The IYPT resulted in improvements in parenting domains such as harsh parenting, inconsistent discipline, positive parenting, responsiveness, and overall parenting skills (Brotman et al., 2003, 2005; Gardner et al., 2006; Gross et al., 2003; Hutchings et al., 2007; Kim, Cain, & Webster-Stratton, 2008; Larsson et al., 2009; Letarte et al., 2010; Linares et al., 2006; Posthumus et al., 2012; Scott et al., 2010; Webster-Stratton, 1984, 1992, 1998; Webster-Stratton & Hammond, 1997; Webster-Stratton, Reid, & Hammond, 2001, 2004; Webster-Stratton et al., 1988). Nonetheless, the overall effectiveness of the IYPT regarding parenting behaviors and characteristics that influence its effectiveness regarding these parenting behaviors remain interesting questions for further research.

#### 4.6. Implications

The results of this meta-analytic review might assist effective decision-making regarding intervention options. Positive results found in this meta-analytic review, combined with the need for evidence-based programs (e.g., Dodge, 2011; Eyberg et al., 2008), imply that the IYPT may be interesting for policymakers, agencies, and practitioners to adopt as an intervention. Homogeneity regarding four out of five outcome constructs indicates that the outcomes of IYPT studies are comparable and that, given the broad range of populations included within individual studies, the IYPT can be used successfully in a diverse range of families. Moreover, positive outcomes regarding different outcome constructs and with the inclusion of a relatively large number of replication studies, emphasize the effectiveness of the IYPT. It seems possible to reduce the potential for child harm as well as to increase the potential for child success, for many kinds of families in diverse contexts.

#### References<sup>1</sup>

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: Research Center for Children, Youth, and Families, University of Vermont.
- Assemany, A. E., & McIntosh, D. E. (2002). Negative treatment outcomes of behavioral parent training programs. *Psychology in the Schools*, 39(2), 209–219.
- \*August, G. J., Realmuto, G. M., Hektner, J. M., & Bloomquist, M. L. (2001). An integrated components preventive intervention for aggressive elementary school children: The Early Risers program. *Journal of Consulting and Clinical Psychology*, 69(4), 614–626. <http://dx.doi.org/10.1037/0022-006X.69.4.614>.
- \*Barrera, M. J., Biglan, A., Taylor, T. K., Gunn, B. K., Smolkowski, K., Black, C., et al. (2002). Early elementary school intervention to reduce conduct problems: A randomized trial with Hispanic and non-Hispanic children. *Prevention Science*, 3(2), 83–94. <http://dx.doi.org/10.1023/A:1015443932331>.
- Bauer, N. S., & Webster-Stratton, C. (2006). Prevention of behavioral disorders in primary care. *Current Opinion in Pediatrics*, 18(6), 654–660. <http://dx.doi.org/10.1097/MOP.0b013e3280106239>.
- Baydar, N., Reid, M. J., & Webster-Stratton, C. (2003). The role of mental health factors and program engagement in the effectiveness of a preventive parenting program for Head Start mothers. *Child Development*, 74(5), 1433–1453. <http://dx.doi.org/10.1111/1467-8624.00616>.
- 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(3), 371–388. <http://dx.doi.org/10.1037/0022-006X>.
- Bennett, K. J., Lipman, E. L., Racine, Y., & Offord, D. R. (1998). Annotation: Do measures of externalising behaviour in normal populations predict later outcome? Implications for targeted interventions to prevent conduct disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 39(8), 1059–1070. <http://dx.doi.org/10.1111/1469-7610.00411>.
- 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(2), 180–189. [http://dx.doi.org/10.1207/s15374424jccp2702\\_5](http://dx.doi.org/10.1207/s15374424jccp2702_5).
- \*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(4), 724–734. [http://dx.doi.org/10.1207/s15374424jccp3404\\_14](http://dx.doi.org/10.1207/s15374424jccp3404_14).
- Brotman, L. M., Gouley, K. K., Huang, K., Rosenfelt, A., O'Neal, C., Klein, R. G., et al. (2008). Preventive intervention for preschoolers at high risk for antisocial behavior: Long-term effects on child physical aggression and parenting practices. *Journal of Clinical Child and Adolescent Psychology*, 37(2), 386–396. <http://dx.doi.org/10.1080/15374410801955813>.
- \*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 Clinical Child and Adolescent Psychology*, 32(2), 246–257. [http://dx.doi.org/10.1207/s15374424jccp3202\\_10](http://dx.doi.org/10.1207/s15374424jccp3202_10).
- Brotman, L. M., O'Neal, C. R., Huang, K., Gouley, K. K., Rosenfelt, A., & Shrout, P. E. (2009). An experimental test of parenting practices as a mediator of early childhood physical aggression. *Journal of Child Psychology and Psychiatry*, 50(3), 235–245. <http://dx.doi.org/10.1111/j.1469-7610.2008.01929.x>.
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, 685–716. <http://dx.doi.org/10.1146/annurev.psych.52.1.685>.
- \*Connolly, L., Sharry, J., & Fitzpatrick, C. (2001). Evaluation of a group treatment programme for parents of children with behavioural disorders. *Child and Adolescent Mental Health*, 6(4), 159–165. <http://dx.doi.org/10.1111/1475-3588.00340>.
- \*Cummings, J. G., & Wittenberg, J. (2008). Supportive expressive therapy–parent child version: An exploratory study. *Psychotherapy: Theory, Research, Practice, Training*, 45(2), 148–164. <http://dx.doi.org/10.1037/0033-3204.45.2.148>.
- De Graaf, I., Speetjens, P., Smit, F., De Wolff, M., & Tavecchio, L. (2008). Effectiveness of the Triple P Positive Parenting Program on parenting: A meta-analysis. *Family Relations*, 57(5), 553–566. <http://dx.doi.org/10.1111/j.1741-3729.2008.00522.x>.
- Dodge, K. A. (2011). Context matters in child and family policy. *Child Development*, 82(1), 433–442. <http://dx.doi.org/10.1111/j.1467-8624.2010.01565.x>.
- \*Drugli, M. B., & Larsson, B. (2006). Children aged 4–8 years treated with parent training and child therapy because of conduct problems: Generalisation effects to day-care and school settings. *European Child & Adolescent Psychiatry*, 15(7), 392–399. <http://dx.doi.org/10.1007/s00787-006-0546-3>.
- \*Drugli, M. B., Larsson, B., & Clifford, G. (2007). Changes in social competence in young children treated because of conduct problems as viewed by multiple informants. *European Child & Adolescent Psychiatry*, 16(6), 370–378. <http://dx.doi.org/10.1007/s00787-007-0609-0>.
- Eisenberg, N., & Fabes, R. A. (1998). Prosocial development. In W. Damon, & N. Eisenberg (Eds.), *Handbook of child psychology. Social, emotional, and personality development*, Vol. 3. (pp. 701–778). New York: Wiley.
- Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, 5(2), 163–183. <http://dx.doi.org/10.1007/s11292-009-9071-y>.
- 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 and Adolescent Psychology*, 37(1), 215–237. <http://dx.doi.org/10.1080/15374410701820117>.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Fossum, S., Mørch, W., Handegard, B. H., Drugli, M. B., & Larsson, B. (2009). Parent training for young Norwegian children with ODD and CD problems: Predictors and mediators of treatment outcome. *Scandinavian Journal of Psychology*, 50(2), 173–181. <http://dx.doi.org/10.1111/j.1467-9450.2008.00700.x>.
- Foster, E. M., Olchowski, A. E., & Webster-Stratton, C. H. (2007). Is stacking intervention components cost-effective? An analysis of the Incredible Years program. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46(11), 1414–1424. <http://dx.doi.org/10.1097/chi.0b013e3181514c8a>.
- \*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 of change. *Journal of Child Psychology and Psychiatry*, 47(11), 1123–1132. <http://dx.doi.org/10.1111/j.1469-7610.2006.01668.x>.

<sup>1</sup> References marked with an asterisk indicate studies included in the meta-analyses.

- Gardner, F., Hutchings, J., Bywater, T., & Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child and Adolescent Psychology*, 39(4), 568–580. <http://dx.doi.org/10.1080/15374416.2010.486315>.
- \*Gross, D., Fogg, L., & Tucker, S. (1995). The efficacy of parent training for promoting positive parent–toddler relationships. *Research in Nursing & Health*, 18(6), 489–499. <http://dx.doi.org/10.1002/nur.4770180605>.
- \*Gross, D., Fogg, L., Webster-Stratton, C., Garvey, C., Julion, W., & Grady, J. (2003). Parent training of toddlers in day care in low-income urban communities. *Journal of Consulting and Clinical Psychology*, 71(2), 261–278. <http://dx.doi.org/10.1037/0022-006X.71.2.261>.
- \*Helfenbaum-Kun, E. D., & Ortiz, C. (2007). Parent-training groups for fathers of Head Start children: A pilot study of their feasibility and impact on child behavior and intra-familial relationships. *Child & Family Behavior Therapy*, 29(2), 47–64. [http://dx.doi.org/10.1300/J019v29n02\\_04](http://dx.doi.org/10.1300/J019v29n02_04).
- Hinshaw, S. P. (2002). Intervention research, theoretical mechanisms and causal processes related to externalizing behavior patterns. *Development and Psychopathology*, 14(4), 789–818. <http://dx.doi.org/10.1017/S0954579402004078>.
- \*Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., et al. (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), 678–682. <http://dx.doi.org/10.1136/bmj.39126.620799.55>.
- Incredible Years (2011). About Incredible Years certification/accreditation and the process. Retrieved 11/10, 2011, from <http://www.incredibleyears.com/Certification/about.asp>
- Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology*, 36(4), 567–589. <http://dx.doi.org/10.1007/s10802-007-9201-9>.
- Kazdin, A. E. (1997). Parent management training: Evidence, outcomes, and issues. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(10), 1349–1356. <http://dx.doi.org/10.1097/00004583-199710000-00016>.
- \*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(9), 1261–1273. <http://dx.doi.org/10.1016/j.ijnurstu.2007.10.002>.
- \*Kratochwill, T. R., Elliott, 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(3), 269–302. <http://dx.doi.org/10.1521/scpq.18.3.269.22574>.
- 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(3), 373–382. <http://dx.doi.org/10.1037/a0015954>.
- \*Larsson, B., Fossom, S., Clifford, G., Drugli, M. B., Handegard, B. H., & Mørch, W. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: Results of a randomized controlled trial. *European Child & Adolescent Psychiatry*, 18(1), 42–52. <http://dx.doi.org/10.1007/s00787-008-0702-z>.
- \*Lavigne, J. V., LeBailly, S. A., Gouze, K. R., Cicchetti, C., Pochyly, J., Arend, R., et al. (2008). Treating oppositional defiant disorder in primary care: A comparison of three models. *Journal of Pediatric Psychology*, 33(5), 449–461. <http://dx.doi.org/10.1093/jpepsy/jsm074>.
- \*Letarte, M., & Normandeau, S. (2008). Parent training program with ADHD: Who benefits? Paper presented at the biennial meeting of the ISSBD.
- \*Letarte, M., Normandeau, S., & Allard, J. (2010). Effectiveness of a parent training program “Incredible Years” in a child protection service. *Child Abuse & Neglect*, 34(4), 253–261. <http://dx.doi.org/10.1016/j.chiabu.2009.06.003>.
- \*Linares, L. O., Montalto, D., Li, M., & Oza, V. S. (2006). A promising parenting intervention in foster care. *Journal of Consulting and Clinical Psychology*, 74(1), 32–41.
- \*Lindsay, G., Davies, H., Band, S., Cullen, M.A., Cullen, S., Strand, S., et al. (2008). *Parenting early intervention pathfinder evaluation. No. DCSF-RW054*. Coventry: University of Warwick.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Lundahl, B., Risser, H. J., & Lovejoy, M. C. (2006). A meta-analysis of parent training: Moderators and follow-up effects. *Clinical Psychology Review*, 26(1), 86–104. <http://dx.doi.org/10.1016/j.cpr.2005.07.004>.
- Lundahl, B., Tollefson, Risser, & Lovejoy, M. C. (2008). A meta-analysis of father involvement in parent training. *Research on Social Work Practice*, 18(2), 97–106. <http://dx.doi.org/10.1177/1049731507309828>.
- McCall, R. B., & Green, B.L. (2004). Beyond the methodological gold standards of behavioral research: Considerations for practice and policy. *Social Policy Report*, 18(2), 1–20.
- 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(4), 527–543. <http://dx.doi.org/10.1007/s10802-006-9031-1>.
- \*McIntyre, L. L. (2008). Parent training for young children with developmental disabilities: Randomized controlled trial. *American Journal on Mental Retardation*, 113(5), 356–368. <http://dx.doi.org/10.1352/2008.113.356-368>.
- McMahon, B., Holly, L., Harrington, R., Roberts, C., & Green, J. (2008). Do larger studies find smaller effects? The example of studies for the prevention of conduct disorder. *European Child & Adolescent Psychiatry*, 17(7), 432–437. <http://dx.doi.org/10.1007/s00787-008-0685-9>.
- Mihalic, S. F., & Irwin, K. (2003). Blueprints for violence prevention: From research to real-world settings – Factors influencing successful replication of model programs. *Youth Violence and Juvenile Justice*, 1(4), 307–329. <http://dx.doi.org/10.1177/1541204003255841>.
- Mrazek, P. J., & Haggerty, R. J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- \*Nilsen, W. (2007). Fostering futures: A preventive intervention program for school-age children in foster care. *Clinical Child Psychology and Psychiatry*, 12(1), 45–63. <http://dx.doi.org/10.1177/1359104507071055>.
- Nixon, R. D.V. (2002). Treatment of behavior problems in preschoolers: A review of parent training programs. *Clinical Psychology Review*, 22(4), 525–546. [http://dx.doi.org/10.1016/S0272-7358\(01\)00119-2](http://dx.doi.org/10.1016/S0272-7358(01)00119-2).
- 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(3), 114–144. <http://dx.doi.org/10.1007/s10567-008-0033-0>.
- \*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(6), 472–477. <http://dx.doi.org/10.1136/adc.87.6.472>.
- Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1(4), 435–450. <http://dx.doi.org/10.1007/s11292-005-3540-8>.
- Posthumus, J. A., Raaijmakers, M.A. J., Maassen, G. H., Van Engeland, H., & Matthys, W. (2012). Sustained effects of Incredible Years as a preventive intervention in preschool children with conduct problems. *Journal of Abnormal Child Psychology*, 40(4), 487–500. <http://dx.doi.org/10.1007/s10802-011-9518-9>.
- \*Raaijmakers, M.A. J., Posthumus, J. A., Maassen, G. H., Van Hout, B., Van Engeland, H., & Matthys, W. (2009). *Preventive intervention for preschoolers with aggressive behavior*. (unpublished manuscript).
- \*Reid, M. J., Webster-Stratton, C., & Hammond, M. (2007). Enhancing a classroom social competence and problem-solving curriculum by offering parent training to families of moderate- to high-risk elementary school children. *Journal of Clinical Child and Adolescent Psychology*, 36(4), 605–620. <http://dx.doi.org/10.1080/15374410701662741>.
- Reyno, S. M., & McGrath, P. J. (2006). Predictors of parent training efficacy for child externalizing behavior problems – A meta-analytic review. *Journal of Child Psychology and Psychiatry*, 47(1), 99–111. <http://dx.doi.org/10.1111/j.1469-7610.2005.01544.x>.
- Scott, S. (2001). Deciding whether interventions for antisocial behaviour work: Principles of outcome assessment, and practice in a multicentre trial. *European Child & Adolescent Psychiatry*, 10(Suppl. 1), 59–70. <http://dx.doi.org/10.1007/s007870170007>.
- \*Scott, S., Spender, Q., Doolan, M., Jacobs, B., & Aspland, H. (2001). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *BMJ: British Medical Journal*, 323(7306), 194. <http://dx.doi.org/10.1136/bmj.323.7306.194>.
- \*Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Crook, C., et al. (2010). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: The SPOKES project. *Journal of Child Psychology and Psychiatry*, 51(1), 48–57. <http://dx.doi.org/10.1111/j.1469-7610.2009.02127.x>.
- Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy*, 27(2), 171–186. [http://dx.doi.org/10.1016/S0005-7894\(98\)80013-X](http://dx.doi.org/10.1016/S0005-7894(98)80013-X).
- Sheeran, P., Aubrey, R., & Kellett, S. (2007). Increasing attendance for psychotherapy: Implementation intentions and the self-regulation of attendance-related negative affect. *Journal of Consulting and Clinical Psychology*, 75(6), 853–863. <http://dx.doi.org/10.1037/0022-006X.75.6.853>.
- \*Spaccarelli, S., Cotler, S., & Penman, D. (1992). Problem-solving skills training as a supplement to behavioral parent training. *Cognitive Therapy and Research*, 16(1), 1–17. <http://dx.doi.org/10.1007/BF01172953>.
- \*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(2), 221–240. [http://dx.doi.org/10.1016/S0005-7894\(98\)80004-X](http://dx.doi.org/10.1016/S0005-7894(98)80004-X).
- 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(3), 475–495. <http://dx.doi.org/10.1007/s10802-007-9104-9>.
- \*Webster-Stratton, C. (1982). Teaching mothers through videotape modeling to change their children's behavior. *Journal of Pediatric Psychology*, 7(3), 279–294. <http://dx.doi.org/10.1093/jpepsy/7.3.279>.
- \*Webster-Stratton, C. (1984). Randomized trial of two parent-training programs for families with conduct-disordered children. *Journal of Consulting and Clinical Psychology*, 52(4), 666–678. <http://dx.doi.org/10.1037/0022-006X.52.4.666>.
- \*Webster-Stratton, C. (1992). Individually administered videotape parent training: “Who benefits?”. *Cognitive Therapy and Research*, 16(1), 31–52. <http://dx.doi.org/10.1007/BF01172955>.
- \*Webster-Stratton, C. (1998). Preventing conduct problems in Head Start children: Strengthening parenting competencies. *Journal of Consulting and Clinical Psychology*, 66(5), 715–730. <http://dx.doi.org/10.1037/0022-006X.66.5.715>.
- Webster-Stratton, C. (2001). *The Incredible Years: Parents and children videotape series: A parenting course (BASIC)*. Seattle, WA: Incredible Years.
- Webster-Stratton, C. (2002). *The Incredible Years: Parents and children videotape series: A parenting course (ADVANCE)*. Seattle, WA: Incredible Years.
- Webster-Stratton, C. (2004). Quality training, supervision, ongoing monitoring, and agency support: Key ingredients to implementing the Incredible Years programs with fidelity. Retrieved 04/13, 2010 from <http://www.incredibleyears.com/Library/paper.asp?nMode=1&nLibraryID=446>
- Webster-Stratton, C. (2009). Affirming diversity: Multi-cultural collaboration to deliver the Incredible Years parent programs. *International Journal of Child Health and Human Development*, 2(1), 17–32.

- \*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*(1), 93–109. <http://dx.doi.org/10.1037/0022-006X.65.1.93>.
- \*Webster-Stratton, C., Kolpacoff, M., & 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*(4), 558–566. <http://dx.doi.org/10.1037/0022-006X.56.4.558>.
- \*Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology, 30*(3), 283–302. [http://dx.doi.org/10.1207/S15374424JCCP3003\\_2](http://dx.doi.org/10.1207/S15374424JCCP3003_2).
- \*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*(1), 105–124. [http://dx.doi.org/10.1207/S15374424JCCP3301\\_11](http://dx.doi.org/10.1207/S15374424JCCP3301_11).
- \*Williford, A. P., & Shelton, T. L. (2008). Using mental health consultation to decrease disruptive behaviors in preschoolers: Adapting an empirically-supported intervention. *Journal of Child Psychology and Psychiatry, 49*(2), 191–200. <http://dx.doi.org/10.1111/j.1469-7610.2007.01839.x>.
- Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*(4), 413–429. <http://dx.doi.org/10.1037/1082-989X.6.4.413>.