

Sustained Effects of Incredible Years as a Preventive Intervention in Preschool Children with Conduct Problems

Jocelyne A. Posthumus · Maartje A. J. Raaijmakers ·
Gerard H. Maassen · Herman van Engeland ·
Walter Matthys

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The present study evaluated preventive effects of the Incredible Years program for parents of preschool children who were at risk for a chronic pattern of conduct problems, in the Netherlands. In a matched control design, 72 parents of children with conduct problems received the Incredible Years program. These families (intervention group) were compared with 72 families who received care as usual (control group). Two years after termination of the intervention, it appeared that observed and self-rated parenting skills were significantly improved in the intervention group. Likewise, in this group, observed child conduct problems showed sustained intervention effects. The decrease in

observed critical parenting mediated the decrease in observed child conduct problems over time. In addition, it appeared that parental influence increased over time.

Keywords Prevention · Conduct problems · Aggressive behavior · Parent management training · Preschoolers

Introduction

Conduct problems are very common in early childhood. However, a high level of conduct problems in young children has been found to be relatively stable over time, and can be seen as a risk factor for the development of conduct disorders (e.g., Coté et al. 2006; Tremblay et al. 2004; Shaw et al. 2005). In addition to the negative developmental consequences of conduct disorders for the individual, such as poor school, interpersonal, and occupational adjustment, substance abuse, delinquency and other psychiatric disorders (Kim-Cohen et al. 2003; Maughan and Rutter 2001), conduct disorders also incur high costs to society (Raaijmakers et al. 2011; Scott et al. 2001a). Moreover, the majority of children and adolescents in mental health services are referred because of severe conduct problems (Kazdin and Weisz 2003).

Early prevention of conduct disorders has become an important goal for authorities in child development and those who provide community mental health services. Hence, intervention programs targeting preschool children with a high level of conduct problems have been developed. Addressing parenting practices is considered a valuable starting point for prevention, since ineffective parenting, consisting of physical punishment, inconsistent discipline and poor responsiveness to the child (Farrington 2005; Webster-Stratton and Taylor 2001), is associated with

Jocelyne A. Posthumus and Maartje A. J. Raaijmakers equally contributed to the study and they therefore share first authorship.

J. A. Posthumus · M. A. J. Raaijmakers · H. van Engeland ·
W. Matthys
Department of Child- and Adolescent Psychiatry, Rudolf Magnus
Institute of Neuroscience, University Medical Centre Utrecht,
Utrecht, The Netherlands

M. A. J. Raaijmakers
Department of Developmental Psychology, University of Utrecht,
Utrecht, The Netherlands

G. H. Maassen
Department of Methodology and Statistics, University of Utrecht,
Utrecht, The Netherlands

J. A. Posthumus
De Waag, Center for Forensic Psychiatry,
Utrecht, the Netherlands

J. A. Posthumus (✉)
Department of Child and Adolescent Psychiatry,
University Medical Centre Utrecht,
Kruisweg 16 bis,
3513 CT, Utrecht, The Netherlands
e-mail: jocelyneposthumus@hotmail.com

the development and persistence of conduct problems (Patterson 1982; Patterson et al. 2002), whereas effective parenting, consisting of praise and the use of appropriate discipline techniques such as time out (Gardner et al. 2006; Webster-Stratton and Taylor 2001) serves as a protective factor (Tremblay et al. 2004).

Behavioral Parent Training (BPT), which positions the parent as the primary agent for change, is proven to be the most effective method in reducing conduct problems, particularly in young children (McCart et al. 2006). One of these BPT's is the Incredible Years Videotape Modeling Program (IY; Webster-Stratton 2001a, b; Webster-Stratton and Hancock 1998) aimed at improving parenting skills in order to reduce children's problem behavior. This program consists of two components; the BASIC component addresses play, praise and rewards, limit setting and handling misbehavior. Parents are taught to use child directed play skills, to use less critical statements and harsh discipline, and to increase the use of positive and consistent strategies. The ADVANCE component (Webster-Stratton 2002) elaborates on the BASIC program and covers topics such as how to cope with upsetting thoughts and depression, communication skills and solving problems with adults and children. ADVANCE has shown to corroborate the effects of the BASIC program (Webster-Stratton 1994).

Originally, the IY program was designed for the treatment of conduct disorders in young children, and it has proven effective in reducing severe problem behavior (e.g., Scott et al. 2001b; Taylor et al. 1998; Webster-Stratton and Hammond 1997; Webster-Stratton et al. 2004). However, whether the IY program is also effective as a preventive intervention is less clear.

Indeed, studies into the effectiveness of the IY parent program as a preventive intervention yield various results, especially with respect to changes in child behavior. On the one hand, there are studies that reported decreases in child problem behavior; some studies reported positive effects either on parent or on teacher rated measures (Hutchings et al. 2007; Patterson et al. 2002), some only on observed child behavior (Brotman et al. 2008, 2009; Webster-Stratton 1998), and some on both measures (Barrera et al. 2002; Webster-Stratton et al. 2001). On the other hand, in several studies the preventive effectiveness of the IY program on child behavior was not demonstrated (Kratochwill et al. 2003; Reid et al. 2007; Scott et al. 2010); no differences between the intervention and control groups on (observed or parent reported) child behavior were found in these studies. Furthermore, most of these studies into the preventive effect of IY have only investigated the effectiveness of the BASIC program, often in samples with relatively low socioeconomic status, conducted follow up for a relatively short period of time, and showed a low attendance rate (e.g., Reid et al. 2007; Scott et al. 2010).

The present study aimed to evaluate the effectiveness of the IY parent program as an intervention to prevent a chronic pattern of conduct problems in preschool children. Our study is novel and contributes to the current literature regarding the preventive effectiveness of the IY parent program in several ways. First, to establish solid preventive effects it is necessary to conduct long term follow-up assessments. Therefore, relatively long follow-up assessments, up to 2 years after termination of the intervention, were conducted to evaluate the preventive effects of the IY program in the current study.

Second, this study evaluates both the BASIC and ADVANCE component of the IY program. Most evaluations of the IY parent program have only investigated the effectiveness of the BASIC program. Furthermore, since we limited the intervention in this study to the program for parents, results found can be attributed only to the IY parent program, in contrast to studies in which additional programs were used.

Third, in addition to parent-rated measures, an observation of parent-child interaction was conducted as a more objective measure of child behavior. Parent-ratings of child behavior are often susceptible to biases, such as parents' mood or expectations of the intervention (Eddy et al. 1998) and observations have been found to be sensitive to change in parent and child behavior as a result of an intervention (Aspland and Gardner 2003; Frick and Loney 2000).

Finally, the present study adds to the existing knowledge by examining mediation mechanisms. Mediation was examined to investigate whether improvements in parenting skills preceded the changes in child behavior, as suggested by Kazdin and Nock (2003). Additionally, bidirectional influences of parenting skills and child behavior over time were studied.

Method

Design

A case control design, in which participants were selected to be in either the intervention group (IG) or control group (CG) based on their place of residence, was used in this study. Randomization was not feasible because of geographical reasons. According to the Standards of Evidence given by the Society for Prevention Research (2005), use of a case control design is permitted "as long as assignment was not by self-selection, but instead by some other factor (for instance geography)". The families to be recruited lived in several different towns and cities in the province of Utrecht, The Netherlands. As motivation to participate is a recurrent problem in intervention studies, especially when families of children with conduct problems are involved (Luk et al. 2001), every effort was made to encourage

families to participate. To avoid low attendance rates due to the location of the training, the IY program was delivered at four different sites which were within 15 km distance from the consenting families' homes and which are also easily accessible, such as community centers. Moreover, the IY program requires at least 6 parents to participate in a parent group to optimize discussion and to foster a sense of support (Webster-Stratton 2001a, b). Consequently, sufficient parents had to live in the same area to form a group. In addition, parents in the control group were blind to their condition, i.e., they were not informed on the fact that the other group received a parent program. The control group instead was told that the study aimed to investigate the development of aggressive behavior in young children and that they would be informed on the study design after completion of the study. CG parents were allowed to use regular services for their child's behavior, i.e., care-as-usual, and were informed about the design of the study retrospectively. Families of the two conditions were matched on the child's gender, level of aggression, IQ, the parents' educational level, stress level, and address density of the place of residence of the family. In a separate study, the performance of a case control design was compared to a randomized study design by simulating hypothetical intervention and control groups in a mathematical software program based on the data in the present study. The Mahalanobis metric was used to assess the distance between families in the IG and CG and pairwise matching was performed. The equivalence of the predefined intervention and control group was compared to the equivalence of the randomized groups, resulting in a more equally balanced distribution of the six key characteristics in our matched predefined groups than in randomization in more than half of the simulated trials, indicating that matching in a case control design was a viable alternative for this study (Raaijmakers et al. 2008). Therefore, in the present study, the same matching procedure was executed. In this study, families were assessed at pre-intervention (PRE), directly after termination of the intervention (POST), one year after termination of the intervention (FU1) and two years after termination of the intervention (FU2).

Participants

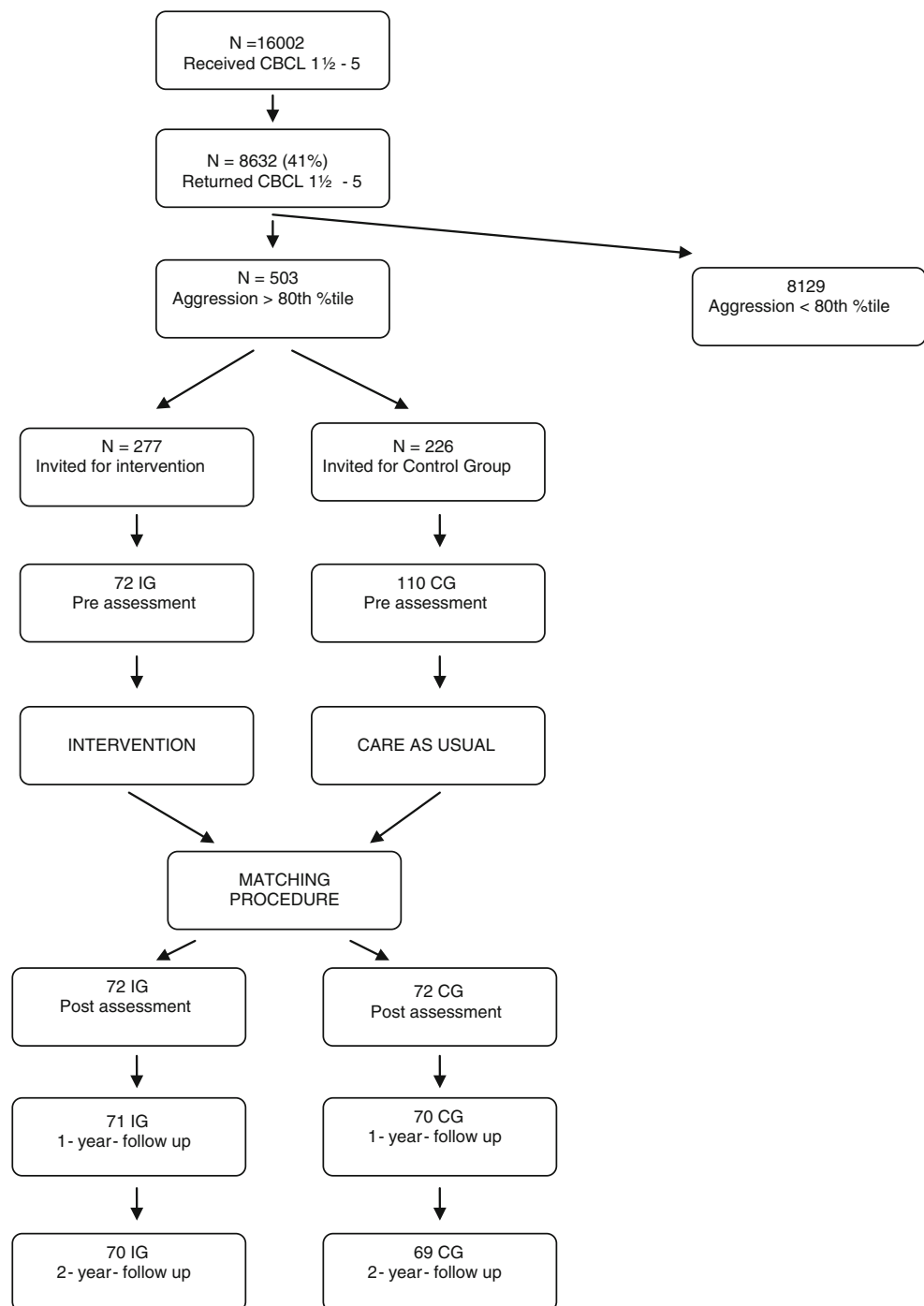
Addresses of families were acquired by the Office for Screening and Vaccination in the province of Utrecht, The Netherlands. Parents of 16002 4-year-old children born either in 2000 or 2001 received a Child Behavior Checklist 1½–5 (CBCL; Achenbach and Rescorla 2000; Dutch version by Verhulst & Van der Ende) by mail. More than half of the parents filled out and returned this questionnaire (see Fig. 1). Children were selected to participate if they scored at or above the 80th percentile of the Aggressive Behavior scale of the CBCL. In total, 503 children scored at

or above the 80th percentile and were considered to show conduct problems. Based on their place of residence, 277 families were selected for the IG and 226 families for the CG. First, rural and urban areas were identified, based on address density data, resulting in eight urban and eight rural areas. Then, those areas were divided between the intervention and control group. The intervention groups were recruited from four urban and four rural areas, and the control group families were recruited from four other urban and four rural areas. Parents were invited to participate by letter and were called maximally 2 weeks later to ask for their response. If parents were interested in participation, two research team members visited the family to explain the procedure of this research project. During this home visit, families who were invited to participate in the intervention received additional information on the IY parent program. Children with an estimated full scale IQ below 80 were excluded from the study. This resulted in 72 families (26% of the selected families) in the IG and 110 families (47% of the selected families) in the CG.

Reasons for non-participation were: 1) experiencing stressors (IG: 10%, CG: 18%, e.g. chronic illness of a family member, pressure of a partner to decline), 2) parents found treatment/the study not relevant (IG: 39%, CG: 42%, e.g. parents indicated they were capable of handling their child's aggressive behavior, lack of parental recognition of the child's aggressive behavior), 3) parents found treatment/the study too demanding (IG: 23%, CG: 9%, e.g. parents did not want to commit themselves to a 3-year study), 4) practical reasons (IG: 23%, CG: 24%, e.g. involvement in other interventions, child diagnosed with another disorder, i.e., 4 autism spectrum disorders, 1 Prader Willi) and 5) parents indicated no reason (IG: 5%, CG: 7%). The Aggressive Behavior score of children whose parents agreed or refused to participate in this study was not significantly different, neither in IG, nor in CG.

Mahalanobis person to person matching was performed after PRE on 72 IG families and 72 CG families. An independent administrator who was not involved in this study carried out the matching procedure. Families lost between POST to FU2 (3 CG and 2 IG) did not differ in their initial level of aggression from those retained. Attrition of these families was due to personal circumstances such as medical conditions of the child or parent, or participation was a too heavy burden for the family.

Characteristics of the IG and CG are presented in Table 1. At selection, the mean percentile of the CBCL aggressive behavior score was the 93rd percentile. Groups did not significantly differ on any of these descriptive characteristics, except for age of the child, $t(71)=2.41$, $p=0.018$; CG children were 2 months older than IG children. All primary caregivers were biological parents, except for one mother in the IG, who was an adoptive parent. Almost all children were Caucasian except for three IG children (one from South

Fig. 1 Flow chart of selection and assessments

America, two from Asia) and one CG child (from South America). At PRE, all children were medication naïve. All families were allowed to use care as usual. Twelve families in IG (17.9%) and 11 CG families (15.7%) received other professional help during the intervention phase. The families received mental health care (7 IG children, 5 CG children), youth care (2 IG children, 5 CG children), educational support (4 IG children, 7 CG children), and community care (3 IG parents, 5 CG parents). In addition, the researchers offered their help in finding adequate mental health services

when needed. Questionnaires and home observations were conducted with the primary caregiver of the child. In the IG, 18% of the primary caregivers were fathers and in the CG, 10% of the primary caregivers were fathers.

Procedure

Prior to PRE, written informed consent was obtained from the participating families. Every assessment consisted of a set of questionnaires, which was mailed to the parents, and

Table 1 Sample characteristics by group

| Measure | IG (<i>n</i> =72) <i>M</i> (<i>SD</i>) | CG (<i>n</i> =72) <i>M</i> (<i>SD</i>) |
|--|---|---|
| Child | | |
| Gender (% boys) | 70.8 | 70.8 |
| Age (months) | 50.3 (3.11) | 51.3 (2.53) |
| IQ | 107.3 (9.87) | 107.5 (11.57) |
| CBCL 1½–5 (raw scores at selection) | | |
| Aggressive Behavior | 21.99 (4.37) | 22.49 (4.69) |
| Parent | | |
| Age (years) | | |
| Primary caregiver | 35.5 (4.84) | 34.1 (5.49) |
| Secondary caregiver | 37.9 (5.12) | 36.8 (4.83) |
| Education (%) | | |
| Primary | – | 2.8 |
| Secondary | 4.2 | 5.6 |
| Intermediate vocational | 29.2 | 31.0 |
| Higher vocational | 38.9 | 31.0 |
| University | 27.8 | 29.6 |

Education denotes the highest educational level of both parents.

a home visit. Home visits consisted of an observation of the primary caregiver and the child playing together. Parents received a monetary reward (€ 25,-) for each assessment. The study was approved by the Medical Ethical Review Committee of the Utrecht University Medical Center.

The Incredible Years Parent Program: BASIC and ADVANCE

In this study, the BASIC and ADVANCE curriculum were delivered in 18 2-hour sessions (11 BASIC and 7 ADVANCE). Eight groups of parents received the intervention in community centers in different towns and cities spread over the province of Utrecht. The parent groups were led by two certified group leaders with parents of 6 to 11 children per group. Parents were encouraged to attend the group together. If parents missed a session, group leaders called them, sent them home assignments, and encouraged parents to come to the next session half an hour earlier in order to discuss the content of the missed session. After termination of the IY program, two booster sessions were offered 3 months and 6 months after termination of the intervention. In the weekly sessions, parents watched approximately 225 brief vignettes of parents and children interacting. After each vignette, the group leader asked questions to stimulate discussion about what parents found particularly (in)effective and to practice alternative responses. Dutch subtitles were used in the video-vignettes. Parents were

encouraged to role-play new skills and to practice these skills at home in order to establish new habits. Before each session, parents read a chapter on the topic of that particular session in the book belonging to the program.

Teaching methods were used within a collaborative setting, in which group leaders established themselves as part of the group, rather than as experts. Parents were empowered due to the group process and the collaborative attitude of the therapist. Group leaders encouraged parents to solve problems in order to ensure that the progress made during the intervention was maintained after program completion.

Treatment Fidelity and Integrity

Treatment fidelity has been demonstrated to be a predictor of positive change (Bellg et al. 2004). Therefore, it is crucial to ensure the intervention program is delivered as originally intended. Six members of the team, with a background in clinical child psychology or child psychiatry, were trained by the program developer during a 3-day workshop. The members of the research team ran two pilot groups at child psychiatry settings to become familiar with the materials and specific techniques, and they received supervision from accredited IY trainers to become certified. Intervention sessions were videotaped and reviewed during weekly meetings of group leaders to ensure that the program was delivered with fidelity. In addition, a quarter of the video taped sessions was peer-reviewed. Finally, the manual of the IY program was used and both parental evaluations as well as checklists for group leaders were filled out after every session.

Measures

Child Behavior Checklist (CBCL 1½–5)

The CBCL 1½–5 (Achenbach and Rescorla 2000) is a parent-rated questionnaire, consisting of 99 items, on which the child is rated on various behavioral and emotional problems. The CBCL 1½–5 consists of 7 subscales in which the items can be clustered, i.e., Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems and Aggressive Behavior. By summing all the item scores, a Total Problems score is computed. Parents circle the answer that fits the behavior of their child; “never”, “sometimes” or “always”. The CBCL is widely used in clinical and research settings because of its demonstrated reliability and validity, ease of administration and applicability to clinical and nonclinical groups (Dutra et al. 2004). To recruit the children, the level of conduct problems was assessed using the Child Behavior Checklist 1½–5 *Aggressive Behavior scale* (Achenbach and Rescorla 2000). This scale contains 19 items like “hits

others”, “does not feel guilty” and “often has temper tantrums”.

Eyberg Child Behavior Inventory (ECBI)

The ECBI (Eyberg and Pincus 1999) is a parent-rated questionnaire used to assess the occurrence of conduct problems in children aged 2 to 16 years. Several studies have demonstrated acceptable reliability and validity of the two scales (e.g., Boggs et al. 1990; Eyberg and Pincus 1999; Rich and Eyberg 2001). The ECBI consists of 36 behavioral items which are rated on two scales; an *Intensity Scale*, which measures the frequency of the problem behavior on a 7-point scale (ranging from ‘never’ to ‘always’; in the present study, $\alpha=0.91$) and a *Problem Scale*, which asks parents to report whether the behavior is perceived to be a problem (yes or no; $\alpha=0.88$).

Dyadic Parent–child Interaction Coding System - Revised (DPICS-R)

The DPICS-R (Eyberg and Robinson 1981; revised 2000) is an observational measure used to assess the quality of parent–child interactions at home, with adequate psychometric properties (Robinson and Eyberg 1981). Parent and child were observed for 20 min while playing with a fixed set of toys at PRE, POST and FU assessments. The observation was videotaped and coded later on. The observation consisted of four five-minute periods; in the first period parent and child played like they would usually do to get used to being videotaped, in the second period the child picked a toy and directed the play session (child directed play, CDI), in the third period the parent picked a toy and directed the play session (parent directed play, PDI), and in the final period the parent had to make the child clean up (clean up, CU). For each period, parenting skills and child behavior were coded separately into 47 categories; 24 for parent behavior (e.g., statements, positive affect) and 23 for child behavior (e.g., physical warmth, smart talk). In this study, parental behavior categories *Critical Statements* and *Labeled Praise* were used. With respect to child behavior, a composite score of the categories Smart Talk, Cry/Whine/Yell, and Physical Negative was used. This composite score was labeled *Conduct Problems* ($\alpha=0.51$). In addition, the category *Comply* was used as a measure of child behavior. A proportional compliance-score was constructed; the number of the child’s comply-scores was divided by the number of parental commands. Trained master-students and trained project staff had to achieve an interrater-reliability of 70% before coding parent and child behaviors into these categories. Coders were blind to condition. The quality of scoring was monitored continuously by having 20% of the

observations checked by a second rater. Double checking the observations revealed a mean interrater-reliability of 80% (SD=5.20, range: 70–96%).

Parent Practices Interview (PPI)

This parent-rated questionnaire (Webster-Stratton 2001) was designed to measure parenting skills or discipline styles of parents of young children. The PPI consists of 15 questions, each with several aspects, asking for a response of the parent to misbehavior, appropriate behavior and several statements. Parents could answer to these questions and respond to the statements on a seven-point Likert-scale, ranging from ‘not (likely) at all’ to ‘always/very likely’. Seven summary scales are extracted from this questionnaire; *Appropriate Discipline* (e.g., actually disciplining the child when it misbehaves, 12 items, $\alpha=0.74$), *Harsh and Inconsistent Discipline* (e.g., threatening, but not punishing, 15 items, $\alpha=0.81$), *Positive Verbal Discipline* (e.g., discussing the problem with the child, 9 items, $\alpha=0.67$), *Monitoring* (e.g., supervision of child activities, 5 items, $\alpha=0.35$), *Physical Punishment* (e.g., slapping or hitting when misbehavior occurs, 6 items, $\alpha=0.87$), *Praise and Incentives* (e.g., giving a hug or compliment, 11 items, $\alpha=0.73$), and *Clear Expectations* (e.g., clear rules about going to bed, 6 items, $\alpha=0.65$). All scales demonstrated acceptable reliability, except for *Monitoring*. Therefore, this scale was excluded from the analyses.

Data Analysis

Since there was a low level of attrition, missing data were not imputed. If a scale score was missing for a family, that scale score of the matched family was removed from the analyses as well. Scale scores were excluded from the analyses if more than 25% of the items of a scale of a measure were missing. Four PPI’s were missing because of that reason; two IG and two CG mothers did not fully fill out that questionnaire. Overall intervention effects were evaluated by means of a repeated measures ANOVA using Helmert contrasts of the time x group interaction (which is tantamount of Helmert contrasts of the IG-CG difference scores across time levels). This contrast compares the mean of the dependent variable at each level of the independent variable (i.e., assessment) with the overall mean of the dependent variable at the subsequent levels of the independent variable. If the intervention effect is present and sustained, this will show in the following way in the Helmert contrasts: PRE versus all later assessments will significantly differ; POST versus all later assessments, as well as FU1 versus FU2, will not significantly differ. Time was entered as a within subject factor and because of matching, group was entered as a within subject factor as

well. A criterion of $p < 0.05$ was used in the analyses. Analyses were performed with SPSS 15.0. Due to technical problems in the video registration of the observations of parent–child interactions, less DPICS data were gathered at PRE than at POST, FU1 and FU2. Moreover, five families were lost from POST to FU2 (3 CG and 2 IG children). As a consequence, 56 pairs of children in the PRE to FU2 DPICS comparisons were analyzed. Mediation is demonstrated by means of structural equation modeling (SEM).

Results

Comparisons at PRE

ANOVA's revealed no significant differences between IG and CG on the parent rated measures at PRE. However, on observed behavior of parents and children several significant differences between IG and CG were found. Parents differed on *Critical Statements*, with IG parents being more critical than CG parents, $F = 3.99$, $df = 1$, $p = 0.050$. Children differed significantly on *Conduct Problems*, $F = 8.68$, $df = 1$, $p = 0.004$. IG children showed more conduct problems than CG children at PRE.

Attendance

Attendance rate was 78%; an average of 14 sessions (out of 18) was attended by at least one of the parents. Groups were attended by couples (43%), single mums (14%), parents who alternated (12%), fathers (6%; while mother stayed at home) and mothers (25%; while father stayed at home). None of the families dropped out of the intervention.

PRE to FU2 Comparisons

Sample sizes, means and standard deviations for IG and CG are presented in Table 2. Results of the repeated measures ANOVA are presented in Table 3.

DPICS: Observed Parenting

The development in *Critical Statements* revealed a pattern that is indicative of a sustained intervention effect. The decrease in *Critical Statements* was significantly larger in the IG than in the CG between PRE and all other moments of assessment, while no significant difference between POST and later moments were ascertained. This means that the effect on *Critical Statements* occurred between PRE and POST. The IG showed a significantly larger increase of *Labeled Praise* at POST, but this effect disappeared between POST and FU1.

DPICS: Observed Child Behavior

Repeated measures ANOVA's revealed that there were no differences between IG and CG on *Comply* between PRE versus later moments of assessment, between POST versus all later moments of assessment and between FU1 and FU2. Thus, no sustained effects on *Comply* was found. A difference emerged between FU1 and FU2 in favor of the CG; the CG showed a larger increase on *Comply* than the IG. Further, the IG showed a larger decrease on *Conduct Problems* between PRE and all later assessments than the CG, while the differences between POST and all later moments of assessment and between FU1 and FU2 did not significantly differ, which is indicative of sustained intervention effects. This means that the effect on *Conduct Problems* occurred between PRE and POST.

PPI: Parent-rated Parenting

Repeated measures ANOVA's revealed several significant differences indicative of sustained effects in parenting skills between the IG and CG. Sustained intervention effects on *Appropriate Discipline*, *Harsh and Inconsistent Discipline* and *Praise and Incentives* were found. All effects pointed in the expected direction, with the IG showing larger improvements in these skills than the CG between PRE and all later moments of assessment.

ECBI: Parent-rated child Behavior

No significant differences between IG and CG were found over time.

Mediation

As a consequence of the results described earlier, the question of whether improvement of parental skills mediated the improvement of child conduct has been constricted to the question of whether the decrease in *Critical Statements* led to a decrease in *Conduct Problems*. More specifically, the question to be answered is whether the improvement in *Critical Statements* (i.e., the change from PRE to POST) would indeed precede the improvement in *Conduct Problems* (i.e., the change from PRE to FU2). Because of the matching of IG with CG, it was not possible to conduct traditional mediation analyses according to guidelines of Baron and Kenny (1986), as is customary in mediation analyses. We also tested a path model, but we were forced to design a model that is based on the use of difference scores, analogous to the earlier reported repeated measures ANOVAs.

With regard to the dependent variable (improvement of *Child Conduct*), difference scores were computed be-

Table 2 Mean scores of outcome measures in IG and CG from PRE to FU2

| | N | PRE | | POST | | FU 1 | | FU2 | |
|------------------------------|----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | | IG M (SD) | CG M (SD) | IG M (SD) | CG M (SD) | IG M (SD) | CG M (SD) | IG M (SD) | CG M (SD) |
| Observed | | | | | | | | | |
| <i>Parenting: DPICS</i> | | | | | | | | | |
| Critical statements | 56 | 8.64 (8.05) | 6.07 (4.74) | 4.70 (5.03) | 5.29 (4.61) | 3.21 (3.27) | 4.41 (4.57) | 3.00 (3.56) | 3.50 (4.02) |
| Labeled praise | 56 | 0.52 (0.85) | 0.45 (0.78) | 1.86 (2.95) | 0.46 (0.71) | 1.07 (1.72) | 0.55 (1.35) | 0.50 (0.92) | 0.43 (0.74) |
| <i>Child Behavior: DPICS</i> | | | | | | | | | |
| Comply | 56 | 40.80 (15.75) | 45.07 (13.90) | 36.46 (15.70) | 39.25 (17.13) | 53.47 (14.71) | 46.31 (11.28) | 48.63 (14.22) | 49.18 (15.21) |
| Conduct problems | 56 | 9.11 (13.62) | 3.88 (5.62) | 3.93 (6.65) | 2.86 (5.04) | 1.91 (4.68) | 0.98 (2.02) | 1.29 (2.19) | 0.80 (1.62) |
| Parent-rated | | | | | | | | | |
| <i>Parenting: PPI</i> | | | | | | | | | |
| Appropriate discipline | 66 | 49.36 (10.36) | 49.86 (8.87) | 54.00 (9.47) | 48.83 (9.21) | 53.06 (7.79) | 49.50 (8.67) | 53.39 (9.26) | 48.55 (7.96) |
| Harsh and incons. discipline | 66 | 45.93 (9.60) | 44.53 (10.04) | 39.28 (7.41) | 43.07 (9.53) | 40.95 (9.73) | 42.67 (10.41) | 39.65 (9.17) | 42.33 (9.84) |
| Positive verbal discipline | 66 | 49.95 (6.21) | 48.62 (5.91) | 51.26 (5.14) | 48.85 (5.91) | 50.73 (6.34) | 48.47 (6.99) | 50.33 (5.31) | 49.07 (6.19) |
| Physical punishment | 66 | 9.57 (4.46) | 10.48 (4.63) | 7.93 (3.93) | 9.90 (4.50) | 7.77 (2.15) | 8.74 (3.59) | 7.43 (1.71) | 8.40 (2.86) |
| Praise and incentives | 66 | 46.65 (7.74) | 45.21 (7.04) | 52.78 (6.74) | 44.14 (6.81) | 50.52 (5.76) | 42.91 (6.96) | 50.85 (6.05) | 43.54 (7.56) |
| Clear expectations | 66 | 21.87 (3.79) | 21.47 (3.79) | 22.20 (3.91) | 21.27 (2.66) | 23.410 (3.85) | 22.19 (4.71) | 24.39 (4.77) | 22.83 (4.23) |
| <i>Child behavior: ECBI</i> | | | | | | | | | |
| Intensity | 53 | 126.70 (25.15) | 128.06 (28.04) | 117.96 (26.92) | 122.60 (29.28) | 120.55 (25.67) | 122.62 (25.34) | 118.34 (24.46) | 118.53 (25.23) |
| Problem | 43 | 12.30 (6.96) | 11.72 (7.96) | 9.93 (6.83) | 10.91 (8.91) | 11.74 (8.17) | 10.70 (7.71) | 10.58 (7.38) | 10.16 (7.76) |

Table 3 Interaction effects of the repeated measures ANOVA with Helmert contrasts

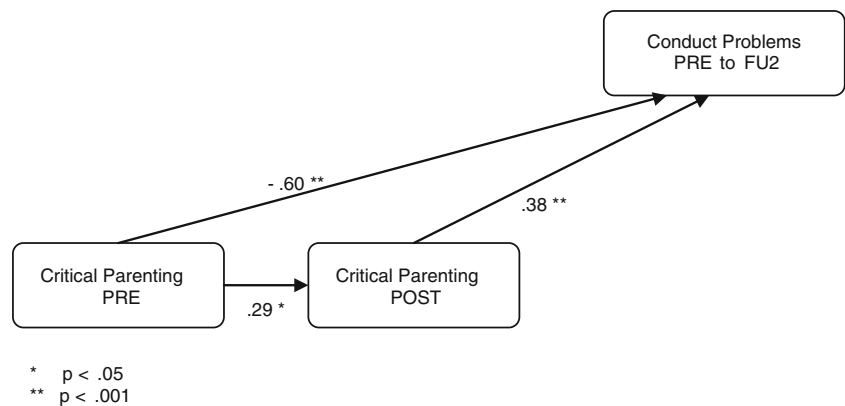
| | PRE versus all later assessments | | POST versus all later assessments | | FU 1 versus FU2 | |
|-----------------------------------|----------------------------------|----------|-----------------------------------|----------|-----------------|----------|
| | <i>F</i> | <i>p</i> | <i>F</i> | <i>p</i> | <i>F</i> | <i>p</i> |
| Observed | | | | | | |
| <i>Parenting: DPICS</i> | | | | | | |
| Critical statements | 6.67 | 0.012 | 0.13 | 0.716 | 0.60 | 0.442 |
| Labeled praise | 8.19 | 0.006 | 10.86 | 0.002 | 2.02 | 0.161 |
| <i>Child Behavior: DPICS</i> | | | | | | |
| Comply | 3.26 | 0.076 | 2.63 | 0.111 | 5.78 | 0.020 |
| Conduct problems | 4.65 | 0.035 | 0.11 | 0.746 | 0.54 | 0.466 |
| Parent-rated | | | | | | |
| <i>Parenting: PPI</i> | | | | | | |
| Appropriate discipline | 12.21 | 0.001 | 0.63 | 0.429 | 1.58 | 0.214 |
| Harsh and inconsistent discipline | 13.63 | 0.000 | 1.39 | 0.244 | 1.05 | 0.310 |
| Positive verbal discipline | 0.49 | 0.484 | 0.58 | 0.451 | 1.49 | 0.230 |
| Physical punishment | 0.40 | 0.530 | 2.77 | 0.101 | 0.00 | 1.00 |
| Praise and incentives | 29.02 | 0.000 | 1.29 | 0.262 | 0.09 | 0.772 |
| Clear expectations | 1.99 | 0.163 | 0.39 | 0.531 | 0.23 | 0.635 |
| <i>Child behavior: ECBI</i> | | | | | | |
| Intensity | 0.10 | 0.756 | 0.99 | 0.324 | 0.28 | 0.598 |
| Problem | 0.12 | 0.732 | 1.57 | 0.217 | 0.23 | 0.631 |

tween PRE and FU2 (PRE was subtracted from FU2) for IG and CG; then, the scores of the CG were subtracted from the scores of the IG for each matched pair. With regard to the independent variable (improvement of *Critical Statements*), Glasnapp (1984) was followed, who advocates inclusion of both components of the change score as separate predictors in a regression analysis. The most important reason is that a straight change score (i.e., equal weights with opposite sign) is less than an optimum predictor. The model used to investigate mediational mechanisms is presented in Fig. 2.

The coefficient from *Critical Statements* at PRE to the difference from PRE to FU2 on *Conduct Problems* was negative, $B=-0.60$, $p<0.001$, while the coefficient from *Critical Statements* at POST to the difference from PRE to

FU2 on *Conduct Problems* was positive, $B=0.38$, $p<0.001$. These results can be expressed in the following formula: $CP(FU2 - PRE) = 0.38*CS(POST) - 0.60*CS(PRE)$. Put verbally, improvement of child conduct shown at FU2, which is expressed as a negative CP(FU2-PRE) score, is positively related to gain of parental skills achieved between PRE and POST, also resulting in a negative value of a weighted difference between the amounts of critical statements uttered at the two moments. These results indicate that decrease in *Critical Statements* due to the IY parent program during the intervention (PRE to POST) led to a decrease in *Conduct Problems* 2 years after termination of the intervention. It appeared that the decrease in *Critical Statements* mediated the decrease in *Conduct Problems*.

Fig. 2 Mediation of critical parenting on child conduct problems



In addition, we conducted longitudinal analyses in order to test the bidirectional influences of parenting skills and child behavior over time. Therefore, a cross-lagged panel model with parenting skills and child behavior measured at four distinct moments in time was designed. In this model, difference scores between the IG and CG were used. Since observed *Conduct Problems* was the only child behavior outcome that improved at FU2 and *Critical Statements* was the only observed parenting skill that improved at FU2, a model with observed *Critical Statements* and *Conduct Problems* from PRE to FU2 was constructed. Note that, when compared with the CG, the IG showed a significantly larger decrease on both *Conduct Problems* and *Critical Statements*. In the initial model, we included stability coefficients of *Critical Statements* and *Conduct Problems*. Further, it was predicted that both parents and children would influence each others behavior, both cross-sectionally and longitudinally. The bidirectional influences could not be assessed cross-sectionally because such a model did not result in a convergent solution. Since it was expected that parental influence on child behavior would be larger than influence of child behavior on parenting skills, in the final model only cross-sectional influences of parenting skills on child behavior were hypothesized. As a result of the IY parent program, we predicted the child's influence to decrease and the parent's influence to increase over time. In order to make the model fit (after considering the modification indices), relations were added. (CFI=0.958; RMSEA=0.057; $\chi^2=14.66$; $df=10$; see Fig. 3).

In this model, the added relations, relative to the hypothesized relations, are represented as dotted lines (Standardized Direct, Indirect and Total Effects are presented in Table 4). Against expectations, the ANOVA's revealed that the IG and CG significantly differed on *Critical Statements* and *Conduct Problems* at PRE. Therefore, the influences of PRE on assessments at subsequent moments are indeed not hypothesized, but added and presented as dotted lines. The difference between IG and CG on *Critical Statements* was relatively stable over time (from POST to FU2), and the difference between IG and CG on *Conduct Problems* appeared to be stable between FU1 and FU2. The cross-sectional influence of the difference between IG and CG on *Critical Statements* on the difference between IG and CG on *Conduct Problems* was moderate at POST, and very weak at FU1 and FU2. The influence of the difference between IG and CG on *Critical Statements* at POST on the difference between IG and CG on *Conduct Problems* at FU1 was weak, while the influence of the difference between IG and CG on *Critical Statements* at FU1 on the difference between IG and CG on *Conduct Problems* at FU2 was moderate. This model indicates that the influence of critical parenting on negative child behavior increased over time, while the influence of the

child's conduct problems on critical parenting remained relatively small.

Discussion

In the present study, the 2 year follow up effects of the Incredible Years program for parents of 4-year old children who were at risk for the development of a chronic pattern of conduct problems was evaluated. As expected, the results showed significant improvements in both observed and parent-rated parenting skills in the IG when compared to the CG, and these improvements were maintained over time. The observation of parenting behavior revealed a sustained decrease in the use of critical statements 2 years after termination of the intervention. The increased use of labeled praise after the intervention, however, disappeared over time. Parents themselves reported sustained increases in appropriate discipline and praise, whereas they reported sustained decreases in their use of harsh and inconsistent discipline. In addition, observed child behavior showed sustained positive intervention effects; children showed less conduct problems 2 years after the intervention. In contrast, parents did not report improvements in child behavior. It should be noted that IG children had significantly higher conduct problem scores at PRE than CG children, and IG parents scored significantly higher on the use of critical statements than CG parents. Thus, there was a greater chance of regression to the mean in the IG. This difference was corrected for by using repeated measures ANOVA's with time x group interaction. However, there still might have been a larger possibility of improvement in the IG. In addition, although the number of drop outs during the intervention phase was zero, the number of cases in some measures at later assessments was relatively low. Therefore, these findings must be interpreted with caution. Further, evidence for parenting practices mediating changes in child's conduct problems was demonstrated; i.e., the decrease in critical parenting during the intervention due to the IY parent program led to a decrease in the child's conduct problems 2 years after termination of the intervention. Additional analyses in which the bidirectional influences of parenting skills and child behavior were investigated revealed that the influence of parenting skills on child behavior increased over time, while the influence of the child's behavior on parenting skills remained weak over time.

Although parents continued to use less criticism towards their children 2 years after termination of the intervention, they did not praise their children as much as they did directly after termination of the intervention. One might speculate that parents did take the child's compliance for granted and did not feel urged to go on praising such behavior. By contrast, persistent use of less criticism might be easier for parents,

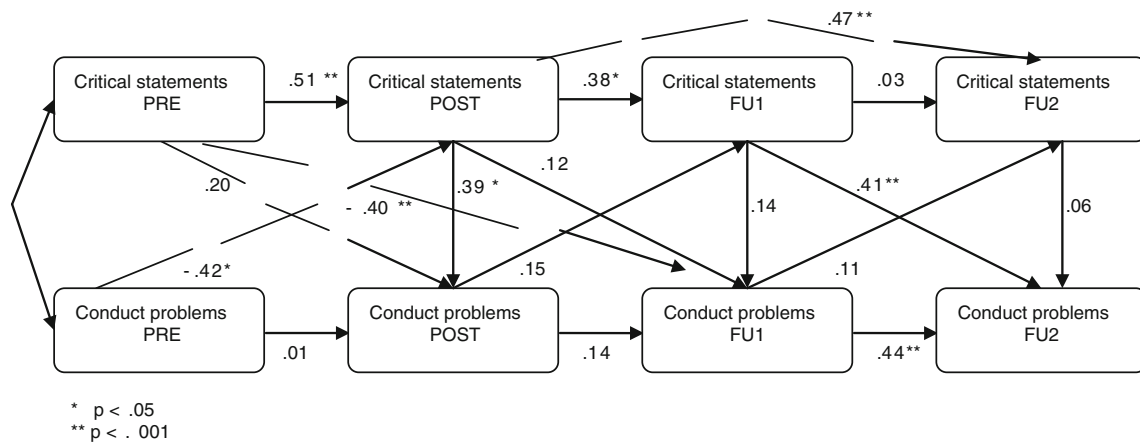


Fig. 3 Bidirectional influences of critical statements and child conduct problems

maybe because they immediately observe the negative effect of criticism on their child’s emotions.

No differences between IG and CG children on parent reported conduct problems were obtained. The discrepancy between parent reported and observed changes in child behavior was reported in earlier studies (Brotman et al. 2008; Gardner et al. 2006). The lack of a parent reported

decrease of the child’s conduct problems might be due to a difference between the two groups in motivation to report these problems. Compared with parents who did not participate in the intervention, parents in the IY parent program learn to observe their child’s behavior and to identify their child’s problems as goals in the IY parent program (Webster-Stratton 1998). One might speculate that, if families do not receive help, they might be more reluctant to acknowledge the child’s conduct problems. By contrast, if families do receive help, parents might be more inclined to report their child’s misbehavior at assessments after termination of the intervention. It is of interest to investigate whether effects on parent-rated child behavior will emerge later on. In order to investigate these ‘sleeper effects’, long-term follow up assessments are required. Another informant which is often used in intervention studies is the teacher. In the present study, teacher ratings were not taken into account, due to the fact that it is mandatory for children to attend school from the age of five onwards in the Netherlands. Most children start attending school at age four. In the present sample, some children were younger than four and most children attended school for only 2 or 3 months at PRE. According to our clinical experience, teachers are often reluctant in reporting behavior problems in 4-year-old children during these first months at school; therefore we did not include teacher reports.

Mediation in BPT research often is examined by studying associations between an improvement in parenting skills and a decrease in children’s problem behavior at the same time. However, to demonstrate that the improvement in parenting skills indeed caused the decrease of problem behavior, the former must be assessed prior to the latter (Kazdin and Nock 2003). We are aware of only one study in which this sequential pattern of changes was demonstrated (DeGarmo et al. 2004). Results of the present study demonstrate an association between improvements in observed parenting during the intervention, and a change of observed child

Table 4 Standardized effects of differences between IG and CG

| | Conduct Problems (DV) | | |
|--------------------------|-----------------------|-------|-------|
| | POST | FU1 | FU2 |
| Critical Statements | | | |
| POST | | | |
| Direct | 0.315 | 0.114 | 0.000 |
| Indirect | 0.000 | 0.099 | 0.269 |
| Total | 0.316 | 0.213 | 0.269 |
| FU1 | | | |
| Direct | 0.000 | 0.140 | 0.408 |
| Indirect | 0.000 | 0.000 | 0.063 |
| Total | 0.000 | 0.140 | 0.471 |
| FU2 | | | |
| Direct | 0.000 | 0.000 | 0.055 |
| Indirect | 0.000 | 0.000 | 0.000 |
| Total | 0.000 | 0.000 | 0.055 |
| Critical Statements (DV) | | | |
| POST | | | |
| Direct | 0.000 | 0.153 | 0.000 |
| Indirect | 0.000 | 0.000 | 0.019 |
| Total | 0.000 | 0.153 | 0.019 |
| FU1 | | | |
| Direct | 0.000 | 0.000 | 0.106 |
| Indirect | 0.000 | 0.000 | 0.000 |
| Total | 0.000 | 0.000 | 0.106 |

behavior over time. Specifically, evidence for a mediating effect of the decrease of critical parenting on the decrease of the child's conduct problems was found. This result is in line with the coercive theory of Patterson (1982), which states that a sequence of interactions based on negative reinforcement maintains aggressive behavior problems in children. This sequence starts with a parent acting aversively towards the child. The child reacts aversively to the parent, and the parent gives in. The child's behavior is thus reinforced and is likely to increase in the future. Although, in contrast to Patterson (1982), the direct influence of the parents' critical statements on the child's conduct problems was not assessed during the observation, the association of the decrease of parental critical remarks from PRE to POST with the decrease of conduct problems from PRE to FU2 is in line with the coercive theory. Furthermore, from the model in which bidirectional influences of parent and child behavior were investigated, it became clear that over time parents gained influence over their child's behavior; the association of using less critical statements by parents with the decrease of child's negative behavior increased over time, while the influence of child behavior on the parental use of criticism remained weak over time. Thus, parents learned to take the lead, which supports mediation processes occurring during the intervention.

The results of this study have to be considered in the light of a number of limitations. First, this study was not a randomized controlled trial. Although matching can be a viable alternative when randomization is not feasible because of geographical reasons, it still lacks the opportunity to control for unobserved variables that might have influenced the results. Specifically, this may have resulted in the inclusion of an IG in which parents were more motivated than parents in the CG due to a higher level of their child's conduct problems. Indeed, although according to the parents the IG and CG did not differ in problem behavior (CBCL Aggressive Behavior), the level of observed child's conduct problems was significantly higher in the IG than in the CG.

Second, although a relatively low inclusion criterion (the 80th percentile on the CBCL aggressive behavior scale) was used in this study, the mean percentile of the aggressive behavior scale was 93, indicating the selection of a group of children at risk for conduct disorders. Nevertheless, it is possible that we included a number of false positives (Offord and Bennett 2002), and this may have been one of the factors that affected the intervention effect (Hill et al. 2004). Accurate identification of children at risk for the development of a chronic pattern of conduct problems is essential for effective prevention interventions, but extremely difficult to obtain (Hill et al. 2004).

Third, the enrollment rate in the intervention group was relatively low. This might be due to the inclusion criterion (80th percentile), but even more by the place of recruitment of the families. Enrollment rates are higher in Incredible

Years studies with a high inclusion threshold, e.g., 77% in a study with the 95th percentile CBCL score as inclusion criterion (Barrera et al. 2002) and 24% in a study in which the inclusion criterion was a score above the median of the ECBI (Patterson et al. 2002). However, in the latter study families were recruited in general practices while in four other studies with high enrollment rates families were recruited at schools (August et al. 2001; Barrera et al. 2002; Reid et al. 2007; Scott et al. 2010). Recruitment of families at schools probably makes participation in a program easier for parents. Parents, for example, are invited to coffee mornings to learn about the study and the parenting program (Scott et al. 2010). When families from one class are invited to participate because of a relatively high behavior problem score of their child, as a group they are probably more inclined to participate than when they would have been invited separately as was the case in the present study.

Fourth, our results might have been biased by the high educational level of the parents who participated in the present study. Therefore, these findings have limited generalizability to lower educated parents. Finally, since the outcome measures used in the present study lack clear cut off points for clinical levels of functioning, it is not possible to draw conclusions about the clinical significance of the findings.

The present study provides evidence for improvements in parenting skills and observed child behavior resulting from the IY parent program used as an indicated preventive intervention. We showed a mediating effect of critical parenting on the child's conduct problems and it appeared that parental influence on child problems behavior increased over time. Although families participating in this study were followed for 2 years, we cannot draw conclusions with respect to the long-term effects of the IY parent program as a preventive intervention yet. Since the present study showed sustained effects on observed conduct problems at 2 year follow up and a mediation effect of parenting skills on the child's misbehavior, we regard the IY parent program as a promising intervention for the prevention of conduct disorders. However, follow up assessments in middle childhood and adolescence are needed.

Acknowledgements We are grateful to the parents and children who participated in this study. We thank Karlijn Steggehuis, Marte van Horst and Nienke Willering for delivering the parent groups and for conducting part of the home visits. We thank Ben van Hout for his contribution to the design of this study. The original research project was funded by ZonMw Prevention (#2620.0001).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: an introductory review. *Child and Adolescent Mental Health, 8*, 136–143.
- 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*, 614–626.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
- Barrera, M., 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*, 83–94.
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, O., et al. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*, 443–451.
- Boggs, S. R., Eyberg, S. M., & Reynolds, L. A. (1990). Concurrent validity of the Eyberg Child Behavior Inventory. *Journal of Clinical Child Psychology, 19*, 75–78.
- 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*, 386–396.
- Brotman, L. M., O'Neal, C. R., Huang, K. Y., 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*, 235–245.
- Coté, S. M., Vaillancourt, T., LeBlanc, J. C., Nagin, D. S., & Tremblay, R. E. (2006). The development of physical aggression from toddlerhood to pre-adolescence: a nationwide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology, 34*, 71–85.
- DeGarmo, D. S., Patterson, G. R., & Forgatch, M. S. (2004). How do outcomes in a specified parent training intervention maintain or wane over time? *Prevention Science, 5*, 73–89.
- Dutra, L., Campbell, L., & Westen, D. (2004). Quantifying clinical judgment in the assessment of adolescent psychopathology: reliability, validity, and factor structure of the Child Behavior Checklist for clinician report. *Journal of Clinical Psychology, 60*, 65–85.
- Eddy, J. M., Dishion, T. J., & Stoolmiller, M. (1998). The analysis of change in children and families: methodological and conceptual issues embedded in intervention studies. *Journal of Abnormal Child Psychology, 26*, 53–69.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg child behavior inventory & sutter-eyberg student behavior inventory – revised*. Psychological Assessment Resources. Odessa, FL: Psychological Assessment Resources.
- Eyberg, S. M., & Robinson, E. A. (1981). Dyadic parent-child interaction coding system: A manual. Unpublished Manuscript. Portland: Oregon Health Sciences University
- Farrington, D. P. (2005). Childhood origins of antisocial behavior. *Clinical Psychology and Psychotherapy, 12*, 177–190.
- Frick, P. J., & Loney, B. R. (2000). The use of laboratory and performance-based measures in the assessment of children and adolescents with conduct disorders. *Journal of Clinical Child Psychology, 29*, 540–554.
- 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*, 1123–1132.
- Glasnapp, D. R. (1984). Change scores and regression suppressor conditions. *Educational and Psychological Measurement, 44*, 851–867.
- Hill, L. G., Lochman, J. E., Coie, J. D., Greenberg, M. T., & The Conduct Problems Prevention Research Group. (2004). Effectiveness of early screening for externalizing problems: issues of screening accuracy and utility. *Journal of Consulting and Clinical Psychology, 72*, 809–820.
- 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 trial. *British Medical Journal, 334*, 678–685.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, 44*, 1116–1129.
- Kazdin, A. E., & Weisz, J. R. (Eds.). (2003). *Evidence-based psychotherapies for children and adolescents*. New York: Guilford Press.
- Kim-Cohen, J., Caspi, A., Moffitt, T. E., Harrington, H., Milne, B. J., & Poulton, R. (2003). Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry, 60*, 709–717.
- Kratochwill, T. R., Elliott, S. N., Loitz, P. A., Sladeczek, I., & Carlson, J. S. (2003). Cojoint consultation using a self-administered manual and videotape parent-teacher training: effects on children's behavioral difficulties. *School Psychology Quarterly, 18*, 269–302.
- Luk, E. S. L., Staiger, P. K., Mathai, J., Wong, L., Birlson, P., & Adler, R. (2001). Children with persistent conduct problems who dropout of treatment. *European Child & Adolescent Psychiatry, 10*, 28–36.
- Maughan, B., & Rutter, M. (2001). Antisocial children grown up. In J. Hill & B. Maughan (Eds.), *Conduct disorders in childhood and adolescence* (pp. 507–552). Cambridge: Cambridge University Press.
- McCart, M. R., Priester, P. E., Davies, W. B., & Azen, R. (2006). Differential effectiveness of behavioral parent training and cognitive-behavioral therapy for antisocial youth: a meta-analysis. *Journal of Abnormal Child Psychology, 34*, 527–543.
- Offord, D. R., & Bennett, K. J. (2002). Prevention. In M. Rutter & E. Taylor (Eds.), *Child and adolescent psychiatry* (pp. 881–899). Oxford: Blackwell.
- Patterson, G. R. (1982). *Coercive family process*. Eugene, OR: Castilia.
- Patterson, J., Barlow, J., Mockford, C., Klimes, I., Pyper, C., & Stewart-Brown, S. (2002). Improving mental health through parenting programmes: block randomized controlled trial. *Archives of Disease in Childhood, 87*, 472–477.
- Raaijmakers, M., Koffijberg, H., Posthumus, J., Van Hout, B., Van Engeland, H., & Matthys, W. (2008). Assessing the performance of a randomized versus a non-randomized study design. *Contemporary Clinical Trials, 29*, 293–303.
- Raaijmakers, M. A., Posthumus, J. A., Van Hout, B. A., Van Engeland, H., & Matthys, W. (2011). Cross-sectional study into the costs and impact on family functioning of 4-year-old children with aggressive behavior. *Prevention Science, 12*, 192–200.

- 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*, 605–620.
- Rich, B. A., & Eyberg, S. M. (2001). Accuracy of assessment: the discriminative and predictive power of the Eyberg Child Behavior Inventory. *Ambulatory Child Health, 7*, 249–257.
- Robinson, E. A., & Eyberg, S. M. (1981). The dyadic parent-child interaction coding system: standardization and validation. *Journal of Consulting and Clinical Psychology, 49*, 245–250.
- Scott, S., Knapp, M., Henderson, J., & Maughan, B. (2001a). Financial costs of social exclusion: follow-up study of antisocial children into adulthood. *British Medical Journal, 323*, 191–194.
- Scott, S., Spender, Q., Doolan, M., Jacobs, B., & Aspland, H. (2001b). Multicentre controlled trial of parenting groups for childhood antisocial behaviour in clinical practice. *British Medical Journal, 323*, 1–7.
- Scott, S., O'Connor, T. G., Futh, A., Matias, C., Price, J., & Doolan, M. (2010). Impact of a parenting program in a high risk, multi-ethnic community: the PALS trial. *Journal of Child Psychology and Psychiatry, 51*, 1331–1341.
- Shaw, D. S., Lacourse, E., & Nagin, D. S. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *Journal of Child Psychology and Psychiatry, 46*, 931–942.
- Society for Prevention Research (2005). *Standards of evidence: Criteria for efficacy, effectiveness and dissemination*. <http://www.preventionresearch.org>
- Taylor, T. K., Schmidt, F., Pepler, D., & Hodgins, H. (1998). A comparison of eclectic treatment with Webster-Stratton's Parents and Children Series in a Children's Mental Health Center: A randomized controlled trial. *Behavior Therapy, 29*, 221–240.
- Tremblay, R. E., Nagin, D. S., Séguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., et al. (2004). Physical aggression during early childhood: trajectories and predictors. *Pediatrics, 114*, 43–50.
- Webster-Stratton, C. (1994). Advancing videotape parent training: a comparison study. *Journal of Consulting and Clinical Psychology, 62*, 583–593.
- Webster-Stratton, C. (1998). Preventing conduct problems in Head Start children: strengthening parenting competencies. *Journal of Consulting and Clinical Psychology, 66*, 715–730.
- Webster-Stratton, C. (2001a). *Parenting practices interview*. Unpublished assessment instrument. (<http://www.son.washington.edu/centers/parenting-clinic/forms.asp>).
- Webster-Stratton, C. (2001b). *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., & Hammond, M. (1997). Treating children with early-onset conduct problems: a comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65*, 93–109.
- Webster-Stratton, C., & Hancock, L. (1998). Training for parents of young children with conduct problems: Content, methods, and therapeutic processes. In C. E. Schaefer & J. M. Briesmeister (Eds.), *Handbook of parent training* (pp. 98–152). NY: John Wiley & Sons.
- Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). *Prevention Science, 2*, 165–192.
- 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*, 283–302.
- Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early conduct problems: intervention outcomes for parent-, child- and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33*, 221–239.