UNIVERSITY OF AMSTERDAM

My dopamine has been busy: Research on gene by environment interactions in child externalizing behavior

Chhangur, R.R.

Publication date
2016
Document Version
Final published version

Link to publication

Citation for published version (APA):
Chhangur, R. R. (2016). *My dopamine has been busy: Research on gene by environment interactions in child externalizing behavior*. [Thesis, fully internal, Universiteit van Amsterdam].

5 INTERVENTION EFFECTIVENESS
OF THE INCREDIBLE YEARS:
NEW INSIGHTS INTO SOCIODEMOGRAPHIC
AND INTERVENTION-BASED MODERATORS

CHAPTER 5

INTERVENTION EFFECTIVENESS OF THE INCREDIBLE YEARS: NEW INSIGHTS INTO SOCIODEMOGRAPHIC AND INTERVENTION-BASED MODERATORS

ABSTRACT   —   **We tested the effectiveness of the preventive behavioral parent training (BPT) Incredible Years (IY) and the effects of previously suggested sociodemographic and intervention-based moderator variables (i.e., initial severity of externalizing behavior, child gender, social economic status, family composition, and number of sessions parents attended), in a large scaled randomized-controlled-trial. Using full intention-to-treat analyses, questionnaire and observation data from 387 parents and children aged 4-8 years (Mage = 6.21, SD = 1.33; 55.30 % boys), across pretest, posttest, and 4-month follow-up, were analyzed, correcting for multiple testing. IY was successful in decreasing parent-reported child externalizing behavior (Cohen's d = .20 at posttest; d = .08 at follow-up), increasing parent-reported positive parent behavior (d = .49; d = .45) and observed behavior (d = .06; d = .02), and decreasing parent-reported negative parenting behavior (d = .29; d = .25). No intervention effects were found for reported and observed child prosocial behavior, observed child externalizing behavior, and observed negative parenting behavior. Out of 40 tested moderation effects (i.e., eight outcomes times five moderators), only three significant moderation effects appeared. Thus, no systematic evidence emerged for moderation of IY effects. The present multi-informant trial demonstrated that many previously suggested moderators may not be as potent in differentiating BPT effects as once thought.**

Negative parenting behaviors and strategies, such as disapproval, inconsistent discipline, harshness, and psychological control have been related to externalizing behavior in children and adolescents (e.g., Bor, Sanders, & Markie-Dadds, 2002; Collins, Maccoby, Steinberg, & Hetherington, Bornstein, 2000; Ge, Brody, Conger, Simons, & Murry, 2002; Karreman et al., 2006; Rothbaum & Weisz, 1994). Negative parenting behaviors have also been found to mediate the relation between more distal family risk factors (e.g., socioeconomic status, parental psychopathology) and child externalizing behavior (Dodge, Coie, & Lynam, 2006; Reid, Patterson, & Snyder, 2002). In contrast, positive parenting behaviors and strategies, such as acceptance, appropriate discipline, responsiveness, and limit setting have been related to child prosocial behavior (Zahn-Waxler, Iannotti, Cummings, & Denham, 1990). The most effective interventions aimed at reducing externalizing and promoting prosocial child behavior have therefore been designed to target both negative and positive parenting behaviors (McCart et al., 2006). Specifically, behavioral parent training programs (BPT) use parents as agents by training them in using parenting strategies that create positive changes in parenting behavior and, through this, indirectly lead to positive changes in child behavior.

Effectiveness of BPT in reducing child externalizing behavior and promoting prosocial behavior has been proven in multiple independent studies, but effect sizes are moderate (McCart et al., 2006; Menting et al., 2013; Reyno & McGrath, 2006). One way to boost effectiveness is by taking into account factors that determine intervention effectiveness. However, our insights into such moderators is limited due to scarcity of studies and several specific methodological issues, such as small sample sizes, low statistical power, and assessment of effects that are limited to questionnaire information. Moreover, the direction of the suggested moderators is unclear and to date the moderators have been studied separately in different studies, or inferred from comparisons of effect sizes between studies. In order to control for possible confounding effects of the different moderators, assessment within a single RCT with sufficient statistical power to do so is required. This will be the scope of the present paper – building not only on parent-reports but also on observational data on parenting and child behaviors. Because we measured the degree of beneficial effects of Incredible Years in a "real world" prevention setting, working through health care institutions, this study can be seen as an effectiveness trial.

BPT is an effective method to reduce child externalizing behavior and promote child prosocial behavior in different populations (for meta-analyses see McCart et al., 2006; Menting et al., 2013; Sandler et al., 2011). However, the mean effect size of BPT in indicated prevention settings is relativeley modest (d =.20), compared to the mean effect size in treatment settings (d = .50) (McCart et al., 2006; Menting et al., 2013; Reyno & McGrath, 2006). In addition, the effectiveness of BPT programs appears to be influenced by sociodemographic moderators (Gardner, Hutchings, Bywater, & Whitaker, 2010; Scott & O'Connor, 2012) and intervention-based moderators (Wilson & Lipsey, 2001). Recent meta-analyses on both the IY program (Menting et al., 2013),

as well as on BPT in both prevention and intervention settings more broadly defined (Leijten, Raaijmakers, Orobio de Castro, & Matthys, 2013; Lundahl, Risser, & Lovejoy, 2006), suggested that specifically initial severity of child externalizing behavioral problems, child gender, socioeconomic status (i.e., SES), family composition (i.e., single parent families vs. two-parent families), and the number of sessions parents attended might be important moderators of the intervention effects of the BPT program IY.

Initial severity of externalizing behavior

Initial severity of child externalizing behavior is one of the strongest predictors of intervention effects (Leijten et al., 2013; Lundahl et al., 2006; Menting et al., 2013). However, moderation has been found in both directions. On the one hand, initial severity might account for a threshold that confers advantage in terms of intervention effectiveness; larger initial severity leaves more room for improvement (Gardner et al., 2010). Also, larger initial severity might be related to increased motivation in parents to change, leading parents to more readily agree to and engage in treatment (i.e., larger treatment adherence). On the other hand, it has also been found that initial severity of externalizing behavior could reduce intervention responsivity (e.g., Kazdin, 1995; Ruma, Burke, & Thompson, 1996). Possibly, this is because more severe levels of externalizing behavior are related to increased numbers of child, parental and environmental risk factors (e.g., comorbid psychopathology, severe child dysfunctioning, parental stress, and parents' perception of failing), that in turn negatively affect parents' motivation and engagement in BPT. Initial severity of externalizing behavior might therefore be a specifically important moderator of effectiveness in a prevention setting, where severity might vary more between families, compared to treatment settings.

Child gender

Child gender might be another moderator of BPT effectiveness. A previous study suggested that the effects of BPT in a prevention setting were stronger for boys than for girls (Gardner et al., 2010, but see review McMahon, Wells, & Kotler, 2008) for conflicting results). However, the influence of child gender as a moderator might partly be due to confounding effects of initial severity (i.e., boys showing more externalizing behavioral problems than girls) (Menting et al., 2013). It might therefore be specifically important to control for initial severity when testing for possible moderating effects of gender.

SES

Although the meta-analysis by Lundahl and colleagues (2006) suggested that BPT is less effective for economically disadvantaged families, a recent meta-analysis showed that—when initial externalizing behavior was controlled for—economically disadvantaged and advantaged families benefited equally from the intervention

efforts directly post intervention (Leijten et al., 2013). SES and initial problem severity are likely to be confounded. Therefore, the unique influence of SES is unknown.

Family composition

It has been suggested that single parents possibly benefit less from BPT programs compared to two-parent families, which may be linked to limited financial resources, fewer coping resources, and/or greater isolation in single parent families (e.g., Griffin, Botvin, Scheier, Diaz, & Miller, 2000). However, findings on single parenting as a moderator of BPT effectiveness in indicative prevention and treatment settings are inconclusive (e.g., Kazdin, 1995; Reyno & McGrath, 2006, but see Fossum, Mørch, Handegård, Drugli, & Larsson, 2009; Gardner et al., 2010) for conflicting results). Single parents and/or parents from a low SES background might be less able to attend sessions due to a lack of social and economic resources to meet preconditions for attending, such as transportation and child care. Again, this moderator might therefore be confounded with other possible moderators (SES and number of sessions attended).

The number of intervention sessions parents attended

Meta-analyses by Wilson and Lipsey (2001) and Menting and colleagues (2013) suggested that a higher amount of sessions parents attend is positively related to effect sizes of BPT in both prevention and treatment settings. BPT programs teach specific parenting techniques and during training sessions parents have opportunities to see how such techniques can be implemented, practiced, and refined. Missing one or more training sessions means missing specific intervention content. The number of sessions parents attend might therefore be relevant for intervention success (i.e., dosage effect). However, the number of sessions attended might be associated with other moderators, specifically SES and family composition. Therefore, it is important to investigate possible confounding effects of different moderators.

Methodological limitations of previous studies

Besides conflicting findings about the direction and unique (vs. confounding) effects of moderators of intervention effectiveness, the reviewed findings should be viewed as preliminary because of several methodological limitations that have plagued previous studies (see Weersing & Weisz, 2002). Specifically, most previous research on preventive BPT programs relied exclusively on parent reports of both parenting and child behavior. However, these reports might be biased and confounded (Sessa, Avenevoli, Steinberg, & Morris, 2001; Stifter, Willoughby, & Towe-Goodman, 2008). By providing a blinded assessment of changes in parenting and child behavior, observations besides questionnaires have important methodological advantages, (Daley et al., 2014; Scott, 2001; Sonuga-Barke et al., 2013). In addition, most intervention studies have a modest sample size. For instance, studies incorporated in the meta-analysis

on IY intervention effects by Menting and colleagues (2013) contained on average 95 families. This is problematic given that small sample sizes lead to an increased risk for both Type 1 error (i.e., incorrectly concluding there is an effect) and Type 2 error (i.e., concluding there is no effect when one actually exists). This is especially the case in moderation analyses where the sample is split up in multiple subgroups (Sullivan & Feinn, 2012). Therefore, a rigorous evaluation trial of preventive BPT is required, including observational assessments of both parenting and child behavior, with a sufficiently large sample size to test the effects of specific sociodemographic and intervention-based moderators. The current study tested moderation (i.e., initial severity of externalizing behavior, child gender, SES, family composition, and number of sessions parents attended) of the effectiveness of the BPT program IY in an indicated preventive context, by assessing parent reports and observational data on both child and parenting behavior, within one multivariate model, controlling for the possible confounding effects of the moderators. This study can mainly be seen as an effectiveness trial as it was conducted in conditions of routine clinical practice.

METHODS

Design

The current study is a randomized controlled indicated prevention trial with two conditions (intervention vs. control) and three measurement waves (pretest, posttest, and follow-up). It was built up in two stages. In stage one, all families with children aged 4-8 in the targeted municipalities were invited for a screening. In stage two, all eligible families were invited to participate in an RCT: the Observational Randomized Trial on Childhood Differential Susceptibility (i.e., The ORCHIDS study). Enrolled families participated in the following three waves: pretest before randomization; posttest immediately after the intervention (i.e., 4 months after pretest procedure); follow-up 4 months after intervention (i.e., 8 months after pretest procedure). Randomization to either control or experimental condition (1:1) occurred after pretest and consent to participate: An independent researcher drew a ticket (which read either control or experimental condition) that was put back afterwards. Both assessors and parents were blind to allocation status at initial assessment.

Screening

Families were screened and recruited through community records via two Dutch regional health care organizations. All families with children aged 4-8 years (N = 20,048) of four (i.e., two large and two small) municipalities received a personalized information letter, including a consent form and the screening questionnaire (i.e., Eyberg Child Behavior Inventory; Eyberg & Pincus, 1999). Families were offered €7.50 for returning the questionnaire within two weeks. A total of 5,876 questionnaires were returned timely (response rate 22.52 %). Children scoring at or above the 75th

percentile of their relative cohort (i.e., sum score of 112 for girls and 120 for boys aged 4 and 5, 107 for girls and 116 for boys aged 6 to 8; 110 for girls and 115 for boys aged 4 and 5, 106 for girls and 112 for boys aged 6 to 8, for the two cohorts respectively) were eligible for participation in the study (N = 1,524). One parent-child dyad per family (N = 1,393) was invited to participate. Parents of either sex and of any ethnic group (mastering the Dutch language) were eligible. Eligible families received an invitation letter to participate in the RCT. One week later parents were individually contacted by a researcher or trained research assistant who briefly explained the study process. We were able to reach approximately 61% (N = 850) of eligible families, of which 46% agreed to participate (see Figure 1 for an overview on the selection-process of participants). The ECBI intensity scores of participating and non-participating children slightly differed (F (1, 1.40) = 6.66, p = .01), in that parents' perceptions of children's externalizing behavior were higher in participating families (M = 3.65; SD = .45) than in families who did not participate (M = 3.58; SD = .46).

Participants

In total, 387 parent-child dyads eventually participated in the RCT. Children were between 4 and 8 years of age at baseline (Mage = 6.31, SD = 1.33), mostly born in The Netherlands (97.4%), and about half of them (55.30%) were boys. Participating parents (91% mothers) were between 23 and 51 years of age at baseline (Mage = 38.10, SD = 4.84), mostly born in The Netherlands (i.e., 86% of mothers, 84% of fathers), and about half of them completed a higher form of education (i.e., higher vocational training or university level educational tracks) (see Tables 1-3 for demographic and descriptive statistics). For descriptive purposes of our sample, parents also reported—on a 3-point scale (1 = *not true* to 3 = *certainly not true*)—on levels of child peer problems (M = .41; SD =.29), conduct problems (M = .45; SD = .37), emotional problems (M = .66; SD = .48), hyperactivity (M = 1.16; SD = .53), and prosocial behavior (M = 1.34; SD = .42) at pretest (i.e, Strength and Difficulties Questionnaire; Goodman, 1997). About a third (28.6%) of participating families received additional (mental health or family) care or help (e.g., mental health care parents or social services) and 8% of children used psychoactive medication (mainly psychostimulants) between pretest and posttest.

Randomization check

Participants in the intervention and control condition did not significantly differ in age (child or parent), gender (child or parent), country of birth (child or parent), parental education level, work status, marital status, religion, parent-reported and observed parenting behavior, and parent-reported child behavior at baseline (ps > .06) (see Tables 1-3 for descriptive statistics). Observed negative child behavior significantly differed between the two conditions (F (1, 1.61) = 5.40, p = .02), indicating that children in the intervention condition scored higher on observed negative behavior (M = .52; SD = .62) compared to children in the control condition (M = .39; SD = .46). This difference was controlled for in all analyses.

Figure 1. Participant Flow

Screening

Recruited families
(n = 20,048 children)

No response
(n = 14,172)

Returned questionnaire timely
(n = 5,876)

Families with more than one eligible child
(n = 78 families)

Not eligible (i.e., ECBI score below 75th percentile of cohort)
(n = 4,405)

Informed about ORCHIDS study
(n = 1,393)

Unable to contact
(n = 543)

Not willing to participate
(n = 463)

Signed consent
(N = 387)

Allocation

Allocated to control
(n = 190)

Allocated to intervention
(n = 197)

Assessment

Observation:
Pretest: n = 190
Posttest: n = 178
Follow-up: n = 178

Questionnaires:
Pretest: n = 189
Posttest: n = 180
Follow-up: n = 180

Observation:
Pretest: n = 197
Posttest: n = 172
Follow-up: n = 181

Questionnaires:
Pretest: n = 196
Posttest: n = 178
Follow-up: n = 182

Dropouts

During the study 28 families dropped out, of which 23 at posttest and 5 at follow-up. Reasons for dropping out were inability to reach parents, (upcoming) divorce of parents, and/or moving house. There was no difference between conditions in the number of families that dropped out of the study (p = .19). When comparing parents that participated in all three waves with parents who dropped out, no significant differences were found at baseline regarding reported and observed measurements (ps > .09). Also, we found no significant differences regarding sociodemographic and intervention variables except for marital status (χ^2 (4, N = 386) = 11.30, p = .02) and mother's education level (χ^2 (8, N = 386) = 21.52, p < .01). Mothers who participated in all three waves were more likely to be married (71% vs. 51% married) and higher educated (81% vs. 51% high educated), compared to parents who dropped out during the study.

Table 1. Sample Demographics

		Total	(N = 387)	E	(n = 197)	C	(n = 190)
				Condition			
Age child (M; SD)		6.31;	1.33	6.30;	1.36	6.32;	1.29
Gender child	Boy	55.30		57.90		52.60	
	Girl	44.70		42.10		47.40	
Country of birth mother	The Netherlands	86.00		84.80		87.40	
	Other	14.00		15.20		12.60	
Country of birth father	The Netherlands	84.00		82.80		85.80	
	Other	16.00		17.20		14.20	
Age parent (M; SD)[a]		38.10;	4.84	38.02;	4.90	38.17;	4.79
Gender parent[a]	Male	8.00		7.60		8.40	
	Female	91.00		92.40		91.60	
Education mother[b]	Low	21.20		23.50		18.90	
	Medium	27.50		27.00		27.90	
	High	50.50		48.00		52.60	
Education father[b]	Low	25.60		25.00		26.30	
	Medium	26.20		28.10		50.30	
	High	45.60		44.90		46.40	
Employment parent[a]							
	Employed	73.60		75.60		71.60	
	Not working:	26.40		24.40		28.40	
	Unemployed	23.30		19.60		26.40	
	Stay at home parent	44.40		43.50		45.50	
	Other	32.30		36.90		28.10	
Marital status parent[a]	Married/ living together	87.0		85.70		88.40	
	Single	8.80		11.20		6.30	
	Other	4.10		3.10		5.30	
Children in family (M;SD)		2.23;	0.82	2.30;	0.87	2.24;	0.78
Religion	Christian	39.15		42.30		36.85	
	Islamic	5.45		5.10		5.75	
	Non	34.20		34.05		33.45	
	Other	21.20		18.55		23.95	

Note. [a]*descriptive statistics participating parent only.* [b]*low = completed middle or high school; medium = completed vocational training; high = completed higher vocational training or university level educational tracks.*

Procedure

At each measurement wave, parent-child interactions were filmed during a structured play situation and parents filled out a digital questionnaire. During pretest researchers or trained research assistants took time to explain the study in more detail, to answer questions, and in turn asked parents to sign the informed consent form. During this wave, pretest questionnaire data were collected, parent-child interactions during structured play situations were videotaped, and saliva samples for genotyping were collected (Chhangur, Weeland et al., 2012). During posttest and follow-up assessment, the observation and questionnaire procedures were repeated. In addition, parents were interviewed by a trained researcher about children's genetic ancestry (i.e., country of birth great-grandparents) and family mental health care

(i.e., use of psychopharmacotherapy children, psychosocial treatment, family care etc.) received during the study. Participating families received €20,- for the first two home visits and €40,- for the third home visit. The Institutional Review Broad in The Netherlands (METC UMC Utrecht, protocol number 11-320/K) approved the study.

Questionnaire measures

Parenting Practice Inventory (PPI)

The PPI measures parenting skills and discipline styles of parents with young children 6-12 years (Webster-Stratton, 2001a). The PPI consists of 15 sections, each containing multiple items, asking for a response of the parent to children's misbehavior, appropriate behavior, and several statements. Parents answered these questions and responded to these statements using different scales. In total, four summary scales were extracted from this questionnaire: harsh and inconsistent discipline (15 items, e.g., 'Threatening but not punishing'), positive verbal discipline (9 items, e.g., 'Discussing the problem with the child'), physical punishment (6 items, e.g., 'Slapping or hitting when misbehavior occurs'), and praise and incentives (11 items, e.g., 'Giving a hug or compliment'). Statements about parenting were excluded, because we were interested in the actual behavior of parents. To assess positive parenting behavior, we combined the dimensions positive verbal discipline and praise and incentives. To assess

Table 2. Descriptive Statistics of Child and Parent Behavior Measures

Note. ªEyberg Child Behavior Inventory; ᵇSum scores; ᶜDyadic Parent-Child Interaction Coding System; ᵈMatson Evaluation of Social Skills with Youngsters; ᵉParenting Practice Inventory.

	Pretest						Posttest						Follow-up					
	Total		Intervention		Control		Total		Intervention		Control		Total		Intervention		Control	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Child behavior																		
Reported externalizing behavior[a]	133	19.27	135	19.92	131	18.41	126	19.38	124	18.96	127	19.68	122	20.10	121	20.83	123	19.39
Observed externalizing behavior[c]	.46	.55	.52	.62	.39	.46	.42	.57	.40	.54	.43	.60	.39	.44	.39	.43	.39	.45
Reported prosocial behavior[d]	.64	.61	.64	.66	.64	.57	.65	.62	.65	.65	.64	.58	.69	.64	.69	.62	.73	.51
Observed prosocial behavior[c]	1.33	.64	1.34	.67	1.32	.61	1.30	.71	1.29	.66	1.30	.75	1.40	.76	1.43	.77	1.37	.75
Parent behavior																		
Reported negative behavior[e]	2.75	.61	2.80	.62	2.70	.59	2.48	.63	2.39	.58	2.57	.66	2.47	.56	2.40	.56	2.54	.56
Observed negative behavior[c]	1.56	1.06	1.54	1.05	1.58	1.07	1.41	.94	1.29	.86	1.53	1.01	1.47	1.04	1.41	1.01	1.53	1.08
Reported positive behavior[e]	4.65	.62	4.70	.63	4.67	.61	4.72	.66	4.88	.65	4.57	.64	4.70	.64	4.86	.61	4.55	.68
Observed positive behavior[c]	2.38	.96	2.43	.95	2.32	.97	2.55	1.05	2.79	1.12	2.30	.93	2.53	1.09	2.80	1.21	2.26	.88

Table 3. Correlations between Sociodemographic and Intervention-based Moderators

*Note. child gender: 0 = boys; 1 = girls; family composition: 0 = living together; 1 = single parent. aonly applicable for intervention group. *p < .05, **p < .01. None of the groups differed significantly on the moderators with independent samples t test or χ² -test.*

	1	2	3	4	5
1. Initial severity of externalizing behavior	-				
2. Child gender (boys %)	-.205**	-			
3. Social economic status	.060	-.043	-		
4. Family composition (single parent %)	.105*	-.058	-.134**	-	
5. Number of session parents attended	.044	-.040	.163*	-.046	-

negative parenting behavior, we combined the scales harsh and inconsistent discipline and physical punishment. Reliability for both scales was satisfactory on all measurements (positive parenting behavior α > .70; negative parenting behavior α > .78).

Eyberg Child Behavior Inventory (ECBI)

The ECBI assesses the occurrence of conduct problems in children aged 2 to 16 years (Eyberg & Pincus, 1999). We used the ECBI intensity scale consisting of 36 items, which measures the frequency of the problem behavior (e.g., 'Acts defiant when told to do something') on a 7-point scale (1 = *never* to 7 = *always*). Reliability of the intensity scale was good for all three measurements (α > .84).

The Matson Evaluation of Social Skills with Youngsters (MESSY)

he MESSY assesses social skills in school-aged children (Matson et al., 1983). The questionnaire consists of 62 items measured on a 5-point Likert scale (1 = *not at all* to 5 = *very much*) assessing the frequency of 2 types (i.e., prosocial and aggressive) of behavior in a range of social situations. In the current study, the scale appropriate social behavior was selected to measure prosocial behavior. The scale consists of 20 items (e.g., 'Sticks up for friends') and reliability was good for all three measurements (α > .88).

Observational measure

Dyadic Parent–child Interaction Coding System (DPICS)

The DPICS is a home observational measure for parent–child interactions, which assesses the quality of the social interaction (Robinson & Eyberg, 1981; Webster-Stratton, 1989). Parent and child were observed for 20 minutes while playing with a fixed set of toys at pretest, posttest, and follow-up. The observation procedure

consisted of 4 five-minute periods: free play (i.e., to get used to being videotaped); child directed play (i.e., child picked a toy and directed the session); parent directed play (i.e., parent picked a toy and directed the session); and clean up (i.e., parent had to make the child clean up). For the last three periods, positive and negative behavior of both parent and child were coded using different categories; 7 for positive parent behavior (i.e., acknowledgement, descriptive question, descriptive comment/ encouragement, unlabeled praise, labeled praise, parent positive affect, and physical positive); 6 for negative parent behavior (i.e., physical intrusive, physical negative, critical statement, negative command, indirect command with no opportunity to comply, direct command with no opportunity to comply); 5 for positive child behavior (i.e., compliance to an indirect command, compliance to a direct command, child positive affect nonverbal, child positive affect verbal, and child psychical warmth); 4 for negative child behavior (i.e., non-compliance to an indirect command, non-compliance to a direct command, cry-whine-yell, and smart talk). The reliability of the parenting scales was α > .60 for positive parenting; α > .67 for negative parenting; α > .49 for positive child behavior; and α > .56 for negative child behavior. This is comparable to previous studies (e.g., Posthumus et al., 2012). Despite the modest reliabilities for the two observed child behavior scales, we still included these variables in our data analyses for transparency reasons (see for a priori hypotheses our protocol paper Chhangur, Weeland et al., 2012).

The observations were coded by trained research assistants who were not involved in the study and who were blind to condition and measurement wave. Monthly calibration meetings were held to prevent observer drift. To provide estimates of interrater reliability, a random 237 of the 1161 observations (20%) were independently coded by two coders. Coders were unaware of which observations were used to assess observer agreement. Intraclass correlation (ICCs) was assessed through a two-way mixed effects model using SPSS 22.0. Interrater reliability was excellent at all measurement waves; ICCs ranged from .96 to .97 for positive parenting; from .85 to .93 for negative parenting; from .82 to .93 for positive child behavior; and from .70 to .83 for negative child behavior.

Intervention: The Incredible Years BASIC program

The IY program is developed to prevent and intervene in the development of child externalizing behavior by building a warm parent-child relationship through increasing positive parenting strategies such as child directed play, social and emotion coaching, praise and incentives and decreasing negative parenting strategies such as being critical and inconsistent (Webster-Stratton, 2001b). IY is a group behavioral parent training program consisting of 15 weekly sessions. The program starts with the focus on positive parenting strategies such as play, praise and incentives, before discussing effective limit setting, ignoring unwanted behavior, and finally time out strategies. During the sessions parents watch video-vignettes of parents and children interacting (in our study Dutch subtitles were used in the vignettes), act in role-plays,

have brainstorm sessions, and exchange experiences and ideas in small groups. After each vignette, the group leader asks questions to stimulate discussion about what parents found particularly (in)effective and to practice alternative responses. Parents are encouraged to role-play new skills in front of the group and in small subgroups during the meeting. Before each session, parents read a book chapter on the topic of that particular session. Additionally, they receive home assignments to practice the discussed skills at home. Parents are also assigned a "buddy" (i.e., another parent of the same group), which they call weekly to check in with and discuss successes and difficulties with the new learned skills. The program uses a collaborative setting, in which group leaders establish themselves as facilitators rather than as experts. Group leaders encourage parents to solve problems and to help each other solve problems in order to ensure maintenance of the intervention effects.

Fourteen IY intervention groups (consisting of 8-15 parents) were delivered across three different Dutch municipalities (i.e., large urban city, medium urban city, and a suburban area). The groups consisted of 14 weekly 2-hour sessions and a "booster" session one month after termination of the program (i.e., total of 15 sessions). Every group was led by two group leaders who had followed a commensurate three-day basic training. All main leaders had a background in clinical child psychology, had experience running IY groups before the study commenced, and were officially certified by The Incredible Years, Inc. (two group leaders got certified during the study). Parents completed weekly satisfaction questionnaires to ensure the session contents address the specific goals parents have. To boost attendance, child-care was arranged for parents who attended the course during office hours. Parents were compensated for travel costs when needed. At the start of the program all parents received the program book and an IY magnet. During the sessions group leaders provided the groups with coffee, tea, and snacks. Candy or stickers (parent's choice) were handed out during the sessions as a reward for active participation. During the sessions on 'tangible rewards' leaders also brought small surprise rewards for parents (worth approximately 1 euro, such as stickers, stamps, or small games, which they had to blindly grab from a "treasure bag"). At the last session, all parents received a certificate, a personal felicitation talk, and a plant as a reminder of the program.

Analyses

Preliminary analyses

The preliminary analyses showed that there were no outliers but that the data were skewed. Therefore, we used maximum likelihood robust (MLR) to deal with non-normally distributed data. Also, two-level MwiN models were run as preliminary analyses to test the nesting of families within intervention groups (i.e., families formed a part of fourteen groups). There was no variance at the group level at pretest, and at posttest variance did not exceed 4% of total variance. Therefore, group-level was not included in the final models.

Main analyses

In the primary analyses, latent growth curve modeling (LGCM) in Mplus (Muthén & Muthén, 2008-2015) was used to assess the development of observed and reported parenting and child behavior across pretest, posttest, and follow-up assessments. LGCM estimates individual growth for each child or parent separately, which is an excellent approach for examining variation in the development of the outcome variables, while controlling for baseline levels. Model fit is considered good if the Root Mean Square Error of Approximation (RMSEA) is < .05 and Confirmatory Fit Index (*CFI*) values are >.95 (Hu & Bentler, 1999). To calculate effect sizes of the intervention effectiveness Cohen's *d* was used, where $d \geq .20$ is considered a small effect, $\geq .50$ as a moderate effect and $\geq .80$ as a large effect. In total, 197 parents were assigned to the IY intervention groups, of whom 44 decided not to participate in the program or never attended a session. We found no differences on any pretest measures (*ps* > .09) between parents who actively participated in the intervention and parents who did not. Therefore, the 44 allocated-to-intervention families who did not attend any sessions were still included in the analyses. In total, we assessed eight separate outcome measures (i.e., four parent-reported and four observed) in the full intention-to-treat analyses. Benjamini-Hochberg False Discovery Rate correction (Benjamini & Hochberg, 2005) was used to correct for chance capitalization.

After assessment of the intervention effectiveness, moderator variables (i.e., initial severity of externalizing behavior, child gender, SES measures by parental education, family composition, and number of sessions parents attended) were assessed using eight multivariate mixture models, controlling for possible confounding effects of the moderator variables. The effects of the moderator variables on the slope of the outcome variables were estimated for the two classes (intervention and control) separately, since the variable 'number of sessions attended' is only relevant in the intervention class.

RESULTS

Intervention integrity

A total of 197 parents were assigned to the IY intervention groups. Active participants attended on average 11.01 (*SD* = 3.69) out of 15 sessions. Of these participants, 74% attended at least 10 sessions and 84% at least half of the sessions. If parents missed a session, group leaders called them to discuss the content and sent them home assignments. If parents missed three subsequent sessions, a home visit was scheduled by the trainers to discuss the missed program content with the parent(s). Besides the IY intervention about a third of the families in the experimental condition (31.30%) received additional (mental health) care or help between baseline and posttest, which was significantly less than families in the control condition

Table 4. Intervention Effects of Reported and Observed Child and Parent Behavior

Note. df = degree of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation. As χ² < df, the CFI is set to 1.0 and RSMEA to .001, which makes it sufficient to read off whether the p value is not significant. p values are corrected with Benjamini-Hochberg False Discovery Rate correction. *p < .05. **p < .01. *** p < 001

Outcome	Intercept B (SD)	Slope B (SD)	Corrected p value	Effect sizes Cohen's d Posttest	Effect sizes Cohen's d Follow-up	χ² (df)	CFI	RMSEA	p value
Child behavior									
Reported externalizing behavior	.085 (.05)	-.125 (.04)**	<.01	0.20	0.08	8.08 (2)	0.98	0.09	
Observed externalizing behavior	.122 (.08)*	-.091 (.07)	.28	0.06	0.02	3.77 (2)	0.96	0.05	
Reported prosocial behavior	.012 (.06)	.019 (.03)	.62	0.08	0.07	0.57 (2)	1.00	<.001	.75
Observed prosocial behavior	.012 (.07)	-.008 (.06)	.90	0.06	0.02	7.12 (4)	0.94	0.05	
Parent behavior									
Reported negative behavior	.073 (.06)	-.175 (.04)***	<.001	0.29	0.25	18.25 (4)	0.95	0.10	
Observed negative behavior	-.087 (.10)	-.063 (.08)	.58	0.06	0.02	6.35 (4)	0.98	0.04	
Reported positive behavior	.055 (.06)	.186 (.04)***	<.001	0.49	0.45	6.41 (2)	0.99	0.08	
Observed positive behavior	.130 (.10)	.280 (.06)***	<.001	0.06	0.02	16.93 (4)	0.96	0.09	

(49.58%) (χ² (1, *N* = 350) = 3.71, *p* = .05). The IY program can only be implemented by officially certified group leaders. To ensure and monitor implementation fidelity group leaders followed a treatment manual for each session (i.e., standardized manuals and session guidelines, check-lists, books, DVDs, handouts, tangible rewards etc.) and completed weekly protocol checklists of standards (e.g., vignettes, brainstorms, and role-plays) to be covered in each session. Checklists showed that on average 70.4% of the standards were executed by trainers. It has been suggested that that positive intervention effects are often obtained with levels of program integrity exceeding 60% (Durlak & DuPre, 2008). Furthermore, co-leaders filled out standardized peer-feedback forms and parents filled out standardized feedback forms. All sessions were videotaped for feedback. Group leaders received ongoing supervision, feedback, and training throughout the study. Treatment integrity of IY overall is very high because of the close monitoring, standardized materials, and comprehensive training manuals (see Webster-Stratton & Hammond, 1997).

Intervention effects

Child externalizing behavior

For parent-reported child externalizing behavior on the ECBI intensity scale, condition proved non-significant at intercept (B_0 = .085, p = .12) but proved significant at slope (B_1= -.125, p =. 001, corrected p = .002) (χ^2 [df = 2, N = 387] = 8.08, CFI = 0.98, RMSEA = .09, d = .08), indicating that reported levels of child externalizing behavior did not differ between control and intervention group at pretest, but that parents in the intervention group reported a significantly larger decrease in child externalizing behavior at posttest and follow-up compared to parents in the control group (see Table 4). For observed child externalizing behavior, after correction for multiple testing, condition proved neither significant at intercept (B_0 = .122, p = .12) nor at slope (B_1= -.091, p = .18, corrected p = .28) (χ^2 [df = 2, N = 382] = 3.77, CFI = 0.96, RMSEA = .05, d = .02), indicating that observed levels of child externalizing behavior between the intervention group and control group did not significantly differ at pretest or over time.

Child prosocial behavior

For parent-reported child prosocial behavior, condition proved neither significant at intercept (B_0 = .012, p = .85) nor slope (B_1 = .019, p = .54, corrected p = .62) (χ^2 [df = 2, N = 387] = 0.57, CFI = 1.00, RMSEA <.001, p = .75, d = .07). Also, for observed child prosocial behavior, condition proved neither significant at intercept (B_0 = .012, p = .85) nor slope (B_1 = .008, p = .90, corrected p = .90) (χ^2 [df = 4, N = 382] = 7.12, CFI = .94, RMSEA = .05, d = .02). Thus, reported and observed levels of child prosocial behavior between the intervention group and control group did not significantly differ at pretest or over time (see Table 4).

Negative parenting behavior

For parent-reported negative parenting behavior condition proved non-significant at intercept (B_0 = .073, p = .23) but proved significant at slope (B_1 = -.175, p <.001, corrected p <.001) (χ^2 [df = 4, N = 387] = 18.25, CFI = .95, RMSEA = .10, d = .25) (see Table 4), indicating that the control and intervention group did not differ on negative parenting behavior at pretest, but that parents in the intervention group reported a significantly stronger decrease of negative parenting behavior over time compared to the control group. For observed negative parenting behavior, condition proved neither significant at intercept (B_0 = -.087, p = .40) nor slope (B_1 = -.063, p = .44, corrected p = .58) (χ^2 [df = 4, N = 382] = 6.35, CFI = .98, RMSEA = .04, d = .02), indicating that change in observed levels of negative parenting behavior did not significantly differ between groups.

Table 5. Moderator Variables of Slopes in Intervention and Control Group

*Note. *p < .05. **p < .01. *** p < .001.*

	Initial severity of problem behavior		Child gender		Number of sessions attended		SES		Family composition	
	Intervention B (SD)	Control B (SD)	Intervention B (SD)	Control B (SD)	Intervention B (SD)	Control B (SD)	Intervention B (SD)	Control B (SD)	Intervention B (SD)	Control B (SD)
Child behavior										
Reported externalizing behavior	-.121 (.06)*	-.031 (.05)	-.010 (.06)	.045 (.05)	-.003 (.00)	-	-.014 (.04)	.035 (.07)	.085 (.09)	
Observed externalizing behavior	-.005 (.10)	-.113 (.11)	.143 (.11)	-.089 (.10)	.000 (.00)	-	-.156 (.08)*	.115 (.14)	-.212 (.24)	
Reported prosocial behavior	.000 (.07)	.035 (.06)	-.059 (.05)	-.019 (.04)	.007 (.00)	-	.007 (.03)	-.007 (.08)	-.077 (.09)	
Observed prosocial behavior	.023 (.09)	.008 (.09)	.203 (.09)*	.015 (.08)	-.003 (.00)	-	-.000 (.00)*	.137 (.11)	-.171 (.10)	
Parent behavior										
Reported negative behavior	-.017 (.06)	-.040 (.06)	.042 (.05)	.034 (.0.5)	-.012 (.00)***	-	.005 (.04)	.087 (.04)*	.061 (.08)	.034 (.09)
Observed negative behavior	-.116 (.12)	.039 (.13)	.037 (.12)	.070 (.13)	-.001 (.01)	-	.000 (.00)	-.094 (.04)*	-.018 (.25)	-.065 (.23)
Reported positive behavior	-.030 (.05)	.012 (.06)	-.009 (.05)	-.152 (.06)	.011 (.00)**	-	-.008 (.04)	-.040 (.04)	.045 (.09)	-.132 (.12)
Observed positive behavior	-.148 (.11)	-.063 (.10)	.058 (.09)	.093 (.08)	.018 (.01)**	-	.000 (.00)	.086 (.06)	-.070 (.15)	-.096 (.15)

Positive parenting behavior

For parent-reported positive parenting behavior condition proved non-significant at intercept (B_0 = .055, p = .38) but significant at slope (B_1 = .186, p <.001, corrected p <.001) (χ^2 [df = 2, N = 387] = 6.41, CFI = .99, RMSEA = .08, d = .45) (see Table 4). Also, for observed positive parenting behavior, condition proved not significant at intercept (B_0 = .130, p = .17) but significant at slope (B_1 = .280, p < .001, corrected p <.001) (χ^2 [df = 4 , N = 382] = 16.93, CFI = .96, RMSEA = .09, d = .02). Thus, parent-reported and observed positive parenting behavior did not differ at pretest, but parents in the intervention group reported and showed a significantly stronger increase of positive parenting behaviors over time compared to parents in the control group.

All significant effects survived correction for multiple testing. In addition, we also performed a "completers-only" analyses (i.e., only including families who attended at least one IY session). The results remained the same as the results of the intention-to-treat analyses (see Supplementary Material, Table 1a).

Moderators of intervention effects

Table 3 shows correlations between the moderator variables. Of these variables, initial severity of externalizing behavior was significantly negatively correlated with child gender (r = -.21, p < .05) and family composition (r = -.21, p < .05), indicating that single

parents and parents of boys reported more initial severity of externalizing behavior. SES was significantly negatively correlated with family composition (r = -.13, p < .01) and positively correlated with number of sessions parents attended the intervention group (r = .16, p < .05), indicating that low SES families included more single parents and attended less intervention sessions when allocated to the intervention group. As planned, these correlations between moderator variables were taken into account to control for their possible mutually confounding effects.

Initial severity of externalizing behavior

Initial severity of child externalizing behavior was a significant predictor of the slope of parent-reported externalizing behavior over time in the intervention group (B_1 = -.121, p = .04) but not in the control group (B_1 = -.031, p = .57). This indicates that parents who reported higher levels of initial severity of child externalizing behavior (at screening), reported a larger effect of the intervention on externalizing child behavior over time (see Table 5). However, comparison of the coefficients of the control and intervention group showed, that the coefficients were not significantly different from each other (t = 1.15, df = 374, p = .25). Initial severity did not influence intervention effects on observed child externalizing behavior, reported and observed child prosocial behavior, reported and observed negative parenting behavior, and reported and observed positive parenting behavior. Thus, no moderation of intervention effects by initial severity was found.

Child gender

Child gender predicted the slope of observed child prosocial behavior over time in the intervention group (B_1 = .203, p = .03) but not in the control group (B_0 = .015, p = .85), indicating that girls in the intervention group showed a larger increase in prosocial behavior compared to boys in the intervention group. However, comparison of the coefficients of the control and intervention group showed, that the coefficients were not significantly different from each other (t = .27, df = 374, p = .79). Child gender did not influence intervention effects on reported and observed child externalizing behavior, reported child prosocial behavior, reported and observed negative parenting behavior, and reported and observed positive parenting behavior. Thus, no moderation of intervention effects by child gender was found.

SES

SES (parental education) was a predictor of the slope of reported negative parenting behavior in the control group. Low SES families in the control group reported more negative parenting over time (B_0 = .087, p = .03) but there was no such effect of SES on slope of reported negative parenting in the intervention group (B_1 = .005, p = ns.). However, comparison of the coefficients of the control and intervention group showed, that the coefficients were not significantly different from each other

(t = 1.45, df = 364, p = .15). Also, we found that SES was a predictor of the slope of observed child externalizing behavior in the control condition (B_0 = -.156, p = .04) but not in the intervention condition (B_1 = .000, p = .00). This indicates that in the control condition higher SES predicted lower levels of child externalizing behavior over time. However, comparison of the coefficient of the control and intervention group showed, that the coefficients were not significantly different from each other (t = 1.56, df = 374, p = .12). In addition, SES was a predictor of the slope of observed child prosocial behavior in the intervention condition ($B1$<.000, p = .000) but not in the control condition (B_0 = - .020, p = ns.). However, comparison of the coefficients of the control and intervention group showed, that the coefficients were not significantly different from each other (t = .33, df = 374, p = .74). Thus, no moderation of intervention effects by SES was found.

Family composition

Family composition was not a significant predictor of the slope of any of the outcomes, indicating that single parent and two-parent families benefited equally from the IY intervention.

The number of intervention sessions parents attended

Number of attended IY sessions predicted the slope of parent-reported negative parenting behavior (B_1 = -.012, p <.001), parent-reported positive parenting behavior (B_1 = .011, p = .001), and observed positive parenting behavior (B_1 = .018, p = .001) in the intervention group. These results indicate that parents who attended more IY sessions reported a higher decrease in negative and higher increase in positive parenting behavior and also showed a larger increase in observed positive parenting behavior than parents who attended less IY sessions. The number of intervention sessions parents attended did not influence intervention effects on reported and observed child externalizing behavior, reported and observed prosocial child behavior, reported positive parenting behavior, and observed negative parenting behavior.

DISCUSSION

Previous research demonstrated that the behavioral parenting training (BPT) Incredible Years (IY) is effective in preventing externalizing behavior. In addition, studies suggested that specific sociodemographic and intervention-based factors (i.e., initial severity of externalizing behavior, child gender, social economic status (SES), family composition, and number of sessions attended) may influence the intervention effectiveness of the program. However, the effects of these moderators were mostly studied in isolation rather than in multivariate analyses, inhibiting insight into "real" moderation effects when different moderators are controlled for each other. Also, these effects were studied in modestly sized samples (e.g., average N = 95; Menting et al.,

2013) that predominantly relied on parental reports of intervention effectiveness. Therefore, in a large randomized-controlled-trial (RCT; 387 parents and their children aged 4-8 years) we tested the unique contribution of multiple moderators within one multivariate model, controlling for their possible confounding effects – based on parent reports as well as observational data. Our results demonstrated that IY was successful in decreasing parent-reported child externalizing behavior (posttest: $d = .20$; follow-up: $d = .08$), in decreasing parent-reported negative parenting behavior (posttest: $d = .29$; follow-up: $d = .25$), and in increasing parent-reported (posttest: $d = .49$; follow-up: $d = .45$) and observed positive parenting behavior (posttest: $d = .06$; follow-up: $d = .02$).

Intervention effects were neither found for observed child externalizing behavior and observed negative parenting behavior nor for reported and observed child prosocial behavior. Therefore, even though a recent meta-analysis found convincing proof that IY is effective in preventing observed child externalizing behavior ($d = .37$ for all 23 studies, $d = .35$ for the four indicative prevention studies; Menting et al., 2013), we did not establish this effect in the present study. In interpreting these findings it is important to note that reliability of the two observed child behavior scales was relatively low and that we adhered to stringent controls for multiple testing. This resulted in low power to detect effects on observed child behavior. We found no effects of IY on child prosocial behavior. This is contradictory to the outcomes of a recent meta-analysis by Menting and colleagues (2013), who found IY to be effective in reducing externalizing behavior as well as increasing child prosocial behavior. This might be partly explained by differences in how prosocial behavior is operationalized and assessed between studies.

Our results illustrate the importance of using multi-informant data when assessing intervention effects. Including both reported and observed data on parent and child behavior gave a more complete picture of changes herein. Specifically, although we found significant intervention effects on most parent-reported measures, on the observed measures we only found a significant effect on positive parenting behavior. It has been argued that parents might justify the time and effort they have invested through attending the intervention by reporting a decrease in child externalizing behavior, without the occurrence of an actual change in such behavior (Leijten, Overbeek, & Janssens, 2012). However, both the recent Menting and colleagues (2013) meta-analysis, including 23 studies with observational measures, and a Dutch study by Posthumus and colleagues (2012) did establish effects on observed child behavior. Among those studies the DPICS is often used as observation instrument, however there is little correspondence between studies in which DPICS categories for parent and child behavior are being used. For instance, some studies used a composite score of the categories child smart talk, Cry/Whine/Yell, and physical negative, excluding (Posthumus et al., 2012) or including non-compliance (Webster-Stratton, Reid, & Hammond, 2001) and destructive behavior (Webster-Stratton, 1998), whereas others used separate categories such as non-compliance or physical negative to index

child externalizing behavior (Eyberg et al., 2001). It might be that previous studies also encountered difficulties in forming a reliable scale for this behavior. In our case, inability to create a reliable scale for observed child behavior combined with small variance might have precluded detection of an intervention effect in the present study. Another possible explanation might be differences between the parenting questionnaire and observation measure. The questionnaire asks parents about child behavior over a longer period of time and across different contexts, whereas the observation is conducted in a period of 20 minutes and is restricted to a play-session. Future research could explore whether different effects can be captured with observations using a more extensive timeframe and/or in different settings (e.g., morning routines or mealtimes).

The discrepancy between specifically our findings on observed positive parenting behavior and observed negative parenting behavior, might also suggest that the increase in positive parenting behavior is the most rapidly evoked and/or most robust intervention induced behavioral change. This seems plausible, considering that the IY program highly invests in strengthening the parent-child relationship by advocating positive parenting strategies (such as regular playtime and praise). Furthermore, it might be easier to stimulate parents to increase the use of novel positive parenting behaviors than to change coercive parent-child interaction patterns. In addition, it may be that, although parents already perceive and report a decrease in their own negative parenting behavior and their child's externalizing behavior, these behaviors have not yet observably changed enough to be detected by the limited observation measure at the time of the follow-up.

The simultaneous inclusion of multiple potential moderators in our analyses proved worthwhile, as moderators that have previously been studied in isolation were found to be correlated. Initial severity of externalizing behavior and number of sessions parents attended were the only moderators that influenced significant IY intervention effects. However, we did not find a consistent pattern of moderation across all outcome measures examined. For example, the number of sessions that parents attended moderated the intervention effects on reported negative parenting behavior and reported and observed positive parenting behavior, but not on any child behavior outcome measures. Out of 40 tested moderation effects (i.e., eight outcomes × five moderators), we eventually found three significant moderation effects. This raises the question how potent these independent moderators are in differentiating the effectiveness of the intervention, and consequently, the true clinical relevance of these moderators. The current findings indicate that IY may be an effective intervention, specifically for reducing perceived externalizing child behavior, across a broader range of child and family subgroups.

However, our findings do not rule out that the intervention effects are indeed different for different children and families and/or are influenced by moderators other than the ones currently examined. For example, the operationalization of parenting may

be culture-bound, in that parenting practices related to negative child behavior may differ by ethnicity. The current study had a homogenous sample for which the number of parents not born in the Netherlands was only 16% and the number of parents not born in a Western-European country was less than 13%. As our sample was a predominantly indigenous Dutch sample we were unable to examine ethnicity as a possible moderator. On the other hand, a previous study on IY, targeting Dutch ethnic minority mothers, showed that ethnic minorities benefited equally from IY compared to Dutch families (Leijten, Raaijmakers, et al., 2015). Furthermore, a more theoretically informed search for moderators may be fruitful. We know that externalizing behavior is a very heterogeneous behavioral cluster and has different etiologies in different children (e.g., Frick, 1998). This suggests that intervention effects may depend on the extent to which a specific intervention addresses the specific factors pertinent to the development of these behaviors in individual children. Thus, interesting moderators to investigate might be those factors that are likely to make children more or less susceptible to specific intervention techniques. Some of these factors indicated by previous studies relate to the neurocognitive domain such as inhibitory control (Lochman et al., 2015; Matthys et al., 2012), to child temperament (e.g., Gallitto, 2015; Scott & O'Connor, 2012), and to children's genetic make-up (Bakermans-Kranenburg & Van IJzendoorn, 2015). More insight into "what works for whom" might also help to tailor interventions and to improve their effectiveness.

Using a full intention-to-treat model, the overall effect sizes of the intervention were small ranging between *d* = .06 and *d* = .49 at posttest and *d* = .02 and *d* = .45 at follow-up. However, the effect size on reported externalizing child behavior at follow-up was comparable to meta-analytical findings on the effectiveness of IY in an indicated prevention context (Menting et al., 2013). Another way to further improve effectiveness of prevention programs is by unraveling the active components why interventions are effective (i.e., mediators). One way to gain more insight into such mechanisms of change, is by conducting micro trials (i.e., small-scale randomized experiments using a brief and focused environmental manipulation, designed to target one specific risk mechanism) focused on discrete parenting intervention elements (Collins, Murphy, & Strecher, 2007; Howe et al., 2010; Leijten, Dishion, et al., 2015).

Our findings have to be interpreted in the light of some limitations. First of all, because of the indicated prevention setting we had a large group of parents who were attributed to the intervention but never participated. We used an intention-to-treat model to estimate more realistic intervention effects of IY in a real-world outreaching prevention setting, including those parents allocated to the intervention that did not participate. However, as such an analysis might lead to a conservative estimation of intervention effects we also conducted "completers-only" analyses that showed similar intervention effects (see Supplementary Material). Secondly, our follow-up was on average only 4 months after the intervention. Therefore, we cannot say whether the improvements in parenting and child behavior remain over a longer period of time. However, previous longitudinal and quasi experimental studies do suggest long

89

INTERVENTION EFFECTIVENESS OF THE INCREDIBLE YEARS: NEW INSIGHTS INTO SOCIODEMOGRAPHIC AND INTERVENTION-BASED MODERATORS

Table 1a. Intervention Effects of Reported and Observed Child and Parent Behavior – Completers Only

Note. df = degree of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation. As χ² < df, the CFI is set to 1.0 and RSMEA to .001, which makes it sufficient to read off whether the p value is not significant. p values are corrected with Benjamini-Hochberg False Discovery Rate correction. *p < .05. **p < .01. *** p < 001.

Outcome	Intercept B (SD)	Slope B (SD)	Corrected p value	χ² (df)	CFI	RMSEA	p value
Child behavior							
Reported externalizing behavior	.109 (.06)	-.145 (.04)***	.00	8.08 (2)	0.97	0.12	
Observed externalizing behavior	.086 (.08)	-.077 (.07)	.36	3.77 (2)	0.95	0.04	
Reported prosocial behavior	-.027 (.07)	.007 (.03)	.84	0.57 (2)	1.00	0.00	
Observed prosocial behavior	-.011 (.07)	.037 (.06)	.61	7.12 (4)	1.00	0.00	.92
Parent behavior							
Reported negative behavior	.113 (.06)	-.224 (.04)***	.00	18.25 (4)	0.93	0.17	.42
Observed negative behavior	-.048 (.11)	-.104 (.09)	.36	6.35 (4)	0.99	0.04	
Reported positive behavior	.050 (.07)	.226 (.04)***	.00	6.41 (2)	0.98	0.11	
Observed positive behavior	.174 (.10)	.327 (.07)***	.00	16.93 (4)	0.96	0.10	

term effects of IY up to adolescence (Jones, Daley, Hutchings, Bywater, & Eames, 2008; Posthumus et al., 2012; Webster-Stratton, Rinaldi, & Reid, 2011). More experimental research is needed to confirm these long term findings, but also to further investigate the longitudinal role of the sociodemographic and intervention-based moderators. Third, IY is a highly protocolled program, which overall yields high treatment integrity (Webster-Stratton & Hammond, 1997). Following IY standard procedure, program integrity is measured using group leader self-reported checklists. However, the use of therapist-reported treatment integrity has been criticized (Perepletchikova & Kazdin, 2005). It might be that other measures of integrity (e.g., observational coding of sessions) would yield lower integrity scores. Fourth, our study also only included a small number of single parent and low SES families, which might cause a power issue to detect possible moderator effects of family composition and low generalizability to other samples (see Table 1). Finally, it is worth mentioning that the reliability of observed child behavior was low (α = .49 to .67), which could have led to less reliable estimations of intervention effects. Nonetheless, for reasons of transparency, we decided to report the analyses because they were conducted to test previously published hypotheses (see for a priori hypotheses in Chhangur, Weeland et al., 2012).

Despite these limitations, our current trial may be considered a major step forward

in terms of its large scale, the use of observational data to establish intervention effects on child and parenting behavior, the use of sophisticated statistical analyses, controlling for multiple testing, and high level of attendance at the intervention meetings and little overall attrition (retaining 93% of participants at follow-up). Moreover, we sought to extend recent work in this area by examining multiple moderators indicated by previous meta-analyses and by investigating the unique effects (i.e., controlling for possible confounding effects of different moderators) of these moderators on the intervention effectiveness. Our result show that previously suggested moderators may not be as potent in differentiating BPT effects as once thought. Based on this approach, IY has proven to be an effective prevention strategy to reduce parent perceived child externalizing behavior in a prevention setting.

FUNDING

ACKNOWLEDGEMENTS

SUPPLEMENTARY MATERIAL

Table 1a.