

NBER WORKING PAPER SERIES

AN ANALYSIS OF THE MEMPHIS NURSE-FAMILY PARTNERSHIP PROGRAM

James J. Heckman  
Margaret L. Holland  
Kevin K. Makino  
Rodrigo Pinto  
Maria Rosales-Rueda

Working Paper 23610  
<http://www.nber.org/papers/w23610>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2017

We are very grateful to Professor David Olds, founder of the NFP, and his team for generously sharing the data and source materials from the NFP Memphis Randomize Control Trial. Years of collaboration and productive discussions with him have made this study possible. We thank Terrance Oey and Willem van Vliet for superb research assistance. We are grateful to Juan Pantano, Sylvi Kuperman, Jorge Luis Garcia, Maryclare Griffin, Andres Hojman, Yu Kyung Koh, Cullen Roberts, Karl Schulze, Naoko Takeda, and Joyce Zhu for helpful comments. An early version of this paper was presented at seminars at the University of Chicago and the Association for Public Policy Analysis and Management 2012 Fall conference. This research was supported by NICHD R37HD065072 and NICHD R01HD054702, and previous support of the Pritzker Children's Initiative. A portion of Holland's time on this project was also supported by an NRSA Institutional Research Training Grant from the Health Resources and Services Administration (T32 PE12002), awarded to the Department of Pediatrics at the University of Rochester. The views expressed in this paper are solely those of the authors and do not necessarily represent the official views of the National Institutes of Health. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by James J. Heckman, Margaret L. Holland, Kevin K. Makino, Rodrigo Pinto, and Maria Rosales-Rueda. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

An Analysis of the Memphis Nurse-Family Partnership Program

James J. Heckman, Margaret L. Holland, Kevin K. Makino, Rodrigo Pinto, and Maria Rosales-Rueda  
NBER Working Paper No. 23610

July 2017

JEL No. C5,H5,I1

**ABSTRACT**

This paper evaluates a randomized controlled trial of the Nurse-Family Partnership (NFP) program conducted in Memphis, TN in 1990. NFP offers home visits conducted by nurses for disadvantaged first-time mothers during pregnancy and early childhood. We test NFP treatment effects using permutation-based inference that accounts for the NFP randomization protocol. Our methodology is valid for small samples and corrects for multiple-hypothesis testing. We also analyze the underlying mechanisms generating these treatment effects. We decompose NFP treatment effects into components associated with the intervention-enhanced parenting and early childhood skills. The NFP improves home investments, parenting attitudes and mental health for mothers of infants at age 2. At age 6, the NFP boosts cognitive skills for both genders and socio-emotional skills for females. These treatment effects are explained by program-induced improvements in maternal traits and early-life family investments. At age 12, the treatment effects for males (but not for females) persist in the form of enhanced achievement test scores. Treatment effects are largely explained by enhanced cognitive skills at age 6. Our evidence of pronounced gender differences in response to early childhood interventions contributes to a growing literature on this topic.

James J. Heckman  
Department of Economics  
The University of Chicago  
1126 E. 59th Street  
Chicago, IL 60637  
and IZA  
and also NBER  
jjh@uchicago.edu

Margaret L. Holland  
Yale School of Nursing  
400 West Campus Drive  
Orange CT 06477  
margaret.holland@yale.edu

Kevin K. Makino  
John A. Burns School of Medicine  
University of Hawaii  
1319 Puhanou Street  
Honolulu, HI 96826  
kkmakino@hawaii.edu

Rodrigo Pinto  
University of California, Los Angeles  
Department of Economics  
8283 Bunche Hall  
Los Angeles, CA 90095  
rodrig@econ.ucla.edu

Maria Rosales-Rueda  
University of California – Irvine  
School of Education  
Irvine CA 92697-5500  
mfrosale@uci.edu

Supplementary materials available at <http://www.nber.org/papers/w23610>:

- data appendix

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Experimental Design and Background</b>	<b>6</b>
<b>3</b>	<b>Inference</b>	<b>12</b>
<b>4</b>	<b>Estimated Treatment Effects</b>	<b>14</b>
<b>5</b>	<b>Understanding the Sources of NFP Treatment Effects</b>	<b>26</b>
5.1	Empirical Strategy to Decompose Treatment Effects . . . . .	28
<b>6</b>	<b>Decomposing Treatment Effects</b>	<b>30</b>
6.1	NFP Effects on Skills at Ages 2 and 6 . . . . .	30
6.2	Decomposition of Treatment Effects at Age 6 . . . . .	31
6.3	Decomposition of Treatment Effects at Age 12 . . . . .	32
<b>7</b>	<b>Summary and Conclusion</b>	<b>43</b>

# 1 Introduction

Home visiting programs are widely used to improve the well-being of disadvantaged pregnant women and their children (Currie and Rossin-Slater, 2015; Howard and Brooks-Gunn, 2009a). In 2010, the United States expanded federal funding for the establishment of the Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV Program) to support state implementation of early childhood home visiting programs. At present, funding for MIECHV Program has been reauthorized, with bipartisan support for the program. Between 2012–2015, the MIECHV program served more than 300,000 disadvantaged families and children in the US.<sup>1</sup>

The Nurse-Family Partnership (NFP) is widely considered the primary MIECHV Program among evidence-based home visiting models. The NFP offers prenatal and early childhood care to disadvantaged, unmarried, first-time mothers. Treatment consists of home visits starting during pregnancy and lasting until two years after birth. The NFP home visits are conducted by professional registered nurses with at least of a Bachelor of Science degree in nursing. They have formal training and follow a detailed curriculum of activities specific to the different stages of pregnancy and child development (Currie and Rossin-Slater, 2015). The NFP aims to improve the long-term success of disadvantaged children. It operates by promoting healthy maternal behaviors and by fostering parenting skills. Prior to being implemented at scale, three randomized controlled trials (RCTs) were performed to assess the impact of the NFP in different populations.<sup>2</sup>

Today, NFP surrogates operate in 43 states across the U.S. and have provided service for more than 200,000 families since 1996. A version of NFP is currently operating in the U.K. and Germany (Robling et al., 2016; Sandner et al., 2017).

Most of the NFP literature focuses on three social experiments targeting poor high-risk

---

<sup>1</sup>Between 2010-2014, the MIECHV Program provided \$1.5 billion to states for implementing home visiting programs. In 2015, a two year extension was approved to continue supporting these programs until 2017. The program is currently up for renewal.

<sup>2</sup>The Memphis randomization protocol is presented in Appendix A. A summary of previous experiments is presented in Appendix B. Appendix K summarizes the findings of previous studies.

pregnant females. The first NFP trial occurred in Elmira (1978) and its sample consists of primary non-Hispanic white mothers. The second trial took place in Memphis (1990) and targeted mainly non-Hispanic blacks. A third intervention occurred in Denver (1994) and overrepresented Hispanic participants. This paper uses data from the Memphis randomized control trial to examine NFP impacts through age 12, and to investigate the mechanisms producing the treatment effects.

There are several previous evaluations of the NFP that examine its effects on a range of domains, including maternal and child health, child development, parenting, and family economic self-sufficiency (Avellar et al., 2011; Daro, 2006; Olds, 2012).<sup>3</sup> However, the literature does not account for features of the randomization protocol (such as stratification), does not control for multiple-hypothesis testing (i.e., control for “cherry picking”) and does not analyze the underlying channels generating treatment effects. Our analysis presents a rigorous examination that addresses these issues. The Memphis randomized control trial is the only NFP intervention that collects detailed item-level data necessary to analyze different channels of influence. In particular, the Memphis trial gathered rich information on mother and child well-being from pregnancy up to age 12. It includes information on maternal behavior, home environments, parenting, children’s cognitive and socio-emotional skills measured at ages 2 and 6, and middle childhood outcomes such as achievement and behavior problems, measured at age 12.

We contribute to the existing NFP literature in three ways. First, our paper is the first to conduct inference that accounts for the features of the NFP randomization protocol and addresses the problem of cherry picking. We use permutation-based inference and implement multiple-hypothesis testing adjustments using the step down procedure of Romano and Wolf (2005). We show that most of the findings from previous papers survive permutation testing. However, many fewer treatment effects survive corrections for multiple-hypothesis testing.

Second, our paper is the first to analyze the effects of the NFP by gender, following

---

<sup>3</sup>Our results broadly corroborate the findings of Olds (2012) while producing new insights on the NFP intervention.

the previous literature that has found gender differences in early-childhood program effects (see [Elango et al., 2016](#); [García et al., 2017b](#); [Heckman et al., 2010](#) and [García, Heckman, and Ziff, 2017](#)). Third, this is the first paper that conducts a mediation analysis of NFP treatment effects. Following the methodology developed by [Heckman et al. \(2013\)](#), we decompose statistically significant treatment effects into interpretable components associated with program-induced changes in children’s early skills and parental investments. We estimate the channels underlying treatment effects at age 6 (using maternal skills and investment effects at the end of the program as mediators) and at age 12 (using the program effects at age 6 as mediators).

We find statistically significant treatment effects of NFP on home environments, parenting attitudes, and maternal mental health for parents of both boys and girls at age 2. At age 6, the program improved cognitive skills for both boys and girls, while it enhanced early socio-emotional skills for girls. These treatment effects arise from program-induced improvements in maternal traits and early-life family investments at age 2. The treatment effects for males persist through age 12. Treated males outperform controls on a range of cognitive achievement scores. We find that 40%–60% of male treatment effects at age 12 can be explained by enhanced cognitive skills measured at age 6. The program has much weaker long-term effects for girls. This finding is consistent with the evidence in [Elango et al. \(2016\)](#) and [García et al. \(2017a\)](#). This paper contributes to the literature on the positive influences of early childhood interventions. We show that parenting matters. Specifically, we find that program-induced improvements in maternal mental health (reduction in maternal anxiety and improvement on mastery/self-control) and program-induced enhancement of parenting skills are strong channels underlying program effects on children’s skills. We also present evidence that the NFP had long-lasting effects on cognitive skills for boys, which contrast with the limited evidence on the persistent effects of preschool interventions on cognition. One possible explanation for the long-term effects of NFP on cognitive skills for boys is the health effects of the home visiting program on birth weight. This explains a large share

of the age 12 program impact on cognition. This finding suggests that early childhood interventions that start before birth have the potential to affect cognitive skill formation for which sensitive periods occur in utero and in the first years of life (see [Nelson, 2012](#)).

The remainder of this paper is organized as follows. Section 2 provides information on the experimental design, background variables and measured outcomes of the NFP Memphis trial. Section 3 explains our inferential approach. Section 4 presents our empirical results. Section 5 describes the methodology used to decompose treatment effects. Section 6 presents our mediation analysis. Section 7 concludes.

## 2 Experimental Design and Background

The NFP Memphis trial was initiated in 1990. The main goals of NFP are: (1) improvement of maternal and fetal health during pregnancy; (2) development of parenting skills; and (3) planning of social and economic aspects of maternal life through counseling services ([Olds, 2002](#); [Olds et al., 1997](#)).

Visits started during pregnancy and continued until the newborn reached the age of 2. During these visits, nurses encouraged mothers to adopt a healthy diet and eliminate the use of tobacco, alcohol, and illegal drugs. Nurses also taught mothers to recognize complications in pregnancy and to administer adequate prenatal care. After delivery, nurses promoted infant health care and good parenting skills. Mothers were instructed on how to interact with their children in order to foster emotional and cognitive development. NFP further assisted mothers in the process of establishing life goals related to work, education and future pregnancies.

Table 1 summarizes the main elements of NFP. We discuss its features in detail. We compare the designs of the different NFP studies in Appendix B.

Table 1: Nurse-Family Partnership Intervention: Memphis Trial

<b>Intervention Goals</b>	Healthy prenatal behaviors Parenting skills Life-planning strategies
<b>Target Population</b>	Low-income first-time mothers During pregnancy to two years of age
<b>Eligibility Criteria</b>	<p><b>Biological Criteria:</b> all mothers must comply with the following criteria</p> <ol style="list-style-type: none"> <li>1. Less than 29 weeks of pregnancy</li> <li>2. No previous live births</li> <li>3. No specific chronic illness affecting fetus</li> </ol> <p><b>Socio-Economic Criteria:</b> at least two of the following indicators</p> <ol style="list-style-type: none"> <li>1. Unmarried</li> <li>2. Less than 12 years of education</li> <li>3. Unemployed</li> </ol>
<b>Sociodemographic</b>	Race: 92% African-American Marital Status: 97% unmarried Income: 85% low-income Age: 64% under 18
<b>NFP Groups</b>	<p><b>Group 1</b> (Sample 166) Free transportation to and from appointments (not followed after birth)</p> <p><b>Group 2</b> (Sample 514) Free transportation and developmental screening at ages 6, 12, 24 months</p> <p><b>Group 3</b> (Sample 230) Home visits by nurse during pregnancy plus Group 2 benefits (not followed after birth)</p> <p><b>Group 4</b> (Sample 228) Regular home visits during child’s infancy plus all the Group 3 benefits</p>
<b>Treatment Groups</b>	In our analysis, Group 2 is our “control group” and Group 4 is our “treatment group.”
<b>Frequency of Visits</b>	Weekly for the first month after enrollment Bi-monthly from second month after enrollment until birth Weekly for the first 6 weeks after birth Bi-monthly from seventh week after birth until child is 20 months Monthly from 21 months of age until child is 24 months
<b>Dosage</b>	Average number of visits during pregnancy: 7 visits Average number of visits during infancy (0-24 months): 26 visits

**Notes:** Basic information on the NFP Memphis Trial. See [Kitzman et al. \(1997\)](#) for a detailed description.



### *Eligibility*

The NFP Memphis trial recruited pregnant women from June 1, 1990 to August 31, 1991 through the Memphis-Shelby County Tennessee Health Department. Eligible mothers satisfied the following biological criteria: (1) less than 29 weeks of pregnancy; (2) no previous live birth; and (3) no chronic illnesses that could contribute to fetal-growth retardation or preterm delivery. They also satisfied two or more of the following socio-economic criteria: (1) unmarried; (2) less than 12 years of education; and (3) unemployed.

### *Sample*

The total sample includes 1,290 invited participants and 1,138 mothers effectively enrolled. The majority of the participants were African-American (92%), unmarried (97%), low income (95%), and under 18 years old (64%). Mothers who participated in the program were more likely to be younger and African-American and less likely to have completed high school compared to mothers who refused to participate ([Kitzman et al., 1997](#)).

### *Treatment*

The mothers who agreed to participate in the trial were randomized into four different groups that differed by treatment:

- Group 1: Round-trip transportation from their homes to prenatal session appointments (sample size: 166).
- Group 2: Developmental screening and referral services when their babies were age 6, 12, and 24 months, in addition to the benefits of Group 1 (sample size: 514).<sup>4</sup>
- Group 3: Nurse visits during pregnancy, one visit when in the hospital and one visit at home after childbirth, in addition to the benefits of Group 2 (sample size: 230).
- Group 4: Nurse home visits during pregnancy and until the child's second birthday, in addition to the benefits of Group 2 (sample size: 228).

---

<sup>4</sup>An infant specialist hired by the research project screened the children for sensory and developmental problems and referred those with suspected problems to other specialists for further evaluation and treatment.

There are no available data on the participants of Groups 1 and 3 after the child’s birth, as those participants were not followed. Our paper only uses data on the participants who were originally assigned to Groups 2 or 4. These treatment groups comply with the selection of treatment groups used in previous evaluations other NFP trials. (See Appendix B.) Henceforth, we term Group 4 as the treatment group and Group 2 as control group. It is worth noting that Group 2 received some benefits in the form of developmental screening.

Table 2 presents a summary of selected baseline characteristics for the treatment and control groups by gender of the child. The table shows that pre-program characteristics are fairly balanced across treatment groups. Some exceptions are that treated mothers have lower income, lower likelihood of being in school, lower employment status of the household head, and higher grandmother support compared to mothers in the control group.<sup>5</sup>

#### *Randomization protocol*

The NFP randomization was performed within strata defined by 5 characteristics: (1) maternal race (African-American vs non-African-American); (2) maternal age ( $< 17$ ,  $17-18$ ,  $> 18$  years ); (3) gestational age at enrollment ( $< 20$ ,  $\geq 20$  weeks); (4) employment status of the head of household; and (5) geographic region of residence.<sup>6</sup> Appendix A provides a detailed description of the randomization protocol.

#### *Attrition*

Post-randomization dropout was rare during the program period: four women in both treatment and control groups refused further participation in the intervention after randomization. There were 487 live births in the control group (out of 514 pregnant mothers) and 214 live births in the treatment group (out of 228 pregnant mothers). There were 27 miscarriages in the control group and 14 in the treatment group. Table E.1 in Appendix E shows the percentage of available data (retention rates) by gender and age of follow-up interviews. The differences in attrition by group are not statistically significant. When the children were

---

<sup>5</sup>In addition, control mothers of male children were slightly taller than their treatment counterparts.

<sup>6</sup>The regions are: Inner City, Bisson, Cawthon and Hollywood.

Table 2: Descriptive Statistics of NFP Baseline Characteristics

	Female Sample					Male Sample				
	Control Mean	Control Std.Dev.	Treatment Mean	Treatment Std.Dev.	<i>p</i> -val	Control Mean	Control Std.Dev.	Treatment Mean	Treatment Std.Dev.	<i>p</i> -val
<i>Background Characteristics</i>										
Maternal Race (White)	0.086	0.281	0.101	0.303	0.656	0.076	0.265	0.112	0.317	0.300
Marital Status (Married)	0.012	0.110	0.009	0.096	0.791	0.025	0.157	0.019	0.136	0.696
Maternal Age	18.32	3.290	18.27	3.531	0.877	17.95	3.130	17.96	2.962	0.970
Years of Education	10.38	1.833	10.15	2.040	0.300	10.25	1.919	10.112	1.997	0.543
Mother in School	0.541	0.499	0.596	0.493	0.333	0.630	0.484	0.533	0.501	<b>0.093</b>
Head of Household Is Employed	0.607	0.490	0.495	0.502	<b>0.054</b>	0.517	0.501	0.519	0.502	0.972
% of Census Tract Below Poverty	32.53	20.23	36.29	21.820	0.128	36.047	22.10	34.25	18.32	0.431
Household Density	0.944	0.490	1.053	0.652	0.123	0.923	0.490	0.983	0.446	0.264
<i>Total Household Income (Past 6 Months)</i>										
Less Than \$3000	0.278	0.449	0.358	0.482	0.141	0.286	0.453	0.355	0.481	0.208
\$3000 - \$6999	0.237	0.426	0.220	0.416	0.732	0.248	0.433	0.224	0.419	0.632
\$7000 - \$10999	0.216	0.413	0.229	0.422	0.788	0.235	0.425	0.178	0.384	0.213
Greater Than \$11000	0.180	0.385	0.083	0.277	<b>0.008</b>	0.134	0.342	0.178	0.384	0.320
Income, No Response	0.090	0.286	0.110	0.314	0.565	0.097	0.296	0.065	0.248	0.311
<i>Region of Residence</i>										
Inner City	0.282	0.451	0.284	0.453	0.958	0.298	0.458	0.280	0.451	0.734
Bisson	0.171	0.378	0.229	0.422	0.220	0.210	0.408	0.206	0.406	0.925
Cawthon	0.229	0.421	0.174	0.381	0.233	0.185	0.389	0.224	0.419	0.410
Hollywood	0.318	0.467	0.312	0.465	0.905	0.307	0.462	0.290	0.456	0.750
<i>Maternal skills</i>										
Maternal IQ (Shipley)	96.43	10.30	96.51	10.46	0.949	96.32	10.19	96.79	10.57	0.700
Maternal Bavolek Score	99.77	7.63	100.93	8.76	0.233	99.57	7.78	100.53	8.44	0.318
Maternal Mental Health	99.73	10.12	99.12	10.53	0.610	100.66	10.03	99.36	10.66	0.288
Self-Efficacy	100.71	9.836	100.55	9.28	0.889	99.44	10.06	98.89	11.05	0.664
Maternal Mastery	100.13	10.202	99.10	10.11	0.375	100.32	10.20	100.14	9.75	0.880
Maternal Psychological Resources	100.15	9.903	99.51	10.43	0.586	100.15	10.19	99.57	11.13	0.644
<i>Maternal Health Characteristics</i>										
Maternal Height (Cmt)	164.30	7.47	164.49	6.46	0.809	164.89	6.98	163.60	6.66	0.109
Pre-Pregnancy Weight (kgr)	62.83	14.13	61.22	12.03	0.273	61.31	15.57	63.39	14.67	0.233
Gestational Age	16.34	5.72	16.31	5.47	0.972	16.75	5.82	17.07	5.83	0.644
<i>Maternal Social Support</i>										
Grandmother Social Support	99.04	10.74	101.22	9.60	<b>0.059</b>	100.99	8.30	101.86	7.91	0.355
Husband/Boyfriend Social Support	99.81	10.20	99.96	9.77	0.891	100.42	9.98	101.07	10.25	0.585
<i>Maternal Risky Behaviors</i>										
Alcohol Consumption (Past 2 wks)	0.033	0.178	0.064	0.246	0.232	0.046	0.210	0.028	0.166	0.389
Smoking (Past 3 days)	0.090	0.287	0.128	0.336	0.303	0.101	0.302	0.093	0.292	0.830
Used Marijuana (Past 2 wks)	0.025	0.271	0.018	0.192	0.805	0.038	0.323	0.112	1.160	0.516
Sexually Transmitted Diseases	0.335	0.473	0.349	0.479	0.801	0.324	0.469	0.402	0.493	0.167

**Notes:** This table presents the statistical description of selected pre-program variables at baseline. The first column of the table gives the variable description. The variables are divided into groups that share similar meanings. The remainder of the table consists of the description of the blocks of variables associated with the whole sample, the female sample and the male sample. Each block has 6 columns: (1) Control mean, (2) Control standard deviation, (3) Treatment mean, (4) Treatment standard deviation and (5) Asymptotic *p*-value associated with the difference in means. Bold *p*-values indicate that the t-statistic between the control and the treatment means is significant at the 10% level. Maternal social support corresponds to standardized indices. Additional baseline tables using samples at years 6 and 12 can be found in Section E of the Appendix.

12 years old, 86% of mothers who had no fetal or child death were interviewed.<sup>7</sup>

### *Baseline Information*

We use race, mother’s age, marital status, family income, employment, education, fertility, delinquency, maternal mental health, family support, and maternal risk behavior as background information.

### *Outcomes*

Data from the NFP Memphis trial were collected at multiple time points including birth (e.g., placenta and birth weight) and when the child was 6 months, 1 year, 2 years, 6 years, and 12 years old. Table 3 summarizes the data available for each period of child development. See Appendix C for a detailed description of each of the NFP measures. Data on home environment (HOME) and parenting skills (Bavolek) were collected at 6 months, 1 year, and 2 years old.<sup>8</sup> Data on socio-emotional skills (CBCL), hospital records<sup>9</sup>, and maternal mental health are available at age 2.<sup>10</sup> Maternal mental health was assessed at age 2 by the Rand Mental Health Inventory and the Rosenberg and Pearlin Scales.<sup>11</sup> Child cognition (KABC, PPVT, WISC-III)<sup>12</sup> and socio-emotional skills (CBCL, MSSB) were measured at age 6.<sup>13</sup> Outcomes at age 12 consist of child’s internalizing and externalizing behaviors (CBCL),

---

<sup>7</sup>Tables E.1–E.3 in Appendix E present a statistical description of retention levels and baseline variables. Table E.1 presents the statistical description of retention levels by gender and time of survey. Table E.2 presents the statistical description of selected pre-program variables after 6 years. Table E.3 investigates the distribution of baseline variables after 12 years.

<sup>8</sup>HOME stands for the Home Observation for Measurement of the Environment. It measures the resources and time allocated to cognitive and emotional stimuli. Bavolek is the Adult-Adolescent Parenting Inventory that measures parental skills and child-rearing attitudes such as abusive and neglecting parenting.

<sup>9</sup>Those include health care usage such as hospitalizations and medical appointments.

<sup>10</sup> Child-Behavior Checklist (CBCL) measures child behavioral and emotional problems reported by the mother ([Association and on DSM-IV., 2000](#)).

<sup>11</sup>These instruments measure maternal psychological skills. The Rosenberg Self Esteem Scale assesses global self-worth ([Rosenberg, 1965](#)). The Pearlin Mastery Scale measures personal mastery over life outcomes ([Pearlin and Schooler, 1978](#)). The Rand Mental Health Inventory captures psychological distress and well-being factors including anxiety, depression, and positive well-being ([Veit and Ware, 1983](#)).

<sup>12</sup>KABC stands for the Kaufman Assessment Battery for Children and focuses on problem solving and language skills. PPVT stands for Peabody Picture Vocabulary Test (PPVT) and tests receptive vocabulary. WISC-III stands for the Wechsler Intelligence Scale for Children (WISC-III) and measures short-term memory and visual perception.

<sup>13</sup>MSSB stands for the McArthur Story Stem Battery which measures storytelling affection and coherence ([Olds et al., 2004](#)). MSSB measures four dimensions of socio-emotional skills: (1) child deregulated aggression; (2) parental warmth and empathy; (3) child emotional integration; and (4) child anxiety.

achievement scores (GPA, TCAP, PIAT),<sup>14</sup> and child mental health. We also examine data on parental participation in welfare.<sup>15</sup>

Table 3: NFP Classification of Outcomes/Mediators

Instrument	Content		Child's Age					
	Subject	Classification	0	0.5	1	2	6	12
Birth Biometrics	Child/Mother	Health	✓					
Hospital Visits	Child	Health	✓					
Non-Abusive Parenting (Bavolek)	Parents	Environment/Parenting		✓	✓	✓		
Home Environment Score (HOME)	Family	Environment/Parenting			✓	✓		
Maternal Mental Health	Mother	Mental Health				✓		
Kaufman Assessment (KABC)	Child	Cognition					✓	
Peabody Picture Vocabulary Test (PPVT)	Child	Cognition					✓	
Wechsler Intelligence Scale (WISC-III)	Child	Cognition					✓	
Child Behavior Checklist (CBCL)	Child	Socio-emotional				✓	✓	
MacArthur Story Stem Battery (MSSB)	Child	Socio-emotional					✓	
Health Injuries	Child	Health						✓
Body Mass Index (BMI)	Child	Health						✓
Peabody Individual Achievement Test (PIAT)	Child	Achievement Scores						✓
Tennessee Comprehensive Assessment (TCAP)	Child	Achievement Scores						✓
Schooling Grades (GPA)	Child	Achievement Scores						✓
CBCL/Youth Self-Report	Child	Mental Health						✓
Welfare Use	Family	Governmental Aid						✓

**Notes:** This table presents a summary of the data available for the analysis of treatment effects in NFP mediators and outcomes. The first column presents the name of the outcome or instrument. The remaining columns specify the time of survey according to the age of the child in participating families. Additional information on the description of the available instruments of the NFP data can be seen at Appendix C.

### 3 Inference

Previous literature uses classical inference that relies on large sample (asymptotic) assumptions to make statistical inference about NFP. Heckman et al. (2010) show that the classical approach may generate biased inference when applied to small sample sizes, if the distribution of the data is skewed or, more generally, non-normal. Permutation-based tests provide a solution to these potential problems.

Permutation tests are often called *distribution-free* as they do not require parametric assumptions about the data-generating process and are valid for any sample size. Permutation

<sup>14</sup> GPA records from grades 1 to 5 were obtained through school records. TCAP stands for the Tennessee Comprehensive Assessment Program, an achievement test administered by Tennessee schools after 3rd grade. PIAT stands for Peabody Individual Achievement Test.

<sup>15</sup>Obtained from the Tennessee administrative records.

tests exploit an exchangeability property that arises from the randomization protocol. If the null hypothesis of no treatment effect is true, then permuting treated and control outcomes within strata (blocks) formed by the variables used in the randomization protocol should not alter the distribution of the observed outcomes. This insight enables us to evaluate the  $p$ -value for the null hypothesis of no treatment effect by using the fraction of the re-sampled data that generates a test statistic greater than the one obtained from the original data. This procedure is called a block-permutation test. In the case of NFP, the baseline variables used in the randomization protocol are: maternal age and race, gestational age at enrollment, employment status of the head of the household, and geographic region. See Appendix D for a detailed discussion of the permutation test we use.

Another concern in evaluating and testing treatment effects is the imbalance of baseline characteristics across treatment and control groups. We address this issue by assuming linearity in background variables in the outcome equations to control for imbalances. We use the approach suggested by [Freedman and Lane \(1983\)](#) which, according to [Anderson and Legendre \(1999\)](#), provides the most reliable results among a range of permutation methods for linear models. See Appendix D.1 for a detailed description of this method. We control for maternal height, household income, grandmother support, maternal parenting attitudes, and whether the mother is currently enrolled in school.

In addition, the large number of outcomes associated with NFP creates the possibility of “cherry picking,” namely the selective reporting of statistically significant treatment effects that may occur by chance. We account for the multiplicity of outcomes by using a multiple-hypothesis testing procedure based on the stepdown method of [Romano and Wolf \(2005\)](#). The method generates  $p$ -values that control for the Family-wise Error Rate (FWER): that is, the probability of rejecting at least one true hypothesis out of a set of multiple hypotheses. Stepdown  $p$ -values are more conservative than single hypothesis  $p$ -values, since they take into account the multiplicity of outcomes. For a detailed description of the stepdown method, see Appendix D.

Table 5 presents a summary of the inference analysis and compares the number of statistically significant outcomes using three methods of inference. The first method is classical inference for single-hypotheses, which relies on large sample assumptions. The second method is the single-hypothesis block-permutation test previously discussed. The third method applies the stepdown method, which accounts for multiple outcomes. We find that correcting for the randomization protocol and the imbalance of baseline variables using permutation tests often produces sharper inferences compared to those obtained from classical inference procedures. As expected, fewer outcomes survive corrections for multiple-hypothesis testing. The next section explain the results of our inference reanalysis in more detail.

## 4 Estimated Treatment Effects

Table 4 presents a synopsis of the overall significance of the NFP treatment effects. The table shows the percentage of outcomes for which treatment effects are statistically significant at different significance levels by gender. These percentages are larger than what would be expected to arise by chance.

Table 4: Percentage of Test Statistics Exceeding Various Significance Levels

Significance Level	Female Sample	Male Sample
Percentage of $p$ -values smaller than 1%	7.6%	2.5%
Percentage of $p$ -values smaller than 5%	22.3%	17.2%
Percentage of $p$ -values smaller than 10%	33.8%	35.7%

**Note:** This table presents the percentage of statistically significant treatment effects based on a selection of 157 outcomes in the NFP study. Among those, 5 outcomes are surveyed at birth; 1 at age 6 months old; 2 at age 1; 51 outcomes are surveyed at age 2; 37 outcomes are surveyed at age 6; 3 outcomes are surveyed at age 9; and 58 outcomes are surveyed at age 12.

Table 5 summarizes the inference results on treatment effects presented in Tables 6-

10. According to Table 5, the NFP intervention significantly improved maternal mental health, home environment and parenting skills. On average, treated boys were healthier at birth and experienced an increase in cognitive abilities by age 6. We find that the NFP intervention generated stronger effects on socio-emotional skills for girls and stronger effects on academic achievement for boys. By age 12, treated boys outperformed controls in math and reading achievement. Treated girls experienced an improvement in cognitive and socio-emotional skills at age 6. At age 12, treated girls had a lower body-mass index (BMI), but the estimated treatment effects are not significantly different from the control group in other measures. We also find that controlling for background characteristics increases the precision of treatment effect estimates, often leading to statistical significance in multiple-hypothesis inference.<sup>16</sup> Tables F.7-F.10 of Appendix F show that our findings hold when correcting for attrition using the inverse probability weighting method.<sup>17</sup> Treatment effect estimates and inferences follow closely the results presented in Tables 6-10. We conclude that the correction for attrition does not play a substantial role in the evaluation of the NFP intervention.

Tables 6-10 display the estimation and inference of treatment effects by outcome category and gender. We divide outcomes into five categories: child health, family environments, child cognition, socio-emotional development, and achievement scores. Outcomes are divided into blocks of variables within each category. Each block contains variables of similar content surveyed at the same period. Each table consists of six columns for each gender. The first four columns display the basic statistics for each block of outcomes: (1) control group mean, (2) conditional difference in mean, (3) conditional effect size, and (4) (classic) asymptotic  $p$ -value. The fifth column presents the one-sided single hypothesis permutation  $p$ -value that accounts for the characteristics of the NFP randomization protocol as described in Section 2. The sixth column shows the adjusted stepdown  $p$ -values that correct for multiple-hypothesis testing. In the rest of this section we discuss each outcome category in more detail.

### *Child Health Outcomes*

---

<sup>16</sup>Tables F.1-F.5 of Appendix F present the unconditional analysis of the conditional treatment effects.

<sup>17</sup>See, e.g., Robins et al. (1994).



Table 5: Summary of Statistically Significant Outcomes by Inference Method

Outcomes	Age	Measures	Mediator Classification	Females			Males			
				Assym.	Perm.	MHT	Assym.	Perm.	MHT	
<i>A. Child's Health</i>										
Birth Biometrics	5	Birth	Health	0	0	0	3	5	5	Significant Effects (+)
Health Injuries	5	12	Health	2	3	0	0	1	0	Mixed Effects
Body Mass Index (BMI)	1	12	Health	1	1	1	0	0	0	Significant Effects (-)
<i>B. Family Environment at Age 2</i>										
Non-Abusive Parenting (Bavolek)	1	2	Parenting	1	1	1	1	1	1	Significant Effects (+)
Home Environment Total Score (HOME)	2	1	Parenting/Environment	1	1	1	1	1	1	Significant Effects (+)
Home Environment Subscores (HOME)	2	6	Parenting/Environment	4	4	4	2	1	0	Mixed Effects
Maternal Mental Health	2	7	Mental Health	5	6	4	1	1	0	Mixed Effects
<i>C. Cognitive Skills at Age 6</i>										
Kaufman Assessment Sub-scales (KABC)	6	8	Cognition	2	2	0	3	2	2	Significant Effects (+)
Kaufman Assessment Global Scales (KABC)	6	3	Cognition	2	3	2	1	1	1	Significant Effects (+)
Peabody Picture Vocabulary Test (PPVT)	6	1	Cognition	0	0	0	1	1	1	Significant Effects (+)
Wechsler Intelligence Scale (WISC-III)	6	1	Cognition	0	0	0	0	0	0	No Effect
Cognitive Skill Factors	6	1	Cognition	0	1	0	1	1	1	Significant Effects (+)
<i>D. Socio-emotional Skills at Age 6</i>										
Child Behavior Checklist (CBCL)	6	6	Behavior Problems	2	2	1	0	0	0	No Effect
MacArthur Story Stem Battery (MSSB)	6	5	Socio-emotional	2	4	2	0	2	0	Mixed Effects
<i>E. Achievement Scores and Behavior at 12</i>										
Reading achievement (PIAT+TCAP+GPA)	12	6	Achievement Scores	1	0	0	2	3	1	Significant Effects (+)
Math achievement (PIAT+TCAP+GPA)	12	3	Achievement Scores	0	0	0	1	3	3	Significant Effects (+)
Internalizing/externalizing (CBCL)	12	3	Behavior Problems	0	0	0	2	2	2	Significant Effects (-)
<i>F. Parental Behavior</i>										
Welfare use	1-12	3	Governmental Aid	2	0	0	3	2	1	Significant Effects (-)
Subsequent births	1-12	4	Fertility	0	0	0	2	2	0	Mixed Effects
Total Number of Significant Effects				25	28	16	24	28	18	

**Note:** This table presents a summary of the treatment effect findings from our inference methods displayed in Tables 6-9. The outcomes/mediators are organized into six categories: (A) Child's Health; (B) Family Environment at Age 2; (C) Cognitive Skills at Age 6; (D) Socio-emotional Skills at Age 6; (E) Achievement Scores and Behavior at 12; (F) Parents Behavior. The first column presents the name of the variable that can be either analysed as an outcome or a mediator. Those include psychological instruments described in Section 2. The second column displays the child's age at time of assessment. The third column displays the number of measures being examined within each outcome/instrument. The fourth column gives the mediator classification. The last columns summarize the findings by gender. The table displays four columns for males and females. The first three columns display the number of statistically significant treatment effects according to the following inference methods: (1) asymptotic p-values, (2) block permutation, and (3) multiple-hypothesis testing (MHT) correction using the stepdown method of Romano and Wolf (2005). The fourth column summarizes whether the estimated average treatment effect is statistically significant and its direction. Significant Effects means that the estimated average treatment effect is statistically significant in at least two subscales within the classification group. Mixed effects means that the estimate of the average treatment effect is statistically significant at least one sub-scale within classification but results are not robust under multiple-hypothesis testing. No effect means that the estimates of the intervention average treatment effects are not statistically significant within the outcome. The signs (+) and (-) denote the expected direction (positive or negative) of estimated treatment effects. TCAP stands for the Tennessee Comprehensive Assessment Program. PIAT stands for Peabody Individual Achievement test.

Table 6 displays the NFP results on child health. The first block shows treatment effects on birth outcomes. Treated boys were relatively healthier at birth: there are strong positive effects on placenta weight, birth weight, head circumference, length and gestational age at delivery. The results are both statistically significant at the 5% significance level and robust to correcting for multiple-hypothesis testing. Treated boys are, on average, 193 grams heavier and were delivered 0.7 weeks later than their control counterparts. In contrast, there are no statistically significant effects for girls at birth. The second block focuses on child health outcomes at age 12. We find statistically significant treatment effects on body mass index (BMI) for girls. Boys have fewer statistically significant results at age 12, which do not survive multiple-hypothesis corrections.

### *Family environment*

Table 7 displays results on family environments and parenting. The first two blocks display treatment effects on parenting beliefs and the home environment for ages 1 and 2. Treatment effects are statistically significant for both genders and robust to application of the stepdown procedure. By age 2, the NFP improved maternal parenting attitudes regarding non-abusive and non-neglecting behaviors (Bavolek Inventory). We find effect sizes of approximately 0.3 standard deviations (SDs) for mothers of both females and males. In addition, the intervention improved home environments by 0.17 SDs for males and by 0.3 SDs for females. This positive impact on home environments is mainly reflected in treatment effects on HOME sub-scales associated with variety in daily stimulation, and provision of appropriate playing material. Additionally, we find statistically significant treatment effects on maternal mental health. Mothers of female children reported less anxiety, and better emotional stability, self-esteem, and mastery skills. The effects range between 0.2 and 0.3 SDs. Mothers of male children also experienced an improvement in their mastery skills. However, this effect does not survive multiple-hypothesis corrections. The bottom block examines the treatment effects on the cost of welfare program participation. The NFP reduced the total cost of participation in AFDC/TANF, SNAP (food stamps), and Medicaid

for mothers with male children at age 12. Reduction in SNAP enrollment survive the multiple hypothesis correction for males. Additionally, one of the NFP program objectives was to improve maternal planning of future pregnancies. Table 8 shows the effects of the program on the probability of having subsequent births after the first child at different periods between ages 0 and 12. The estimated treatment effects are imprecise and do not survive multiple-hypothesis testing correction.

### *Child Cognitive Outcomes*

Table 9 displays the NFP results on cognition and achievement for children at ages 6 and 12. The first block analyzes the Kaufmann Assessment Battery for Children (KABC) inventory measured at age 6. In particular, it displays the subtests that are part of the mental processing composite (sequential and simultaneous processing). NFP increased IQ for both genders. However, only the results for boys are robust to controlling for multiple-hypothesis testing. The second block displays outcomes on some cognitive-related subscales from the KABC, and the results are robust for both boys and girls. Also, treated boys scored 0.25 SDs higher in the PPVT (Peabody Picture Vocabulary Test) than their control counterparts.

To summarize effects on child cognition at age 6, we estimate two factor scores. The first factor captures both cognition and achievement, which is derived using the three instruments KABC, PPVT and WISC-III. The second factor captures a pure measure of cognitive skills that comes from the KABC mental processing composite. The treatment effect sizes are larger for boys than girls. For instance, NFP increased cognitive skills by 0.27 SDs for boys versus 0.12 SDs for girls.

### *Achievement Scores*

The bottom part of Table 9 presents the treatment effects on achievement outcomes at age 12 (reading and math). The results, again, are stronger for treated boys than treated

girls, and are robust to the application of the stepdown procedure. Treated boys scored 0.24 SDs higher in the language section of an achievement test: the Tennessee Comprehensive Assessment Program (TCAP). Also, NFP improved math achievement scores exclusively for boys. Treatment effect sizes range between 0.15 and 0.22 SDs. In general, Table 9 documents a wide variety of effects on the cognitive skills of boys and girls. These results extend previous evaluations, which found this effect only for children born to mothers with low psychological resources (Kitzman et al., 2010; Olds, 2012)

### *Child Socio-Emotional Outcomes*

Table 10 displays the NFP results on socio-emotional development. The upper blocks investigate the treatment effects on psychosocial functioning based on the Child Behavior Checklist (CBCL) at ages 2 and 6. The results are statistically significant only for treated girls, who had fewer affective and attention problems at age 2 (effect size: 0.24–0.33 SDs) and fewer conduct and attention problems at age 6 (effect size: 0.27 SDs). Also, children’s socio-emotional skills were assessed using the MacArthur Story Stem Battery (MSSB) surveyed at age 6. We find that treated girls improved pro-social skills (warmth and empathy) by 0.36 SDs and decreased aggressive behavior by 0.18 SDs. The bottom block of Table 10 considers socio-emotional outcomes at age 12. Treated boys experienced less internalizing disorders and fewer school absences (effect sizes of approximately 0.2 SDs). NFP treatment effects differ by gender. Socio-emotional effects are stronger for girls while treatment effects on academic achievement are statistically significant for boys.

Our results are consistent with the existing literature on home visiting programs that have shown positive effects on children’s socio-emotional development. For example, using reports from teachers at age 5 from the Preparing for Life (PFL) program, Doyle (2017) found reductions in problematic behaviors but not improvement in pro-social skills. The effects sizes ranged between 0.31 and 0.5 SDs. For the case of the NFP, results on socio-emotional development at age 6 are found mainly for girls, who experienced a reduction in

conduct problems and aggression and an improvement in pro-social skills with effect sizes between 0.18 and 0.36 SDs.

Table 6: Child Health Outcomes

Outcome Description	Female Sample			Male Sample		
	Control Mean	Basic Statistics Difference in Means	Effect Size	Asymp. $p$ -value	Block Permutation Single $p$ -value	Stepdown $p$ -value
<i>Birth Outcomes for Child</i>						
Placenta Weight	682.995	-1.826	-0.011	0.534	0.362	0.599
Birth Weight	3055.224	-126.604	-0.232	0.963	0.893	0.893
Head Circumference	33.262	-0.007	-0.004	0.513	0.254	0.595
Length	49.665	0.184	0.069	0.291	0.285	0.588
Gestational Age at Delivery	39.119	-0.491	-0.217	0.925	0.805	0.887
<i>Child Health Outcomes (Year 12)</i>						
Any Injuries Since Last Interview	0.164	-0.052	-0.149	0.119	<b>0.085</b>	0.230
# Hospitalizations for Injuries Since Last Interview	0.009	-0.011	-0.112	0.156	0.213	0.388
Total # Injuries Since Last Interview	0.197	-0.079	-0.181	<b>0.067</b>	<b>0.028</b>	0.113
Hospitalized Since Last Interview	0.042	-0.036	-0.183	<b>0.050</b>	<b>0.050</b>	0.182
Have Chronic Condition/Health Problem	0.197	-0.002	-0.004	0.488	0.692	0.692
Standardized Child BMI	1.121	-0.255	-0.294	<b>0.013</b>	<b>0.005</b>	<b>0.025</b>
	0.224	-0.048	-0.122	0.176	0.124	0.425
	0.010	-0.012	-0.117	0.142	<b>0.078</b>	0.341
	0.268	-0.044	-0.085	0.264	0.257	0.672
	0.039	0.053	0.295	0.974	0.890	0.981
	0.361	0.068	0.144	0.858	0.811	0.986
	0.797	0.209	0.235	0.959	0.907	0.907

**Note:** The first column provides the outcome description. Our results are presented in six columns for each gender. The first column (Control Mean) of each result set shows the unconditional mean for the control group. When factor scores were computed, we set the mean in the control group to zero. The second column (Difference in Means) gives the conditional difference in means between the treatment group and the control group. As mentioned in Section 2, the control group stands for the original treatment group 2 of the NFP experiment and the treatment group stands for the original group 4. The third column (Effect Size) presents the conditional effect size for the respective group. The fourth column (Asymp.  $p$ -value) provides the asymptotic  $p$ -value for the one-sided single hypothesis test associated with the  $t$ -statistic for the conditional difference in means between treatment and control groups. The fifth column (Block Permutation – Single  $p$ -value) presents the one-sided restricted permutation  $p$ -values for the single-hypothesis testing based on the  $t$ -statistic associated with the treatment indicator in the Freedman and Lane (1983) regression as described in Section 3. By restricted permutation we mean that permutations are done within strata defined by the baseline variables used in the randomization protocol: maternal age and race, gestational age at enrollment, employment status of the head of the household, and geographic region. The covariates used in the Freedman and Lane (1983) regression are: maternal height, household income, grandmother support, maternal parenting attitudes and mother currently in school. Finally, the last column (Block Permutation – Stepdown) provides  $p$ -values that account for multiple-hypothesis testing based on the Stepdown procedure of Romano and Wolf (2005). Blocks of outcomes that are tested jointly are separated by lines. The selection of blocks of outcomes is done on the basis of their meaning. Outcomes that share similar meaning are grouped together. Female maternal outcomes allude to mothers whose first child is a girl. Likewise, male maternal outcomes allude to mothers whose first child is a boy. See Table F.1 in Appendix F for the unconditional analysis.

Table 7: Family Environment

Outcome Description	Female Sample			Male Sample			Block Permutation Stepdown <i>p</i> -value		
	Control Mean	Basic Statistics Difference in Means	Effect Size	Asymp. <i>p</i> -value	Control Mean	Basic Statistics Difference in Means		Effect Size	Asymp. <i>p</i> -value
<i>Home Environment, Parenting (Year 1)</i>									
Family Environment (HOME Score)	0.000	0.354	0.354	<b>0.003</b>	0.000	0.208	0.208	<b>0.051</b>	<b>0.041</b>
Non-Abusive Parenting Attitudes (Bavolek)	0.000	0.289	0.288	<b>0.012</b>	0.000	0.273	0.273	<b>0.0150</b>	<b>0.006</b>
<i>Home Environment, Parenting (Year 2)</i>									
Family Environment (HOME Score)	0.000	0.302	0.301	<b>0.010</b>	0.000	0.169	0.169	<b>0.092</b>	<b>0.075</b>
Non-Abusive Parenting Attitudes (Bavolek)	0.000	0.371	0.370	<b>0.003</b>	0.000	0.316	0.316	<b>0.006</b>	<b>0.006</b>
<i>Home Environment, sub-scales (Year 1)</i>									
Emotional/Verbal Responsivity of Mother	11.591	0.333	0.189	<b>0.070</b>	10.236	0.284	0.171	<b>0.091</b>	0.207
Avoidance of Restriction and Punishment	7.461	-0.026	0.583	0.807	1.112	-0.039	-0.027	0.580	0.767
Organization of Environment	3.967	0.007	0.478	0.320	3.198	0.095	0.095	0.228	0.340
Provision of Appropriate Play Material	11.266	0.391	0.223	<b>0.037</b>	1.149	0.118	0.074	0.286	0.233
Maternal Involvement with Child	7.735	0.335	0.237	<b>0.031</b>	5.991	0.148	0.104	0.207	0.340
Oppor. For Variety in Daily Stimulation	3.259	0.249	0.218	<b>0.041</b>	2.643	0.305	0.267	<b>0.019</b>	0.107
<i>Home Environment, sub-scales (Year 2)</i>									
Emotional/Verbal Responsivity of Mother	13.046	-0.078	-0.052	0.650	8.591	-0.269	-0.179	0.914	0.853
Avoidance of Restriction and Punishment	2.061	0.299	0.171	<b>0.097</b>	4.286	0.053	0.027	0.416	0.400
Maternal Involvement with Child	6.150	0.108	0.248	0.193	7.368	0.177	0.141	0.534	0.743
Organization of Environment	6.742	0.176	0.187	<b>0.071</b>	5.175	-0.009	-0.011	0.534	0.850
Provision of Appropriate Play Material	8.161	0.785	0.428	<b>0.000</b>	6.754	0.483	0.287	<b>0.012</b>	0.290
Oppor. For Variety in Daily Stimulation	4.786	0.501	0.400	<b>0.001</b>	4.322	0.486	0.407	<b>0.001</b>	<b>0.007</b>
<i>Maternal Mental Health (Year 2) - Factor Scores</i>									
Anxiety	0.000	-0.247	-0.247	<b>0.030</b>	0.000	-0.038	-0.038	0.382	0.457
Depression	0.000	-0.129	0.151	0.094	0.000	-0.062	-0.062	0.311	0.717
Positive Well-Being	0.000	0.101	0.209	0.344	0.000	-0.136	-0.136	0.853	0.887
Emotional Stability	0.000	0.207	0.055	<b>0.048</b>	0.000	0.050	0.049	0.347	0.634
Overall Mental Health	0.000	0.210	0.210	<b>0.051</b>	0.000	-0.014	-0.014	0.544	0.637
Self-Esteem	0.000	0.313	0.313	<b>0.007</b>	0.000	0.073	0.073	0.291	0.716
Mastery	0.000	0.286	0.286	<b>0.016</b>	0.000	0.198	0.198	<b>0.062</b>	0.244
<i>Welfare (Child Ages 1 - 12 Years)</i>									
AFDC/TANF	2744.043	-166.474	-0.066	0.300	2743.386	-485.836	-0.188	<b>0.055</b>	0.101
Food Stamp	2996.965	-313.891	-0.192	<b>0.056</b>	3263.273	-357.933	-0.232	<b>0.030</b>	<b>0.093</b>
Medicaid	3543.761	-262.590	-0.158	<b>0.097</b>	3823.048	-271.631	-0.183	<b>0.070</b>	0.122

**Note:** The first column provides the outcome description. Our results are presented in six columns for each gender. The first column (Control Mean) of each result set shows the unconditional mean for the control group. When factor scores were computed, we set the mean in the control group to zero. The second column (Difference in Means) gives the conditional difference in means between the treatment group and the control group. As mentioned in Section 2, the control group stands for the original treatment group 2 of the NFP experiment and the treatment group stands for the original group 4. The third column (Effect Size) presents the conditional effect size for the respective group. The fourth column (Asymp. *p*-value) provides the asymptotic *p*-value for the one-sided single hypothesis test associated with the *t*-statistic for the conditional difference in means between treatment and control groups. The fifth column (Block Permutation – Single *p*-value) presents the one-sided restricted permutation *p*-values for the single-hypothesis testing based on the *t*-statistic associated with the treatment indicator in the Freedman and Lane (1983) regression as described in Section 3. By restricted permutation we mean that permutations are done within strata defined by the baseline variables used in the randomization protocol: maternal age and race, gestational age at enrollment, employment status of the head of the household, and geographic region. The covariates used in the Freedman and Lane (1983) regression are: maternal height, household income, grandmother support, maternal parenting attitudes and mother currently in school. Finally, the last column (Block Permutation – Stepdown) provides *p*-values that account for multiple-hypothesis testing based on the Stepdown procedure of Romano and Wolf (2005). Blocks of outcomes that are tested jointly are separated by lines. The selection of blocks of outcomes is done on the basis of their meaning. Outcomes that share similar meaning are grouped together. Female maternal outcomes allude to mothers whose first child is a girl. Likewise, male maternal outcomes allude to mothers whose first child is a boy. See Table F.2 in Appendix F for the unconditional analysis.

Table 8: Maternal Cumulative Subsequent Births

Outcome Description	Female Sample				Male Sample			
	Control Mean	Basic Statistics Difference in Means	Basic Statistics Effect Size	Block Permutation Single $p$ -value	Control Mean	Basic Statistics Difference in Means	Basic Statistics Effect Size	Block Permutation Single $p$ -value
<i>Cumulative Subsequent Births (Years 0 - 12)</i>								
Subsequent Children Birth (Years 9 - 12)	0.259	0.045	0.077	0.470	0.397	-0.050	-0.075	0.373
Subsequent Children Birth (Years 6 - 9)	0.344	0.159	0.276	0.961	0.459	0.035	0.051	0.797
Subsequent Children Birth (Years 2 - 6)	0.884	-0.005	-0.006	0.871	1.027	-0.141	-0.176	<b>0.056</b>
Subsequent Children Birth (Years 0 - 2)	0.298	-0.067	-0.151	0.239	0.315	-0.116	-0.253	<b>0.018</b>

**Note:** The first column provides the outcome description. Our results are presented in six columns for each gender. The first column (Control Mean) of each result set shows the unconditional mean for the control group. When factor scores were computed, we set the mean in the control group to zero. The second column (Difference in Means) gives the conditional difference in means between the treatment group and the control group. As mentioned in Section 2, the control group stands for the original treatment group 2 of the NFP experiment and the treatment group stands for the original group 4. The third column (Effect Size) presents the conditional effect size for the respective difference in means between treatment and control groups. The fourth column (Asymp.  $p$ -value) provides the asymptotic  $p$ -value for the one-sided single hypothesis test associated with the  $t$ -statistic for the conditional difference in means between treatment and control groups. The fifth column (Block Permutation - Single  $p$ -value) presents the one-sided restricted permutation  $p$ -values for the single-hypothesis testing based on the  $t$ -statistic associated with the treatment indicator in the Freedman and Lane (1983) regression as described in Section 3. By restricted permutation we mean that permutations are done within strata defined by the baseline variables used in the randomization protocol: maternal age and race, gestational age at enrollment, employment status of the head of the household, and geographic region. The covariates used in the Freedman and Lane (1983) regression are: maternal height, household income, grandmother support, maternal parenting attitudes and mother currently in school. Finally, the last column (Block Permutation - Stepdown) provides  $p$ -values that account for multiple-hypothesis testing based on the Stepdown procedure of Romano and Wolf (2005). Blocks of outcomes that are tested jointly are separated by lines. The selection of blocks of outcomes is done on the basis of their meaning. Outcomes that share similar meaning are grouped together. Female maternal outcomes allude to mothers whose first child is a girl. Likewise, male maternal outcomes allude to mothers whose first child is a boy. See Table F.3 in Appendix F for the unconditional analysis.



Table 9: Cognitive Abilities and Achievement Outcomes

Outcome Description	Female Sample			Male Sample			Block Permutation Single Stepdown <i>p</i> -value					
	Control Mean	Basic Statistics Difference in Means	Effect Size	Asymp. <i>p</i> -value	Block Permutation Single Stepdown <i>p</i> -value	Control Mean		Basic Statistics Difference in Means	Effect Size	Asymp. <i>p</i> -value		
<i>Kaufman Assessment Battery for Children (Year 6)</i>												
Gestalt Closure	8.981	0.235	0.078	0.276	0.163	0.430	9.775	-0.385	-0.133	0.836	0.647	0.647
Hand Movements	9.282	0.402	0.186	<b>0.082</b>	<b>0.025</b>	0.147	9.287	0.211	0.099	0.233	0.421	0.737
Matrix Analogies	8.632	0.148	0.087	0.257	0.271	0.536	8.478	0.316	0.199	<b>0.071</b>	0.122	0.428
Number Recall	9.423	0.397	0.138	0.144	0.109	0.383	8.952	0.954	0.400	<b>0.003</b>	<b>0.004</b>	<b>0.030</b>
Photo Series	6.967	0.434	0.216	<b>0.053</b>	0.216	0.216	6.774	0.050	0.024	0.431	0.578	0.779
Spatial Memory	8.434	0.204	0.084	0.266	0.295	0.457	8.526	0.278	0.115	0.204	0.170	0.439
Triangles	8.868	0.378	0.163	0.104	0.155	0.457	9.120	0.126	0.095	0.344	0.160	0.464
Word Order	9.693	0.002	0.001	0.497	0.300	0.300	9.191	0.751	0.293	<b>0.017</b>	<b>0.008</b>	<b>0.050</b>
<i>Kaufman Assessment Battery for Children (Year 6)</i>												
Nonverbal	89.203	2.035	0.230	<b>0.049</b>	<b>0.045</b>	<b>0.090</b>	89.244	1.421	0.152	0.136	0.175	0.221
Sequential Processing	96.582	1.696	0.132	0.161	<b>0.067</b>	0.117	94.507	3.768	0.320	<b>0.011</b>	<b>0.010</b>	<b>0.023</b>
Simultaneous Processing	88.844	1.928	0.191	<b>0.080</b>	<b>0.075</b>	<b>0.075</b>	89.919	0.709	0.071	0.303	0.216	0.216
<i>WISC-III, PPVT for Children (Year 6)</i>												
Wechsler Intelligence Scale (WISC-III)	96.256	1.061	0.059	0.324	0.336	0.336	90.657	1.778	0.104	0.223	0.279	0.279
Peabody Picture Vocabulary Test (PPVT)	83.299	1.773	0.163	0.108	0.165	0.289	82.466	2.646	0.252	<b>0.039</b>	<b>0.008</b>	<b>0.016</b>
<i>Child Cognition (Year 6) - Factor Scores</i>												
Cognition + Achievement (KABC, PPVT, WISC-III)	0.000	0.110	0.110	0.208	<b>0.084</b>	<b>0.084</b>	0.000	0.185	0.185	<b>0.088</b>	<b>0.083</b>	<b>0.083</b>
Cognitive Skills (Mental Processing KABC)	0.000	0.119	0.119	0.186	<b>0.082</b>	0.113	0.000	0.273	0.273	<b>0.024</b>	<b>0.022</b>	<b>0.030</b>
<i>Reading Achievement for the Child (Year 12)</i>												
Average Reading Grade (Grades 1 - 5)	2.703	0.032	0.042	0.380	0.220	0.368	2.327	0.071	0.096	0.251	0.104	0.200
TCAP % Language (School Years 1 - 5, Grd 3+)	50.854	0.137	0.006	0.484	0.189	0.370	38.063	5.243	0.240	<b>0.058</b>	<b>0.009</b>	<b>0.035</b>
TCAP % Reading (School Years 1 - 5, Grd 3+)	41.607	0.146	0.007	0.480	0.163	0.351	34.912	1.611	0.076	0.305	<b>0.099</b>	0.237
PIAT Total Reading (Derived Score)	90.246	0.715	0.075	0.295	0.358	0.417	89.292	1.356	0.101	0.222	0.111	0.148
PIAT Reading Comprehension (Derived Score)	88.307	-0.196	-0.022	0.563	0.576	0.576	87.585	2.272	0.195	<b>0.077</b>	<b>0.037</b>	0.112
PIAT Reading Recognition (Derived Score)	94.221	2.276	0.189	<b>0.095</b>	0.147	0.347	92.456	0.379	0.026	0.422	0.171	0.171
<i>Math Achievement for the Child (Year 12)</i>												
Average Math Grade (Grades 1 - 5)	2.634	0.036	0.044	0.371	0.287	0.405	2.368	0.123	0.162	0.127	<b>0.054</b>	<b>0.054</b>
TCAP % Math (School Years 1 - 5, Grd 3+)	46.935	1.566	0.066	0.326	0.217	0.369	40.346	3.392	0.153	0.159	<b>0.0304</b>	<b>0.076</b>
PIAT Mathematics (Derived Score)	87.188	-0.125	-0.013	0.538	0.758	0.758	86.316	2.247	0.223	<b>0.054</b>	<b>0.043</b>	<b>0.077</b>

**Note:** The first column provides the outcome description. Our results are presented in six columns for each gender. The first column (Control Mean) of each result set shows the unconditional mean for the control group. When factor scores were computed, we set the mean in the control group to zero. The second column (Difference in Means) gives the conditional difference in means between the treatment group and the control group. As mentioned in Section 2, the control group stands for the original treatment group 2 of the NFP experiment and the treatment group stands for the original group 4. The third column (Effect Size) presents the conditional effect size for the respective group. The fourth column (Asymp. *p*-value) provides the asymptotic *p*-value for the one-sided single hypothesis test associated with the *t*-statistic for the conditional difference in means between treatment and control groups. The fifth column (Block Permutation - Single *p*-value) presents the one-sided restricted permutation *p*-values for the single-hypothesis testing based on the *t*-statistic associated with the treatment indicator in the Freedman and Lane (1983) regression as described in Section 3. By restricted permutation we mean that permutations are done within strata defined by the baseline variables used in the randomization protocol: maternal age and race, gestational age at enrollment, employment status of the head of the household, and geographic region. The covariates used in the Freedman and Lane (1983) regression are: maternal height, household income, grandmother support, maternal parenting attitudes and mother currently in school. Finally, the last column (Block Permutation - Stepdown) provides *p*-values that account for multiple-hypothesis testing based on the Stepdown procedure of Romano and Wolf (2005). Blocks of outcomes that are tested jointly are separated by lines. The selection of blocks of outcomes is done on the basis of their meaning. Outcomes that share similar meaning are grouped together. Female maternal outcomes allude to mothers whose first child is a girl. Likewise, male maternal outcomes allude to mothers whose first child is a boy. See Table F.4 in Appendix F for the unconditional analysis. TCAP stands for the Tennessee Comprehensive Assessment Program, an achievement test administered by Tennessee schools after 3rd grade. PIAT stands for Peabody Individual Achievement test.

Table 10: Socio-Emotional Outcomes

Outcome Description	Female Sample				Male Sample							
	Control Mean	Basic Statistics Difference in Means	Effect Size	Asymp. $p$ -value	Block Permutation Single $p$ -value	Stepdown $p$ -value	Control Mean	Basic Statistics Difference in Means	Effect Size	Asymp. $p$ -value	Block Permutation Single $p$ -value	Stepdown $p$ -value
<i>Child Behavior Checklist (Year 2) - Factor Scores</i>												
Affective Problems	0.000	-0.336	-0.336	<b>0.002</b>	<b>0.004</b>	<b>0.016</b>	0.000	0.163	0.163	0.893	0.848	0.946
Anxiety Problems	0.000	-0.191	-0.191	<b>0.057</b>	0.217	0.217	0.000	-0.029	-0.029	0.411	0.556	0.858
Pervasion Developmental Problems	0.000	-0.262	-0.262	<b>0.013</b>	<b>0.054</b>	0.125	0.000	0.084	0.084	0.747	0.642	0.885
Attention Deficit Hyperactivity Disorder	0.000	-0.239	-0.239	<b>0.027</b>	<b>0.019</b>	<b>0.062</b>	0.000	0.078	0.078	0.725	0.753	0.929
Oppositional Defiant Problems	0.000	-0.224	-0.223	<b>0.036</b>	<b>0.059</b>	0.103	0.000	0.113	0.113	0.822	0.873	0.873
<i>Child Behavior Checklist (Year 6) - Factor Scores</i>												
Affective Problems	0.000	-0.063	-0.063	0.318	0.436	0.712	0.000	-0.107	-0.108	0.193	0.176	0.530
Anxiety Problems	0.000	-0.085	-0.085	0.241	0.473	0.669	0.000	0.108	0.108	0.787	0.871	0.871
Somatic Problems	0.000	0.083	0.083	0.732	0.856	0.856	0.000	0.061	0.061	0.674	0.463	0.679
Attention Deficit Hyperactivity Problems	0.000	-0.269	-0.269	<b>0.017</b>	<b>0.057</b>	0.200	0.000	-0.053	-0.053	0.341	0.293	0.654
Oppositional Defiant Problems	0.000	-0.032	-0.032	0.402	0.339	0.679	0.000	-0.103	-0.103	0.222	0.271	0.658
Conduct Problems	0.000	-0.269	-0.269	<b>0.013</b>	<b>0.002</b>	<b>0.012</b>	0.000	-0.017	-0.017	0.447	0.435	0.758
<i>MacArthur (Year 6) - Factor Scores</i>												
Dysregulated Aggression	0.000	-0.040	-0.040	0.372	<b>0.099</b>	0.197	0.000	-0.179	-0.179	0.109	<b>0.056</b>	0.228
Warmth and Empathy	0.000	0.360	0.360	<b>0.004</b>	<b>0.006</b>	<b>0.025</b>	0.000	-0.097	-0.097	0.767	0.557	0.936
Emotional Integration	0.000	-0.045	-0.045	0.633	0.813	0.813	0.000	0.022	0.022	0.437	0.566	0.856
Performance Anxiety	0.000	-0.037	-0.037	0.381	<b>0.075</b>	0.210	0.000	0.051	0.051	0.637	0.811	0.811
Aggression	0.000	-0.182	-0.182	<b>0.066</b>	<b>0.001</b>	<b>0.006</b>	0.000	-0.151	-0.151	0.147	<b>0.082</b>	0.297
<i>Internalizing, Externalizing, Absences (Year 12)</i>												
Presence of Internalizing Disorders	0.239	-0.037	-0.037	0.253	0.310	0.649	0.403	-0.098	-0.098	<b>0.062</b>	<b>0.037</b>	<b>0.071</b>
Presence of Externalizing Disorders	0.182	-0.025	-0.025	0.310	0.592	0.831	0.187	0.085	0.085	0.943	0.840	0.840
Average # of Absences (School Years 1 - 5)	10.186	0.399	0.399	0.652	0.681	0.681	11.803	-1.798	-1.798	<b>0.034</b>	<b>0.026</b>	<b>0.074</b>

**Note:** The first column provides the outcome description. Our results are presented in six columns for each gender. The first column (Control Mean) of each result set shows the unconditional mean for the control group. When factor scores were computed, we set the mean in the control group to zero. The second column (Difference in Means) gives the conditional difference in means between the treatment group and the control group. As mentioned in Section 2, the control group stands for the original treatment group 2 of the NFP experiment and the treatment group stands for the original group 4. The third column (Effect Size) presents the conditional effect size for the respective group. The fourth column (Asymp.  $p$ -value) provides the asymptotic  $p$ -value for the one-sided single hypothesis test associated with the  $t$ -statistic for the conditional difference in means between treatment and control groups. The fifth column (Block Permutation - Single  $p$ -value) presents the one-sided restricted permutation  $p$ -values for the single-hypothesis testing based on the  $t$ -statistic associated with the treatment indicator in the Freedman and Lane (1983) regression as described in Section 3. By restricted permutation we mean that permutations are done within strata defined by the baseline variables used in the randomization protocol: maternal age and race, gestational age at enrollment, employment status of the head of the household, and geographic region. The covariates used in the Freedman and Lane (1983) regression are: maternal height, household income, grandmother support, maternal parenting attitudes and mother currently in school. Finally, the last column (Block Permutation - Stepdown) provides  $p$ -values that account for multiple-hypothesis testing based on the Stepdown procedure of Romano and Wolf (2005). Blocks of outcomes that are tested jointly are separated by lines. The selection of blocks of outcomes is done on the basis of their meaning. Outcomes that share similar meaning are grouped together. Female maternal outcomes allude to mothers whose first child is a girl. Likewise, male maternal outcomes allude to mothers whose first child is a boy. See Table F.5 in Appendix F for the unconditional analysis.

## 5 Understanding the Sources of NFP Treatment Effects

We decompose the NFP treatment effects on pre-pubescent outcomes into interpretable components associated with intervention-induced changes in early-age skills. Such decompositions are sometimes called mediation analyses in the statistics literature.<sup>18</sup> We follow the approach of Heckman et al. (2013) who adopt a linear approximation of the outcome equation:<sup>19</sup>

$$Y_{i,d} = \kappa + \underbrace{\alpha_d^m \theta_{i,d}^m}_{\text{measured skills}} + \underbrace{\beta_d \mathbf{X}_i}_{\text{pre-program var.}} + \underbrace{\alpha_d^u \theta_{i,d}^u}_{\text{unmeasured skills}} + \underbrace{\epsilon_i}_{\text{error term}}, \quad d \in \{0, 1\}, \quad (1)$$

where  $Y_{i,d}$  stands for the counterfactual outcome of agent  $i$  when the treatment is set to  $d \in \{0, 1\}$ :  $d = 1$  for treated and  $d = 0$  for control. Variables  $(\theta_{i,d}^m, \theta_{i,d}^u)$  represent the skills that can affect counterfactual outcome  $Y_{i,d}$  (e.g., achievement test scores) of measured and unmeasured components (measured skills are those that can be proxied by psychological instruments collected by NFP surveys).  $\mathbf{X}_i$  stands for observed pre-intervention variables that we control for. Parameters  $\alpha_d^u, \alpha_d^m, \beta_d$  denote vectors of coefficients that map changes in skills and pre-program variables into changes in counterfactual outcomes. Variable  $\epsilon_i$  is an unobserved zero-mean error term that is assumed to be statistically independent of all covariates.

It is useful to rewrite Equation (1) as:

$$Y_{i,d} = \tau_d + \alpha_d^m \theta_{i,d}^m + \beta_d \mathbf{X}_i + \tilde{\epsilon}_{i,d}; \quad d \in \{0, 1\}, \quad (2)$$

where  $\tau_d = \kappa + \alpha_d^u \mathbf{E}(\theta_{i,d}^u)$ ,  $d \in \{0, 1\}$ , and  $\tilde{\epsilon}_{i,d} = \epsilon_i + \alpha_d^u (\theta_{i,d}^u - \mathbf{E}(\theta_{i,d}^u))$ .

Treatment effects arise as a consequence of changes in measured skills  $\theta_{i,1}^m, \theta_{i,0}^m$  or changes in

<sup>18</sup>See VanderWeele (2015).

<sup>19</sup>See Appendix G for a detailed description of this model.

the mapping between these skills and the outcomes of interest, that is, changes in  $\alpha_1^m, \alpha_0^m$  or  $\tau_1, \tau_0$ . We test whether the map between measured skills and outcomes is the same for treated and control children in Tables I.3 and I.4 of In Appendix I. We do not reject the hypotheses associated with these restrictions.<sup>20</sup> Thus, we can simplify Equation (2) to:

$$Y_{i,d} = \tau_d + \alpha \theta_{i,d}^m + \beta \mathbf{X}_i + \tilde{\epsilon}_{i,d}; \quad d \in \{0, 1\}. \quad (3)$$

Let  $D_i \in \{0, 1\}$  to denote the treatment indicator. The observed outcome  $Y_i$  and observed skills  $\theta_i^m$  for participant  $i$  are given by  $Y_i = D_i \cdot Y_{i,1} + (1 - D_i) \cdot Y_{i,0}$ , and  $\theta_i^m = \theta_{i,1}^m D_i + \theta_{i,0}^m (1 - D_i)$ , respectively. In this notation, Equation (3) can be expressed as a linear regression equation:

$$Y_i = \tau_0 + \tau D_i + \alpha \theta_i^m + \beta \mathbf{X}_i + \tilde{\epsilon}_i \quad (4)$$

$$(5)$$

where  $\tilde{\epsilon}_i = \tilde{\epsilon}_{i,1} D_i + \tilde{\epsilon}_{i,0} (1 - D_i)$ , and  $\tau = \tau_1 - \tau_0$ . Parameter  $\tau = \tau_1 - \tau_0$  captures the effect of NFP on outcome  $Y$  that is due to the average change of unmeasured skills. This is the mean unexplained or residual effect. The coefficients  $\alpha$  map how changes in measured skills  $\theta_i^m$  affect the observed outcome  $Y_i$ . The ordinary least squares estimates of  $\alpha$  are unbiased if measured skills  $\theta_i^m$  and the error term  $\tilde{\epsilon}_i$  are uncorrelated. A sufficient condition for unbiasedness is that measured and unmeasured skills are statistically independent conditioned on  $\mathbf{X}, D$ .<sup>21</sup>

Appendix I presents robustness tests of the independence relation. We perform these tests using measures available in the data, in the same fashion as Heckman et al. (2013). We decompose the outcome conditional-mean treatment effects of the NFP program into

---

<sup>20</sup>We follow the inference method described in Heckman et al. (2013).

<sup>21</sup>The linearity assumption enables us to require a weaker condition: mean independence.

changes associated with measured and unmeasured skills by

$$\underbrace{E(Y|D = 1, \mathbf{X}) - E(Y|D = 0, \mathbf{X})}_{\text{conditional treatment effect}} = \underbrace{(\tau_1 - \tau_0)}_{\text{unmeasured skills}} + \underbrace{\alpha E(\boldsymbol{\theta}_1^m - \boldsymbol{\theta}_0^m)}_{\text{measured skills}}. \quad (6)$$

Equation (6) expresses treatment effects as a consequence of skill enhancements that are affected by the program.

Candidates for mediators are skills that are statistically significantly influenced by the program ( $E(\boldsymbol{\theta}_1^m - \boldsymbol{\theta}_0^m) \neq 0$ ) and that are related with the outcomes ( $\alpha \neq 0$ ). The next subsection describes our estimation methodology for decomposing NFP treatment effects.

## 5.1 Empirical Strategy to Decompose Treatment Effects

We estimate a factor model in which measured skills are expressed as factors  $\boldsymbol{\theta}_i^m$  using the range of psychological instruments available in the NFP data. We extract factors to identify the skills that are used to explain final outcomes. Factor models are particularly useful for summarizing large sets of response variables (i.e., psychological instruments) by a small number of latent variables (i.e., skills). Unobserved skills are estimated through a weighted average of relevant observed measures. Factor models outperform other arbitrary indexes by reducing the measurement error associated with factor estimation (Gorsuch, 1983).

We use a three-step procedure employed by Heckman et al. (2013). The first step consists of estimating a linear measurement system in which latent skills are a function of measures, that is, observed item-level data on psychological instruments. In the second step, we use the parameters estimated in the previous step to forecast skills. These forecasts are factor scores obtained by the Bartlett (1937) method. In the third step, we explain later outcomes in terms of earlier skills. Specifically, we use the computed factor scores as covariates in the linear regression represented by Equation (4). See Appendix H for a detailed explanation of the three-step procedure used to secure estimates.

A benefit of our application of the three-step procedure is that it clearly distinguishes

the effect of the intervention on skills and the effect of the experimentally induced changes in these skills on later outcomes. Our standard errors and  $p$ -values are computed using bootstrapping, as explained in Appendix H.

The NFP psychological instruments target well-specified personality traits (see Appendix C for a description). We incorporate this fact into our methodology by adopting a measurement system that is based on dedicated measures. By this we mean that each observed measure is linked to a single factor.<sup>22</sup> We allow the latent factors to be correlated. This approach is termed confirmatory factor analysis (Gorsuch, 1983). Similar to Heckman et al. (2013), we correct for the measurement error that arises from using an estimated factor, i.e., factor scores, instead of the true factor.

We conduct two sets of analyses. We first examine if child skill improvements at age 6 are mediated by variables measured at age 2. Next, we examine if child outcomes at age 12 are mediated by skills measured at age 6. The choice of potential mediators corresponds to those intermediate skills where significant treatment effects are found in our inferential analysis.<sup>23</sup> At age 2, we investigate five classes of mediators: (1) non-abusive parenting attitudes measured by the Bavolek Inventory; (2) home investments measured by the HOME inventory; (3) maternal anxiety from the Rand Mental Health Inventory<sup>24</sup>; (4) maternal self-esteem measured by the Rosenberg Scale; and (5) maternal mastery measured by the Pearlin Scale. At age 6, we examine three broad psychological instruments as mediators: (1) child’s cognitive skills measured by the KABC mental processing composite; (2) child’s socio-emotional skills measured by the Child Behavior Checklist scales of attention and conduct problems; (3) child’s warmth, empathy and aggression measured by the MacArthur Story Stem Battery.

---

<sup>22</sup>Heckman et al. (2013) show that the existence of at least three measures for each latent skill guarantees identification.

<sup>23</sup>We perform the analysis in this way since we do not have measures of both skills and investments over time. We only observe parental investments/skills at age 2 and child’s skills at age 6 and 12. For instance, we observe parental investments at age 2 (HOME score) but no proxy of parental investments at age 6. We have child’s cognitive skills at age 6 (K-ABC), but not at ages 2 or 12.

<sup>24</sup>The Rand Mental Health Inventory captures psychological distress and well-being factors. Its subscales include anxiety, depression, and positive well-being (Veit and Ware, 1983).

## 6 Decomposing Treatment Effects

We consider two channels generating NFP treatment effects. The first channel considers maternal investments and home environments as mediators. In this mediation study, we address the question of how parental investments respond to interventions and how such investments affect child skills. The second channel considers how program-induced enhancements of early skills translate to improvements in later skills.

Section 6.1 describes the mediators used in our analysis. Section 6.2 examines whether program improvements in child skills at age 6 are mediated by the effects on health at birth and home investments at age 2. In Section 6.3, we assess if the enhancements of child outcomes measured at age 12 are mediated by program effects on cognitive and socio-emotional skills at age 6.

### 6.1 NFP Effects on Skills at Ages 2 and 6

Figures 1–2 present the kernel densities of the factor scores used as mediators in our analysis. We also display the  $p$ -value for the single hypothesis inference of no mean treatment effects as described in Section 3.

Figure 1 shows the density of mediators at birth and age 2. It shows that NFP participation significantly increased birth weight of treated boys. At age 2, the NFP intervention significantly improved the quality of home environment for both boys and girls ( $p$ -values: 0.07 and 0.00, respectively). It also significantly reduced abusive/neglecting maternal attitudes ( $p$ -values: 0.01 for females and 0.00 for males). The maternal characteristics for mothers of females also improved: maternal anxiety decreased ( $p$ -value: 0.03) and maternal self-esteem increased ( $p$ -value: 0.00). Maternal mastery skills for mothers of both boys and girls were positively influenced as well.

Figure 2 shows the density of skills at age 6. For boys, the NFP intervention enhanced cognition and reduced aggression ( $p$ -values: 0.02 and 0.08, respectively). For girls, the

program reduced attention deficits, conduct problems, and aggression ( $p$ -values: 0.06, 0.00, and 0.00, respectively). NFP also increased cognitive and warmth or empathy skills (pro-social skills) for girls ( $p$ -values: 0.08 and 0.01, respectively).

## 6.2 Decomposition of Treatment Effects at Age 6

A large body of evidence corroborates the importance of early parental investments in promoting cognitive and socio-emotional development later in life (Almond and Currie, 2010; Cunha and Heckman, 2007; Heckman, 2008). We present evidence on these claims by examining how early inputs impact skills measured at age 6. These skills are: cognition, warmth/empathy, attention problems, conduct problems, and aggression. We only decompose skills whose treatment effects are statistically significant using the methods explained in Section 5. Therefore, the variables used for conducting the decompositions differ by gender. Figures 3 and 4 decompose the treatment effects on these variables into components associated with changes on factors evaluated at birth and age 2.<sup>25</sup>

We find that cognition at age 6 is enhanced through home environment and parenting practices for boys and girls. Specifically, the program-improved home investments at age 2 explain 35% of the treatment effect on cognition at age 6 for girls ( $p$ -value 0.03). The corresponding number for boys is 22% ( $p$ -value 0.05). Similarly, the enhanced parenting practices explain 14% of the treatment effect on cognition for girls ( $p$ -value 0.08). For boys, parenting practices explain 11% ( $p$ -value 0.05). Other contributions to cognition are birth weight and maternal anxiety.<sup>26</sup> Birth weight gains explain 14% of the cognitive treatment effect for boys ( $p$ -value: 0.06). For girls, maternal anxiety explains 25% ( $p$ -value: 0.08).

We find that the intervention-induced changes in home environments account for 21% of the treatment effect on female warmth/empathy ( $p$ -value: 0.01).<sup>27</sup> Improvements in home environments also account for 16% of the reduction in female aggression problems ( $p$ -value:

---

<sup>25</sup>The  $p$ -values are computed using the bootstrap method.

<sup>26</sup> Birth weight, a measure that summarizes prenatal investments and fetal development, is usually associated with childhood development (Breslau et al., 1994; Currie and Moretti, 2005).

<sup>27</sup>This controls for birth weight, gender and a range of pre-program baseline characteristics.



0.09). Program enhancement of parenting practices at age 2 explains 9% of the reduction in attention problems ( $p$ -value: 0.05) and 11% of the improvement in female warmth/empathy ( $p$ -value: 0.02). For boys, parenting practices explain 8% of the reduction in aggression ( $p$ -value: 0.09).

Maternal skills also influence child development. For girls, the decrease in maternal anxiety explains 14% of the reduction in conduct problems ( $p$ -value: 0.06). The contribution of maternal self-esteem is negative and accounts for 29% of the treatment effect on female warmth and empathy. This result is consistent with research that argues that improvements in self-esteem may increase selfishness (Burr and Christensen, 1992). Conversely, the mother’s maternal mastery explains 29% of the improvement in female warmth/empathy ( $p$ -value: 0.09).

### 6.3 Decomposition of Treatment Effects at Age 12

Figures 5–7 decompose statistically significant treatment effects at age 12 into treatment-induced changes of skills at age 6. We find that changes in male cognitive abilities at age 6 play a substantial role in explaining achievement scores at age 12. The program’s effects on cognition for boys at age 6 explain 41% of the treatment effects on the Tennessee Comprehensive Assessment Program (TCAP) test for language and 66% for math ( $p$ -value: 0.07).<sup>28</sup> The improvement of cognitive skills also explains 46% of the treatment effect on PIAT reading comprehension ( $p$ -value 0.04) and 51% of the treatment effect on PIAT math scores ( $p$ -value 0.04) for boys. Remarkably, boosts in cognition explain 68% of the treatment effect in the average math GPA for boys ( $p$ -value: 0.08).

Male cognitive gains also explain treatment effects on class absenteeism and internalizing problems. Improvements in cognition account for 24% of the of the reduction in the average number of day absences between the first and the fifth years of schooling ( $p$ -value 0.06). Cognitive gains also explain 17% of the male treatment effect on internalizing behavior ( $p$ -

---

<sup>28</sup>This achievement test was administered at grades 3 and above. In the control group, the mean percentile test score for reading was 38th percentile, and for math 40.3th percentile.

value: 0.06) and 14% of the results on male anxiety/depression at age 12 ( $p$ -value: 0.03). The intervention also reduced aggression at age 6, which accounts for 12% of the reduction in the likelihood of being anxious/depressed ( $p$ -value: 0.09 see Figure 6).<sup>29</sup>

The program decreased the likelihood of being overweight for females. The reduction in conduct problems fostered by the program at age 6 explains 19% of the treatment effect on standardized female BMI ( $p$ -value: 0.06). When we decompose the treatment effects on female risky behavior at age 12, we find that this treatment effect is largely unexplained by our mediators (not displayed in the figures).<sup>30</sup>

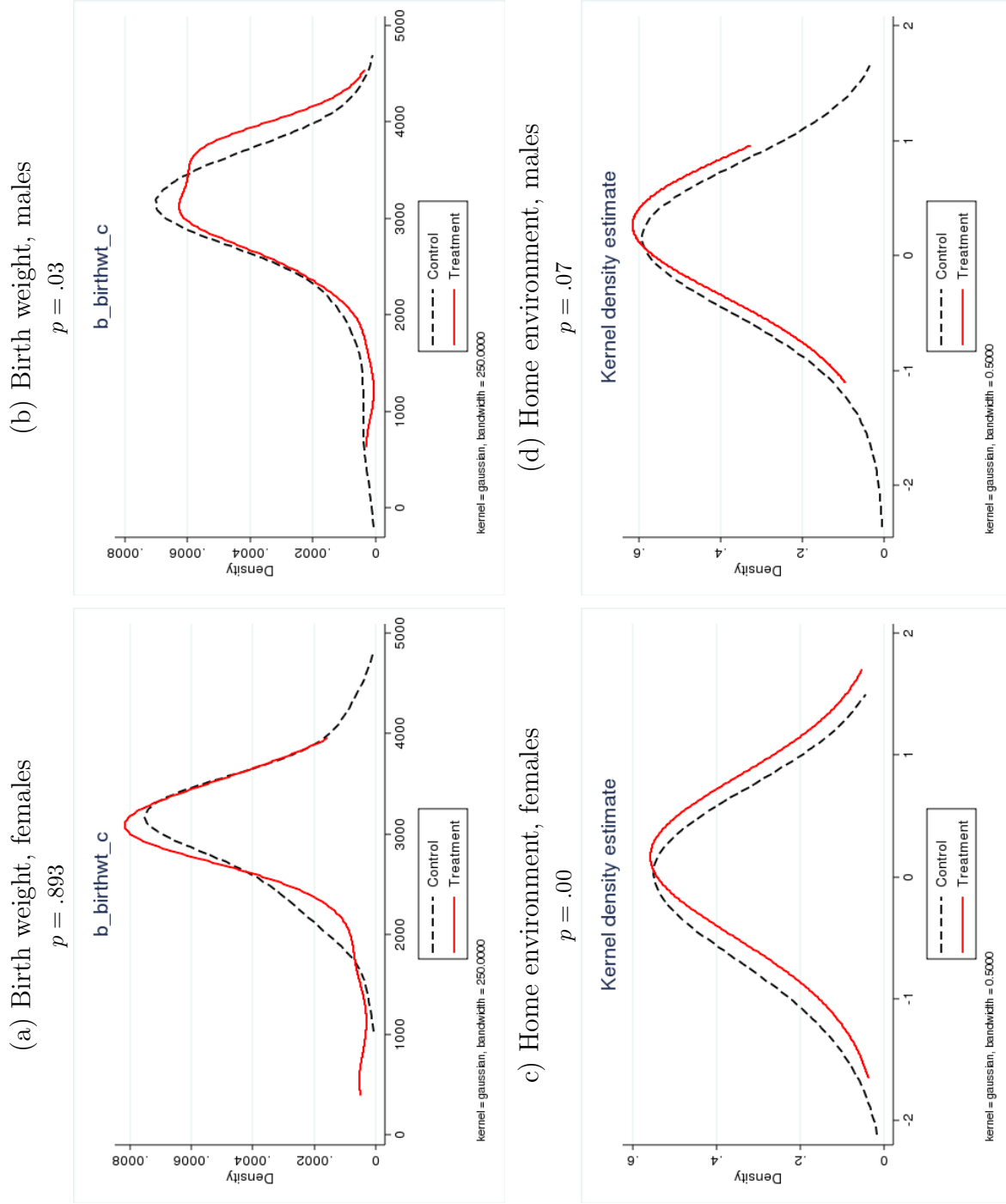
The differential treatment effects on cognition and achievement by gender follow the pattern of treatment effects found on health at birth, which are stronger for boys. Epidemiological and medical literature has argued that boys are more susceptible to prenatal influences than girls (Kraemer, 2000; Schore, 2017), which could explain why we find effects of the program on birth weight only for boys.

---

<sup>29</sup>Other related results are shown in Tables H.1–H.4 in the Appendix H

<sup>30</sup>They are shown in Table H.3 in the Appendix H

**Figure 1:** Kernel Densities of Factor Scores - 2 years skills

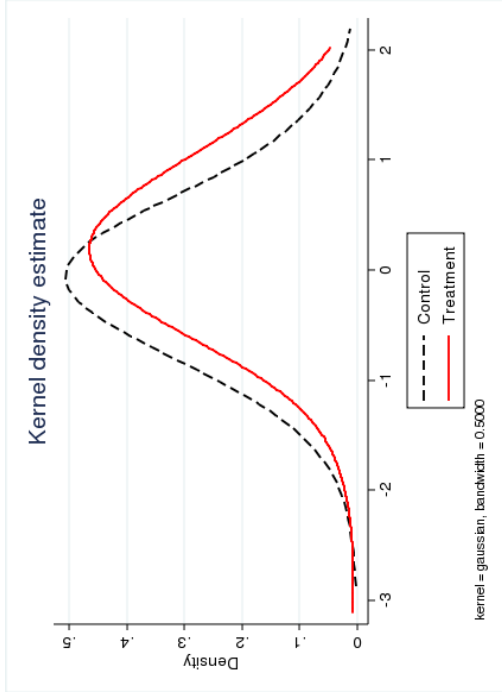


**Notes:** Kernel density functions based on [Bartlett \(1937\)](#) using a normal kernel. Numbers on the top of each figure correspond to one-sided FL permutation  $p$ -values testing the equality of factor scores means as explained in [Section 3](#) and presented in [Tables 6](#) and [7](#).

Figure 1: Kernel Densities of Factor Scores - 2 years skills, Cont'd

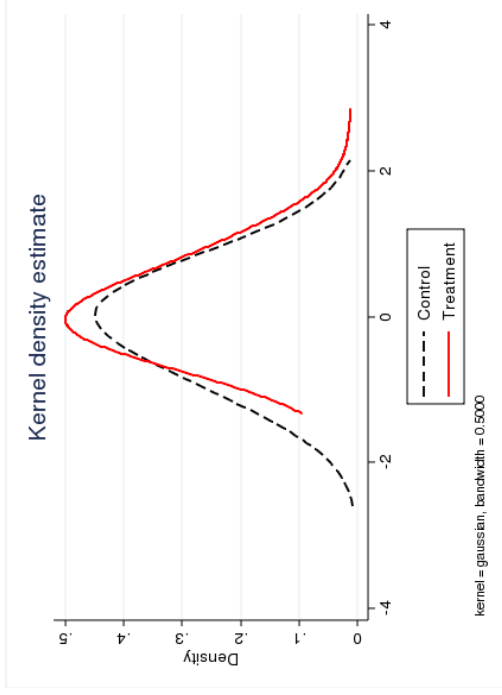
(e) Non-abusive parenting attitudes, females

$p = .01$



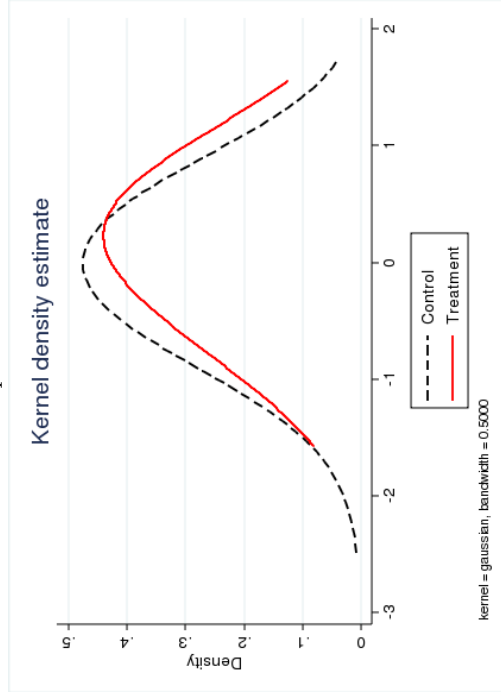
(f) Non-abusive parenting attitudes, males

$p = .00$



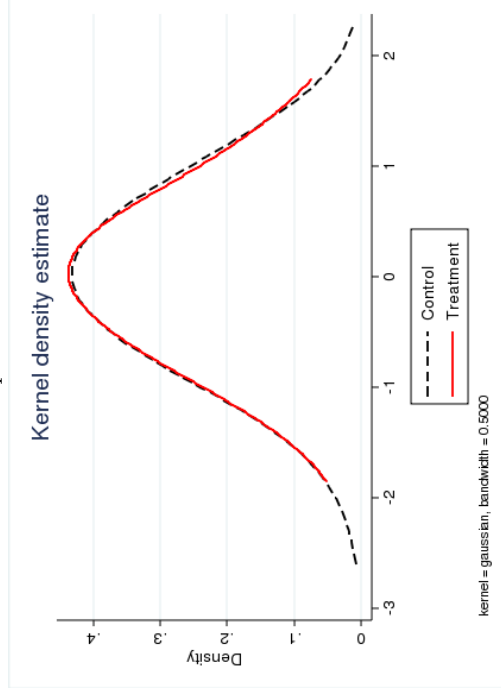
(g) Maternal less anxiety, females

$p = .03$



(h) Maternal less anxiety, male

$p = .46$

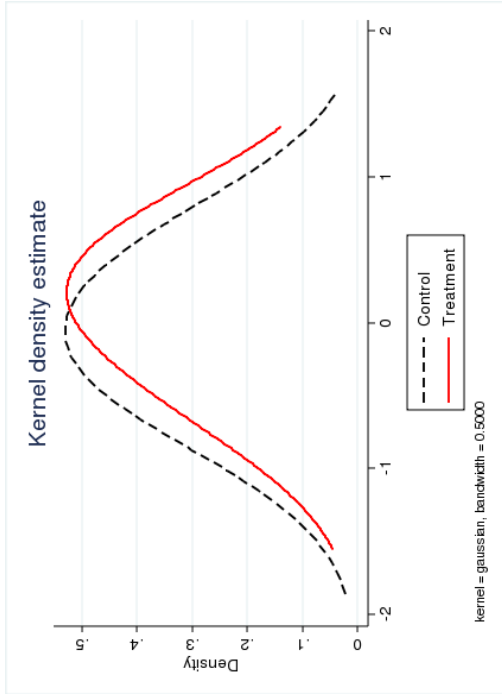


**Notes:** Kernel density functions based on [Bartlett \(1937\)](#) using a normal kernel. Numbers on the top of each figure correspond to one-sided FL permutation  $p$ -values testing the equality of factor scores means as explained in [Section 3](#) and presented in [Table 7](#).

**Figure 1:** Kernel Densities of Factor Scores - 2 years skills, Cont'd

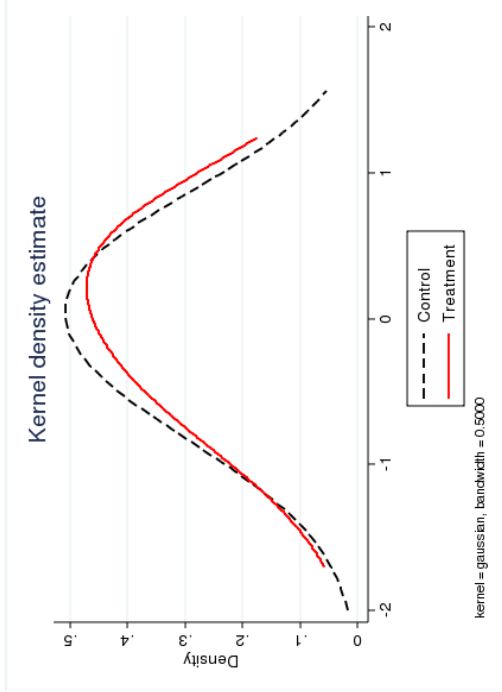
(i) Maternal self-esteem, females

$p = .00$



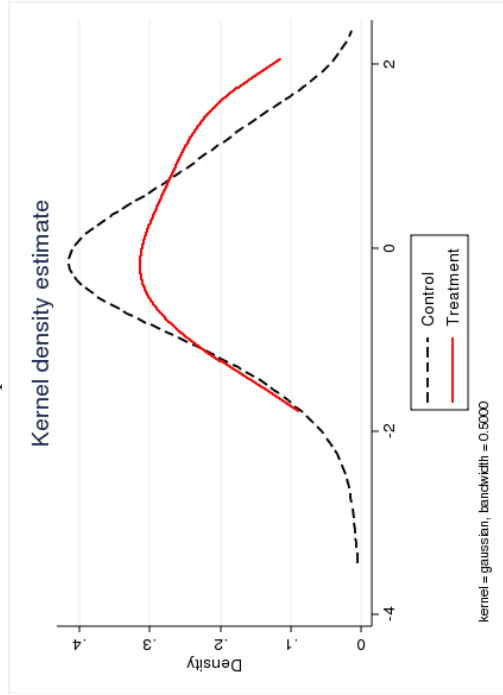
(j) Maternal self-esteem, males

$p = .45$



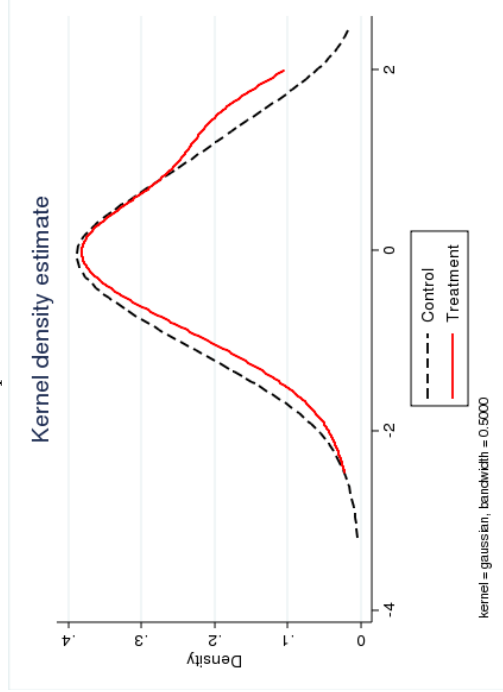
(k) Maternal mastery, females

$p = .01$



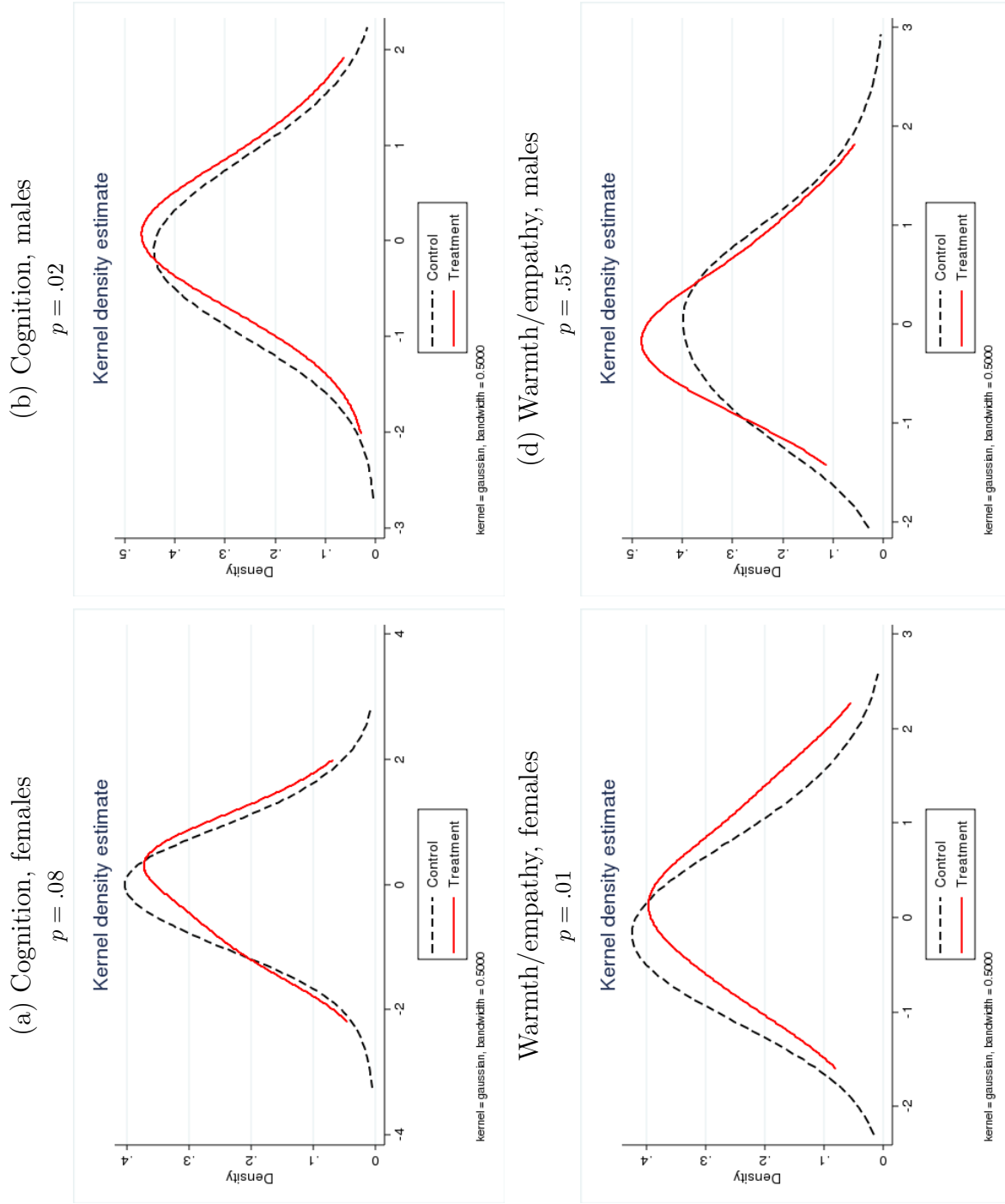
(l) Maternal mastery, males

$p = .08$



**Notes:** Kernel density functions based on [Bartlett \(1937\)](#) using a normal kernel. Numbers on the top of each figure correspond to one-sided FL permutation  $p$ -values testing the equality of factor scores means as explained in [Section 3](#) and presented in [Table 7](#).

**Figure 2:** Kernel Densities of Factor Scores - 6 years skills

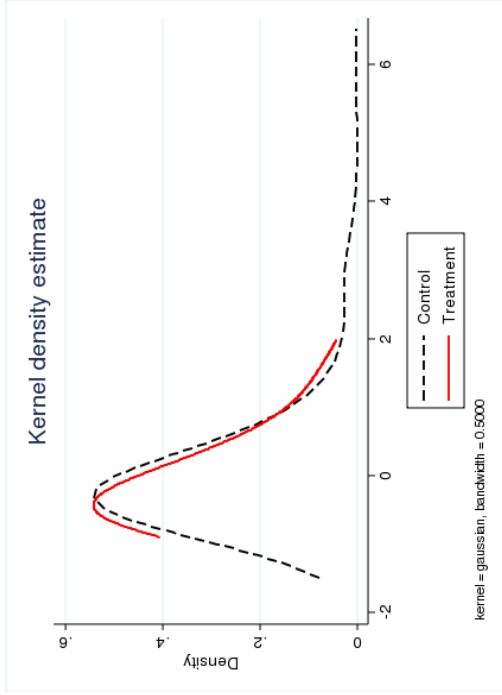


**Notes:** Kernel density functions based on [Bartlett \(1937\)](#) using a normal kernel. Numbers on the top of each figure correspond to one-sided FL permutation  $p$ -values testing the equality of factor scores means as explained in [3](#) and presented in [Tables 9](#) and [10](#).

**Figure 2:** Kernel Densities of Factor Scores - 6 years skills, Cont'd

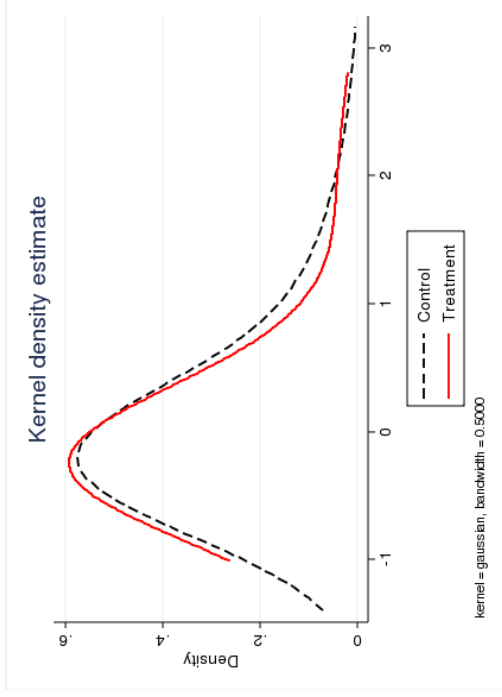
(e) Aggression, females

$p = .00$



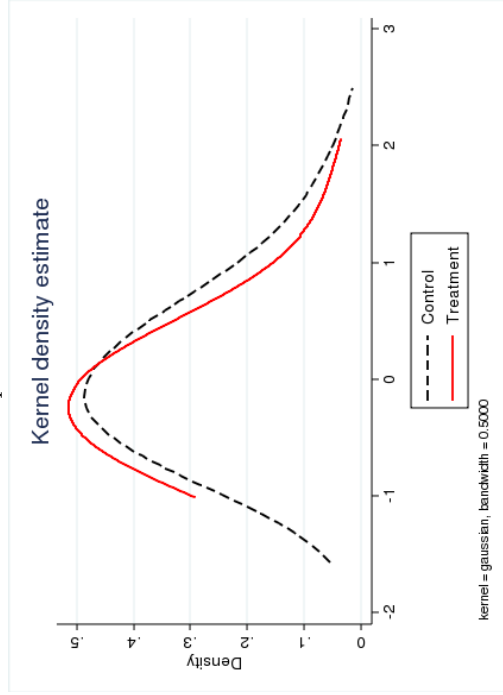
(f) Aggression, males (g)

$p = .08$



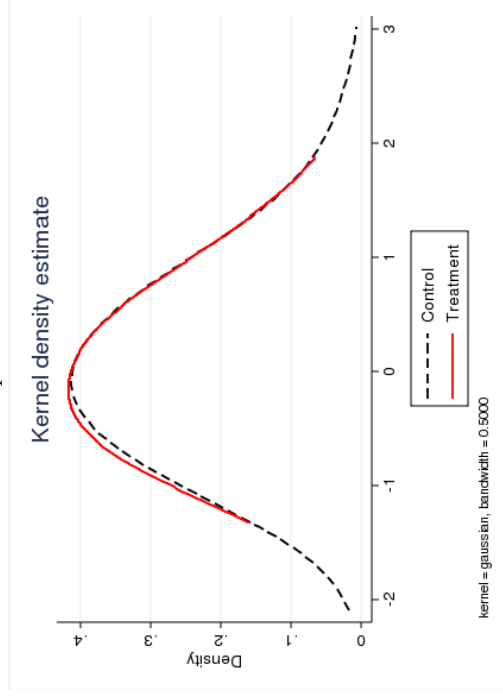
Attention problems, females

$p = .006$



(h) Attention problems, males

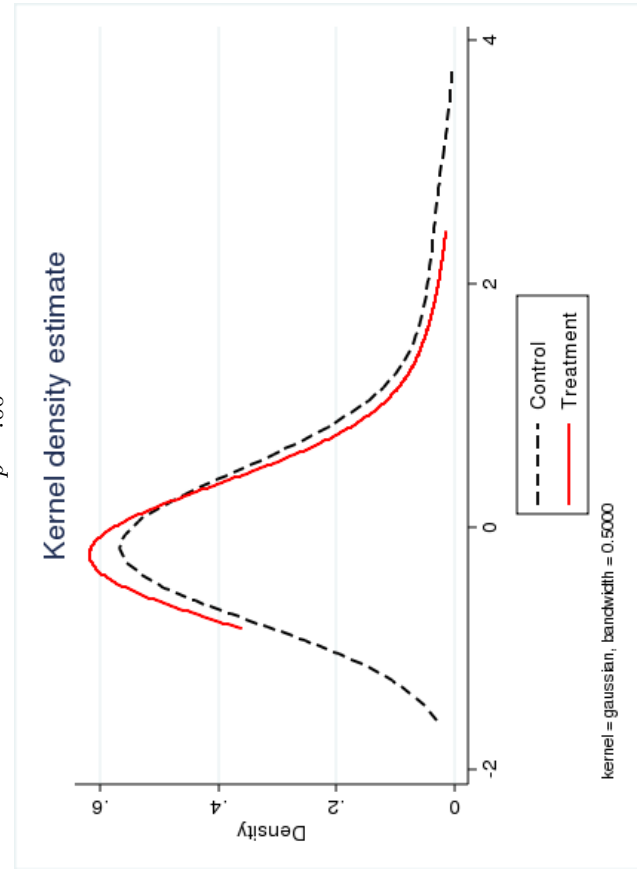
$p = .29$



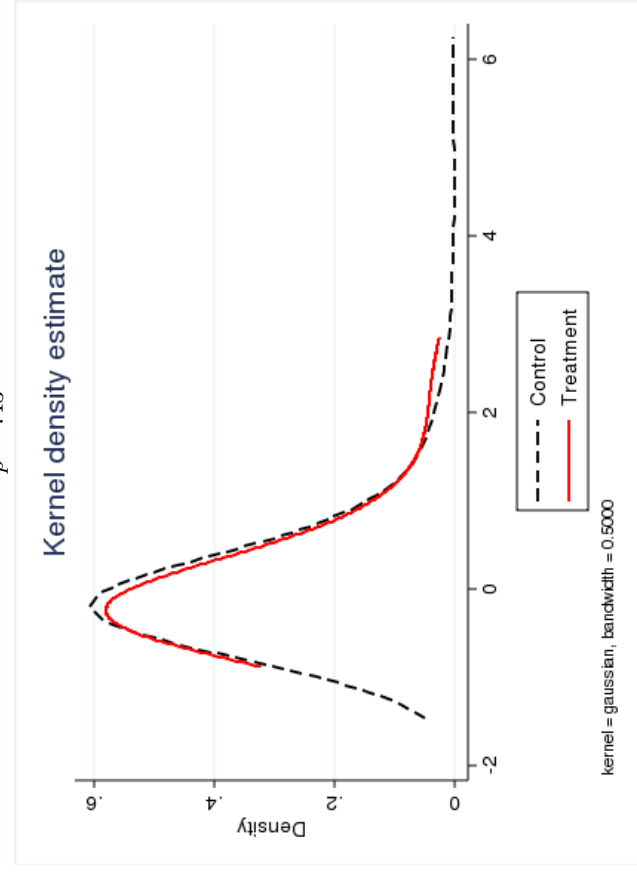
**Notes:** Kernel density functions based on [Bartlett \(1937\)](#) using a normal kernel. Numbers on the top of each figure correspond to one-sided FL permutation  $p$ -values testing the equality of factor scores means as explained in [3](#) and presented in [Table 10](#).

**Figure 2:** Kernel Densities of Factor Scores - 6 years skills, Cont'd

(i) Conduct problems, females  $p = .00$



(j) Conduct problems, males  $p = .43$



**Notes:** Kernel density functions based on [Bartlett \(1937\)](#) using a normal kernel. Numbers on the top of each figure correspond to one-sided FL permutation  $p$ -values testing the equality of factor scores means as explained in [3](#) and presented in [Table 10](#).



Figure 3: Decomposition of Treatment Effects on Female Outcomes - Age 6

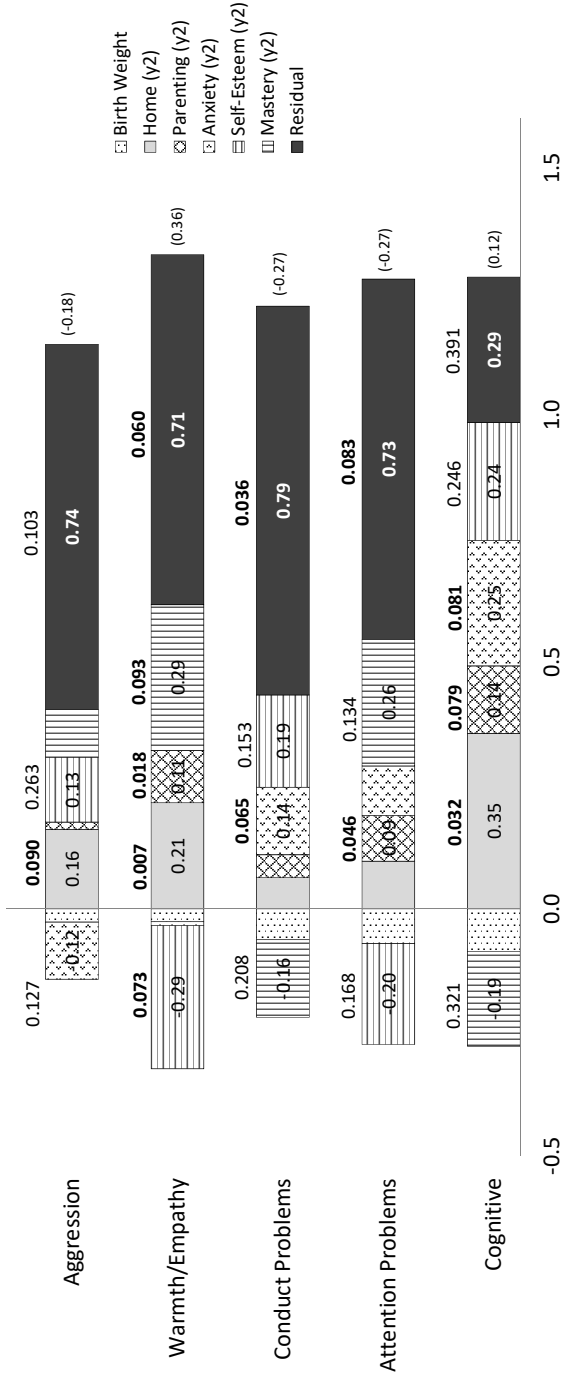
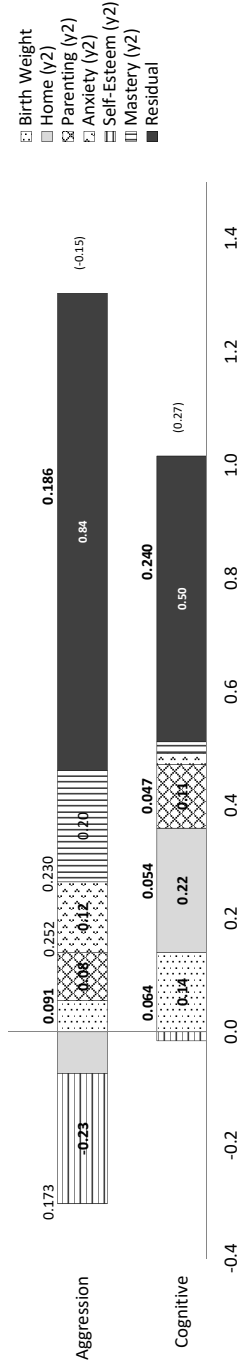


Figure 4: Decomposition of Treatment Effects on Male Outcomes - Age 6



**Notes:** These bar graphs show the fractions of the overall treatment effect explained by changes in the mediators as well as the unexplained portion (residual). The names of the outcomes are located on the left side of the bar graphs. Note that the fractions do not add up to 1 because some of the mediators contribute such a small and statistically insignificant fraction that they have not been shown. Values to the left of the vertical line are negative and values to the right of the vertical line are positive; bold values are significant at the 10% level. Values above the bars are the  $p$ -values associated with the mediation fraction, and values inside the bars are the mediation fractions themselves. Numbers in parentheses to the right of the bars are the total treatment effects. Negative values correspond to the mediator working in the opposite direction of the overall treatment effect, while positive values correspond to the mediator working in the same direction as the overall treatment effect. The mediators are the home environment, parenting attitudes, maternal anxiety, maternal self-esteem, and maternal mastery, all of which were measured when the child was 6 years old, as well as birth weight. The following controls were used: maternal race, maternal age, maternal height, gestational age, household density, region, employment status of household head, grandmother support, randomization wave, income category, mother currently in school, and maternal parenting attitudes.

Figure 5: Decomposition of Treatment Effects on Male Achievement Outcomes - Age 12

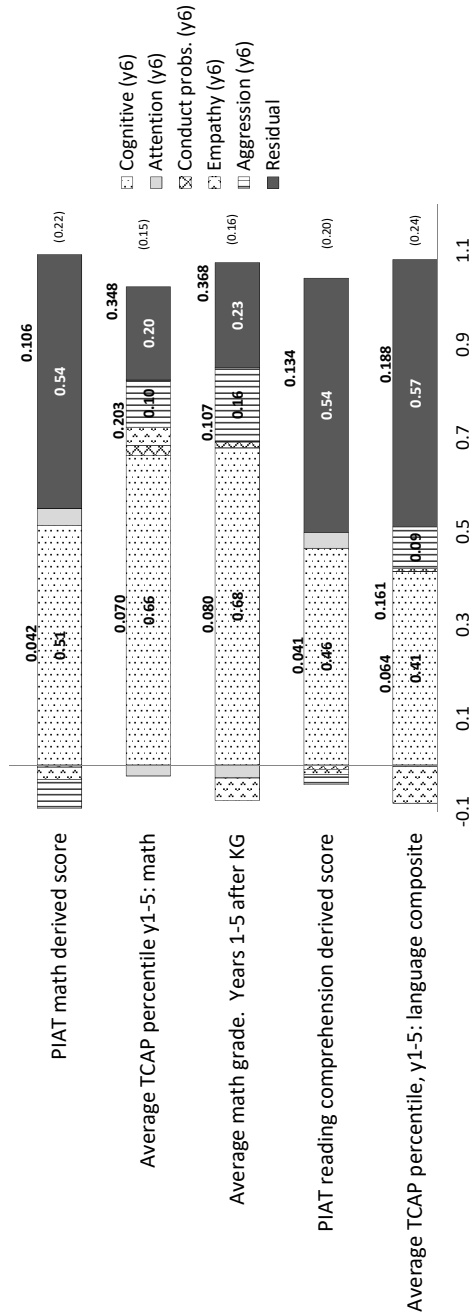
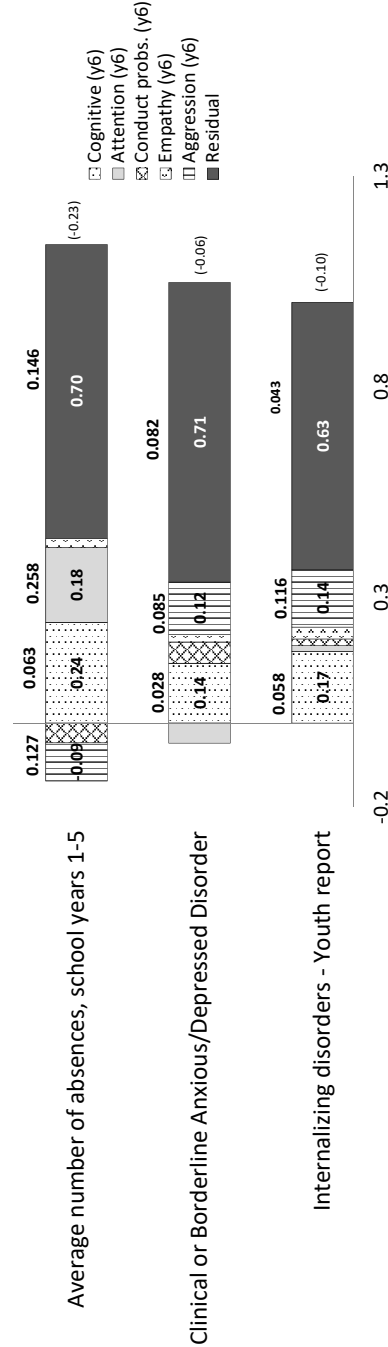
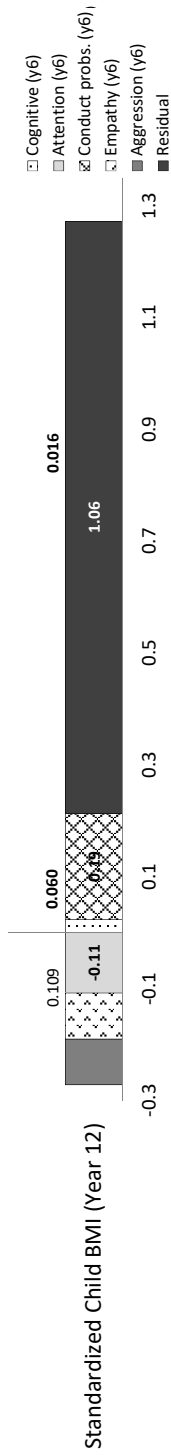


Figure 6: Decomposition of Treatment Effects on Male Non-cognitive Outcomes - Age 12



**Notes:** These bar graphs show the fractions of the overall treatment effect explained by changes in the mediators as well as the unexplained portion (residual). The names of the outcomes are located on the left side of the bar graphs. Note that the fractions do not add up to 1 because some of the mediators contribute such a small and statistically insignificant fraction that they have not been shown. Values to the left of the vertical line are negative and values to the right of the vertical line are positive; bold values are significant at the 10% level. Values above the bars are the *p*-values associated with the mediation fraction, and values inside the bars are the mediation fractions themselves. Numbers in parentheses to the right of the bars are the total treatment effects. Negative values correspond to the mediator working in the opposite direction of the overall treatment effect, while positive values correspond to the mediator working in the same direction as the overall treatment effect. The mediators are cognition, attention problems, conduct problems, aggression and warmth and empathy, all of which were measured when the child was 6 years old. The following controls were used: maternal race, maternal age, maternal height, gestational age, household density, region, employment status of household head, grandmother support, randomization wave, income category, mother currently in school, maternal parenting attitudes and the age at the 12 year interview.

Figure 7: Decomposition of Treatment Effects on Female Outcomes - Age 12



**Notes:** These bar graphs show the fractions of the overall treatment effect explained by changes in the mediators as well as the unexplained portion (residual). The names of the outcomes are located on the left side of the bar graphs. Note that the fractions do not add up to 1 because some of the mediators contribute such a small and statistically insignificant fraction that they have not been shown. Values to the left of the vertical line are negative and values to the right of the vertical line are positive; bold values are significant at the 10% level. Values above the bars are the  $p$ -values associated with the mediation fraction, and values inside the bars are the mediation fractions themselves. Numbers in parentheses to the right of the bars are the total treatment effects. Negative values correspond to the mediator working in the same direction as the overall treatment effect, while positive values correspond to the mediator working in the opposite direction of the overall treatment effect. The mediators are cognition, attention problems, conduct problems, aggression and warmth and empathy, all of which were measured when the child was 6 years old. The following controls were used: maternal race, maternal age, maternal height, gestational age, household density, region, employment status of household head, grandmother support, randomization wave, income category, mother currently in school, maternal parenting attitudes and the age at the 12 year interview.

## 7 Summary and Conclusion

This paper analyzes the impacts through age 12 of the Nurse-Family Partnership Program as implemented in Memphis. The Nurse-Family Partnership (NFP) is the most cited home visiting program in the US ([Howard and Brooks-Gunn, 2009a](#)) and is one of the evidence-based home visiting models supported by the Maternal, Infant and Early Childhood Home Visiting (MEICHV) program. Currently, NFP surrogates have provided services for more than 200,000 families in 43 states across the U.S.

We contribute to the existing literature in the NFP in several ways. We perform inference accounting for the special features of the NFP randomization protocol and addressing selective reporting of statistically significant outcomes (i.e., “cherry picking”). We show that most of the results of previous papers based on large sample asymptotic inference survive small sample permutation tests. However, fewer treatment effects survive corrections for multiple-hypothesis testing. Nonetheless, there are strong effects for boys that are sustained until age 12.

In addition, this is the first paper that formally examines the mechanisms underlying NFP treatment effects. Following the methodology developed by [Heckman et al. \(2013\)](#), we decompose statistically significant treatment effects into interpretable components associated with NFP-induced changes in children’s early skills and parental investments. We estimate the channels underlying treatment effects at age 6 (using maternal skills and investments effects at the end of the program as mediators ) and at age 12 (using the program effects at age 6 as mediators).

We find statistically significant treatment effects of NFP on birth weight for boys as well as improvements in the home environment, parenting attitudes, and maternal mental health for parents of both boys and girls at age 2.<sup>31</sup> At age 6, the program improves cognitive skills for both boys and girls, while it enhances early socio-emotional skills for girls. These

---

<sup>31</sup>Our result of treatment effects on birth weight and other health at birth outcomes for boys contrasts with recent findings of no effects on similar outcomes in the Family Nurse Partnership program in the U.K. ([Robling et al., 2016](#)).

treatment effects arise from program-induced improvements in maternal traits and early-life family investments at age 2. The treatment effects for males persist through age 12, but they do not persist for girls. Treated males outperform controls on a range of achievement scores. We find that 40% to 60% of male treatment effects at age 12 can be explained by enhanced cognitive skills measured at age 6. The program has much weaker long-term effects for girls.

Our results are consistent with the evidence from well-known home visiting interventions like the Infant Health and Development Project (IHDP), and current nurse visitation programs like Preparing For Life (Ireland) (García et al., 2017a; Doyle, 2017; Besharov et al., 2011; Klebanov et al., 2001; Brooks-Gunn et al., 1994). Table 11 compares the results from the present study with the evidence from these two interventions.

Home visiting programs are widely advocated policy tools for promoting early child development by serving both disadvantaged parents and children at the same time. Several program models have existed since the 1960's, and the Nurse-Family Partnership is one of the most widely used and evaluated models. Overall, several home visiting programs have shown positive effects on improving parenting practices and the home environment,<sup>32</sup> which are in line with our findings on those domains (Chen and Chan, 2016; Filene et al., 2013; Nievar et al., 2010).<sup>33</sup> However, in a meta-analytic study of home visiting programs, Chen and Chan (2016) argue that few studies have looked at the effects of these programs on maternal mental health. We provide some evidence that the NFP improved maternal mastery and self-esteem and reduced maternal anxiety with effect sizes between 0.2-0.3 SD. Regarding child development outcomes, Filene et al. (2013) argue that previous home visiting interventions have shown mixed effects on child cognitive skills and the effect sizes vary significantly among the studies that have found effects (between 0.11-0.38 SDs). In this paper, we present evidence that the NFP had strong effects on cognitive skills at age six with effect sizes of 0.27 SDs for boys and 0.12 SDs for girls. These cognitive gains persist in program impacts on achievement

---

<sup>32</sup>For example, Filene et al. (2013) show that average effects sizes in home visiting evaluations range between 0.13 and 0.33 SDs.

<sup>33</sup>For reviews about home visiting program evidence see Chen and Chan (2016); Filene et al. (2013); Howard and Brooks-Gunn (2009b); Nievar et al. (2010); Paulsell et al. (2010); Sweet and Appelbaum (2004).

test scores for boys at age 12. [Filene et al. \(2013\)](#) also demonstrate that the evidence of home visiting programs on birth outcomes is mixed. We find that the NFP improves these outcomes only for boys. Studies that review the evidence from home visiting programs argue that the existence of mixed effects across interventions can be explained because programs differ in the characteristics of the target population, goals, curriculum, services provided, frequency, the use of nurses versus paraprofessionals, and evaluation methods ([Chen and Chan, 2016](#); [Filene et al., 2013](#); [Nievar et al., 2010](#); [Paulsell et al., 2010](#)).

This paper provides evidence that some of estimated NFP program effects for boys are sustained after the intervention ends. The more beneficial long-term effects for boys relative to girls is consistent with a large and growing body of evidence ([Elango et al., 2014](#); [García et al., 2017a,b](#) and [García et al., 2017](#)). We also examine the underlying channels generating the estimated treatment effects and show that enhancements in the skills in early childhood and better parenting can explain a sizeable portion of NFP effects found in late childhood.

Table 11: Comparison of Memphis NFP study with, IHDP, and PFL interventions

Program Characteristics	IHDP	PFL	NFP
Targeted Population	Low birth weight and pre-term Birth-Age 3	Disadvantaged/Low SES Pregnancy-Age 5	Disadvantaged/Low SES Pregnancy-Age 2
Age	1990-1991	1985	2008-2010
Birth Cohort	742 children	985 children	233 children
Total Sample Size	514 Control	608 Control	118 Control
Control Sample	228 Treated	377 Treated	115 Treated
Treated Sample	Home visiting	Home visiting	Home visiting
Treatment			
<b>Program Effects</b>	Full sample	Full sample	Girls Boys
Home Environment/Parenting (age 2)	Sig. effect (+)	Mixed effects	Sig. effect (+)
Cognitive skills (age 6)	Mixed effects	Sig. effect (+)	Sig. effect (+)
Behavior problems (age 6)	No sig. effect	Mixed effects	Sig. effect (-)
Achievement scores around (age 12)	Mixed effects	N.A	No sig. effect
			Sig. effect (+)

Sources: IHDP (Infant Health Development Program): [Besharov et al. \(2011\)](#); [Brooks-Gunn et al. \(1994\)](#); [Klebanov et al. \(2001\)](#) PFL (Preparing For Life): [Doyle \(2017\)](#). NFP (Nurse-Family Partnership): Our paper. Significant effects mean that the estimated average treatment effect is statistically significant in at least two instruments or subscales. The signs (+) and (-) denote the expected direction (positive or negative) of estimated treatment effects. Mixed effects means that either the estimate of the average treatment effect is statistically significant in one instrument but not in others, or that the treatment effect is only statistically significant for a subgroup. No Effects means that the estimates of the intervention average treatment effects on that type of outcome are not statistically significant.

## References

- Almond, D. and J. Currie (2010). Human capital development before age five. Technical report, National Bureau of Economic Research.
- Anderson, M. J. and P. Legendre (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation* 62(3), 271–303.
- Association, A. P. and A. P. A. T. F. on DSM-IV. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Publishing, Inc.
- Avellar, S., D. Paulsell, et al. (2011). Lessons learned from the home visiting evidence of effectiveness review. Technical report, Mathematica Policy Research.
- Bartlett, M. S. (1937, July). The statistical conception of mental factors. *British Journal of Psychology* 28(1), 97–104.
- Besharov, D., P. Germanis, C. Higney, and D. Call (2011). Infant health and development program. *Assessments of twenty-six early childhood evaluations*. Maryland School of Public Policy Welfare Reform Academy.
- Breslau, N., J. DelDotto, G. Brown, S. Kumar, S. Ezhuthachan, K. Hufnagle, and E. Peterson (1994). A gradient relationship between low birth weight and iq at age 6 years. *Archives of Pediatrics and Adolescent Medicine* 148(4), 377.
- Brooks-Gunn, J., C. M. McCarton, P. H. Casey, M. C. McCormick, C. R. Bauer, J. C. Bernbaum, J. Tyson, M. Swanson, F. C. Bennett, D. T. Scott, et al. (1994). Early intervention in low-birth-weight premature infants: Results through age 5 years from the infant health and development program. *Jama* 272(16), 1257–1262.
- Burr, W. and C. Christensen (1992). Undesirable side effects of enhancing self-esteem. *Family Relations*, 460–464.
- Chen, M. and K. L. Chan (2016). Effects of parenting programs on child maltreatment prevention: a meta-analysis. *Trauma, Violence, & Abuse* 17(1), 88–104.
- Cunha, F. and J. J. Heckman (2007). Identifying and estimating the distributions of *Ex Post* and *Ex Ante* returns to schooling: A survey of recent developments. *Labour Economics* 14(6), 870–893.
- Currie, J. and E. Moretti (2005). Biology as destiny? short and long-run determinants of intergenerational transmission of birth weight.
- Currie, J. and M. Rossin-Slater (2015). Early-life origins of life-cycle well-being: Research and policy implications. *Journal of Policy Analysis and Management* 34(1), 208–242.
- Daro, D. (2006). Home visitation assessing progress, managing expectations. *Chapin Hall Center for Children*.



- Doyle, O. (2017). The first 2,000 days and child skills: Evidence from a randomized experiment of home visiting.
- Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2016). Early childhood education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 2, Chapter 4, pp. 235–297. Chicago: University of Chicago Press.
- Elango, S., A. Hojman, and S. Kuperman (2014). Perry cohort differences.
- Filene, J. H., J. W. Kaminski, L. A. Valle, and P. Cachat (2013). Components associated with home visiting program outcomes: A meta-analysis. *Pediatrics* 132(Supplement 2), S100–S109.
- Freedman, D. and D. Lane (1983, October). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics* 1(4), 292–298.
- García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2017a). The life-cycle benefits of an influential early childhood program. Unpublished.
- García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2017b). Quantifying the life-cycle benefits of a prototypical early childhood program. Unpublished.
- García, J. L., J. J. Heckman, and A. L. Ziff (2017). Gender differences in the effects of early childhood education. Unpublished.
- Gorsuch, R. (1983). *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 768–806.
- Heckman, J. J. (2008). Role of income and family influence on child outcomes. *Annals of the New York Academy of Sciences* 1136(Reducing the Impact of Poverty on Health and Human Development: Scientific Approaches), 307–323.
- Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010, August). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics* 1(1), 1–46.
- Howard, K. and J. Brooks-Gunn (2009a). The role of home-visiting programs in preventing child abuse and neglect. *The Future of Children*, 119–146.
- Howard, K. S. and J. Brooks-Gunn (2009b). The role of home-visiting programs in preventing child abuse and neglect. *The Future of Children* 19(2), 119–146.
- Kitzman, H., D. Olds, R. Cole, C. Hanks, E. Anson, K. Arcoleo, D. Luckey, M. Knudtson, C. Henderson Jr, and J. Holmberg (2010). Enduring effects of prenatal and infancy home visiting by nurses on children: follow-up of a randomized trial among children at age 12 years. *Archives of Pediatrics and Adolescent Medicine* 164(5), 412.

- Kitzman, H., D. Olds, C. Henderson, C. Hanks, R. Cole, R. Tatelbaum, K. McConnochie, K. Sidora, D. Luckey, D. Shaver, et al. (1997). Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing. *Jama* 278(8), 644–652.
- Klebanov, P. K., J. Brooks-Gunn, and M. C. McCormick (2001). Maternal coping strategies and emotional distress: Results of an early intervention program for low birth weight young children. *Developmental Psychology* 37(5), 654.
- Kraemer, S. (2000). The fragile male. *Clinical Medicine NetPrints*, 1.
- Nelson, C. A. (2012, October). The effects of early life adversity on brain and behavioral development. Technical report, The Dana Foundation.
- Nievar, M. A., L. A. Van Egeren, and S. Pollard (2010). A meta-analysis of home visiting programs: Moderators of improvements in maternal behavior. *Infant mental health Journal* 31(5), 499–520.
- Olds, D. (2002). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science* 3(3), 153–172.
- Olds, D., H. Kitzman, R. Cole, and J. Robinson (1997). Theoretical foundations of a program of home visitation for pregnant women and parents of young children. *Journal of Community Psychology* 25(1), 9–25.
- Olds, D., H. Kitzman, R. Cole, J. Robinson, K. Sidora, D. Luckey, C. Henderson Jr, C. Hanks, J. Bondy, and J. Holmberg (2004). Effects of nurse home-visiting on maternal life course and child development: age 6 follow-up results of a randomized trial. *Pediatrics* 114(6), 1550.
- Olds, D. L. (2012). Improving the life chances of vulnerable children and families with prenatal and infancy support of parents: the nurse-family partnership. *Psychosocial Intervention* 21(2), 129–143.
- Paulsell, D., S. Avellar, E. S. Martin, P. Del Grosso, et al. (2010). Home visiting evidence of effectiveness review: Executive summary. Technical report, Mathematica Policy Research.
- Pearlin, L. I. and C. Schooler (1978). The structure of coping. *Journal of health and social behavior*, 2–21.
- Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89(427), 846–866.
- Robling, M., M.-J. Bekkers, K. Bell, C. C. Butler, R. Cannings-John, S. Channon, B. C. Martin, J. W. Gregory, K. Hood, A. Kemp, J. Kenkre, A. A. Montgomery, G. Moody, E. Jones-Owen, K. Pickett, G. Richardson, Z. E. S. Roberts, S. Ronaldson, J. Sanders, E. Stamuli, and D. Torgerson (2016, January). Effectiveness of a nurse-led intensive home-visitation programme for first-time teenage mothers (Building Blocks): A pragmatic randomised controlled trial. *The Lancet* 387(10014, 9-15), 146–155.

- Romano, J. P. and M. Wolf (2005, March). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* 100(469), 94–108.
- Rosenberg, M. (1965). *Society and the adolescent self-image*, Volume 11. Princeton university press Princeton, NJ.
- Sandner, M., T. Cornelissen, T. Jungmann, and P. Herrmann (2017). Hceo working paper series.
- Schore, A. N. (2017). All our sons: The developmental neurobiology and neuroendocrinology of boys at risk. *Infant Mental Health Journal* 38(1), 15–52.
- Sweet, M. A. and M. I. Appelbaum (2004). Is home visiting an effective strategy? a meta-analytic review of home visiting programs for families with young children. *Child development* 75(5), 1435–1456.
- VanderWeele, T. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.
- Veit, C. T. and J. E. Ware (1983). The structure of psychological distress and well-being in general populations. *Journal of consulting and clinical psychology* 51(5), 730.