Parent–child interactions and Child Outcomes: Evidence from Randomized Intervention

By Jun Hyung Kim, Wolfgang Schulz, Tanja Zimmermann, and Kurt Hahlweg

Labour Economics (2018)

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 350, Winter 2021

1. Introduction

- This study reports empirical evidence of a causal effect of parent- child interactions during preschool years on adolescent children's behavioral outcomes.

- We focus on two aspects of parent-child interactions—parents' positive engagement with a child and parental discipline.

- Psychologists have long recognized that parents' harsh discipline and poor monitoring lead to negative child outcomes, 1 but economists have only recently begun to study the role of parent-child interactions in the human capital development of children.

- Empirical investigations to date suggest that harshness and warmth during parenting predict adolescents' behavioral outcomes, and explain a portion of the variation in sib- lings' income (Björklund et al., 2010; Dooley and Stewart, 2007; Fiorini and Keane, 2014).

- This study contributes to literature on child human capital development by providing evidence that parent-child interactions play a causal role in behavioral skill development.

- We also show that parent-child interactions can be improved by the intervention we study.

- Well-known, effective early childhood interventions such as Perry Preschool Study, Carolina Abecedarian Project, and Nurse-Family Partnerships involve both home-visit and parenting components (Elango et al., 2015). Current findings suggest that parent-child interactions are mediation channels in these programs.

- We show that the effectiveness of early child- hood intervention is not necessarily limited to low-SES households.

- We argue that extant models of parent-child interactions in economics are inadequate at explaining the link between improved discipline and child outcomes. We propose an alternative framework that accords with psychological theory and empirical findings.

2. Parent–child interaction models in psychology and economics

- We review extant models of parent-child interactions during child- hood from psychology and economics, including implications to this study.

- Psychologists Gerald Patterson and others developed a model in which coercive and inconsistent parenting during childhood leads to
- delinquency in adolescence (Dishion and Patterson, 2015; Granic and Patterson, 2006; Patterson, 1982; Patterson et al., 1992).

- Consider a child who refuses to comply with a parent's effort to modify the child's behaviors.

- In families in which children develop antisocial behaviors, parent-child interactions are characterized by two patterns.

- In the first, a parent uses coercive discipline strategies such as scolding and threatening, and the child responds with hostility.

- The interaction escalates, resulting in feelings of anger and contempt for both.

2.1. A model of parenting skill

- The distinction of the current model from other principal-agent models is that the parent does not have full control over how compensation is realized.

- The parent announces parenting policy $b \in \{0, 1\}$, but the utility transfer the child receives is a Bernoulli random variable B with probability $q(e, b) \equiv Pr(B = 1 | e, b)$, where q increases in e and b.

- For simplicity, assume $q(e, 0) = q(0, b) = 0$ so nothing is transferred to the child if the child chooses no effort or the parent does not announce $b = 1$.

- The child observes the parent's announcement b and solves.

$$\max_{e \in \{0,1\}} -v \cdot e + B \cdot q(e, b)$$

- Specifically, if the parent announces $b = 1$, the child's choice becomes

$$\max\{-v + B \times q(1,1), 0\}.$$

- Parenting skills determine a parent's ability to respond to a child's behavior by compensating for the child's effort when the child chooses $e = 1$.

- If inconsistent or over-reactive, the parent does not always succeed with rewarding pro-social behaviors or disciplining anti-social behaviors appropriately, even if that was the intention.

- These considerations imply $q\ (1,1) < 1$, and $q\ (1,1)$ is low when a parent uses coercive parenting.

- The probability that a parent responds as intended to a child's behavior, $q\ (1,1)$, is interpreted as parenting skill.

- The parent's problem is announcing parenting b to the child, such that

$$\max_{b \in \{0,1\}} -c \cdot b + e^*(b) \times A$$

where $e\ *\ (\ b\ )$ is the optimal response of the child.

3. Experimental design and background

3.1. Background

3.2. Comparison to other parenting interventions

- A strength of this study is that it links parenting intervention in early childhood to adolescent outcomes.

- Based on the most recent meta- analysis of 101 Triple P programs from Sanders et al. (2014), the typical follow-up period is 6 months after intervention, and no other Triple P implementation administered follow-up measures beyond 1 year after intervention.

- The follow-ups suggest that Triple P, on average, has a positive effect on child behaviors and parenting practices.

- They found that greater effect sizes associated with more severe child problems at baseline and using a targeted approach relative to a universal one, though significant effect sizes were also found when using a universal approach.

- The Triple P program can be understood in the context of early child- hood intervention literature since it addresses improvements to children's outcomes by altering early childhood environments.

- High-quality early childhood interventions influence earnings, health, and criminal behaviors of recipients (Elango et al., 2015).

- These findings are based on interventions that feature both direct involvement with children and home-visitation components that encourage parent participation.

- Thus, it is difficult to distinguish direct effects of the program and indirect effects through parents. The current study isolates the parent channel.

3.3. Experimental design

- An intervention was implemented from 2001 to 2002 in Braunschweig, a small, urban city in Germany with a population of 250,000.
- Program staff members presented the program to all 33 preschools in the city, in which administrators of 23 preschools showed interest.
- Seventeen of 23 preschools were selected randomly to participate due to resource constraints.
- The intervention did not target at-risk or low-SES households, unlike most other early childhood interventions.
- The intervention excluded parents who could not understand German, so participants were likely homogeneous regarding cultural backgrounds.
- All parents received information about the program at the preschool and decided whether to participate before randomization. To be eligible, parents had to have a child aged 2.6 to 6 years old.

3.4. Program implementation

- Interventions consisted of 4 weekly training sessions, 2 h each.
- Parents were taught explanations for their children's problematic behaviors, techniques to cope with the behaviors, and supportive strategies for child development.
- These strategies were reinforced through role- playing sessions.
- Subsequent to the group sessions, the parents had the opportunity to hold 4 weekly individual telephone sessions, each 15–20 min long, during which progress, questions, and difficulties that arose regarding the Triple P Training could be discussed.
- These sessions stabilized the implemented strategies and supported generalization to future problems.
- Five licensed trainers led 28 groups of parents during the sessions, usually at the participating preschools. Trainers were trained and tested by the Triple P certification agency. During the program, trainers received supervision weekly to maintain quality.

3.5. Interview procedure and measurement

- During the 10-year follow-ups, all families involved initially were informed of the current project in writing.

- They were asked by phone whether they were willing to participate. Data collection was conducted using a combination of interviews and a written and electronically standardized survey.

- Interviews were conducted during home visits or at the Technical University of Braunschweig. Parents (in 94% of cases, the mothers) and children were interviewed concurrently but in separate locations.

- The interviews were conducted by two interviewers, one for the parent and one for the child (at least one of them with a Master's degree in Psychology, training to become a clinical psychologist [Ger- man: Psychological psychotherapist]).

- Since some questions were sensitive, children were questioned by an interviewer of the same gender. For more sensitive questions, such as problematic behaviors, the respondent used a tablet to answer electronically standardized surveys so that the interviewer could not observe the answers.

3.6. Sample profile

- Attrition was 11% at the 10-year follow-up.
- We focus on parenting behaviors of mothers since in most cases, only a mother attended the program.
- There was 1 single-father family, which was excluded from analysis. The sample included 234 families in 17 preschools.
- Table 1 shows descriptive statistics of baseline characteristics. Pre-treatment characteristics were balanced between intervention and control groups.
- The mean child age was 4, and child gender was split evenly.
- The average mother's age was 35, and about half of the mothers graduated from high school.
- Five percent of intervention and 10% of control group mothers were not married at the baseline. None of the t -tests rejected the null of equal means between intervention and control groups, suggesting the sample was balanced.

Table 1: Baseline Sample Characteristics

	Sample size	Int'vn Mean	Ctrl Mean	*p*-value
Child's age	278	4.03	4.13	0.43
Child's gender(Female)	278	0.48	0.49	0.85
Number of siblings	278	1.04	1.19	0.31
Mother's age	278	35.05	35.62	0.50
Mother's work hours	278	2.78	3.52	0.16
Mother is HS grad	278	0.54	0.55	0.93
Single mother	278	0.05	0.11	0.23
Household income	271	4870.79	4327.96	0.17
CBCL Ext.	275	0.05	-0.10	0.31
CBCL Int.	275	0.08	-0.15	0.22
Mother's harsh discipline	275	-0.05	0.11	0.21
Mother's positive engagement	275	-0.02	0.05	0.53
Father's harsh discipline	200	-0.03	0.08	0.61
Father's positive engagement	200	-0.03	0.07	0.53
(IQ) KABC-mental	267	0.02	-0.05	0.74
(IQ) KABC-sequential	267	0.03	-0.06	0.58
(IQ) KABC-simultaneous	270	0.01	-0.05	0.78
(IQ) KABC-achievement	269	-0.01	0.00	0.94
Recalled parenting-own father	247	0.01	-0.01	0.89
Recalled parenting-own mother	267	0.00	-0.01	0.88

Note: The *p*-value is based on a two-tailed mean *t*-test between intervention and control groups using a wild cluster bootstrap-*t* test with 99,999 replications. The Child Behavior Checklist (CBCL) measured a child's problematic behaviors based on a parent's report, in this case, a mother's. Income is monthly household income based on DEM from 2001 to 2002, when the exchange rate was approximately 1 DEM $\approx$ 0.54 USD. Recalled parenting measured parenting behaviors of a responding mother (i.e., maternal grandparents of a respondent's children). It was constructed by predicting factor scores from measures of warmth, control, and punishment that a mother received. Higher scores mean less warmth, more control, and more punishment received by a mother.

- To assess the representativeness of the sample, Table 2 compares the current sample with a sample of mothers in the nationally representative German Socioeconomic Panel (SOEP) who would have been eligible for the program.

- The high school and college graduation rates, including technical colleges, are much higher, and the share of single mothers is much lower in the current sample.

- Seventeen percent of mothers in the sample work full-time, but only 12% in SOEP do.

- Therefore, although average monthly income is smaller in the sample, it is plausible that the study sample is comprised of higher SES households. Table B.11 is a correlation table of all outcomes assessed during the study.

- The correlation between externalizing and internalizing behaviors was positive at 0.43. The correlation between wellbeing and social relationships was also positive at 0.60.

- Negative behaviors and quality of life (i.e., wellbeing and social relationships) correlated negatively, with values between -0.64 and -0.38.

Table 2: Sample Comparison

	SOEP		Triple P	
	Mean	SD	Mean	SD
Mother's age	35.18	4.94	35.18	4.99
Child is female	0.51	0.50	0.50	0.50
Household income	5090	2330	4685	1620
Work hourr/day	2.51	3.05	3.06	3.17
Working full time	0.12	0.32	0.17	0.38
Working part time	0.41	0.49	0.39	0.49
Single mother	0.17	0.38	0.07	0.26
Graduated high school	0.29	0.45	0.54	0.50
Graduated college	0.16	0.37	0.41	0.49
Sample size	479		271	

Note: A SOEP sample was selected from wave S (2002) of sample E and F, which are nationally representative and independently sampled from other samples in SOEP. Female respondents who are not immigrants and gave birth between 1995 and 2000 were included in the sample, so they approximate the eligible population were Triple P available in their area. Net monthly household income was converted to Deutsche Marks for 2002. High school graduation equaled 1 if respondent's last degree attained was upper secondary or attended college. Full-time work status was defined as working 37 hours or more per week, and part-time status fewer than 37 hours. The Triple P sample was collected from 2001 to 2002. High school graduation was measured using "Abitur " status, which corresponds to high school in the United States. Full-time status was defined as working 7 hours or more per day, and part-time status fewer than 7 hours. Household income was measured in Deutsche Marks in the year it was answered, 2001 or 2002. For both samples, college graduation equaled 1 if the respondent completed technical college or university.

4. Econometric framework and outcome measures

- Randomization was at the preschool level, so each preschool was treated as a cluster. The standard model for evaluation of a randomized experiment describes observed outcome $Y\ i\ k$ of participant $i\ k\ \in I\ k$ by

$$Y_{i_k} = Z_k Y_{i_k}(1) + (1 - Z_k) Y_{i_k}(0) \qquad (1)$$

where $k\ \in \{1\ ,\ \ldots\ ,N\ K\ \}$ is an index for cluster and $I\ k\ =\ \{1\ k\ ,\ \ldots\ ,N\ k\ \}$ a set of participants in cluster k .

- $Z\ k\ =\ 1$ if cluster k was assigned to the intervention group and $Z\ k\ =\ 0$ otherwise.

- Intervention status was the same for all participants in the same cluster. $(\ Y\ i\ k\ (0)\ ,Y\ i\ k\ (1))$ are potential out- comes for participant $i\ k$ in cluster k.

- We test the null hypothesis of no intervention effect, which is equivalent to counterfactual outcomes having the same distribution: $Y\ i\ k\ (0)\ d\ =\ Y\ i\ k\ (1)$, where d = denotes equality in distribution. The intervention effect was estimated using ordinary least squares:

$$Y_{i_k} = a + \beta Z_k + \gamma X_{i_k} + \epsilon_{i_k} \qquad (2)$$

- The goal was to estimate β, the coefficient for assignment status Z k .
- $X\ i\ k$ is a vector of control variables.
- $\epsilon i\ k$ is an individual-specific error term that might correlate within each cluster k, but is assumed to be independent across k .
- To account for clustered error, confidence intervals and p -values were calculated using the wild cluster bootstrap method with 99,999 replications, which maintains cluster structure in each bootstrap sample (see Cameron and Miller, 2015; Davidson and MacKinnon, 2010).

5. Results and discussion

5.1. Intent-to-treat

- We present the intervention effect on parenting behaviors during each of the follow-ups 1 through 4 and plot the effect in Figures 1 and 2 separately for parental discipline and positive engagement.

$$parenting_{i_{k,t}} = a_t + \beta_t Z_{i_k} + \gamma_t X_{i_k} + \epsilon_{i_k,t} \qquad (3)$$

- Fig. 1 shows that changes to parenting behaviors were persistent for at least 4 years after intervention. Improvements to discipline behaviors were immediate after intervention, and an increase in positive engagement behaviors was significant 3 years after intervention.

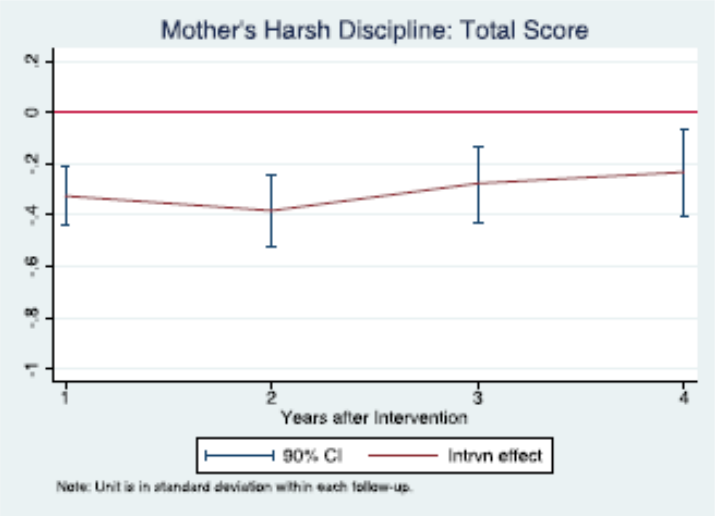
- Analyzing components of discipline separately in Fig. 2 , verbosity, over-reactivity, and laxness were lower in the intervention group, and the reduction in verbosity was most consistent.

- Increases to positive engagement and re- ductions to laxness suggest that improvements to discipline style did not occur at the expense of overall discipline use.

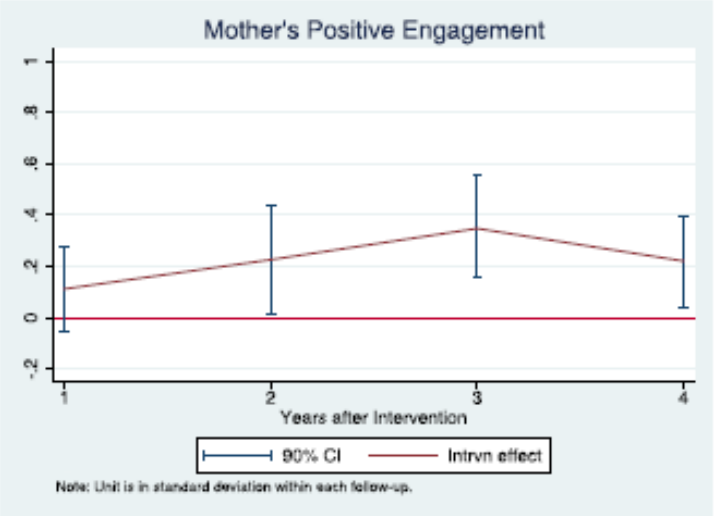
- Table 3 shows intent-to-treat estimates in standard deviations in the sample.
- The intervention effect on externalizing behaviors was -0.31, significant at 10%.
- The effect on internalizing behaviors was positive, contrary to expectations, but the magnitude was small and statistically non-significant.
- The effect on quality of life was 0.161, and the effect on child subjective wellbeing was 0.2, both significant at 10%.
- The effect on a child's social relationships was non-significant.

Figure 1: Parenting Style by Mother and Father

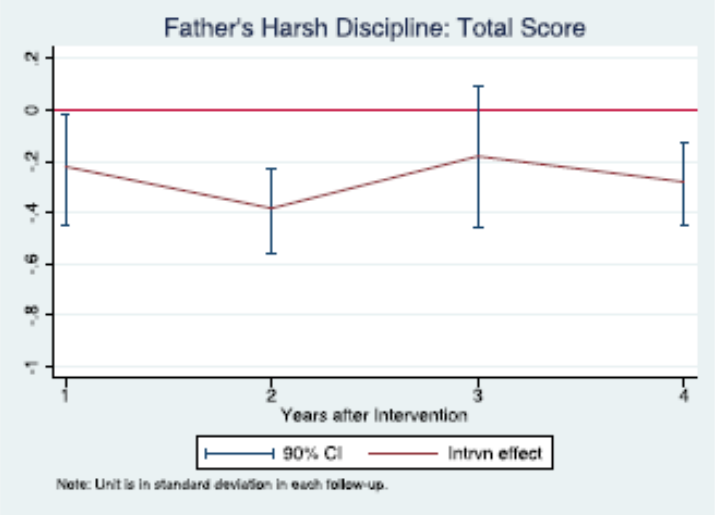
(a)



(b)



(c)



(d)

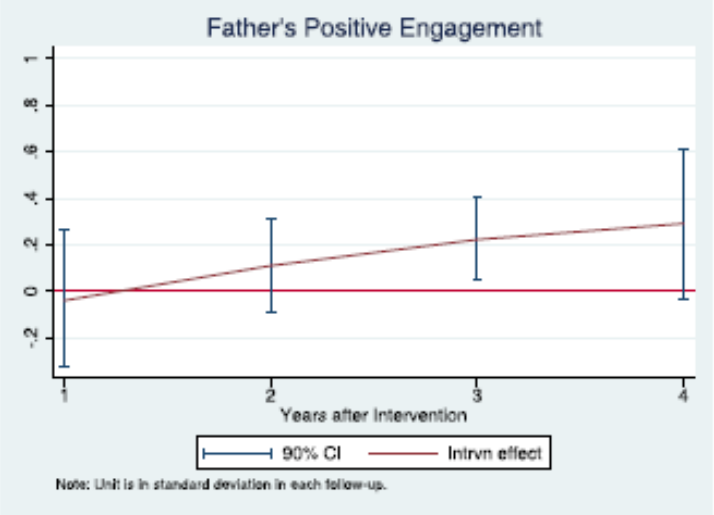
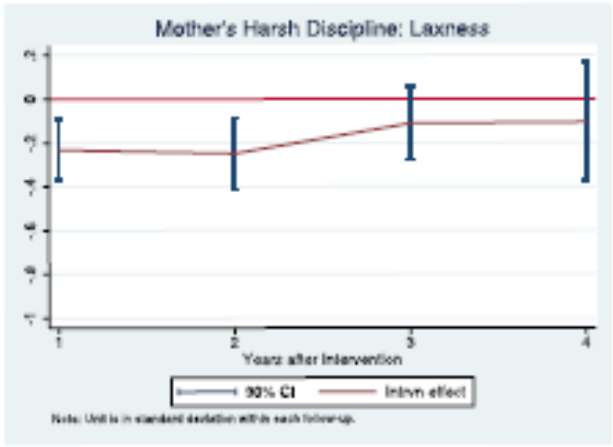
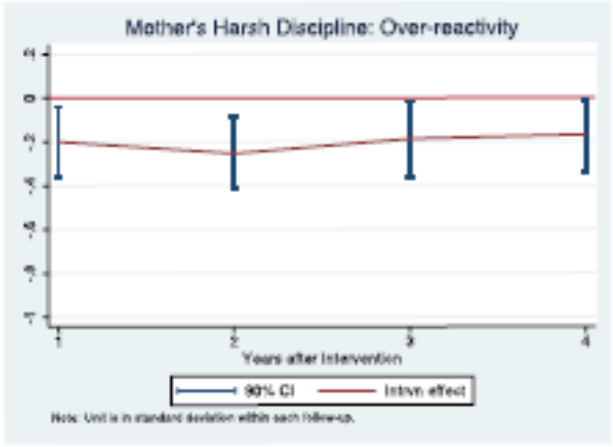


Figure 2: Harmful Discipline Sub-scales by Mother and Father

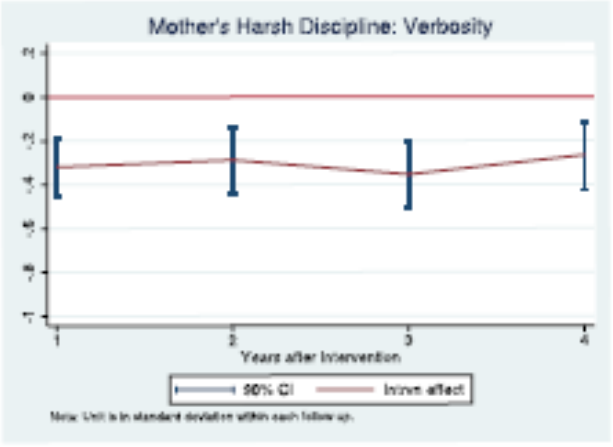
(a)



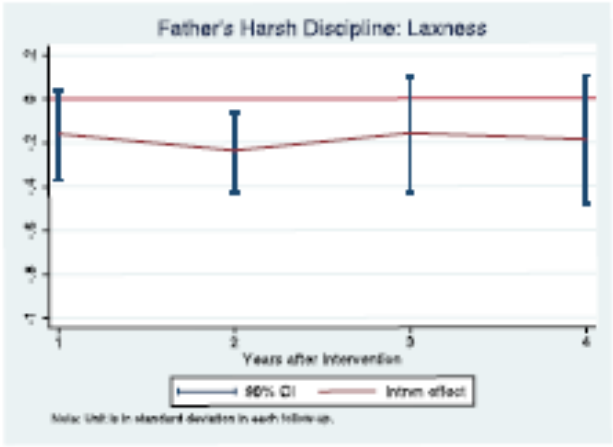
(b)



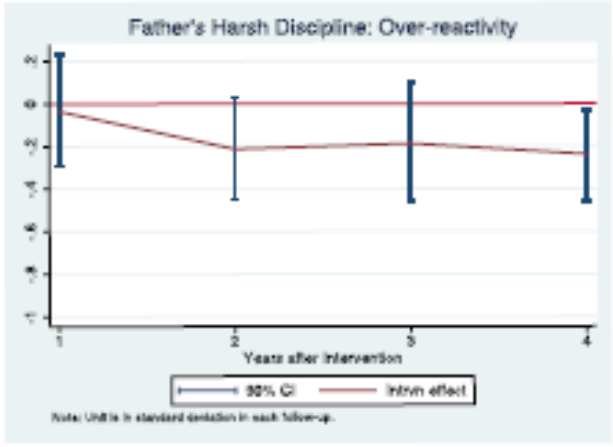
(c)



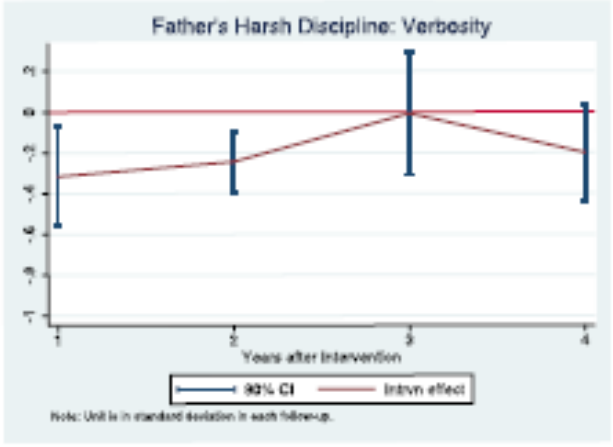
(d)



(e)



(f)



- Table 4 shows the robustness of estimates using various controls.
- The first three columns show that when only demographic variables are included, the effect on wellbeing is significant for both the pooled and female samples.
- The middle three columns show results with only three control variables of baseline skill and parenting measures.
- The effect on externalizing behaviors is significant at 10% for the pooled sample. This effect is now significant for boys, while not so in Table 3.
- The last three columns show that none of the effects is significant when no controls are included, though the magnitudes of the estimates are like those in other models.
- We conclude that the estimates are consistent across models with various controls.

Table 3: Intent-to-Treat

Variable name	Ctrl mean	Intervn effect	Std. error	*p*-Value	R^2
Pooled sample (N = 231)					
Negative behavior total score	-0.004	-0.120	0.157	0.276	0.105
Externalizing behavior	0.135	-0.310	0.151	**0.095**	0.111
Internalizing behavior	-0.085	0.033	0.156	0.427	0.176
Quality of life total score	-0.063	0.161	0.098	**0.081**	0.123
Wellbeing	-0.090	0.200	0.110	**0.069**	0.156
Social relationship	-0.005	0.067	0.076	0.199	0.074
Female sample (N = 111)					
Negative behavior total score	0.119	-0.191	0.199	0.221	0.159
Externalizing behavior	0.045	-0.302	0.183	0.115	0.319
Internalizing behavior	0.260	-0.044	0.205	0.421	0.149
Quality of life total score	-0.226	0.160	0.162	0.176	0.159
Wellbeing	-0.323	0.301	0.151	**0.043**	0.175
Social relationship	-0.048	-0.050	0.181	0.400	0.159
Male sample (N = 120)					
Negative behavior total score	-0.109	0.031	0.159	0.425	0.090
Externalizing behavior	0.212	-0.216	0.180	0.148	0.119
Internalizing behavior	-0.380	0.175	0.138	0.130	0.138
Quality of life total score	0.077	0.092	0.142	0.272	0.104
Wellbeing	0.109	0.064	0.143	0.328	0.104
Social relationship	0.033	0.092	0.153	0.292	0.073

Table 4: Intent-To-Treat with different control variables

Variable name	Effect	*p*-Value	R^2	Effect	*p*-Value	R^2	Effect	*p*-Value	R^2
Pooled sample									
Negative behavior total score	-0.024	0.445	0.026	-0.070	0.346	0.030	0.010	0.523	0.000
Externalizing behavior	-0.221	0.167	0.040	-0.290	**0.081**	0.050	-0.203	0.162	0.009
Internalizing behavior	0.091	0.719	0.084	0.094	0.712	0.016	0.127	0.797	0.003
Quality of life total score	0.130	0.153	0.027	0.085	0.261	0.031	0.088	0.242	0.002
Wellbeing	0.203	**0.091**	0.067	0.100	0.257	0.029	0.131	0.192	0.004
Social relationship	0.001	0.494	0.004	0.031	0.374	0.026	0.001	0.495	0.000
Demographic var.	O			X			X		
Baseline outcomes	X			O			X		
N	245			239			245		
Female sample									
Negative behavior total score	-0.152	0.280	0.031	-0.127	0.287	0.058	-0.086	0.343	0.002
Externalizing behavior	-0.235	0.166	0.049	-0.271	0.131	0.091	-0.202	0.164	0.009
Internalizing behavior	-0.076	0.376	0.030	0.029	0.558	0.010	0.015	0.535	0.000
Quality of life total score	0.189	0.149	0.026	0.147	0.221	0.042	0.135	0.212	0.004
Wellbeing	0.348	**0.034**	0.073	0.213	0.141	0.042	0.206	0.142	0.008
Social relationship	-0.047	0.600	0.031	0.021	0.456	0.046	0.005	0.487	0.000
Demographic var.	O			X			X		
Baseline outcomes	X			O			X		
N	115			113			115		
Male sample									
Negative behavior total score	0.086	0.715	0.032	-0.013	0.468	0.015	0.090	0.719	0.002
Externalizing behavior	-0.203	0.198	0.037	-0.292	**0.090**	0.025	-0.200	0.213	0.009
Internalizing behavior	0.218	0.946	0.045	0.152	0.866	0.036	0.211	0.944	0.014
Quality of life total score	0.070	0.326	0.019	0.047	0.381	0.032	0.054	0.349	0.001
Wellbeing	0.092	0.304	0.027	0.037	0.408	0.028	0.074	0.325	0.001
Social relationship	0.012	0.472	0.011	0.029	0.437	0.024	0.000	0.500	0.000
Demographic var.	O			X			X		
Baseline outcomes	X			O			X		
N	130			126			130		

5.2. Connection to the theoretical model

5.3. Mediation analysis

- Figs. 3–5 show decompositions of intervention effects using linear mediation analysis.

- The intervention was successful at changing a parent's behavior and a child's outcomes, but we are able to argue that the intervention effect is channeled through changes to a parent's behaviors only if experimentally induced changes to a parent's behaviors correlate with experimentally induced changes to a child's outcomes.

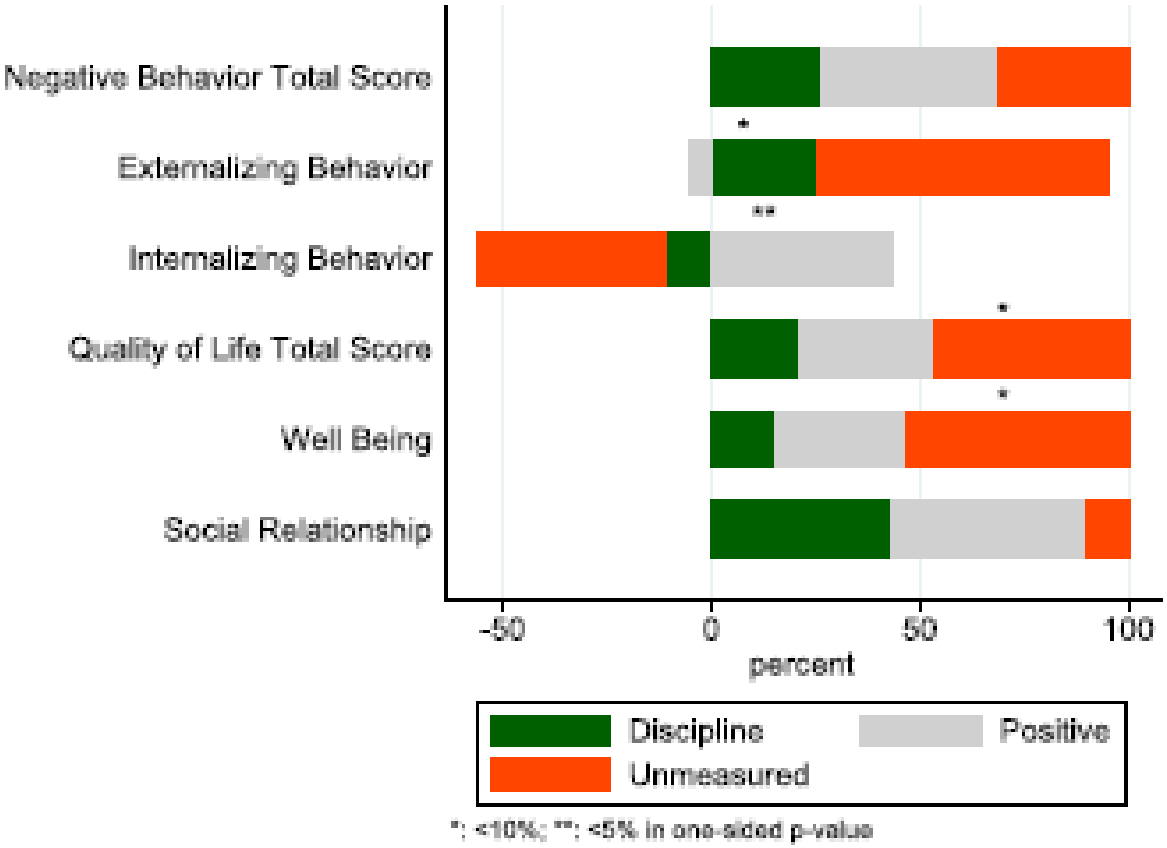
- For each outcome, the intervention effect is decomposed into change in discipline, change in positive engagement, and unobserved channels.

- By unobserved channels we mean all experimentally induced changes not captured by measured parenting changes.

- Mediation analysis tests the validity of theoretical models that un-derlie interpretation of data.

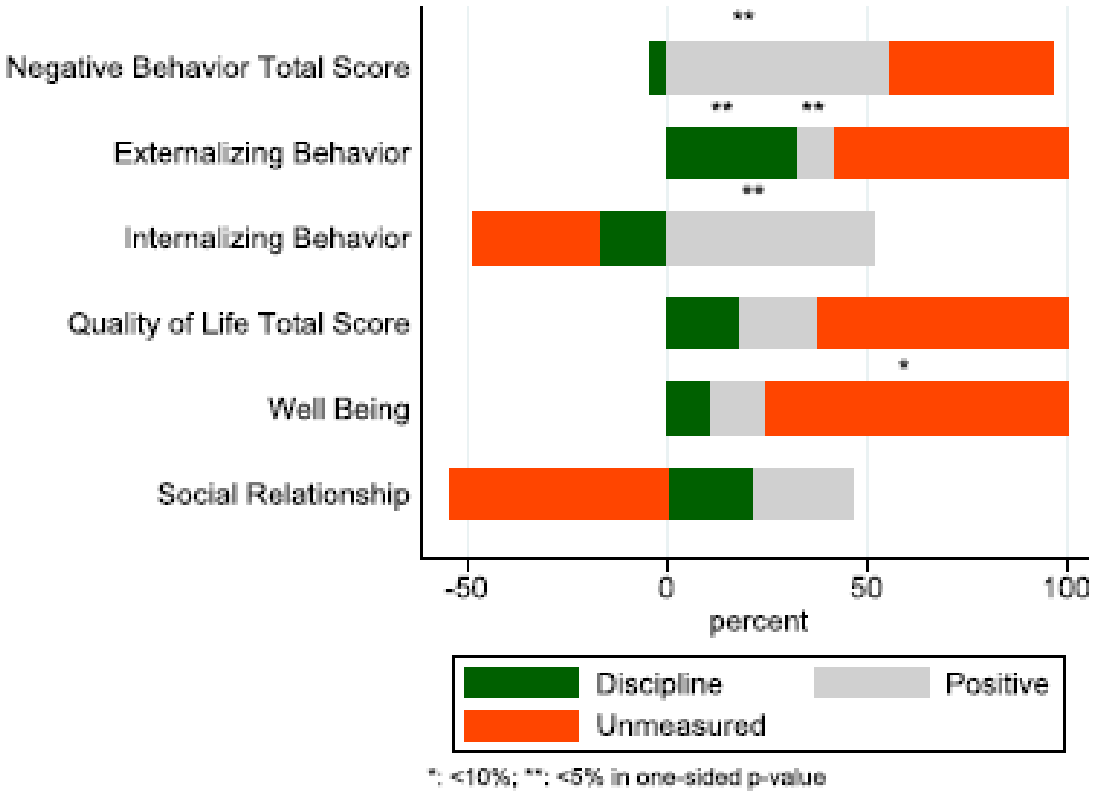
- Each model controls for the same baseline characteristics as the ITT model used to estimate Table 3.
- Fig. 3 shows mediation analysis estimates in the pooled sample. Improvements to discipline and increases to positive engagement both explain intervention effects.
- Lack of results for internalizing behaviors is due to the unmeasured channel and increases in positive engagement contributed to reductions of internalizing behaviors.
- Discipline and positive engagement both played roles in reducing externalizing behaviors for girls.
- For boys, improvements to discipline contributed to reductions in externalizing and internalizing behaviors.

Figure 3: Mediation analysis for pooled sample



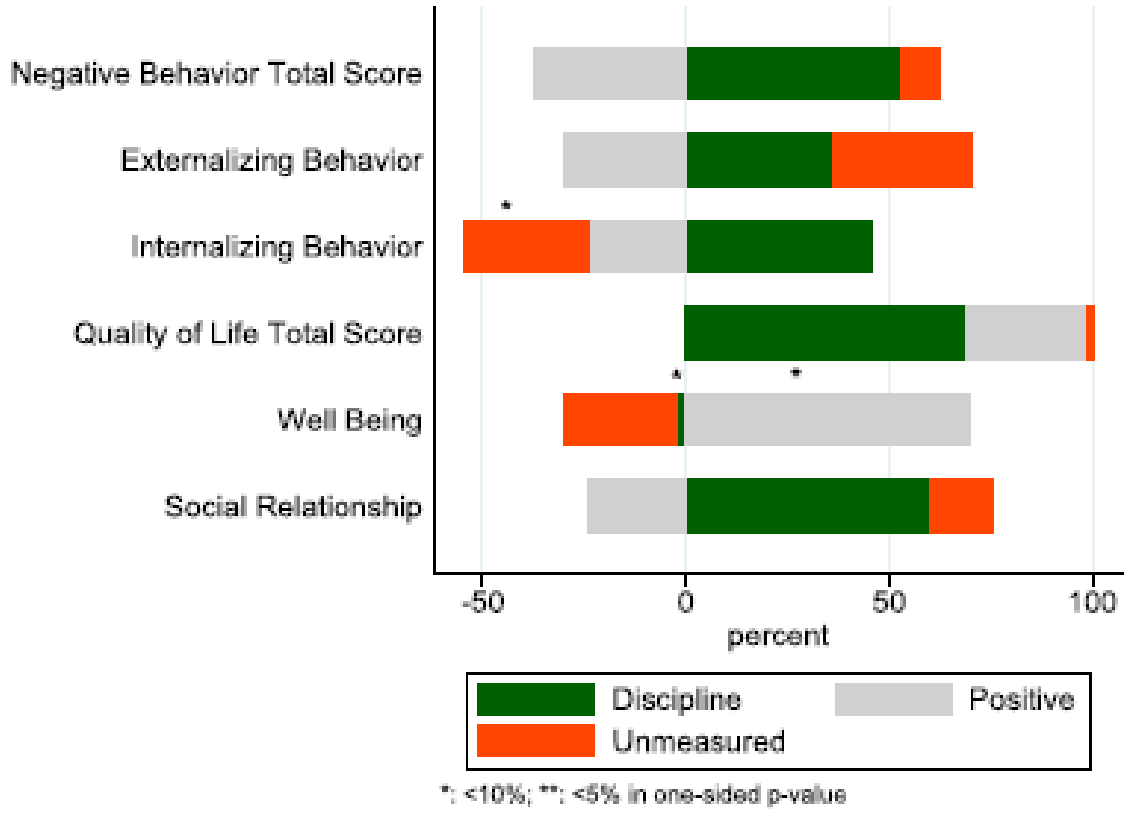
Notes: p -values are based on right-side, one-tailed t -tests, which were calculated from wild cluster bootstrap of t -statistics with 99,999 replications. The stars in the figure indicate statistical significance of each component against the null hypothesis of zero contribution. See Table B.14 for p -values from mediation analysis.

Figure 4: Mediation analysis for female sample



Notes: p -values are based on right- side, one-tailed t -tests, which were calculated from wild cluster bootstrap of t -statistics with 99,999 replications. The stars in the figure indicate statistical significance of each component against the null hypothesis of zero contribution. See Table B.14 for p -values from mediation analysis.

Figure 5: Mediation Analysis for Male Sample



Notes: p -values are based on right- side, one-tailed t -tests, which were calculated from wild cluster bootstrap of t -statistics with 99,999 replications. The stars in the figure indicate statistical significance of each component against the null hypothesis of zero contribution. See Table B.14 for p -values from mediation analysis.

- This study reports evidence of the effect of parent-child interactions during early childhood on behavioral outcomes during early adolescence.
- The evidence derives from evaluating a randomized intervention that targets parents of preschool children and provides education and training on non-harsh discipline methods and positive engagement.
- The intervention reduced externalizing behaviors and improved well- being of children 10 years after the intervention, during a child's early adolescence.
- We investigate heterogeneous effects by gender, account for non-compliance in the intervention group, and adjust for attrition.
- The effects were greater for girls and remain robust when accounting for non-compliance and attrition.
- Mediation analyses suggest that experimentally induced changes in parenting are channels through which interventions improved child outcomes.

- If these models are valid, we should observe that experimentally induced improvements to parenting meaningfully explains intervention effects on behavioral outcomes.

- Mediation analysis estimated a linear model:

$$E\left[Y_1 - Y_0\right] = \underbrace{\mathbf{a}^1 E\left[DISC_1 - DISC_0\right]}_{\text{change in discipline style}} + \underbrace{\mathbf{a}^2 E\left[SUPP_1 - SUPP_0\right]}_{\text{increase in positive engagement}} + \underbrace{\tau_1 - \tau_0 + \left(\mathbf{b}_1 - \mathbf{b}_0\right) X_{K'}}_{\text{other factors}} \quad (4)$$

5.4. Accounting for non-compliance

- Intent-to-treat estimates provide information on the effect of providing an intervention but is not the effect of participating in the program itself due to the presence of non-compliance in the intervention group.

- This section investigates the nature of non-compliance in the sample, assessing whether findings are robust to non-compliance.

- Table 5 shows baseline characteristics of compliers, non-compliers, and those in the control group.

- Since the compliance decision was made after randomization, there is evidence of self-selection.

- Mean mother age was 35.47 for compliers, higher than 33.59 for non-compliers. Average household income was also higher for compliers.

- These mean differences are significant at 10%. Average household income was higher for compliers, but baseline externalizing behaviors were also higher than for non- compliers.

Figure 5: Baseline sample characteristics of compliers and non-compliers

	Compliers	Noncompliers	Control	Two-sided p-values	
	(1)	(2)	(3)	(1)-(2)	(2)-(3)
Sample size	144	41	93		
Child age	4.03	4.00	4.13	0.90	0.46
Child female	0.47	0.54	0.49	0.52	0.67
# siblings	1.08	0.88	1.19	0.16	0.12
Mother age	35.47	33.59	35.62	**0.09**	**0.04**
Mother work hour	2.53	3.63	3.52	0.10	0.85
Mother HS	0.56	0.49	0.55	0.59	0.64
Single mother	0.04	0.07	0.11	0.51	0.55
Net income	5040.15	4304.88	4327.96	**0.06**	0.93
CBCL (ext)	0.13	-0.22	-0.10	**0.07**	0.44
CBCL (int)	0.09	0.05	-0.15	0.87	0.41
Harsh discipline (mother)	-0.05	-0.03	0.11	0.94	0.42
Positive engagement (mother)	-0.09	0.19	0.05	0.10	0.29
Harsh discipline (father)	-0.04	0.03	0.08	0.75	0.76
Positive engagement (father)	0.01	-0.27	0.07	0.20	0.21
IQ (KABC ment.)	0.09	-0.24	-0.05	**0.09**	0.43
IQ (KABC seq.)	0.04	0.01	-0.06	0.92	0.78
IQ (KABC simu.)	0.15	-0.46	-0.05	**0.01**	0.12
IQ (KABC achiv.)	0.07	-0.28	0.00	0.13	0.35
Recalled parenting (father)	-0.04	0.18	-0.01	0.27	0.48
Recalled parenting (mother)	-0.09	0.35	-0.01	**0.02**	**0.06**

Note: p -values are based on mean t -tests between respective groups using bootstrap- t tests with 99,999 replications. Income was monthly household income based on DEM from 2001 to 2002, when the exchange rate was approximately 1 DEM ≈0.54 USD. Recalled parenting behaviors measured parenting behaviors of the parents of a responding mother (i.e., maternal grandparents of a respondent's children).

- Differences are also observable in outcomes.
- Table 6 shows that, on average, the children of non-complier mothers reported greater adolescent externalizing behaviors than those of complier mothers, but less than children of control group mothers.
- Children of non-complier mothers reported greater internalizing behaviors, lower wellbeing, and lower social relationships than children in other groups.
- Although the children of non-complying mothers engaged in less problematic behaviors at the baseline, 10 years later, they were overtaken by the children of complier mothers in terms of mean externalizing behaviors.
- Table 7 shows that children in the non-compliance group reported less externalizing behaviors, greater internalizing behaviors, and about the same wellbeing as children in the control group.
- Coefficient estimates for β were smaller than 0.2, and the null hypothesis of $\beta = 0$ was not rejected for all outcomes other than externalizing behaviors.

Figure 6: Mean outcome of complier group, non-complier group, and control group

	Compliers	Non-compliers	Ctrl group
Negative behavior total score	-0.026	0.109	-0.004
Externalizing behavior	-0.085	-0.012	0.135
Internalizing behavior	0.017	0.124	-0.085
Quality of life total score	0.085	-0.171	-0.063
Wellbeing	0.092	-0.129	-0.090
Social relationship	0.053	-0.191	-0.005

Notes: Negative behavior total score is the sum of externalizing and internalizing behaviors sub-scales, based on the Youth Self Report (YSR). Quality of life total score is the sum of wellbeing and social relationship sub-scales, based on a German instrument for health-related quality of life for children (i.e., KINDL). These measures are from the 10-year follow-ups, collected from 2012 to 2014.

Figure 7: Non-compliers versus control group

(N=109)	β est.	p-value	Std.Error
Negative behavior total score	-.057	0.399	0.138
Externalizing behavior	-.289	**.049**	0.139
Internalizing behavior	0.110	0.300	0.191
Quality of life total score	-.022	0.460	.186
Wellbeing	0.00	0.499	0.229
Social relationship	-.057	0.342	.138

Notes: p -values are based on one-tailed t -tests, which are based on the wild cluster bootstrap of t -statistics with 99,999 replications. p -values under 0.1 are in bold. Negative behavior total score is the sum of externalizing and internalizing behavior sub-scales, based on the Youth Self Report (YSR). Quality of life total score is the sum of wellbeing and social relationships subscales, based on a German instrument for health-related quality of life for children (i.e., KINDL). These measures are from 10-year follow-ups of intervention, collected from 2012 to 2014. Control variables include demographics measured at the baseline, child behavior and cognitive skill measures at baseline, and parenting variables measured at baseline, and a mother's parenting received from her own mother.

- We discuss the two-stage least squares estimates, in which randomized assignment status was used as an instrument for program participation.
- We interpret these estimates as the effect of the intervention on compliers.
- As expected from descriptive statistics, Table 8 shows that the two-stage least squares estimates were greater than intent-to-treat estimates.
- The effect on externalizing behaviors was -0.43, and the effect on child wellbeing was 0.28.
- Corresponding ITT estimates were -0 . 31 and 0.2, respectively.
- For girls, the effect on externalizing behaviors was similar to the effect in the pooled sample, and the effect on wellbeing was greater at 0.44.
- The effect on externalizing behaviors was no longer significant, with $p = .11$.
- For boys, all effects were greater than for those in the pooled sample, but none were significant and all were smaller than effects for the female sub-sample.

Figure 8: Two-stage least squares

Variable name	TSLS estimate	*p*-value
Pooled sample (N=231)		First stage *F*: 298.514
Negative behavior total score	-0.170	0.267
Externalizing behavior	-0.430	**0.066**
Internalizing behavior	0.050	0.421
Quality of life total score	0.220	**0.082**
Wellbeing	0.280	**0.069**
Social relationship	0.090	0.200
Female sample (N=111)		First stage *F*: 98.833
Negative behavior total score	-0.280	0.213
Externalizing behavior	-0.440	0.110
Internalizing behavior	-0.060	0.419
Quality of life total score	0.230	0.180
Wellbeing	0.440	**0.046**
Social relationship	-0.070	0.391
Male sample (N=120)		First stage *F*: 201.872
Negative behavior total score	0.040	0.422
Externalizing behavior	-0.280	0.135
Internalizing behavior	0.230	0.118
Quality of life total score	0.120	0.274
Wellbeing	0.080	0.333
Social relationship	0.120	0.287

Notes: p -values are based on one-tailed t -tests, which are based on the wild cluster bootstrap of t -statistics with 99,999 replications. p -values under 0.1 are in bold. Negative behavior total score is the sum of externalizing and internalizing behavior sub-scales, based on the Youth Self Report (YSR). Quality of life total score is the sum of wellbeing and social relationships subscales, based on a German instrument for health-related quality of life for children (i.e., KINDL). These measures are from 10-year follow-ups of intervention, collected from 2012 to 2014. Control variables include demographics measured at the baseline, child behavior and cognitive skill measures at baseline, and parenting variables measured at baseline, and a mother's parenting received from her own mother.

5.5. Accounting for attrition and non-response

- Table 9 repeats the intent-to-treat analyses in Section 5.1 with IPW applied.
- Overall, estimates were qualitatively similar to the main analyses.
- The effect on externalizing behaviors remained robust to IPW, and the magnitude changed little in both the pooled sample and gender subsamples.
- The effect on child wellbeing also remained robust, though its magnitude declined from 0.2 to 0.156 in the pooled sample.

Figure 9: Intent-To-Treat (IPW)

Variable name	Ctrl mean	Intervn effect	Standard error	*p*-value
Pooled sample (N=224)				
Negative behavior total score	-0.004	-0.103	0.154	0.293
Externalizing behavior	0.135	-0.308	0.152	**0.094**
Internalizing behavior	-0.085	0.070	0.145	0.330
Quality of life total score	-0.063	0.126	0.101	0.128
Wellbeing	-0.090	0.156	0.104	**0.094**
Social relationship	-0.005	0.048	0.093	0.315
Female sample (N=111)				
Negative behavior total score	0.119	-0.183	0.210	0.234
Externalizing behavior	0.045	-0.307	0.195	0.126
Internalizing behavior	0.260	-0.025	0.197	0.449
Quality of life total score	-0.226	0.125	0.184	0.255
Wellbeing	-0.323	0.261	0.168	**0.075**
Social relationship	-0.048	-0.069	0.218	0.386
Male sample (N = 113)				
Negative behavior total score	-0.109	0.053	0.165	0.379
Externalizing behavior	0.212	-0.208	0.183	0.158
Internalizing behavior	-0.380	0.205	0.141	**0.097**
Quality of life total score	0.077	0.069	0.136	0.306
Wellbeing	0.109	0.046	0.139	0.369
Social relationship	0.033	0.063	0.142	0.330

Notes: p -values are based on one-tailed t -tests, which are based on the wild cluster bootstrap of t -statistics with 99,999 replications. p -values under 0.1 are in bold. Negative behavior total score is the sum of externalizing and internalizing behavior sub-scales, based on the Youth Self Report (YSR). Quality of life total score is the sum of wellbeing and social relationship sub-scales, based on a German instrument for health-related quality of life for children (i.e., KINDL). These measures are from the 10-year follow-ups of intervention, collected from 2012 to 2014. Control variables included demographics measured at baseline, child behavior and cognitive skill measures at baseline, and parenting variables measured at baseline, and a mother's parenting received from her own mother.

5.6. The father's role

6. Conclusion