The Lasting Effects of Early-Childhood Education on Promoting the Skills and Social Mobility of Disadvantaged African Americans and Their Children

Jorge Luis García

Clemson University, Institute of Labor Economics, and National Bureau of Economic Research

James J. Heckman

University of Chicago and University of Southern California

Victor Ronda

University of Chicago

This paper demonstrates the long-term intra- and intergenerational benefits of the HighScope Perry Preschool Project, which targeted disadvantaged African American children. We use newly collected data on the original participants through late middle age and on their children into their midtwenties. We document long-lasting improvements in the original participants' skills, marriage stability, earnings, criminal behavior, and health. Beneficial program impacts through the child-rearing years translate into better family environments for their children, leading to intergenerational gains. Children of the original participants have higher levels of education and employment, lower levels of criminal activity, and better health than children of the controls.

This research was supported by the Buffett Early Childhood Fund and the National Institutes of Health's Eunice Kennedy Shriver National Institute of Child Health and Human Development under award R37HD065072 and the National Institute of Aging under awards R01AG042390 and R01AG053343. This research was also supported in part by the Leonard

Electronically published May 26, 2023

Journal of Political Economy, volume 131, number 6, June 2023.

https://doi.org/10.1086/722936

I. Introduction

This paper analyzes newly collected life-cycle panel data on the original participants of the pioneering HighScope Perry Preschool Project (PPP) social experiment through late midlife and on their children into their midtwenties. We use longitudinal data based on multiple surveys and administrative criminal records. PPP aims to promote the social mobility of disadvantaged African American children, and it is successful. It also has substantial beneficial intergenerational effects. Gains in cognition are sustained through late midlife, contradicting claims about cognitive fade-out of PPP and other early-childhood programs. Enriched early-childhood education programs are promising vehicles for promoting social mobility within and across lifetimes.

A program created 60 years ago is relevant today because it influences the design of current and proposed early-childhood education programs. The populations it was designed to serve are still substantial. At least 30% of current Head Start programs are based on it (Elango et al. 2016). About 10% of African American children born in the 2010s satisfy the eligibility criteria for PPP.[1] Commonalities over time and across cultures and ethnic groups in the process of child development make our conclusions relevant to other contexts (see WHO Multicentre Growth Reference Study Group 2006; Fernald et al. 2017; Ertem et al. 2018). Our study provides general lessons for policies that foster child development.

It is well documented that PPP improved the life-cycle outcomes of its original participants through age 40 (e.g., Heckman et al. 2010a; Elango et al. 2016; Heckman and Karapakula 2021). We show that positive impacts on the original participants persist through their child-rearing years. These gains led to better environments for their children, who are more likely than children of the first generation of control participants to grow up in stable two-parent households. Their parents have higher average

D. Schaeffer Center for Health Policy and Economics at the University of Southern California. The views expressed in this paper are solely ours and do not necessarily represent those of the funders or the official views of the National Institutes of Health. We thank the researchers of the HighScope Educational Research Foundation's Perry Preschool Project—especially Alejandra Barraza and Lawrence Schweinhart—for access to study data and source materials. We also thank Melissa Dell and three anonymous referees for constructive comments. This paper supersedes Heckman and Karapakula (2019a), an unpublished manuscript that presents a preliminary analysis of the intergenerational data analyzed in this paper. Heckman and Karapakula (2019a) has been withdrawn. Instructions for requesting access to the data used in this paper and code for replicating the tables and figures in the main text and appendix can be found in the Harvard Dataverse: https://doi.org/10.7910/DVN/WUUZJ3. This paper was edited by Melissa Dell.

[1] This is the percentage of males and females born in households satisfying PPP's eligibility criteria. We calculate it using the US Census Bureau's 2010 and 2015 American Community Surveys.

earnings, less engagement with the criminal justice system, and better executive functioning (cognition), socioemotional skills, and health.

PPP did not directly treat the children of the original participants. Nonetheless, it generated positive intergenerational externalities. Children of treated participants are 17 percentage points less likely to have been suspended from school during their K–12 years compared with children of control participants. They are also 9 percentage points more likely to be in good health through young adulthood, 26 percentage points more likely to be employed, and 8 percentage points less likely to be divorced. There are pronounced impacts by sex. Children of male treated participants are 18 percentage points less likely to have been arrested through young adulthood compared with children of male control participants. Our estimates are statistically significant and robust when we use multiple estimation strategies and inferential procedures designed to address methodological challenges inherent in PPP and many other social experiments (Bruhn and McKenzie 2009). We apply rigorous small-sample inferential methods in recognition of PPP's sample size and to counter the undocumented but often repeated claim that "Perry's samples are too small."

This paper proceeds in the following way. Section II briefly describes the literature related to our study. It clarifies our contribution relative to other work on the intergenerational impact of early-childhood education and to recent studies of PPP. Section III describes the program, our data, and our methodology. Section IV presents impacts on the original participants of PPP, which include improvements in the environments in which their children were raised. Section V presents our intergenerational estimates. Section VI concludes.

II. Related Literature

Little is known about the intergenerational impacts of early-childhood education. Rossin-Slater and Wüst (2020) and Barr and Gibbs (2022) are exceptions. The latter paper exploits differential timing in preschool availability in Denmark during the period 1933–60 and studies its intergenerational impact on educational attainment at age 25. Barr and Gibbs (2022) study the intergenerational impact of Head Start programs available in the 1970s using a similar design. They analyze education, teenage pregnancy, and youth criminality. Both studies find beneficial intergenerational impacts.

Our study breaks new ground because it is based on experimental data and collects detailed information about the long-term life-cycle outcomes of the original participants and the outcomes of their children. We study intergenerational outcomes of the children across the life cycle from early life (e.g., special education and school suspension) to young adulthood (e.g., employment and marriage stability).

Recent studies of the HighScope PPP.—A companion paper, García et al. (2021), monetizes the treatment effects of the program we study through age 54. It reports a cost-benefit ratio of 6.0 after adjusting for the distortion generated by the taxes required to fund the program. That paper focuses on the cost-benefit analysis of the program. It does not analyze the life-cycle patterns that we document in this paper. It also does not analyze in detail the age 54 outcomes of the original participants discussed in this paper, especially those related to newly collected measures of skills. Its analysis includes a crude monetization of some of the intergenerational treatment effects in this paper. It does not present treatment effects on the newly collected intergenerational outcomes. Heckman et al. (2010a) and Heckman and Karapakula (2021) study the impact of PPP on its original participants through age 40. Both studies develop identification, estimation, and inference methods especially suited for tackling the challenges inherent to PPP's design and implementation.

Heckman and Karapakula (2019b) is the working paper version of Heckman and Karapakula (2021). We use some of the methods developed and tested in that paper. Heckman et al. (2010b) and Heckman, Pinto, and Savelyev (2013) are related studies. Heckman et al. (2010b) provide estimates of the internal rate of return of PPP using extrapolations informed by original-participant data through age 40. Heckman, Pinto, and Savelyev (2013) develop and apply a mediation framework to document that the short-term impact of PPP on socioemotional skills largely explains the long-term impacts on age 40 outcomes such as employment and crime. Conti, Heckman, and Pinto (2016) report treatment effects on health outcomes of the original participants at ages 27 and 40. They build on the framework of Heckman, Pinto, and Savelyev (2013) to provide dynamic mediation analyses of these health outcomes. None of these studies use the intergenerational data studied in this paper.

III. Program, Data, and Methods

A. *The HighScope PPP*

The HighScope PPP was a high-quality early-childhood education program.[2] Its participants were born in the late 1950s and early 1960s. Its curriculum was designed to foster development of cognitive and socioemotional skills. Children were active learners who planned, executed, and reflected on activities guided by teachers. Children made choices and

[2] We refer interested readers to Weikart, Bond, and McNeil (1978) and Heckman et al. (2010a) for extensive details on PPP and its rounds of data collection and to Kautz et al. (2014) and Elango et al. (2016) for a broad discussion of PPP and its relationship with other influential early-education and social programs.

solved problems. Teachers gave them feedback (Schweinhart, Barnes, and Weikart 1993), a form of reinforcement learning (Dehaene 2021).[3]

Participants lived in the catchment area served by the Perry Elementary School in Ypsilanti, Michigan. In-school surveys, referrals, and canvassing identified an initial pool of participants. Eligibility criteria were based on IQ scores and socioeconomic status. A pool of 123 disadvantaged African American children was randomized into the program (treatment group) or not (control group). Treatment group children received 2 years of 2.5-hour preschool sessions during weekdays starting at age 3. They also received weekly teacher home visits during the 2-year treatment period. Control group children did not receive any treatment because there were no treatment substitutes available in the area where they lived. The program was implemented before Head Start and indeed influenced its design and creation. Comparing the treatment and control groups allows us to identify program impacts compared with no treatment in any other program. For early-childhood programs implemented later, control group parents enroll their children in alternative preschools of varying quality. Identification of clearly defined treatment effects thus requires additional assumptions to control for choices of other options (Heckman et al. 2000; Kline and Walters 2016; García et al. 2018).

Weikart, Bond, and McNeil (1978) report that every family that received an offer to participate in PPP accepted it. We thus estimate the average treatment effect for program eligibles. Heckman et al. (2010a) report that 15% of African American females and 17% of African American males satisfied PPP's eligibility criteria at the time of its implementation. After participants were randomized, the status of a few participants was swapped. This reassignment potentially compromised the randomization protocol and resulted in an imbalance of baseline characteristics (see table 1). Failure in implementation of randomization protocols is not rare in social experiments. This failure can have sizable empirical consequences (Bruhn and McKenzie 2009). We are up front about these issues and adjust our estimates accordingly.[4]

[3] Barnett (1996) reports a total program cost per participant of $21,151 (in 2017 US dollars) over the 2-year life of the program, which ranks PPP in the lower end among programs of its type regarding implementation cost (Elango et al. 2016).

[4] The randomization protocol was as follows: (1) Participant status of the younger siblings was the same as that of their older siblings. (2) Those remaining were ranked by their baseline IQ score with odd- and even-ranked subjects assigned to separate groups (we do not know the pairings). (3) Some individuals initially assigned to one group were swapped between groups to balance gender and mean socioeconomic index scores, with average IQ scores held more or less constant. This generated a minor imbalance in family background variables. (4) A coin toss randomly selected one group as the treatment group and the other as the control group. (5) Some individuals provisionally assigned to treatment whose mothers were employed at the time of the assignment were swapped with control individuals whose mothers were not employed. The reason for this swap was that it was difficult for working mothers to participate in home visits assigned to the treatment group.

TABLE 1
Original Participants, Sample Sizes, and Unadjusted Mean Differences

	Pooled		Male		Female	
	Control	Treatment-Control	Control	Treatment-Control	Control	Treatment-Control
	A. Baseline, Age 3					
IQ	78.54	1.03	77.85	1.37	79.58	.46
Socioeconomic index	8.62	.17	8.65	.24	8.57	.09
Mother works	.31	*−.22*	.28	*−.22*	.35	*−.23*
Mother's age	28.66	.92	28.63	.84	28.71	1.01
Sample size	65	−7	39	−6	26	−1
	B. Age 54 Follow-Up					
Sample size, observed	50	2	30	−1	20	3
With children	41	−1	22	−2	19	1
Total children	104	6	56	−4	48	10
Child's age	28.11	.18	25.75	1.42	30.85	*−1.58*
Sample size, not observed:						
Deceased	9	−3	4	0	5	−3
Other reasons	6	−6	5	−5	1	−1
	C. Fertility, Age 54					
No children	.07	.03	.10	.03	.02	.02
Children	2.08	.04	1.87	−.07	2.40	.12
>5 children	.04	.04	.03	.04	.05	.04
Age when child born	21.80	1.25	22.82	1.98	20.63	.67
	D. Parenting					
Not out of wedlock when child born	.20	.03	.23	.07	.16	−.01
No cohabitation with new partner while child grew up	.39	.11	.41	.09	.37	.13
Fraction of years married through child's age 10	.13	*.19*	.09	*.21*	.18	.16
Read daily to child	.13	*.13*	.14	.11	.11	.14
	E. Skills and Health (Latent Variables),[a] Age 54					
Executive functioning	−.14	*.36*	−.21	*.56*	−.05	.14
Positive personality	−.20	*.42*	−.22	*.37*	−.19	*.46*
Grit	−.08	.16	−.17	.15	.01	.15
Openness to experiences	−.24	*.32*	−.23	.25	−.26	*.39*
Health	−.14	.15	−.13	.15	−.15	.16
	F. Education at Age 54					
High school graduation	.46	*.31*	.55	.15	.37	*.48*
College graduation	.20	−.15	.14	−.04	.26	−.26
	G. Average Employment and Earnings of Parents through Child's Age 10					
Fraction of years employed	.44	*.16*	.47	*.16*	.40	*.17*
Earnings (thousands of 2017 dollars)	18.39	*8.58*	22.55	*11.53*	13.57	*6.65*

TABLE 1 (*Continued*)

	Pooled		Male		Female	
	Control	Treatment-Control	Control	Treatment-Control	Control	Treatment-Control
	H. Crime, Age 54					
Days in jail	71.15	*−35.52*	119.18	*−73.83*	15.53	10.37
Misdemeanor arrests	.90	*−.60*	1.45	*−.95*	.26	−.16
Felony arrests	.80	*−.60*	1.40	*−1.00*	.11	−.11

Note.—Panel A summarizes basic variables and the sample size at baseline of the original participants. Panel B summarizes the sample size of the original participants observed and not observed in the age 54 follow-up. The first part of panel C summarizes fertility variables for all the original participants observed in the age 54 follow-up. The second part of panel C and panels D–H summarize variables at the original participant level for those observed in the age 54 follow-up who report having children, using information on up to their five eldest children. For sample size rows, column header "Control" indicates the number of observations in the control group. For outcome rows, "Control" indicates the control group mean for variables at the original-participant level (panels A, D, F, and H and first part of panel C) and the control group mean in the within original-participant average across up to their five eldest children for variables at the child-of-original-participant level (panels D and G and second part of panel C). Columns labeled "Treatment-Control" are constructed analogously to the columns labeled "Control" for treatment-control differences. We italicize "Treatment-Control" entries for outcome rows when their permutation *p*-values are less than .10. The null hypothesis for each difference is that it is less than or equal to zero for all outcomes except crime outcomes. For crime outcomes, the null hypothesis for each difference is that it is greater than or equal to zero. Table A.1 presents variable definitions and construction details. Standard errors and alternative estimates for the mean differences in this table are given in table 2.

[a] Latent variables are constructed using the method described in sec. III.D and the measures in table A.1.

B. *Age 54 Follow-Up*

Panel A of table 1 gives the sample size and baseline characteristics of the study. Original or first-generation participants were followed in multiple rounds of data collection through age 54. In this paper, we use data from the age 54 follow-up, in which information on their adult children was collected.[5] We supplement these data with earlier waves to form panel observations. Panel B of table 1 provides the sample size of original participants in the age 54 follow-up: 83% of the 123 original participants were surveyed, 12% were not surveyed because they were deceased, and 5% were not surveyed for other reasons. Combining survey questions with criminal (police and court) administrative records, we observe marriage, earnings, and criminal histories from enrollment to age 54.[6]

[5] Table A.1 provides definitions and details for the outcomes of the original participants. It shows that missing-data rates owing to item nonresponse are minimal. Our empirical strategy accounts for missing data.

[6] Criminal outcomes are self-reported in salient studies of early-childhood education (e.g., Garces, Thomas, and Currie 2002; Deming 2009). Our use of administrative data is an advantage relative to previous works. It eliminates potential biases owing to nonclassical measurement error in the reporting of sensitive outcomes (for a recent discussion, see Millimet and

The first part of panel C of table 1 summarizes fertility information of the 102 participants surveyed in the age 54 follow-up. Eighty-one of these participants report having children. The treatment-control difference in the number of children is small and statistically insignificant. This evidence rules out experimentally induced fertility as an important consideration. The program had minimal impacts on childbearing. The original participants are asked only about their first five children. This does not result in a major loss of information because only a small fraction of first-generation participants report having more than five children. Information losses due to not observing children yet to be born are also a minor issue in the age 54 follow-up, as the vast majority of the original participants are likely to have completed childbearing and adoption.

Table A.2 compares the sample of participants with children in the age 54 follow-up with the sample of participants without children. No consistent statistical differences are found between the two samples, although this comparison is not precise because the sample of those without children is very small (nine control group and 12 treatment group participants).

C. *Analysis Sample*

Our main analysis sample includes 41 first-generation control and 40 first-generation treatment participants. The sample includes original treatment and control participants who report having children in the age 54 follow-up. They have 104 and 110 children, respectively, who constitute the sample of children that we use to assess intergenerational impacts. We conduct analyses at the first-generation participant level because only first-generation participants were randomized. Accordingly, when we analyze the child sample, we need to account for its origin.[7]

Parmeter 2022). Original participants who followed up at age 54 provided their consent for us to search their criminal records. The records were collected by searching the electronic systems of the Michigan State Police Law Enforcement Information Network, county district and circuit records, the Detroit Recorder's Court, and the Federal Court in Detroit. Records of a few additional criminal incidents were obtained by searching the county social services records. Records for the subjects living out of state were requested from state criminal information offices. Information on juvenile offenses was obtained from the county's juvenile court records. Prison records were collected through the Michigan Department of Corrections. We harmonized the information across these sources to create the variables summarized in panel H of table 1. The rest of the variables that we use are based on self-reports. While less sensitive, this self-reported information could be subject to recall bias. This is a caveat of our marriage and earnings variables. Our education variables are less prone to recall issues—we use a simple question regarding the highest degree ever obtained.

[7] Our main sample of children consists of the biological children of original participants. A small number of original participants report information on adoptees and stepchildren. They report a total of 10 adopted children (four in the control group and six in the treatment group) and 17 stepchildren (seven in the control group and 10 in the treatment group). We do not include adopted children or stepchildren in the main analysis because we do not observe important information on their parental origin and age of adoption. The exclusion of adoptees and stepchildren is a minor issue. The treatment-control difference in the number of adopted

We construct intergenerational outcomes as follows. Let $\mathcal{I}$ index first-generation participants and $\mathcal{J}$ index outcomes. Define $Y_{i,j}^{c(i)}$ as the outcome $j \in \mathcal{J}$ of child $c(i)$ of first-generation participant $i \in \mathcal{I}$. The mean outcome j for the children of i is

$$\bar{Y}_{i,j}^{c} \coloneqq \frac{1}{\#\mathcal{C}_i} \sum_{c \in \mathcal{C}_i} Y_{i,j}^{c(i)}, \tag{1}$$

where $\mathcal{C}_i$ indexes the children of first-generation participant i.[8] We define $\bar{Y}_{i,j}^{c}$ as outcome j for each first-generation participant; $\bar{Y}_{i,j}^{c}$ is the outcome for the "average child" of i.

D. *Measurement Framework*

Let $Y_{i,j}^{1}$ denote outcome $Y_{i,j}$ when first-generation participant $i \in \mathcal{I}$ is assigned to treatment status ($D_i \coloneqq 1$), and let $Y_{i,j}^{0}$ denote outcome $Y_{i,j}$ when first-generation participant $i \in \mathcal{I}$ is assigned to control status ($D_i \coloneqq 0$). The observed outcome is thus $Y_{i,j} = Y_{i,j}^{1} D_i + Y_{i,j}^{0}(1 - D_i)$ (see Quandt 1958, 1972). When analyzing child outcomes, we treat average child outcomes as treatment and control outcomes for the original participants, dropping the $c(i)$ and c superscripts for notational simplicity.

For outcome $j \in \mathcal{J}$, we consider three estimators of the average treatment effect $\mathbb{E}[Y_{i,j}^{1} - Y_{i,j}^{0}]$. The first is the (unadjusted) treatment-control *mean difference*. We pool first-generation treatment and control participants and estimate the coefficients in the model

$$Y_{i,j} = \gamma_j + \delta_j D_i + \varepsilon_{i,j}, \tag{2}$$

where $\varepsilon_{i,j}$ is an error term with $\mathbb{E}[\varepsilon_{i,j} | D_i = d_i] = 0$, γ_j is an estimator of the control mean, and δ_j is the mean-difference estimator; δ_j identifies the average treatment effect assuming that treatment is randomized without compromises and that attrition is random.

A second estimator—regression-adjusted mean difference (ordinary least squares [OLS])—is used to address the randomization compromises and attrition patterns described in section III. We construct this estimator including the baseline variables in panel A of table 1 in addition to

children and stepchildren is small and statistically insignificant. The average number of adoptees and stepchildren in the control group is 0.22. The treatment-control average difference in adoptees and stepchildren is 0.09 (permutation p-value > .10). Table A.2 compares the sample of original participants who report having adopted children or stepchildren with those who do not. Table A.7 compares the outcomes of children analyzed in the main paper with the outcomes of adoptees and stepchildren and finds slight differences. Table A.9 presents our main intergenerational estimates including adoptees and stepchildren. Estimates barely change when these children are added into the sample.

[8] "#" denotes cardinality.

first-generation participant sex as covariates in equation (2). We know the qualitative features of the randomization failure but not details about individual participants. We know that baseline variables are only partially balanced across the treatment and control groups. OLS identifies the average treatment effect under the assumption of conditional random assignment to treatment and attrition conditional on baseline covariates.

Our third estimator is a more general mean-difference adjustment used in Heckman and Karapakula (2021). It is an augmented inverse-probability weighting estimator (AIPW) adapted to the sampling protocol of PPP. It weights equation (2) by the inverse probability of being treated and having attrited (recall that the reasons for attrition are being dead, not being interviewed in the age 54 follow-up or not having children). AIPW imputes (missing) counterfactual outcomes for each first-generation participant based on the same baseline variables used as covariates when computing the OLS estimates. AIPW is useful for its double-robustness property. It provides a consistent estimator of the average treatment effect if either the weighting scheme or the (imputed) equation (2) is correctly specified.[9]

We also present Lee (2009) bounds for average treatment effects, a supplemental set of results accounting for compromises in randomization. This method is appropriate for contexts with (conditional) randomized assignment to treatment and sample selection generated by attrition. We refer readers to the source paper for details. Lee's bounds require two assumptions: (i) there is (conditional) randomized assignment to treatment and (ii) treatment affects attrition uniformly across the sample (i.e., the probability of being attrited should either increase or decrease as a function of treatment status for all individuals). The first assumption is plausible in our context, as we condition on variables unbalanced owing to the compromises in the randomization protocol. The evidence in panel B of table 1 is consistent with the second assumption. First-generation treatment group participants were more likely to be followed up with at age 54 (either because of death or for other reasons).

Creating factors.—We observe sets of survey items designed to measure different skills. For each set of skill measurements, we use factor analysis to create a one-dimensional, interpretable aggregate of each set of items. Factor analysis summarizes the covariability among the observed items (for a review, see Borghans et al. 2008). It reduces the dimension of the data by creating one latent factor variable from multiple items in each category. It accounts for measurement error. Table A.1 lists the items that we use for estimation. Following standard practice (Gorsuch 1983; Thompson 2004),

[9] The identification proofs for the three estimators that we use are standard, and we omit them for brevity. Heckman and Karapakula (2021) and the appendix of García et al. (2021) provide detailed proofs.

we assume that each item is associated with at most one skill.[10] For estimation, we assume that the measurement system is the same across treatments and controls. This hypothesis is tested and not rejected in Heckman, Pinto, and Savelyev (2013). We standardize each latent factor variable to have an in-sample mean zero and standard deviation one. We use the same procedure to produce a health latent variable.

Inference.—Outcomes are reported so that a positive point estimate indicates a beneficial treatment effect. Crime outcomes of the original participants are an exception. For "positive outcomes," we test whether the treatment effect is less than or equal to zero, outcome by outcome. For the crime outcomes of the original participants, we test whether the treatment effect is greater than or equal to zero. We report one-sided tests because most outcomes of Perry are beneficial in this and other studies. For our baseline AIPW estimator, we present several p-values. Our baseline p-value is permutation based because it is especially suited for analyzing small samples such as ours. We also present bootstrap standard errors for all of the estimators considered. All of our inference is clustered at the first-generation participant level.[11]

IV. Impact on the Original Participants

PPP had an impact on the socioemotional skills of the original treatment participants. Heckman, Pinto, and Savelyev (2013) document that this impact translated into improvements in labor market, crime, and health outcomes through age 40.[12] In this section, we update that analysis. This motivates the source of the intergenerational externalities that we report below.[13] We show that the impact on the skills, marriage, earnings, crime, and health outcomes of the original participants persists through their child-rearing years up to their late midlife years. Their improved outcomes produce better home environments for their children.

[10] This is called a "dedicated factor model."

[11] We follow standard procedures when computing standard errors and p-values. Our standard errors are the standard deviation of the empirical bootstrap distribution of each estimator. Analytic p-values are asymptotic and robust to heteroskedasticity and arbitrary correlation within first-generation participants (e.g., Liang and Zeger 1986). They do not account for sampling variation in preliminary estimation stages (e.g., construction of weights in the AIPW estimator). Permutation p-values are calculated as in Lehmann and Romano (2006, 831–37). They are especially suited for settings with small sample size. All bootstrap p-values are calculated as in Hansen (2021, 262–305). They account for sampling variation in all estimation stages. This accounting introduces minor additional variance.

[12] This impact is documented by Heckman et al. (2010a), Heckman, Pinto, and Savelyev (2013), Conti, Heckman, and Pinto (2016), and Heckman and Karapakula (2021).

[13] The evidence discussed in this section is based on first-generation participants who have children. We present evidence for the full sample of original participants in figs. A.2 and A.3*b*. The results for the original participants presented in this section barely change when we add original participants without children into the analysis sample.

We start by describing the treatment-control mean differences for the original participants in table 1 and figure 1. We then discuss robustness of our estimates to multiple estimation strategies and inferential procedures. Panel E of table 1 summarizes newly collected data at age 54 on skills of the original participants. It shows that PPP has a long-lasting impact on both cognitive and socioemotional skills. PPP increases the skills of the pooled group of male and female original participants by 0.2–0.4 standard deviations. Male original participants drive this impact. However, the average treatment-control differences are positive for all skills for both males and females.

Ours is the first paper to document the impact of high-quality early education on skills at late midlife. The long-lasting impact on executive functioning challenges the often-repeated claim of "fade-out" in the treatment effects on skills, specifically on cognition. Previous research claims that the impact of early-childhood education on cognitive-test scores disappears (fades out) shortly after the end points of interventions (Protzko 2015; Hojman 2016; Bailey et al. 2020). Some authors argue that the fade-out in cognition (and also socioemotional skills) is real and not simply a measurement artifact (Bailey et al. 2017, 2020). These studies are all based on short-run follow-ups. Our estimates dispute this claim. Our measure of executive functioning is based on well-established tests that measure cognition (Raven and Stroop tests).

Panel E of table 1 also describes another relevant life-cycle outcome: health. We summarize health using a latent factor (see sec. III.D). Examples of items underlying this factor include waist-to-hip ratio, high total cholesterol, and chronic severe pain (for a complete list of items, see table A.1). These items are part of the newly collected data at age 54. The mean difference of the latent variable is not precisely estimated. However, figure A.3*a* shows estimates of the treatment and control distributions of the health latent variable. Treatment shifts the distribution rightward. Treatment group participants are 15 percentage points more likely to be healthier than 80% of individuals in the pooled treatment and control sample (permutation p-value = .06). Table A.3 presents treatment effects on the individual items forming the health latent variable. Other studies document impacts on adult health of early-childhood education up to age 30 or 40 (Campbell et al. 2014; Conti, Heckman, and Pinto 2016). Our analysis confirms that health impacts persist up to age 54. This finding is new. Positive forecasts of the long-term health impact of early-childhood education are thus justified.[14]

Panels *A*–*C* of figure 1 display the average evolution of life-cycle outcomes of the original participants who reported having children in the

[14] Examples of these forecasts are in García and Heckman (2020) and García et al. (2020, 2021).

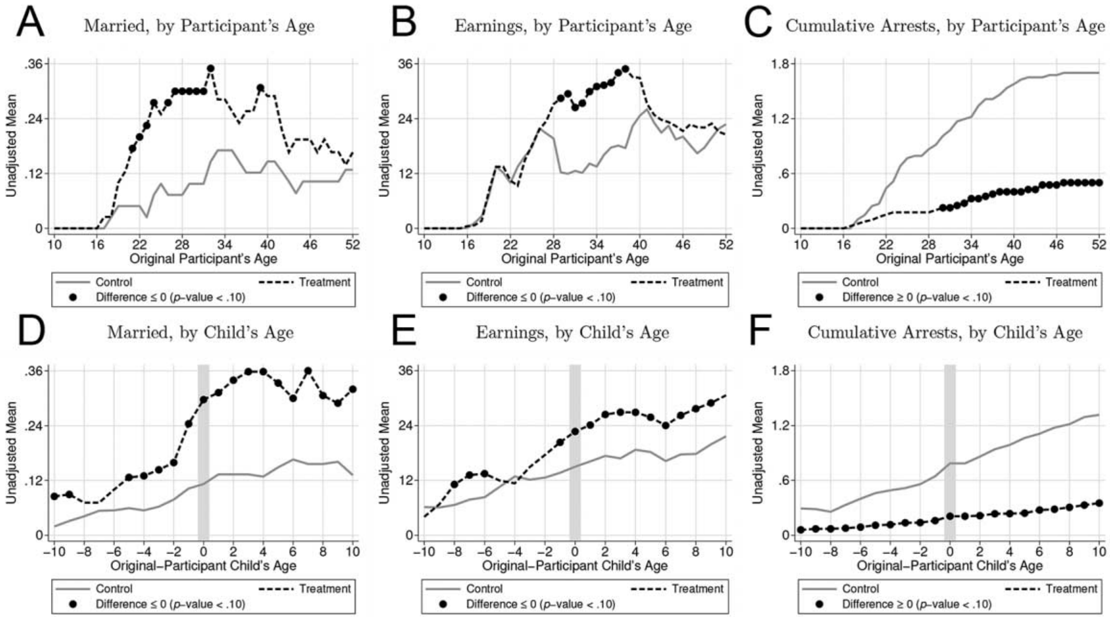

FIG. 1.—Original-participant marriage, earnings, and crime by their age and by their children's age. *A*, Control group and treatment group unadjusted means of a married-status indicator by age of the original participants who reported having children. We mark the treatment group mean when the unadjusted treatment-control mean difference has a permutation p-value less than .10. The null hypothesis for the difference is that it is less than or equal to zero. *B*, Analogous in format to *A* but for annual earnings in thousands of 2017 US dollars. *C*, Analogous in format to *A* but for cumulative violent misdemeanor and felony arrests. *C*, Null hypothesis for the difference is that it is greater than or equal to zero. *D–F*, Analogous in format to *A–C* but they are plotted by age of the children of original participants. For *D–F*, the outcomes are first averaged within original participants across up to five eldest children before constructing control and treatment means.

age 54 follow-up. The improvements in marriage stability, earnings, and criminal behavior of the original treatment participants are substantial. At age 30, they are more than 10 percentage points more likely to be married, have $10,000 higher average annual earnings, and accumulate approximately one fewer average arrest.[15] Panels *D–F* show the means of the same variables by child treatment status and age.[16] The improvements of the original treatment participants imply that their children are more than 15 percentage points more likely to be born to married parents than children of control group participants. They are also born to parents who make on average almost $10,000 per year more and have a lower average of cumulative arrests. The advantage of children of the original treatment participants builds up years before they are born and persists throughout their childhoods.

Panels D–H of table 1 reinforce the evidence in figure 1. They show that, on average, children of the original treatment group were more likely to grow up in two-parent stable environments compared with children of the original controls. They were read to more often while growing up.[17] Their parents had greater skills, were employed a larger fraction of time, had more education and earnings, and engaged less in criminal behavior when they were growing up.

Table 2 reports robustness checks for the estimates presented in table 1. Columns 1 and 2 replicate the mean-difference estimates in table 1 for reference. It provides the corresponding bootstrapped standard errors. We then show the OLS estimates of the treatment effect and the Lee (2009) bounds. The OLS estimates align with the mean differences. The bounds are tight. Panel B of table 2 presents AIPW estimates, which most comprehensively address the randomization compromises described in section III.A. We also provide alternative p-values. The several checks verify that the treatment-control differences are robust to different estimation strategies and remain statistically significant under several inferential procedures. Table 2 also provides estimates and inference for two summary measures of the outcomes in figure 1—fraction of years married and earnings between ages 21 and 40, which are the ages in which the original participants do most of their parenting. The results confirm substantial and significant treatment-control differences. Treatment increased the parenting

[15] The life-cycle profile of marriage stability during the child-rearing years in fig. 1*D* summarizes various relationship aspects described in panel D of table 1, for which we also present treatment effects in table 2: not having children out of wedlock, not cohabiting with new partners while children grow up, and fraction of years married while children grow up.

[16] Panels *D–F* of fig. 1 are based on the average child. Figure A.1 is analogous in format to fig. 1*D*–1*F* but based only on the first child of original participants. Figure 1*D*–1*F* and figure A.1 display very similar patterns.

[17] This finding is consistent with Bauer and Schanzenbach (2016), who find that participants of Head Start improve their parenting skills when becoming adults. Those authors do not analyze intergenerational outcomes.

and economic resources that the original participants provided to their children.[18]

V. Intergenerational Outcomes

We now turn to the outcomes of the children of the original participants. We analyze the eight outcomes displayed in figure 2.[19] These are outcomes for the average child, constructed using the formula in equation (1). We analyze children of all ages when examining school suspension, special education, arrests, and health. We consider only children aged 19 or older when analyzing teenage parenthood. For years of education, employment, and divorce, we consider children aged 23 and older. The children of the original participants are on average 28 years old when information on them is reported at the age 54 follow-up. Most of them satisfy all of the age cutoffs imposed. All first-generation participants who report having children have at least one child satisfying all of the age cutoffs except for two.[20]

[18] Likely, the impact of the program on the human capital of the original treatments leads them to improve their marital prospects—human capital assortative mating is a well-documented phenomenon. See Eika, Mogstad, and Zafar (2019) for recent documentation, although different measures of assortative mating show different patterns (e.g., Chiappori, Costa Dias, and Meghir 2020; Gihleb and Lang 2020). Thus, children of the original treatments are likely to have grown up with two parents with higher human capital than the children of the original controls. The greater marriage stability of treated participants is indirect evidence of the improved human capital of the couple. We do not have the appropriate information to test this directly. In the age 54 follow-up, we observe only the employment status and education of the current partners of the original participants (21 partners of original treatments and 19 partners of original controls). Multiple issues arise when analyzing these outcomes (e.g., selection into cohabitation and marriage, uncertainty about whether partners in the age 54 follow-up are parents of the children of the original participants, and small sample size). Estimates in these selected samples are unreliable. We speculate that improvements in the skills of the original treatments improve their parenting and the parenting of their partners through assortative mating. These are joint mechanisms explaining the intergenerational treatment effects. We presume that these mechanisms persist intergenerationally. Children of the original treatments have greater human capital and thus better marital prospects than children of the controls. This is reflected in their greater marriage stability during their midtwenties.

[19] We construct these outcomes using the survey questions in table A.4. This table presents the nine questions that were asked to the original participants about their children.

[20] This section provides a basic description of the data analyzed. Table A.5 provides additional details on variable definitions and observations. Table A.6 displays the sample sizes of first- and second-generation participants after imposing each age cutoff. Imposing age cutoffs has minimal consequences for the sample sizes. When analyzing each outcome, we lose a couple of observations per outcome due to item nonresponse in the interviews. Item nonresponse is very minor. For two outcomes, we do not have item nonresponse cases (never arrested and in good health), for two outcomes we have one case (never suspended from school and never a teen parent), for one outcome we have two cases (years of education), for two outcomes we have three cases (never in special education and never divorced), and for one outcome we have four cases (employed). Our estimators account for item nonresponse as yet another source of attrition.

TABLE 2
Robustness of Estimated Treatment Effects and Standard Errors for the Original Participants

	(1)	(2)	(3)	(4)	(5)	(6)
	A. Basic Estimates					
	Mean Difference		Adjusted Mean Difference (OLS)		Lee (2009) Bounds	
	Estimate	Standard Error	Estimate	Standard Error	Lower	Upper
Age when child born	1.245	1.150	1.332	1.282	.959	1.423
Not out of wedlock when child born	.030	.094	−.040	.100	.013	.035
No cohabitation with new partner while child grew up	.110	.110	.061	.115	.099	.121
Fraction of years married through child's age 10	.188	.077	.159	.086	.174	.195
Fraction of years married, ages 21–40	.156	.069	.156	.080	.135	.164
Read daily to child	.125	.086	.107	.089	.107	.131
Executive functioning	.356	.193	.348	.197	.320	.401
Positive personality	.418	.174	.369	.203	.418	.471
Grit	.158	.197	.165	.241	.113	.158
Openness to experiences	.321	.194	.284	.183	.271	.351
Health	.153	.219	.153	.274	.133	.153
High school graduation	.312	.103	.334	.110	.307	.329
College graduation	−.145	.071	−.146	.075	−.166	−.144
Fraction of years employed through child's age 10	.162	.074	.156	.079	.154	.174
Average earnings (thousands of 2017 dollars) through child's age 10	8.584	4.615	8.170	4.798	7.018	9.145
Average earnings (thousands of 2017 dollars), ages 21–40	8.624	4.031	8.241	4.417	6.464	9.327
Days in jail	−35.521	27.267	−31.652	27.360	−42.790	−34.735
Misdemeanor arrests	−.602	.272	−.735	.330	−.640	−.596
Felony arrests	−.599	.281	−.601	.275	−.599	−.594

	B. Estimates from Preferred Estimator: AIPW					
			p-Values			
					Bootstrap	
	Estimate	Standard Error	Analytic	Permutation	Simple	Studentized
Age when child born	1.482	1.335	.103	.096	.101	.112
Not out of wedlock when child born	−.066	.107	.793	.739	.684	.853
No cohabitation with new partner while child grew up	.049	.129	.309	.359	.307	.331
Fraction of years married through child's age 10	.141	.091	.023	.052	.054	.056
Fraction of years married, ages 21–40	.138	.092	.016	.044	.048	.053
Read daily to child	.089	.091	.114	.197	.152	.116
Executive functioning	.354	.215	.031	.043	.030	.029
Positive personality	.461	.207	.003	.010	.032	.010
Grit	.096	.262	.325	.330	.362	.325
Openness to experiences	.204	.215	.100	.164	.143	.135
Health	.121	.312	.291	.299	.330	.337
High school graduation	.333	.125	.000	.004	.008	.008
College graduation	−.130	.091	.970	.942	.953	.970
Fraction of years employed through child's age 10	.145	.089	.024	.039	.076	.022
Average earnings (thousands of 2017 dollars) through child's age 10	8.879	5.561	.024	.041	.078	.027
Average earnings (thousands of 2017 dollars), ages 21–40	8.462	5.258	.019	.020	.081	.030
Days in jail	−33.219	28.841	.083	.120	.154	.035
Misdemeanor arrests	−.803	.340	.002	.006	.017	.003
Felony arrests	−.611	.314	.008	.005	.032	.005

Note.—Panel A presents treatment effect estimates and standard errors of the average treatment effect for the first-generation participant outcomes summarized in table 1 using the mean difference and the adjusted mean difference OLS estimators explained in sec. III.D. It also presents the Lee (2009) bounds. We include treatment effect estimates for a summary variable for the marriage and earnings longitudinal outcomes (the average between ages 21 and 40). Panel B presents treatment effect estimates, standard errors, and *p*-values based on our preferred estimator (AIPW) for the outcomes in panel A. The AIPW estimator and *p*-values are explained in sec. III.D. The standard errors are bootstrapped and clustered at the first-generation participant level. The null hypothesis for each treatment effect is that it is less than or equal to zero for all outcomes except for the crime outcomes. For the crime outcomes, the null hypothesis for each treatment effect is that it is greater than or equal to zero.

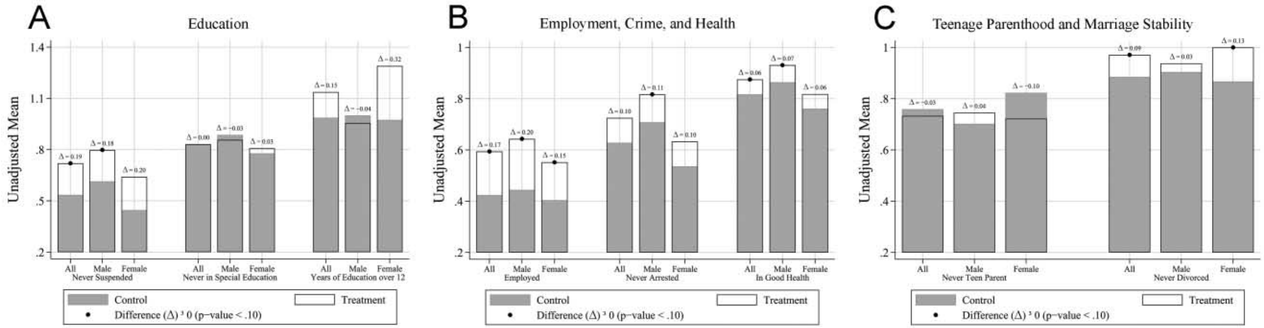

FIG. 2.—Outcomes of the second generation (children) by sex of the first generation (original-participant parents). *A*, Unadjusted mean for never suspended, never in special education, and years of education for the children of the original participants (intergenerational outcomes). To simplify this panel, we display years of education as the number of years after the 12th year of education. The unadjusted means are displayed by treatment status of the original participants. On top of each bar, we display the corresponding treatment-control unadjusted mean difference (Δ). We mark the difference when its corresponding treatment effect estimate using our preferred AIPW estimator (explained in sec. V.A) has a permutation *p*-value less than .10. The null hypothesis for the treatment effect is that it is less than or equal to zero. The outcomes are defined as within original-participant averages across up to five eldest children, as explained in section III.C. The unadjusted means by treatment status are then calculated. *B*, *C*, Analogous in format to *A* but for the outcomes labeled.

A. Adjusted Mean Differences

Figure 2 displays unadjusted treatment-control mean differences. They are sizable. The children of the original treatment participants are more likely to never have been suspended from school, be employed, never have been arrested, be in good health, and never have been divorced. They also accumulate more years of education. We adjust these differences and provide inference in column 1 of table 3. Given the robustness of the estimates in section IV, we focus our discussion on results based on AIPW and permutation-based inference, which most comprehensively address randomization compromises and small sample size.[21] AIPW also accounts for factors preventing us from observing second-generation outcomes. These factors include death and any other reason for not observing first-generation participants in the age 54 follow-up. They also include not observing second-generation outcomes for first-generation participants who do not have children.

PPP has a beneficial intergenerational impact that is consistent with its impact on the first generation. High-quality early-childhood education programs such as PPP improve the early-life socioemotional skills of children. This translates into long-term impacts on labor market, crime, and health outcomes (for a survey, see Elango et al. 2016). School suspension is an indirect measure of early-life socioemotional skills, and PPP has a sizable impact on them for the second generation. The impact on health through young adulthood and longer-term outcomes such as employment and relationship stability (never have been divorced) are also sizable. For crime, the impact is much stronger for men, and we discuss this in section V.B.

We examine "employment" to further interpret our estimates. We compare the second-generation impacts with the first-generation impacts of PPP and Head Start (a Federal early-childhood education program targeted toward disadvantaged families like PPP and founded in its wake).[22] We estimate that PPP increases the second-generation probability of

[21] Table A.8 shows that the results in table 3 are robust to using OLS, Lee (2009) bounds, and alternative inferential procedures. When using the OLS and AIPW estimators, we residualize the child outcomes $Y_{i,j}^{c(i)}$ from age, age squared, and sex to account for age variability and sex of the children at the time of the age 54 follow-up. We residualize before computing the outcome for the average child using the formula in eq. (1). We residualize all outcomes and impose the age cutoffs discussed in the main text of this section. Table A.10 shows that our intergenerational estimates remain virtually unchanged when not using cutoffs or residualization.

[22] Head Start is of relatively high quality but varies in the effectiveness of the services offered across the United States. Walters (2015) finds that variation in these services or inputs largely explains differences in Head Start's short-term effects. The inputs include center-based care, home visiting, the HighScope curriculum modeled after PPP, and class size. Walters (2015) documents that Head Start centers that combine center-based care and home visiting (like PPP) are the most effective; he does not investigate long-term effects.

TABLE 3
INTERGENERATIONAL TREATMENT EFFECTS BY SEX OF THE ORIGINAL PARTICIPANTS AND THEIR CHILDREN

	First Generation								
	Pooled			Male			Female		
	Second Generation								
	Pooled (1)	Male (2)	Female (3)	Pooled (4)	Male (5)	Female (6)	Pooled (7)	Male (8)	Female (9)
Never suspended from school	*.169*	*.196*	.119	*.224*	.114	.163	.092	*.311*	.057
	[.029]	[.064]	[.170]	[.041]	[.276]	[.125]	[.251]	[.032]	[.398]
Never in special education	−.051	.050	−.100	−.068	−.093	−.058	−.028	*.251*	−.160
	[.758]	[.347]	[.880]	[.721]	[.731]	[.646]	[.621]	[.075]	[.988]
Years of education	.084	*.704*	.093	−.165	.331	−.131	.436	*1.231*	.410
	[.409]	[.045]	[.423]	[.619]	[.250]	[.534]	[.190]	[.029]	[.270]
Employed	*.258*	*.228*	*.217*	*.299*	.280	*.365*	*.200*	.155	.007
	[.006]	[.072]	[.056]	[.022]	[.113]	[.033]	[.079]	[.241]	[.482]
Never arrested	.088	.082	−.015	*.176*	.214	−.005	−.036	−.104	−.029
	[.129]	[.240]	[.560]	[.032]	[.114]	[.519]	[.607]	[.739]	[.576]
In good health	*.092*	*.170*	*.135*	*.102*	*.202*	.088	.077	.124	*.203*
	[.090]	[.054]	[.053]	[.047]	[.038]	[.156]	[.264]	[.239]	[.079]
Never teen parent	−.061	−.045	−.083	−.026	−.060	−.066	−.111	−.025	−.107
	[.758]	[.653]	[.758]	[.570]	[.618]	[.640]	[.836]	[.586]	[.758]
Never divorced	*.076*	*.088*	.044	.042	*.074*	−.037	*.123*	.108	*.159*
	[.074]	[.051]	[.275]	[.325]	[.012]	[.639]	[.016]	[.165]	[.035]

NOTE.—This table presents treatment effect estimates for our eight second-generation outcomes by sex of the first-generation participant (parent) and second-generation participant (child) using our AIPW estimator. We present each estimate's permutation p-value in brackets. We italicize the treatment effect estimates when their permutation p-values are less than .10. The null hypothesis for each treatment effect is that it is less than or equal to zero.

employment by 25.8 (standard error = 11.5) percentage points. Table 3 shows that the treatment effect estimate is similar for second-generation male and female participants. Heckman and Karapakula (2021) report an age 40 first-generation impact of PPP on employment of 26.6 percentage points for men ($p = .02$) and −1.6 percentage points for women ($p = .50$). Our results indicate that the first-generation impacts of PPP spill over into the second generation. The first-generation impact spills over to male and female second-generation participants. The intergenerational impact of PPP on employment is also larger than the first-generation impact of Head Start on the probability of not being idle during young adulthood—7.1 (standard error = 3.8) percentage points (Deming 2009). PPP has a larger second-generation impact than the first-generation impact of Head Start.

Column 1 of table 3 shows that while not all estimates are statistically significant at the 10% level, there is a general pattern of positive treatment effects. At the 10% level, we detect significant treatment effects on four out of the eight outcomes we study when analyzing the pooled sample of male and female second-generation participants. Table A.8 shows that this conclusion holds when using alternative inferential procedures. For two reasons, these results are unlikely to be a consequence of cherry picking. First, we analyze *all* of the second-generation outcomes observed. If all eight treatment effects were zero, we would reject the null hypothesis of no treatment effect for 10% of outcomes by chance using a 10% significance level. The F-statistic for the joint null hypothesis of no treatment effect for the eight outcomes is 2.21 ($p = .04$). Second, our outcomes are interpretable categories of independent interest. Correcting p-values for multiple-hypothesis testing for such diverse categories of treatment effects would lump together very different outcomes and lack any behavioral justification.

B. *Gender Differences*

Impacts of early-childhood education on long-term outcomes are usually found to be greater for boys than for girls (Elango et al. 2016). Educational outcomes are the exception to this rule.[23] The first-generation impact of PPP is consistent with these findings. Table 3 shows a greater intergenerational impact on second-generation male children than on second-generation female children. For instance, we reject the null hypothesis that the treatment effect is less than or equal to zero using a significance level of 10% for

[23] See Elango et al. (2016) for a documentation of gender differences in the impact of several early-childhood education programs. Explanations for the gendered impacts include the following: Baker, Gruber, and Milligan (2008, 2015) establish a harmful impact of lower-quality universal childcare. Kottelenberg and Lehrer (2014) localize this negative impact as arising only from boys. Their results indicate that boys are less resilient than girls and that putting them in lower-quality environments instead of keeping them at home hurts them; they are consistent with literature supporting greater vulnerability of boys to adverse environments. Golding and Fitzgerald (2017) and Schore (2017) discuss the potential reasons for this greater vulnerability. They are also consistent with literature documenting that boys develop later than girls and thus benefit from an enriched environment (Lavigueur, Tremblay, and Saucier 1995; Masse and Tremblay 1997; Nagin and Tremblay 2001; Bertrand and Pan 2013). Autor et al. (2019) show that boys are more affected than girls by household economic shocks. Supplementing boys' environment with high-quality early-childhood education is thus more beneficial for them than it is for girls. García et al. (2018, 2019) is an exception in that they find that high-quality early-childhood education favors girls more than boys. The authors document that in their context there is more scope of improvement in households of girls relative to boys, and thus there is a greater benefit for girls. The greater scope of improvement for girls relative to boys results from fathers being more likely to stay together with mothers and provide for their children when a boy (rather than a girl) is born (Dahl and Moretti 2008).

five out of the eight outcomes. For second-generation female participants, we reject the null for only two out of eight outcomes.

For crime, we find a substantial positive impact on never being arrested for children of first-generation male participants. The impact on second-generation male children drives this result. PPP reduced criminal activity of first-generation male participants. The intergenerational impact on their sons is consistent with recent studies in economics and sociology finding that parental incarceration (most incarcerated individuals are men) leads to a significant intergenerational increase in behavior issues and teen crime (Haskins 2014; Murray et al. 2014; Turney and Haskins 2014; Dobbie et al. 2018). These results are primarily for disadvantaged individuals, making the comparison to our study relevant. The second-generation crime impact is also consistent with studies in other fields documenting that early-life environments determine young adult criminal activity (Henry et al. 1999; Wright et al. 1999; Piquero and Moffitt 2005; Belsky et al. 2020). Section V.C further discusses the intergenerational transmission of criminal outcomes in the sample.

C. *Contextualization of Intergenerational Treatment Effects*

Though scarce, the literature on the intergenerational impact of high-quality preschool summarized in panel A of table 4 allows us to contextualize our estimates. Rossin-Slater and Wüst (2020) study the intergenerational impact of high-quality preschool in Denmark. They exploit availability for children born between 1935 and 1957 to women who were born between 1955 and 1987. They argue that most beneficiaries were disadvantaged. They estimate the intent-to-treat or reduced-form effect of preschool availability in the municipality where the mothers resided when they were between 3 and 7 years old. They find an intergenerational impact on years of education of 0.06 ($p < .01$), from a control group mean of 12.13. Our estimate for the full sample is 0.08 ($p = .41$), from a control group mean of 12.99. While their estimate is an intent-to-treat effect and our estimate is an average treatment effect, the alignment may be due to large take-up among disadvantaged populations. We find a large impact of 0.70 ($p = .05$) for male children and a smaller and insignificant impact of 0.093 ($p = .42$) for female children. Rossin-Slater and Wüst (2020) do not report results by gender.

Barr and Gibbs (2022) estimate the intergenerational impact of Head Start, using an approach similar to that of Rossin-Slater and Wüst (2020). They exploit availability of Head Start in a mother's county of birth when she was 4 or 5 years old to estimate an intent-to-treat or reduced-form effect. Mothers were born between 1960 and 1964. Barr and Gibbs (2022) provide estimates for disadvantaged mothers, whose take-up may also have been high. They find a negative impact of -0.23 ($p < .01$) on ever being arrested, convicted, or put on probation at age 20 or older for male children, from a mean of 0.40. For female children, the impact is small and insignificant. Our estimates are also driven by male children. They are smaller in magnitude than theirs. For male children, we find an impact of −0.08 (p = .24) from a mean of 0.59. Crime is an activity performed mainly by men. When limiting the sample to male children of original male participants, our estimate is −0.21 (p = .11). Our estimates thus quantitatively and qualitatively align with those of the other studies in the literature aiming to identify causal intergenerational impacts of high-quality preschool. Our study solidifies previous evidence given the advantages already discussed.

A major result in this and previous studies of PPP is its effectiveness in reducing the criminal activity of its original participants. We investigate the intergenerational relationship of this outcome in panel B of table 4. We report the slope estimate of a regression of an indicator of a child ever being arrested up to their midtwenties (analyzed in this section) on an indicator of parental arrest. We provide estimates by treatment status and limit the sample to male children of the original male participants, given that crime is primarily a male activity.[24] The estimate of 0.32 for the control group indicates an expected positive correlation. This estimate is larger than the closest estimate in the literature of 0.04 reported by Dobbie et al. (2018). Our larger correlation in the control group is sensible given that these authors analyze a US representative sample pooling advantaged and disadvantaged parent-child pairs and including males and females, whereas we use a sample selected to be disadvantaged. The estimate for the treatment group is −0.06.[25] Although the correlations do not have a causal interpretation, their difference across experimental groups suggests that PPP is effective at breaking the intergenerational transmission of criminal activity.

The control group estimate in panel B of table 4 provides an estimate of the male-male intergenerational transmission of the probability of being arrested for disadvantaged individuals. This relationship can be estimated in nonexperimental settings.[26] Treatment decreases the probability of

[24] For this exercise, we use an indicator of whether the parent has any arrest up to age 22. This increases the applicability of the predictions described below, which may be applied in samples with follow-ups before midlife. The age 22 indicator and an indicator of any misdemeanor or felony arrests based on the variables of panel H in table 1 have a correlation of 0.67. For original male participants with male children, treatment decreases the probability of any arrest up to age 22 by 0.24 (p = .09) from a control group rate of 0.40.

[25] We reject the null hypothesis that the treatment-control difference in the intergenerational relationship of −0.37 is greater than or equal to zero. The permutation p-value of the test is .03.

[26] García and Heckman (2023) report intergenerational relationships for other outcomes. Their male-male estimates are 0.77 (p = .00) for years of education and 0.23 (p = .00) for being in good health. Their corresponding female-female estimates are 0.63 (p = .01) and −0.13 (p = .67). Outcomes of the original participants and their children are defined as in table A.5.

TABLE 4
Comparison with Other Causal Intergenerational Estimates and Intergenerational Relationship Estimates

	(1)	(2)	(3)	(4)
	A. Intergenerational Treatment Effects of Preschool			
	Years of Education, Age 25 (Rossin-Slater and Wüst 2020)	Years of Education, Midtwenties (This Study)	Ever Arrested, Convicted, or Put on Probation, 20 or Older (Barr and Gibbs 2022)	Ever Arrested, Midtwenties (This Study)
Pooled	.06	.08	−.13	−.09
	{12.13}	{12.99}	{.28}	{.37}
Males	NA	.70	−.23	−.08
		{12.08}	{.40}	{.59}
Females	NA	.09	−.03	.02
		{13.52}	{.17}	{.21}
Parameter	Intent to treat	Average treatment effect	Intent to treat	Average treatment effect
Relevant population	Children of disadvantaged Danish women—women were born between 1935 and 1957; children were born between 1955 and 1987	Children of disadvantaged African American individuals born in Ypsilanti, Michigan—one parent of each child participated in PPP and was born between 1957 and 1962; children were born between 1973 and 1992	Children of mothers whose mother had less than high school education—mothers were born between 1960 and 1964	Children of disadvantaged African American individuals born in Ypsilanti, Michigan—one parent of each child participated in PPP and was born between 1957 and 1962; children were born between 1973 and 2008
Empirical design	High-quality preschool available in municipality where mother resided when she was between 3 and 7 years old	One parent randomly assigned to high-quality preschool (PPP)	Head Start available in mother's birth county when she was 4 or 5 years old	One parent randomly assigned to high-quality preschool (PPP)

	B. Intergenerational Relationships of Crime Outcomes: Child Criminal Outcomes Correlated with Parental Criminal Outcomes			
	Correlation (Dobbie et al. 2018)	Causal (Dobbie et al. 2018)	Correlation (This Study, Control Group)	Correlation (This Study, Treatment Group)
	.04	.184	.32	−.06
	{.09}	{.237}	{.57}	{.40}
Child population	Children born between 1980 and 1984 in the United States (National Longitudinal Survey of the Young 1997)	Socioeconomically disadvantaged children residing in Sweden whose parents were involved in a criminal trial between 1997 and 2004	Male children of disadvantaged African American males born in Ypsilanti, Michigan—male parents were control participants of PPP and were born between 1957 and 1962; male children were born between 1973 and 2008	Male children of disadvantaged African American males born in Ypsilanti, Michigan—male parents were treatment participants of PPP and were born between 1957 and 1962; male children were born between 1973 and 2008
Estimand	Coefficient from a regression of an indicator of criminal conviction of child between ages 15 and 17 on an indicator of parental incarceration before child turns 16; regression is estimated in sample of parent-child pairs	Same as column 1 except that parental incarceration is instrumented using a judge-leniency instrument	Coefficient from a regression of an indicator of child ever being arrested up to the midtwenties on an indicator of whether parent was arrested up to age 22; regression is estimated in sample of pairs of fathers and the average of male children	Coefficient from a regression of an indicator of child ever being arrested up to the midtwenties on an indicator of whether parent was arrested up to age 22; regression is estimated in sample of pairs of fathers and the average of male children

Note.—Column 1 of panel A displays the intergenerational treatment effect of high-quality preschool for children of disadvantaged women in Denmark on years of education at age 25, taken from col. 1 of table A6 in Rossin-Slater and Wüst (2020). Column 2 displays the closest estimates from our study, taken from table 3. Column 3 displays the intergenerational treatment effect of Head Start for children of disadvantaged women in the United States on an indicator of ever arrested, convicted, or put on probation at age 20 or older, taken from tables 2 (pooled) and 3 (males and females) of Barr and Gibbs (2022). Column 4 displays the closest estimates from our study, taken from table 3. Column 1 of panel B displays OLS estimates from a regression of an indicator of children's criminal outcomes (conviction between ages 15 and 17) on their parents' criminal outcomes (incarceration before own child turns 16). Column 2 displays estimates of the same regressions as col. 1, based on Swedish administrative records and instrumenting parental incarceration using a judge-leniency instrument. Column 3 displays the closest estimate to col. 1 based on the data used throughout this paper. We limit the sample to the control original male participants with male children. Column 4 is analogous in format to col. 3 when limiting the sample to the original treatment participants. In cols. 1, 2, and 4 of panel A, we display the mean of the control group in curly brackets. In col. 3, we display the full-sample mean in curly brackets. In panel B, we display the mean of the children's dependent variable for children whose parents had any incarceration or arrest in curly brackets. In col. 2, the mean reported is for the full sample, as opposed to the sample of the socioeconomically disadvantaged. NA = not applicable.

being arrested for treatment parents by 24 percentage points. If the relationship estimated in the control group were causal, the predicted intergenerational impact of treatment on the probability of being arrested would be $-0.24 \times 0.32 = -0.08$. We also consider an alternative calculation based on a causal estimate of the intergenerational transmission for disadvantaged individuals reported in Dobbie et al. (2018) and described in panel B of table 4.[27] This would yield a predicted intergenerational impact of $-0.24 \times 0.18 = -0.04$. Either prediction provides a lower bound for the actual estimate of the intergenerational impact of PPP (–0.21; see table 3). Our intergenerational estimate is larger than the predictions likely because, as documented in section IV, treatment not only decreases the probability of being arrested for the original treatments but also improves their skills, labor market prospects, and marriage stability. The impacts on these multiple mediators suggest a greater intergenerational impact than that predicted by the one-to-one intergenerational transmission of the probability of being arrested.

VI. Summary

The HighScope PPP was a pioneering early-childhood education program designed to promote the social mobility of disadvantaged African American children. The foundational principles of the program guide current practice and are incorporated in at least 30% of current Head Start programs (Elango et al. 2016). Using newly collected data, we examine its impact on the original participants through age 54 and on their adult children. We find substantial and lasting positive effects for the original treatment group participants on cognition and beneficial personality traits, contradicting claims on fade-out that are based on relatively short-term follow-ups. We also document long-lasting impacts on health using a rich set of measures that include overall health and cardiovascular indicators. The first-generation treatment group participants have more stable adult home lives in terms of marriage and divorce and higher earnings in the child-rearing years.

These benefits promote intergenerational mobility for their children. The children of treatment group participants are less likely to be enrolled in special education programs and have fewer school suspensions than the children of control group participants. They are more likely to be employed and in good health. They are much less likely to engage in crime. We find important differences in impacts by gender. The male children of

[27] Though the definition of crime outcomes in Dobbie et al. (2018) differs from the definition of crime outcomes in our study, we consider their empirical results to be a good approximation for this exercise. The size of the samples we analyze does not allow us to reliably explore causal estimates of this relationship.

the original male treatment group participants receive the greatest benefits, consistent with a literature on the adverse effects of disadvantaged environments on boys (Autor et al. 2019). García and Heckman (2023) estimate that application of Perry to the currently eligible disadvantaged African American children would reduce the black-white prime-age earnings gap by 42%. Given the commonality of the process of child development around the world, our findings generalize broadly.

References

Autor, D., D. Figlio, K. Karbownik, J. Roth, and M. Wasserman. 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Econ. J. Appl. Econ.* 11 (3): 338–81.

Bailey, D. H., G. J. Duncan, F. Cunha, B. R. Foorman, and D. S. Yeager. 2020. "Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions." *Psychological Sci. Public Interest* 21 (2): 55–97.

Bailey, D., G. J. Duncan, C. L. Odgers, and W. Yu. 2017. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions." *J. Res. Educ. Effectiveness* 10 (1): 7–39.

Baker, M., J. Gruber, and K. Milligan. 2008. "Universal Childcare, Maternal Labor Supply, and Family Well-Being." *J.P.E.* 116 (4): 709–45.

———. 2015. "Non-cognitive Deficits and Young Adult Outcomes: The Long-Run Impacts of a Universal Child Care Program." Working Paper no. 21571, NBER, Cambridge, MA.

Barnett, W. S. 1996. *Lives in the Balance: Age 27 Benefit-Cost Analysis of the High/Scope Perry Preschool Program.* Ypsilanti, MI: High/Scope.

Barr, A., and C. Gibbs. 2022. "Breaking the Cycle? Intergenerational Effects of an Anti-poverty Program in Early Childhood." *J.P.E.* 130 (12): 3253–85.

Bauer, L., and D. W. Schanzenbach. 2016. "The Long-Term Impact of the Head Start Program." Report, Hamilton Project, Brookings Inst., Washington, DC.

Belsky, J., A. Caspi, T. E. Moffitt, and R. Poulton. 2020. *The Origins of You: How Childhood Shapes Later Life.* Cambridge, MA: Harvard Univ. Press.

Bertrand, M., and J. Pan. 2013. "The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior." *American Econ. J. Appl. Econ.* 5 (1): 32–64.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel. 2008. "The Economics and Psychology of Personality Traits." *J. Human Resources* 43 (4): 972–1059.

Bruhn, M., and D. McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Econ. J. Appl. Econ.* 1 (4): 200–232.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343 (6178): 1478–85.

Chiappori, P.-A., M. Costa Dias, and C. Meghir. 2020. "Changes in Assortative Matching: Theory and Evidence for the US." Working Paper no. 26932, NBER, Cambridge, MA.

Conti, G., J. J. Heckman, and R. Pinto. 2016. "The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour." *Econ. J.* 126 (596): F28–F65.

Dahl, G. B., and E. Moretti. 2008. "The Demand for Sons." *Rev. Econ. Studies* 75 (4): 1085–120.

Dehaene, S. 2021. *How We Learn: Why Brains Learn Better Than Any Machine . . . for Now.* New York: Viking.

Deming, D. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Econ. J. Appl. Econ.* 1 (3): 111–34.

Dobbie, W., H. Grönqvist, S. Niknami, M. Palme, and M. Priks. 2018. "The Intergenerational Effects of Parental Incarceration." Working Paper no. 24186, NBER, Cambridge, MA.

Eika, L., M. Mogstad, and B. Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *J.P.E.* 127 (6): 2795–835.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman. 2016. "Early Childhood Education." In *Economics of Means-Tested Transfer Programs in the United States*, vol. 2, edited by R. A. Moffitt, 235–97. Chicago: Univ. Chicago Press.

Ertem, I. O., V. Krishnamurthy, M. C. Mulaudzi, Y. Sguassero, H. Balta, O. Gulumser, B. Bilik, et al. 2018. "Similarities and Differences in Child Development from Birth to Age 3 Years by Sex and across Four Countries: A Cross-Sectional, Observational Study." *Lancet Global Health* 6 (3): e279–e291.

Fernald, L. C., E. Prado, P. Kariger, and A. Raikes. 2017. *A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries*. Washington, DC: World Bank.

Garces, E., D. Thomas, and J. Currie. 2002. "Longer-Term Effects of Head Start." *A.E.R.* 92 (4): 999–1012.

García, J. L., F. Bennhoff, J. J. Heckman, and D. E. Leaf. 2021. "The Dynastic Benefits of Early Childhood Education." Working Paper no. 29004, NBER, Cambridge, MA.

García, J. L., and J. J. Heckman. 2020. "Early Childhood Education and Life-Cycle Health." *Health Econ.* 3 (S1): 1–23.

———. 2023. "Policies to Promote Social Mobility." *Ann. Rev. Econ.* 15:349–88.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados. 2020. "Quantifying the Life-Cycle Benefits of an Influential Early Childhood Program." *J.P.E.* 128 (7): 2502–41.

García, J. L., J. J. Heckman, and A. L. Ziff. 2018. "Gender Differences in the Benefits of an Influential Early Childhood Program." *European Econ. Rev.* 109:9–22.

———. 2019. "Early Childhood Education and Crime." *Infant Mental Health J.* 40 (1): 141–51.

Gihleb, R., and K. Lang. 2020. "Educational Homogamy and Assortative Mating Have Not Increased." *Res. Labor Econ.* 48:1–26.

Golding, P., and H. E. Fitzgerald. 2017. "Psychology of Boys at Risk: Indicators from 0–5." *Infant Mental Health J.* 38 (1): 5–14.

Gorsuch, R. L. 1983. *Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Hansen, B. E. 2021. *Econometrics*. Princeton, NJ: Princeton Univ. Press.

Haskins, A. R. 2014. "Unintended Consequences: Effects of Paternal Incarceration on Child School Readiness and Later Special Education Placement." *Soc. Sci.* 1:141.

Heckman, J., N. Hohmann, J. Smith, and M. Khoo. 2000. "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *Q.J.E.* 115 (2): 651–94.

Heckman, J. J., and G. Karapakula. 2019a. "Intergenerational and Intragenerational Externalities of the Perry Preschool Project." Working Paper no. 25889, NBER, Cambridge, MA.

———. 2019b. "The Perry Preschoolers at Late Midlife: A Study in Design-Specific Inference." Working Paper no. 25888, NBER, Cambridge, MA.
———. 2021. "Using a Satisficing Model of Experimenter Decision-Making to Guide Finite-Sample Inference for Compromised Experiments." *Econometrics J.* 24 (2): C1–C39.
Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz. 2010a. "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program." *Quantitative Econ.* 1 (1): 1–46.
———. 2010b. "The Rate of Return to the HighScope Perry Preschool Program." *J. Public Econ.* 94 (1/2): 114–28.
Heckman, J., R. Pinto, and P. Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *A.E.R.* 103 (6): 2052–86.
Henry, B., A. Caspi, T. E. Moffitt, H. Harrington, and P. A. Silva. 1999. "Staying in School Protects Boys with Poor Self-Regulation in Childhood from Later Crime: A Longitudinal Study." *Internat. J. Behavioral Development* 23 (4): 1049–73.
Hojman, A. P. C. 2016. "Three Essays on the Economics of Early Childhood Education Programs." PhD thesis, Univ. Chicago.
Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans. 2014. "Fostering and Measuring Skills: Interventions That Improve Character and Cognition." OECD Educ. and Soc. Program Report, Org. Econ. Cooperation and Development, Paris.
Kline, P., and C. R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Q.J.E.* 131 (4): 1795–848.
Kottelenberg, M. J., and S. F. Lehrer. 2014. "The Gender Effects of Universal Child Care in Canada: Much Ado about Boys." Manuscript, Dept. Econ., Queen's Univ., Kingston, ON.
Lavigueur, S., R. E. Tremblay, and J.-F. Saucier. 1995. "Interactional Processes in Families with Disruptive Boys: Patterns of Direct and Indirect Influence." *J. Abnormal Child Psychology* 23 (3): 359–78.
Lee, D. S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Rev. Econ. Studies* 76 (3): 1071–102.
Lehmann, E. L., and J. P. Romano. 2006. *Testing Statistical Hypotheses.* Berlin: Springer.
Liang, K.-Y., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73 (1): 13–22.
Masse, L. C., and R. E. Tremblay. 1997. "Behavior of Boys in Kindergarten and the Onset of Substance Use during Adolescence." *Archives General Psychiatry* 54 (1): 62–68.
Millimet, D. L., and C. F. Parmeter. 2022. "Accounting for Skewed or One-Sided Measurement Error in the Dependent Variable." *Polit. Analysis* 30 (1): 66–88.
Murray, J., C. C. Bijleveld, D. P. Farrington, and R. Loeber. 2014. *Effects of Parental Incarceration on Children: Cross-National Comparative Studies.* Washington, DC: American Psychological Assoc.
Nagin, D. S., and R. E. Tremblay. 2001. "Analyzing Developmental Trajectories of Distinct but Related Behaviors: A Group-Based Method." *Psychological Methods* 6 (1): 18–34.
Piquero, A. R., and T. E. Moffitt. 2005. "Explaining the Facts of Crime: How the Developmental Taxonomy Replies to Farrington's Invitation." In *Integrated Developmental and Life-Course Theories Offending*, edited by David P. Farrington, 51–72. London: Routledge.
Protzko, J. 2015. "The Environment in Raising Early Intelligence: A Meta-analysis of the Fadeout Effect." *Intelligence* 53:202–10.

Quandt, R. E. 1958. “The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes.” *J. American Statis. Assoc.* 53 (284): 873–80.

———. 1972. “A New Approach to Estimating Switching Regressions.” *J. American Statis. Assoc.* 67 (338): 306–10.

Rossin-Slater, M., and M. Wüst. 2020. “What Is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention.” *American Econ. J. Appl. Econ.* 12 (3): 255–86.

Schore, A. N. 2017. “All Our Sons: The Developmental Neurobiology and Neuroendocrinology of Boys at Risk.” *Infant Mental Health J.* 38 (1): 15–52.

Schweinhart, L. J., H. V. Barnes, and D. P. Weikart. 1993. *Significant Benefits: The HighScope Perry Preschool Study through Age 27.* Ypsilanti, MI: High/Scope.

Thompson, B. 2004. *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications.* Washington, DC: American Psychological Assoc.

Turney, K., and A. R. Haskins. 2014. “Falling Behind? Children’s Early Grade Retention after Paternal Incarceration.” *Soc. Educ.* 87 (4): 241–58.

Walters, C. R. 2015. “Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start.” *American Econ. J. Appl. Econ.* 7 (4): 76–102.

Weikart, D. P., J. T. Bond, and J. T. McNeil. 1978. *The Ypsilanti Perry Preschool Project: Preschool Years and Longitudinal Results through Fourth Grade.* Ypsilanti, MI: High/Scope.

WHO (World Health Organization) Multicentre Growth Reference Study Group. 2006. “Assessment of Sex Differences and Heterogeneity in Motor Milestone Attainment among Populations in the WHO Multicentre Growth Reference Study.” *Acta Paediatrica* 95 (suppl. 450): 66–75.

Wright, B. R. E., A. Caspi, T. E. Moffitt, and P. A. Silva. 1999. “Low Self-Control, Social Bonds, and Crime: Social Causation, Social Selection, or Both?” *Criminology* 37 (3): 479–514.