I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15672

Parenting Promotes Social Mobility Within and Across Generations

Jorge Luis García
James J. Heckman

OCTOBER 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15672

Parenting Promotes Social Mobility Within and Across Generations

Jorge Luis García
Clemson University

James J. Heckman
University of Chicago and IZA

OCTOBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Parenting Promotes Social Mobility Within and Across Generations*

This paper compares early childhood enrichment programs that promote social mobility for disadvantaged children within and across generations. Instead of conducting a standard meta-analysis, we present a harmonized primary data analysis of programs that shape current policy. Our analysis is a template for rigorous syntheses and comparisons across programs. We analyze new long-run life-cycle data collected for iconic programs when participants are middle-aged and their children are in their twenties. The iconic programs are omnibus in nature and offer many services to children and their parents. We compare them with relatively low-cost more focused home-visiting programs. Successful interventions target both children and their caregivers. They engage caregivers and improve the home lives of children. They permanently boost cognitive and non-cognitive skills. Participants in programs that enrich home environments grow up with better skills, jobs, earnings, marital stability, and health, as well as reduced participation in crime. Long-run monetized gains are substantially greater than program costs for iconic programs. We investigate the mechanisms promoting successful family lives for participants and find intergenerational effects on their children. A study of focused home-visiting programs that target parents enables us to isolate a crucial component of successful programs: they activate and promote parenting skills of child caregivers. The home-visiting programs we analyze produce outcomes comparable to those of the iconic omnibus programs. National implementation of the programs with long-run follow up that we analyze would substantially shrink the overall US Black-White earnings gap.

JEL Classification: J18, J13, J24, J31, D13

Keywords: skills, social mobility, inequality, human development

Corresponding author:
James J. Heckman
Center for the Economics of Human Development and Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
USA

E-mail: jjh@uchicago.edu

* The research in this paper is supported by the Buffett Early Childhood Fund, the National Institutes of Health's Eunice Kennedy Shriver National Institute of Child Health and Human Development under award number R37HD065072, and the National Institute of Aging under award numbers R01AG042390 and R01AG053343. Additional funds were received from an anonymous donor to the Center for the Economics of Human Development. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Institutes of Health. We thank Orla Doyle for providing us with the estimates for Preparing for Life presented in Section 5. We thank Alison Baulos, Orla Doyle, Becky Harles, Lyra Li, and Haihan Tian for helpful comments and research assistance.

Abstract

This paper compares early childhood enrichment programs that promote social mobility for disadvantaged children within and across generations. Instead of conducting a standard meta-analysis, we present a harmonized primary data analysis of programs that shape current policy. Our analysis is a template for rigorous syntheses and comparisons across programs. We analyze new long-run life-cycle data collected for iconic programs when participants are middle-aged and their children are in their twenties. The iconic programs are omnibus in nature and offer many services to children and their parents. We compare them with relatively low-cost more focused home-visiting programs. Successful interventions target both children and their caregivers. They engage caregivers and improve the home lives of children. They permanently boost cognitive and noncognitive skills. Participants in programs that enrich home environments grow up with better skills, jobs, earnings, marital stability, and health, as well as reduced participation in crime. Long-run monetized gains are substantially greater than program costs for iconic programs. We investigate the mechanisms promoting successful family lives for participants and find intergenerational effects on their children. A study of focused home-visiting programs that target parents enables us to isolate a crucial component of successful programs: they activate and promote parenting skills of child caregivers. The home-visiting programs we analyze produce outcomes comparable to those of the iconic omnibus programs. National implementation of the programs with long-run follow up that we analyze would substantially shrink the overall US Black-White earnings gap.

JEL Codes: J18; J13; J24; J31; D13.
Keywords: skills; social mobility; inequality; human development.

Jorge Luis García
John E. Walker Department of Economics
Clemson University
309-C Wilbur O. and Ann Powers Hall
Clemson, SC 29634
Phone: +1 864-656-3481
E-mail: jlgarci@clemson.edu

James J. Heckman
Center for the Economics
of Human Development and
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Phone: +1 773-702-0634
E-mail: jjh@uchicago.edu

I see only too well that the care one takes of us in infancy shapes our feelings, our customs, and our faith. Voltaire (1785)

1. Introduction

Recent evidence on social mobility has stimulated interest in policies to promote it (e.g., Chetty and Hendren, 2018a,b; Chetty et al., 2020). Some researchers claim that place of residence during childhood is an important determinant of social mobility. An older, better-documented, and more often-replicated body of literature emphasizes the role of family influence, primarily that of the mother (e.g., Becker and Tomes, 1979, 1986; Leibowitz, 1974). The two approaches to promoting social mobility are not necessarily at odds given the powerful force of sorting into neighborhoods by family characteristics that is a pervasive feature of modern societies (Heckman and Landersø, 2022). As noted by Alfred Marshall (1890):

General ability depends largely on the surroundings of childhood and youth. In this, the first and far more powerful influence is that of the mother.

This paper contributes to this literature by recognizing the fundamental role of the family and its environment. We study programs that enrich family life and the early lives of children. The current literature is unclear about how best to supplement family life. Some advocate income transfers (e.g., Duncan and Le Menestrel, 2019). Superficially, this sounds like the right approach given that many define child disadvantage by family income. But disadvantage has many aspects. It can also be defined in terms of parental characteristics such as education, mental health, parenting style, or quality of home life (Hertzman and Bertrand, 2007). It might equally well be measured by the quality of parent-child interactions, which are known to foster child development (Inhelder and Piaget, 1972; Vygotsky, 1980).

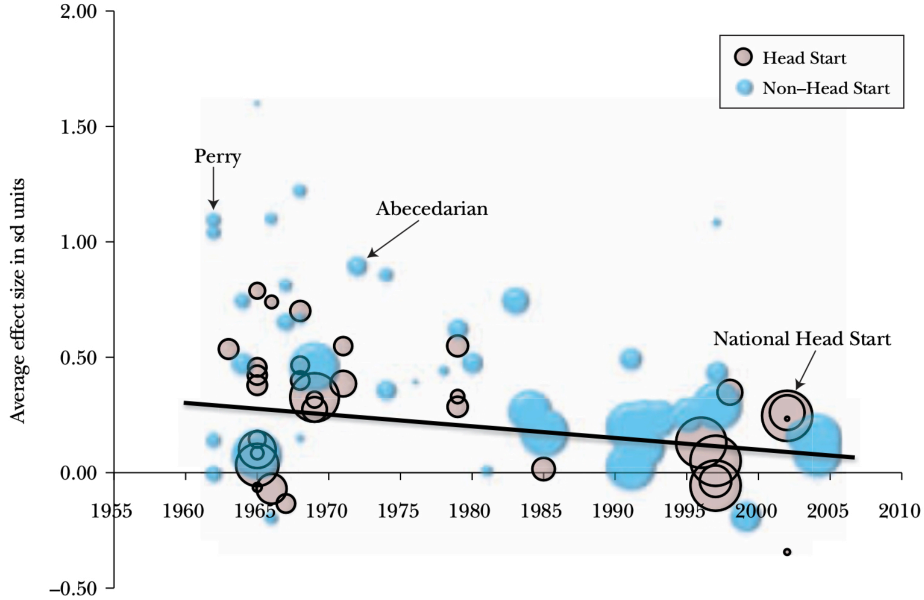
Income has many competing uses. Enhancing it likely has smaller impacts on child development than equally expensive interventions that target specific aspects of child development (Del Boca et al., 2014). Many early childhood education programs target child learning and play activities, as well as parental childrearing skills. They promote attachment of parents with their children.

We critically examine influential programs that supplement the family lives of children and stimulate them. When possible, we examine their impacts on *long-run* child development, avoiding exclusive reliance on the short-term follow-up studies that dominate the recent literature on child development. After comparing omnibus programs with multiple components, we isolate a central component of them that promotes child development: parenting. We move beyond the crude meta-analyses that dominate discussion in the recent literature (e.g., Duncan et al., 2022; Duncan and Magnuson, 2013). Those exercises compare programs that differ greatly in terms of target populations, interventions administered, and measures used to gauge success. They typically compare the incomparable without rigorous methodological standardizations or any understanding of the mechanisms producing successful child development.

Figure 1 is typical of that literature. It compares a hodgepodge of programs organized by date originated without any attempt at standardizing the populations studied. It is based on measures taken during early childhood, shortly after the end of the programs, and does not consider their lifetime consequences. It does not assess the quality of the programs studied, the quality of the reported estimates in terms of methodology, replicability, or comparability in measures used, the quality of the investigators reporting results, the quality of the programs themselves, or the autonomy of the evaluators from the originators of the programs.

This paper avoids these problems and focuses on general principles and long-run evidence where possible. Our approach differs from that taught in schools of education, and promoted by many child-development psychologists and their followers in economics. That approach treats programs as stand alone affairs, and does not search for common develop-

Figure 1. Average Impact of Early Childcare Programs at End of Treatment



Source: Duncan and Magnuson (2013). **Note (directly from source):** This figure shows the distribution of 84 program-average treatment effect sizes for cognitive and achievement outcomes, measured at the end of each program's treatment period, by the calendar year in which the program began. Reflecting their approximate contribution to weighted results, "bubble" sizes are proportional to the inverse of the squared standard error of the estimated program impact. There is a weighted regression line of the effect size by calendar year.

mental mechanisms across them. In this view of policy evaluation, the search is on for the "best" program to be advocated for implementation. The *What Works and What Does Not?* archive is founded on this principle.[1] "Meta-analysis" is built on this approach. Treatment effects from diverse programs, assessed using diverse measures on diverse populations, are "synthesized" forcing comparisons of incomparables. In this approach, statistics substitutes for science.

Our view of policy analysis is fundamentally different. Development is a life-cycle process. We search for *mechanisms* that are universal across time and environments. Such mechanisms are transportable and can guide policy everywhere. Child development is a common dynamic process across eras, cultures, and ethnic groups (Ertem et al., 2018; Fernald et al., 2017; WHO and de Onís, 2006). Policies that build on this commonality have

[1] See Washington State Institute for Public Policy (2015).

the greatest transportability and durability. We ask how to bolster these mechanisms—not to recommend a specific policy off the shelf, but to have a template for assessing and developing successful policies appropriate for targeted populations. It is unlikely that any specific program successful in one context can be transported without modification to another context. The literature develops tools that model the impacts of context and allows analysts to account for it. Long-run studies are central to this approach, as are recently developed approaches that can reliably forecast long-run future outcomes for newly collected samples of program participants (e.g., García et al., 2020).

High-quality programs targeted at socioeconomically disadvantaged participants are socially efficient in the sense of producing net social benefits (i.e., benefits in excess of social costs). We present a harmonized primary analysis of the HighScope Perry Preschool (Perry) and Carolina Abecedarian (ABC) Projects, two prototypical and highly influential omnibus programs that influenced the creation of Head Start and its curricula and continue to shape current policy. We go well beyond an undigested collation of treatment effects without any exploration of mechanisms or social optimality that has become the standard in the literature.[2] Instead, we search for common mechanisms that generate successful lives.

Section 2 briefly sketches an influential framework for studying child development that accounts for program context. It builds on a previous survey by Heckman and Mosso (2014). It is a framework for investigating the mechanisms of child development and measuring skills. Successful programs improve early life skills. Such improvements become building blocks of skills for children at subsequent ages through investment and *dynamic complementarity* (Cunha et al., 2010). The latter is an important concept that justifies the notion that "skills beget skills." Early investment makes later investment more productive in a cumulative fashion. It is valuable because it builds the stock of later-life skills that make later investments productive.

[2]Examples include Bailey et al. (2020), Bailey et al. (2017), Duncan and Magnuson (2013), and Duncan et al. (2022).

We examine the direct impact of programs on parental investment, including parent-child interactions. Successful programs improve the home environments in which children grow up. Impacts on skills and parental investment enhance education and reduce criminal activity. These benefits foster stable labor income and marital lifecycle profiles for male participants, especially during their childrearing years. For female participants, education is a main mediator of midlife outcomes. Participants in these programs grow up to become parents who provide stable environments for their children. They provide more material resources, provide more stable home environments, and engage more in parenting (e.g., they read more often to their children). The enhanced environments produce program impacts that transmit across generations.

Section 3 conducts a primary analysis of the iconic Perry and ABC programs. We document that these successful programs improve parent-child interactions. We synthesize in comparable formats previous evidence from these programs. Perry and ABC remain relevant today. Roughly 30% of Head Start programs are based on the curriculum of Perry, while another 38% of Head Start programs are based on ABC (Walters, 2015). Major preschool initiatives like Educare are modeled after ABC (Yazejian et al., 2015, 2017). García et al. (2021) report that, currently, 10% of African-American children would be eligible for Perry. A similar percentage would likely be eligible for ABC. Apart from their current relevance, the principles underlying them are universal. We analyze Perry and ABC to understand why they are successful. The documented commonalities in the development process give force to our study. We analyze newly collected data on the outcomes of the original participants of ABC when they are 45 years old and on the outcomes of their children when they are in their twenties. We harmonize newly collected data on the outcomes of the original participants of Perry and their children to provide a cohesive analysis of the two programs. The evidence on Head Start, largely based on Perry and ABC, supports our evidence on these two programs. It is also evidence in favor of their scalability.

Our focus on general principles circumvents common criticisms of analyses of Perry and

ABC. Examples include concerns about the small sample sizes of the original interventions, changes over time in the American family, the changing education of women, and the age of the program. Longevity of the participants in follow-up samples is often treated as a liability and not as an opportunity to study long-run outcomes (e.g., Duncan et al., 2022). The argument is that programs are "old" and the world has changed. It ignores the literature investigating mechanisms that enable analysts to account for context and to control for changes in causal factors by conditioning on them.

These trite and shopworn criticisms are based on fundamental misconceptions of science. Good science seeks understanding of mechanisms and not tabulations of treatment effects or quests for a "best" program to replicate. Scientifically based studies control for differences in environments over time that affect program outcomes. Any program chosen to be replicated in a new context needs to be adapted to the circumstances of any potential application. Study of mechanisms facilitates adaptation. The mechanisms generating successful early childhood education programs are universal. They promote family life. Successful programs activate these mechanisms. The search is on for the best ways to do so.

The efficacy of Perry and ABC in boosting outcomes across the life-cycle, and their cost-effectiveness, make the consequences of national implementation of these programs a question of policy relevance. A central feature of most studies of early childhood enrichment programs is that follow-ups are often short-term relative to the full lifetimes that they target. This is much less so for Perry or ABC, but, even for these programs, data are not yet available on the full lifetimes of participants. To forecast benefits accruing in remaining but as yet unrealized lifetimes, it is necessary to make predictions. One approach borrows uncritically from biometrics and uses fitted in-sample relationships to forecast out-of-sample outcomes (e.g., Athey et al., 2019).

This approach is fraught with danger unless the causal factors generating life-cycle outcomes are properly accounted for. Typically, application of the biometric approach uses non-experimental observational data on older cohorts for which the missing outcomes for later ages are measured. This approach has two limitations: (i) the non-experimental estimates used to make forecasts using tests at one age to predict future earnings are subject to selection bias; and (ii) cohort effects may also bias such forecasts. García et al. (2020) develop methods for circumventing these problems. In Section 4, we apply this approach and estimate that, after nationwide program implementations of either Perry or ABC, the overall observed Black-White earnings gap would shrink substantially. We describe this approach and compare its forecasts to those from other approaches. The forecasts based on the approach of García et al. (2020) are accurate when compared to the data. In our example, forecasts based on other, widely used approaches, underestimate realized benefits.

A recurrent feature of successful programs is enhancement of home environments and improvement of parent-child interactions. This is true even in the absence of a formal home-visiting component. Energized and motivated children attending center-based programs stimulate parent-child interactions. In Section 5, we ask if improvements in home environments can be achieved solely through cheaper home-visiting and parent-focused programs rather than the more expensive omnibus programs. A growing literature demonstrates the effectiveness of these programs, which are being implemented worldwide. While most program participants in the available programs are young, measured impacts on skills are similar to those found for the most successful preschool programs. The available long-term evidence indicates strong impacts on cognitive and non-cognitive skills, education, and labor income through age 31 (Gertler et al., 2021; Walker et al., 2022). We analyze the current literature, distill its common elements, and explain why home-visitation and parent-focused programs are promising alternatives to center-based preschools. This section of our study is of interest in its own right as it provides low-cost alternatives to more expensive center-based programs, at a fraction of their cost (5 to 10 percent) and, many, with relatively low-skill requirements for the home visitors—although some programs required professional degrees and extensive training. This portion of our study is of scientific interest because it isolates a mechanism that appears to be highly effective and consistent with the insights of the pioneers of child development.

A significant practical problem with all early childhood education programs is that, while benefits appear to be greatest for the most disadvantaged children (Cornelissen et al., 2018; Gray-Lobe et al., 2021; Havnes and Mogstad, 2015), take up is much greater for the children of the advantaged (e.g., Hermes et al., 2021). Outreach is an important issue that needs to be actively addressed if these programs are to be successfully implemented on a wide scale. Outreach closes the gap between disadvantaged and non-disadvantaged parents in terms of the enrollment of their children; it is equity-enhancing, given that disadvantaged children benefit the most from early childhood education programs. Section 6 concludes.

We emphasize at the outset that this paper offers a template for rigorous syntheses of the evidence with an eye to application to policy. Gaps in the available data prevent full realization of our research program. We show the promise of a mechanism-based approach to the study of early childhood programs and how it produces new insights that are relevant for public policy.

2. A Framework for Interpreting Impacts of Policies Promoting Social Mobility

Following the literature, we use the technology of skill formation (Cunha and Heckman, 2007) to organize ideas and interpret evidence.[3] Vectors of skills at age $a + 1$, $\boldsymbol{S}_{a+1}$, are produced by investments, $\boldsymbol{I}_a$, broadly defined to include money and time invested in children including time spent in interactions with caregivers; by parental attributes, $\boldsymbol{P}_a$, which include quality of spoken language, general knowledge, etcetera; by neighborhoods, $\boldsymbol{N}_a$, through social interactions and peer effects; and by governmentally provided programs, $\boldsymbol{G}_a$. At age a, $\boldsymbol{S}_a$, enhances the productivity of other investments. More skilled children generally benefit more from investment (Heckman and Mosso, 2014). The stock of skills at age $a + 1$

[3] See Heckman and Mosso (2014) for a recent survey.

is generated by the following relationship

$$\boldsymbol{S}_{a+1} = \boldsymbol{F}^{(a)}\left(\boldsymbol{S}_a, \boldsymbol{I}_a, \boldsymbol{N}_a, \boldsymbol{P}_a, \boldsymbol{G}_a\right). \tag{1}$$

Defining inputs in a positive-productivity way so that $\boldsymbol{F}^{(a)}$ is increasing in all of its arguments at each age a (for each increase of a factor at age a, holding other inputs constant, $\boldsymbol{S}_{a+1}$ increases) and cross productive (inputs are complementary; i.e., the technology is supermodular). The technology $\boldsymbol{F}^{(a)}$ is age specific. At certain ages, some investments may be more productive than others corresponding to critical and sensitive periods (Cunha and Heckman, 2007; Heckman and Mosso, 2014). Some investments may only be productive at certain ages. Skills $\boldsymbol{S}_a$ may complement investment. Certain skills may only emerge at later ages (e.g., Steinberg, 2014).

Technology (1) is a valuable interpretive device. It has been estimated in a series of papers.[4] A mounting body of evidence supports the hypothesis of dynamic complementarity (i.e., $\frac{\partial^2 \boldsymbol{F}^{(a)}}{\partial \boldsymbol{S}_a \partial \boldsymbol{I}_a'} > 0$, which is increasing in a).[5] This technology is also a vehicle for comparison of mechanisms across studies and for understanding how parental environments, neighborhoods, and government policies affect productivity of investment and shape the accumulation of skills.

The technology is often joined with a model of measurements of skills:

$$\boldsymbol{M}_a = \Phi^a\left(\boldsymbol{S}_a, \boldsymbol{\tau}_a\right), \tag{2}$$

where $\boldsymbol{\tau}_a$ are other factors that explain measurements $\boldsymbol{M}_a$. Behaviors, $\boldsymbol{B}_a$ (e.g., attending

[4]Examples of these papers include: Cunha and Heckman (2008), Cunha et al. (2010), Agostinelli and Wiswall (2016), Pavan (2016), Del Boca et al. (2019, 2014), Mullins (2019), Attanasio et al. (2020), Caucutt et al. (2020), Chaparro et al. (2020), Del Bono et al. (2022).

[5]Heckman and Mosso (2014) discuss some evidence that $\frac{\partial^2 \boldsymbol{F}^{(a)}}{\partial \boldsymbol{S}_a \partial \boldsymbol{I}_a'} < 0$ at early ages, but the cross partial is increasing over time and becomes positive at later ages.

school, showing up on time), depend on skills and incentives, $\boldsymbol{R}_a$:

$$\boldsymbol{B}_a = \Psi^a \left(\boldsymbol{S}_a, \boldsymbol{E}_a \left(\boldsymbol{R}_a \right) \right), \tag{3}$$

where $\boldsymbol{E}_a$ are factors like effort that affect behaviors and are affected by incentives (e.g., desire to please). One can think of behaviors as a special class of measurements because they are manifestations of $\boldsymbol{S}_a$, among other factors. Systems (2) and (3) facilitate comparisons of outcomes and behaviors across environments. Conditioning on $\boldsymbol{\tau}_a$ and $\boldsymbol{E}_a(\boldsymbol{R}_a)$ allows for meaningful comparisons across individuals, studies, and time.

To this framework, the literature adds parental preferences over skills, lifetime outcomes of children, and information sets on which there is a large literature (e.g., Attanasio et al., 2019; Caucutt and Lochner, 2020; Cunha and Heckman, 2007; Del Boca et al., 2019, 2014). Heckman and Mosso (2014) summarize the literature, but there have been substantial developments.[6] Starting with Becker and Tomes (1979, 1986), parental preferences toward children have been shown to play important roles. As Attanasio et al. (2019) and Cunha et al. (2022) discuss, parental information about patterns of child development play an important role in shaping investments. Coherent models of the household include intertemporal budget constraints. Credit rationing plays an important role in many models, as parents cannot readily access their future income.[7]

These considerations motivate policies to enhance and shape parental information sets, as well as policies that transfer income to circumvent credit constraints (Barr et al., 2022; Deshpande, 2016; Duncan and Le Menestrel, 2019). These alternative theoretical considerations have important policy counterparts. Missing to date are comprehensive models that encompass the system (1), (2), and (3) and are useful for policy design. This produces a disjoint literature that focuses on one aspect or another, leading to a variety of programs

[6] Examples include Agostinelli and Wiswall (2016), Attanasio et al. (2020), Caucutt and Lochner (2020), Caucutt et al. (2020), Chaparro et al. (2020), Del Boca et al. (2019), Doepke and Zilibotti (2017, 2019), Mullins (2020), Seror (2022).

[7] See, e.g., Becker and Tomes (1979), Caucutt et al. (2020), and Cunha et al. (2006).

with different emphases. This paper is a step in the direction of a comprehensive model but much remains to be done.

3. Two Early Childhood Education Programs that Promote Social Mobility

We focus on two prototypical and influential programs. Rather than reporting a meta-analysis of the many dots in Figure 1, we make deeper and more systematic comparisons. We present a side-by-side analysis of the two programs instead of simply reporting estimated treatment effects for selected outcomes. We study how they affect life outcomes of participants on a number of important dimensions and the mechanisms through which these benefits are obtained. Technology (1) and the models of measurement (2) and behavior (3) provide our analytical framework.

We conduct primary analyses of the HighScope Perry Preschool (Perry) and the Carolina Abecedarian (ABC) Projects. Perry was conducted in Ypsilanti, Michigan. ABC was conducted in Chapel Hill, North Carolina. They randomized 123 (Perry) and 114 (ABC) children to treatment and control groups. Both programs implemented eligibility criteria to enroll similarly disadvantaged children. All Perry participants were Black. ABC's eligibility criteria was such that only one child participant was not Black. Participants were representative of a broader disadvantaged population in the US. García et al. (2021) report that the percentage of male and female children eligible today would be 10%.

3.1 Background and Data Analyzed

In Perry, treatment-group children received two years of 2.5-hour preschool sessions during school-year weekdays starting at age three. They also received weekly teacher home visits during the two-year treatment period. In ABC, treatment-group children received five years of eight-hour center-based care during weekdays starting at birth. They did not receive home visits. ABC was more intensive than Perry during weekdays but also throughout the year. It operated 50 weeks a year, while Perry operated 30 weeks a year. Control-group children of Perry did not receive any treatment, and no treatment substitutes were available in the area where they lived. Up to 75% of the control-group children of ABC attended alternative formal childcare arrangements. Most of these alternative arrangements began when children were three years old. Below, we clarify the interpretation of estimates of ABC impacts in light of the fact that control-group childcare arrangements varied greatly.

Perry enrolled five cohorts of participants between the years 1962 and 1967. ABC enrolled four cohorts of participants between the years 1972 and 1976. Between birth and preschool age, ABC provided basic care for the treated children. After that, its curriculum was very similar to that of Perry. The curricula of the programs were designed to foster the development of cognitive and non-cognitive skills. Children were active learners who planned, executed, and reflected on activities. Teachers, who were trained to implement early childhood curricula, guided them. Children solved exercises. Teachers gave them feedback. Both programs were high-quality. Appendix A1 provides details, including comprehensive descriptions of eligibility criteria. See also Table 4.

All parents of child participants who were initially offered participation in the randomization protocols of Perry and ABC accepted the offer (García et al., 2020; Weikart et al., 1978). Therefore, we can identify average treatment effects for the eligible population. Let $\boldsymbol{x}$ be baseline characteristics and $\mathcal{B}$ the set of eligibility criteria. $\mathcal{B} \subseteq \mathcal{X}$, the support of $\boldsymbol{x}$. $\mathcal{B}$ is defined separately for each program. The two programs targeted disadvantaged children. For example, Heckman et al. (2010) document that Perry's eligibility criteria targeted approximately 15% of the children born in the US during the 1960s. García et al. (2020) report similar results for ABC.

After randomization, the treatment status of a few participants was swapped to facilitate program participation. Reassignment compromised the randomization protocol and caused an imbalance of baseline characteristics (see Table 1). A few cases attrited. Multiple studies implement strategies to deal with randomization compromises and attrition when estimating treatment effects (Heckman and Karapakula, 2021; Heckman et al., 2022). These corrections barely affect point estimates and inference (e.g., García et al., 2021, 2018). We thus present unadjusted mean differences, which facilitates clarity and replicability. All inference is permutation-based and accounts for the small sample sizes of both programs.

Follow-ups with the original participants were conducted when they were 19, 27, 40, and 54 in Perry and ages 21, 30, and 45 in ABC. A follow-up was conducted to collect health outcomes with the ABC participants around age 34. Multiple studies analyze the data from these follow-ups.[8] Recently, García et al. (2021) analyze new data collected in a follow-up with the participants of Perry when they were in their late midlives (age 54). They combine observations from retrospective surveys observed throughout adulthood and administrative data to construct longitudinal life-cycle education, labor income, and crime outcomes up to late midlife. This paper emphasizes the new results from the Perry late midlife data and new midlife data from a follow-up with participants of ABC when they were around 45 years old.[9] Both midlife follow-ups ask the original participants about their children. These data allow us to analyze intergenerational outcomes. They also allow us to construct very similar outcomes for participants and their children across the two programs—Table 1 provides details.

3.2 Parameters of Interest

For individual $i \in \mathcal{I}$, let $Y_{i,j}^{0,\mathrm{h}}$ be the potential outcome $j \in \mathcal{J}$ in the no-treatment counterfactual when the child stays at home and does not receive a formal childcare alternative to the program. $Y_{i,j}^{0,\mathrm{c}}$ is analogously defined for the case when the child receives a formal

[8]In economics, studies of Perry include Heckman et al. (2010), Heckman et al. (2010), and Heckman et al. (2013). Studies of ABC include García et al. (2018), García et al. (2019), García et al. (2020), and García and Heckman (2021). Campbell et al. (2014) analyze health outcomes of ABC. Conti et al. (2016) analyze health outcomes of both Perry and ABC. Studies in other fields are listed in these references.

[9]The new midlife ABC data are analyzed in García (2022), including the intergenerational outcomes discussed in Section 3.5.

Table 1. Baseline Characteristics, Outcomes, and Fertility: Original Participants of Perry and ABC

	Perry			ABC		
	Control Mean	Mean Difference (MD)	MD p-value	Control Mean	Mean Difference (MD)	MD p-value
***Panel a.* Baseline**						
IQ (Perry) or Mother's IQ (ABC)	78.54	1.03	0.387	83.49	1.83	0.399
Socioeconomic Index	8.62	0.17	0.530	21.82	-1.93	**0.089**
Mother Does not Work¶	0.69	0.22	**0.002**	0.39	-0.22	**0.010**
Mother's Year of Birth	1959.97	0.03	0.950	1974.35	-0.15	0.674
***Panel b.* Midlife Skills†**						
Cognitive	0.00	0.48	**0.005**	0.00	0.34	**0.031**
Non-Cognitive	0.00	0.50	**0.011**	0.00	0.47	**0.031**
***Panel c.* Midlife Education‡**						
High-School Graduate	0.52	0.20	**0.021**	0.53	0.20	**0.025**
College Graduate	0.05	0.02	0.453	0.09	0.21	**0.007**
Panel d.* Midlife Outcomes						
Married	0.25	0.09	**0.082**	0.42	0.01	0.486
Labor Income (2021 USD)	16,298.91	7,826.94	**0.018**	37,527.95	13,044.70	**0.098**
Household Labor Income (2021 USD)	25,121.43	13,243.21	**0.007**	37,247.62	14,632.67	**0.071**
Accumulated Days (Perry) or Times (ABC) in Jail or Prison	1,326.71	-380.83	0.237	0.14	-0.12	**0.027**
Never Arrested (Perry) or Accumulated Arrests (ABC)	0.46	0.18	**0.039**	0.61	0.26	0.151
Physical Health	0.00	-0.02	0.553	0.00	0.28	**0.096**
Mental Health	0.00	0.31	**0.072**	0.00	0.20	0.111
***Panel e.* Midlife Fertility†**						
Any Children	0.80	-0.01	0.878	0.89	-0.03	0.748
Age at Onset	22.63	0.87	0.469	21.93	2.23	0.122
Number of Children	2.42	0.15	0.727	2.31	-0.19	0.524
> 5 Children	0.07	0.02	0.727	0.00	0.02	0.928
***Panel f.* Sample Sizes**						
Original Participants at Baseline	65	-7		57	2	
Original Participants at Midlife Follow-up	50	2		45	6	

Note: Panels a. and e. present the control-group mean and treatment-control mean difference for the outcome in the label for the Perry Preschool (Perry) and Carolina Abecedarian (ABC) projects. For each treatment-control mean difference (MD), we present the permutation *p*-value associated with the null hypothesis that such mean difference is equal to 0. We bold *p*-values when they are less than 0.10. Panels b. to d. are analogous in format to Panels a. and e. The null hypothesis in these latter panels is that the mean difference is less than or equal to 0. ¶The difference between treatment-group mothers in ABC and Perry is that ABC provided full-day childcare and Perry did not. †Based on identical variables observed at age 54 for Perry and 45 for ABC. ‡Based on identical variables of *completed* years of education for both Perry and ABC. ∗For Perry, marriage is the fraction of years married between ages 20 and 40; labor income is the average earnings from labor income between ages 20 and 40; household labor income is the previous variable in addition to average spouse's labor income between ages 20 and 40 (if married); accumulated days in prison and never arrested are observed up to age 54. For ABC, marriage is an indicator of whether an individual is married at age 45; labor income is measured at age 45; household income is the previous variable in addition to spouse's labor income at age 45 (if married); times in jail and accumulated arrests are measured at age 30. For Perry, physical health is a latent variable of measures describing prevalence and intensity of diabetes, stroke, heart disease, self-rated health, body-mass index, and waste-to-hip ratio at age 54. For ABC, an analogous variable is constructed using information at age 34. For Perry, mental health is a latent variable of measures describing depression and anti-social behavior at age 54. For ABC, an analogous variable is constructed using information at age 45.

childcare alternative. The potential outcome under no-treatment status is

$$Y^0_{i,j} = (1 - V_i) \cdot Y^{0,\text{h}}_{i,j} + V_i \cdot Y^{0,\text{c}}_{i,j}, \tag{4}$$

where V_i indicates whether a control-group participant attended alternative formal childcare. The potential outcome if treated is $Y^1_{i,j}$. Observed outcomes are

$$Y_{i,j} = (1 - D_i) \cdot Y^0_{i,j} + D_i \cdot Y^1_{i,j}, \tag{5}$$

where D_i is an indicator of treatment status (Quandt, 1958, 1972).

We estimate the average treatment effect for those eligible to participate:

$$\Delta_j := \mathbb{E}_{\boldsymbol{x} \in \mathcal{B}} \left[Y^1_{i,j} - Y^0_{i,j} \right]. \tag{6}$$

For Perry, $V_i = 0$ for all participants. For ABC, $V_i = 1$ for 75% of the participants and $V_i = 0$ for the rest. For Perry, Δ_j is the average effectiveness of the program relative to staying at home. For ABC, it is the average effectiveness of the program relative to the perceived next best (by the parents of control-group child participants) quality of all available alternatives, including staying at home.[10,11]

3.3 Long-Run Impacts

Table 1 summarizes our treatment-effect estimates. We show the extent to which program impacts are sustained over the life cycle and thus move well beyond the end-of-program focus of Figure 1 and much of the literature. We discuss mechanisms generating treatment effects. We start by analyzing midlife skills. Previous work shows that short-term impacts

[10] It is thus a local average treatment effect (Imbens and Angrist, 1994), see Heckman and Vytlacil (2007).

[11] García et al. (2018) use observational methods to refine Δ_j and address, separately, the effectiveness of ABC relative to staying at home and to alternative childcare arrangements. They find that those two comparisons are similar to Δ_j. Their estimates are plausible. The disadvantaged environments in which the control-group children grew-up were similar to the environments in the childcare alternatives, which were generally lower quality according to the documentation in García et al. (2020).

on cognitive and especially non-cognitive skills ($\boldsymbol{S}_a$) are the building blocks of long-term impacts (Conti et al., 2016; Heckman et al., 2013).

The Fade Out of Fadeout. Panel b. of Table 1 shows that both programs have a long-run impact on cognition, as measured by well-established tests (Raven and Stroop), which we observe for both programs (i.e., treatment-effect estimates are based on the same measures). These tests were collected at age 54 in Perry and age 45 in ABC. For each program, we summarize diverse measures (test items) forming a latent variable. We standardize the latent variable by subtracting the control-group mean and dividing it by the control-group standard deviation. Perry increased cognition at age 54 by half of a standard deviation (p-value = 0.01). ABC increased cognition at age 45 by one third of a standard deviation (p-value = 0.031). These are new findings for harmonized data on the impact of high-quality early education on late midlife cognition.

These long-lasting impacts on cognition for both programs contradict the frequently repeated refrain about "fadeout" in the treatment effects on skills, specifically cognition. Previous research claims that the impact of early childhood education on cognitive-test scores disappears (fades out) shortly after the interventions (Hojman, 2016; Protzko, 2015). Some authors claim that the fadeout in cognition (and socio-emotional skills) is real and not only a measurement artifact (Bailey et al., 2017, 2020).

These studies are based on short-run follow-ups. The evidence presented here refutes this claim for comparable (across programs) cognitive and non-cognitive skill measures. We construct a latent variable of non-cognitive skills for Perry participants, using the same procedure used to construct the cognitive latent variable. We use reverse-coded ratings (by a household member) and self-ratings of how reserved, critical, disorganized, and anxious participants are to create a positively graduated measure. The constructed latent variable can be interpreted as a measure of "positive personality." For ABC, we construct an indicator of positive personality using similar binary ratings. We standardize it using the control-group mean and standard deviation. Perry and ABC increase positive personality by half a standard deviation, with p-values of 0.01 and 0.02 respectively.

Panels c. and d. show that the long-term impact of the programs goes well beyond just enhancing cognitive and non-cognitive skills. We measure completed education at midlife. Both programs significantly increase the high-school graduation rate by 20 percentage points, from a control-group base rate of around 50%. They also significantly increase earnings from labor income during adulthood and decrease criminal behavior. Perry decreases the likelihood of ever being arrested by 18 percentage points (p-value = 0.04) from a control-group rate of 46%. ABC decreases the average number of times in jail or prison by 0.12 (p-value = 0.03) from a control-group average of 0.14.

We also construct latent variables for physical and mental health. Examples of measures in which the former latent is based on are prevalence of diabetes, stroke, and heart disease, as well as self-rated health, body-mass index, and waste-to-hip ratio. The latter is based on measures of depression, anxiety, and anti-social behavior. Perry improves mental health by 0.31 standard deviations (p-value = 0.07), from a control-group mean of 0. ABC improves physical health by 0.28 standard deviations (p-value = 0.096) from a control-group mean of 0.

The outcomes in Table 1 represent broad categories. They show sustained program impacts generating marriage, labor, and law-abiding stability across the life-cycle. Section 3.8 shows that these factors drive the estimates of social efficiency of investment in these programs.

3.4 Mechanisms

Using technology (1), Heckman et al. (2013) show that Perry's enhanced non-cognitive skills, measured at the end of the program and throughout childhood, mediate the shorter-term impact on achievements tests (e.g., California Achievement Tests at ages 8 and 14) and long-term impact on outcomes such as employment, criminal behavior, and drug use. Conti et al. (2016) show that the same is true for treatment effects on the health outcomes of both Perry and ABC. We expand understanding of mechanisms producing treatment effects by investigating how home environments were shaped by Perry and ABC and how this shaping impacted the development of skills fostered by the programs.

For Perry, we revisit the original data archive and recover observations of the Parental Attitude Research Instrument (PARI, Loewenstein, 1973). These data have never previously been analyzed. PARI was collected from the mothers of the original participants by staff conducting home visits when the children were between three and five years old. It measures the quality of parental investment and parenting. Specifically, it scores the interactions of mothers with their children.[12] It does not score availability of material resources. For ABC, we use four annual observations of the Home Observation Measurement of the Environment (HOME, Leventhal et al., 2004; Linver et al., 2004), collected from the mothers of the original participants when their children were 6, 18, 30, 42, and 54 months old. The majority of items of the HOME inventory measure the quality of parent-child interactions. A minority of its items measure the availability of material resources available for the children in the household. PARI is comparable to the vast majority of items of the HOME inventory which measure learning stimulation, parental and verbals skills, and parental warmth.

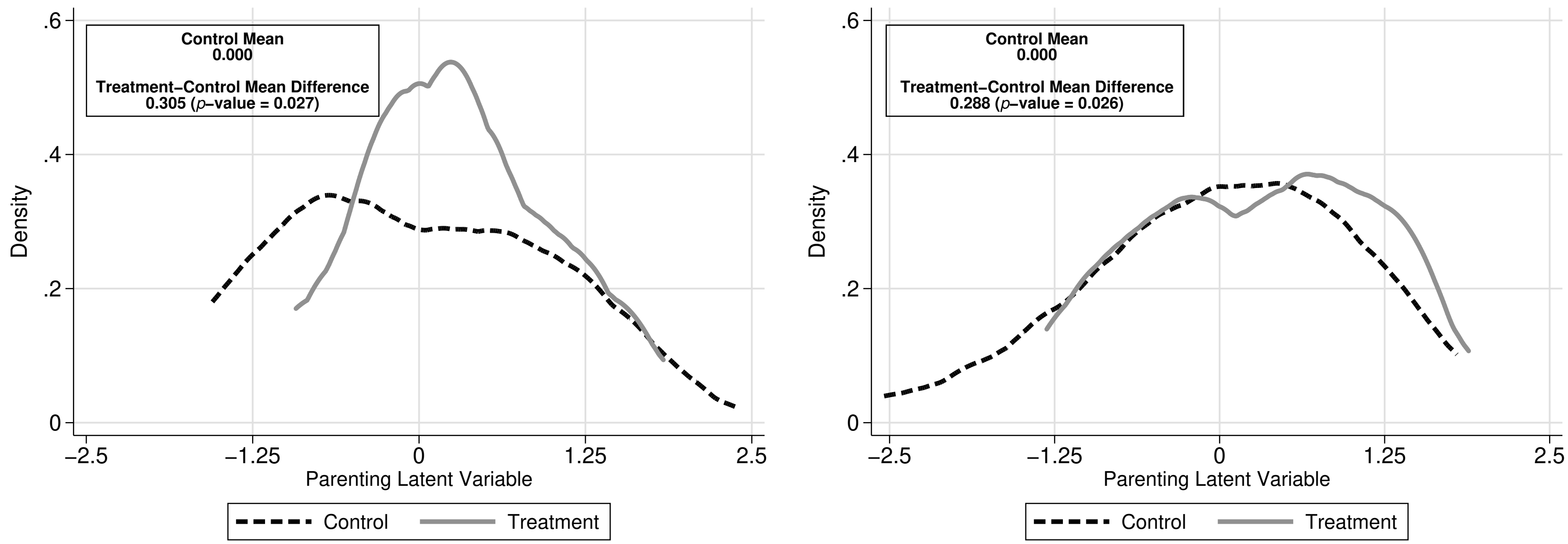
For each program, we form a latent variable based on items of PARI or HOME, using the same methodology as in Section 3.3. We standardize the latent variables by subtracting their control-group mean and dividing it by their control-group standard deviation. Figures 2a and 2b show the distributions of our measures by treatment status for each program. Perry and ABC enhance our measures of parenting or parental investment by an average of 0.3 (p-value = 0.027) and 0.3 (p-value = 0.026), respectively. These findings bolster our interpretation of Perry and ABC as policies targeting disadvantaged families. Perry and ABC enhance the environments in which participants grow up during childhood. Perhaps,

[12] Examples of items in PARI include “encouraging verbalization,” “fostering dependency,” “seclusion of the mother,” and “marital conflict.”

Figure 2. Parenting Received by the Original Participants of the Perry and ABC

(a) Parenting Distribution, Perry

(b) Parenting Distribution, ABC



Note: Panel (a) shows the probability density function of a latent variable describing the parental investment (parenting) received by the original participants of the Perry Preschool Project (Perry) by treatment status. We also display the control-group mean and the treatment-control mean difference in the index together with the permutation p-value for this difference. The null hypothesis for the difference is that it is less than or equal to 0. Panel (b) is analogous in format to Panel (a) for the parental investment received by the original participants of the Carolina Abecedarian Project (ABC).

more importantly, the programs improve the interactions of child participants with their caretakers, which last long after the program ends.

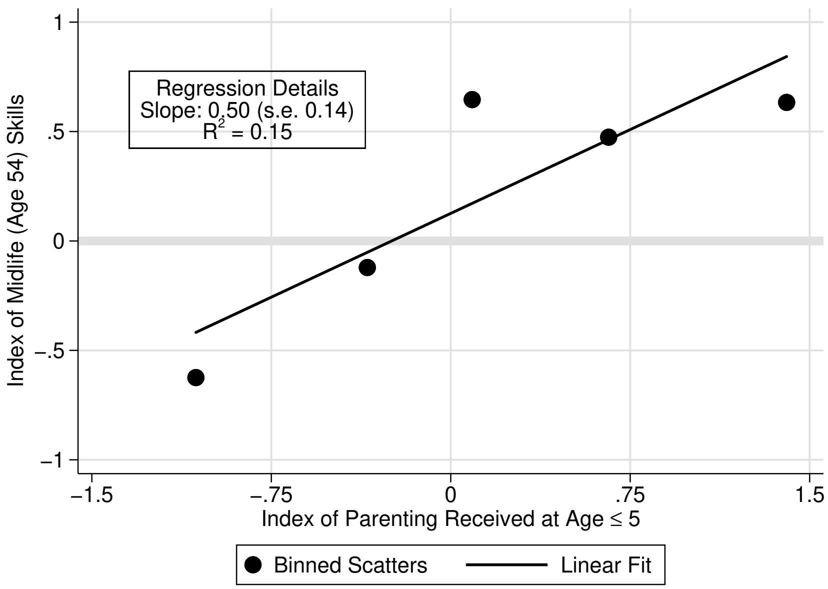
The traditional literature focuses on credentials of program staff and curricula and overlooks the role of the family in bolstering child development. Previous research links the impacts of Perry and ABC on skills to its impacts on long-term outcomes (e.g., education, labor income, and crime; Heckman et al., 2013). Our evidence on the impact of these programs on parenting and parental investment gives a better understanding of the mechanisms underlying program success. It provides evidence of the connection between home environments and skill formation.

We investigate the extent to which the better parent-child interactions generated by the program are building blocks for its impact on skills. For each program, Figure 3 displays the binned relationship between the average of the measures of midlife cognitive and non-cognitive skills previously described and the measures of parenting plotted in Figures 2. We plot the estimated linear relationship between the two measures, which fits the observed binned relationship well. Our measure of parental investment explains at least 15% of the variation of the average of the two skills. Remarkably, these measures are based on data collected fifty years apart. For the two programs, a one-standard-deviation increase in our post-treatment measure of parenting, collected when the original participants were at most 5 years old, is associated with an increase of half a standard deviation in the average of midlife skills.

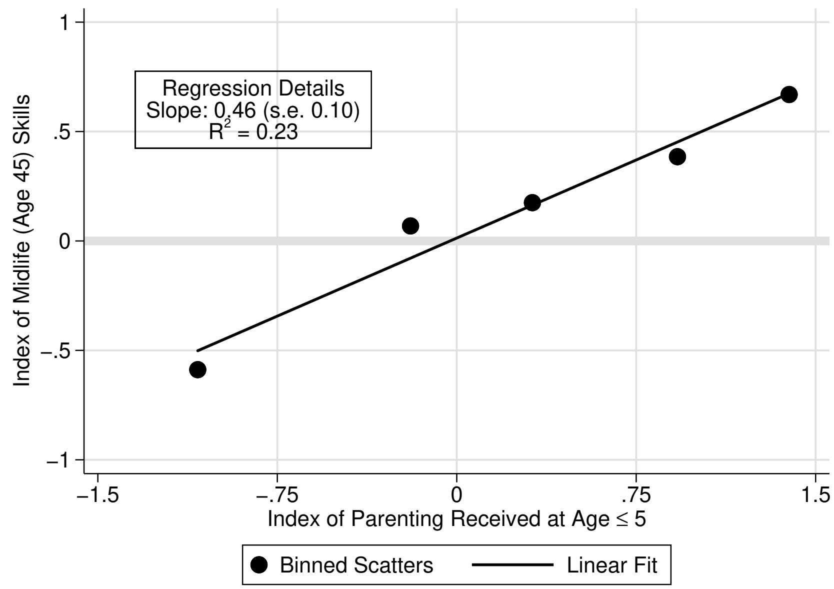
A formal quantification of the degree to which the treatment effect on parenting mediates the treatment effect on midlife outcomes is presented in Figure 4. We focus on midlife skills at age 54 as displayed in Figure 3a and decompose the treatment effect of Perry on this average of cognitive and non-cognitive skills into the treatment effect on parenting displayed in Figure 2a and treatment effects on early-life cognitive and non-cognitive skills. We use the mediation methodology of Heckman et al. (2013) and the measures of early-life skills that

Figure 3. Parenting Received by the Original Participants of Perry and ABC and their Adult Skills

(a) Midlife Skills and Parenting, Perry



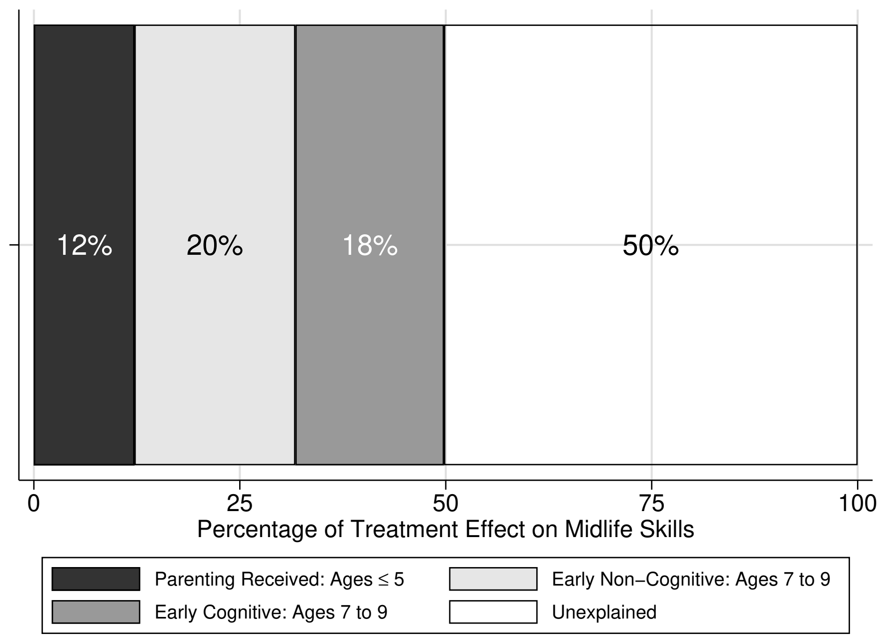
(b) Midlife Skills and Parenting, ABC



Note: Panel (a) displays the linear relationship between the latent variable of parental investment received by the original participants of the Perry Preschool Project summarized in Figure 2a and an average of their midlife cognitive and non-cognitive skills measured at age 54, together with the corresponding description of the linear-regression. Panel (b) is analogous in format to Panel (a) for the original participants of Carolina Abecedarian Project, whose midlife skills are measured at age 45. The number of bins in the scatterplots is calculated using the procedure in Cattaneo et al. (2019).

these authors use.[13] We confirm their finding that impacts on early-life skills mediate the impacts on long-term outcomes. We explain the sources of impacts of programs like Perry. Parenting received by children before age 5, which Perry, and, more generally, successful programs improve, is a relevant mediator (e.g., the treatment effect on parenting received at age 5 or before explains 12% of the treatment effect on midlife skills).

Figure 4. Decomposition of Treatment Effect on Midlife Skills into Treatment Effects on Parenting Received and Early-Life Skills, Perry



Note: This figure displays a decomposition of the treatment effect on the average of cognitive and non-cognitive midlife skills at age 54 for the participants of the Perry Preschool Project displayed in Figure 3a into treatment effects on the parenting that they received (described in Figure 2a), as well as treatment effects on their early-life (ages 7 to 9) non-cognitive and cognitive skills. The decomposition is estimated using the methodology for mediation analysis in Heckman et al. (2013). The skill measures are described in Sections 3.4 and 5.

[13] For non-cognitive skills, we use an index of reverse-coded measures of externalizing behavior observed between ages 7 and 9. Heckman et al. (2013) use an index of academic motivation in addition to the index of externalizing behavior. They show that externalizing behavior is the main measure of non-cognitive skills explaining long-term outcomes. We thus omit academic behavior. For cognitive skills, we combine three measures of the Stanford-Binet IQ score observed between ages 7 and 9. Section 5 provides details on these skill measures as well as treatment-effect estimates for them.

3.5 Intergenerational Impacts

The long-lasting improvements of the original participants' skills, marriage stability, earnings, criminal behavior, and health of both Perry and ABC occurred during their childrearing years. These impacts translate into better family environments for their own children during their childrearing years. García et al. (2021) document that the more skilled and educated original treatment-group members of Perry who become parents are less likely to have children out of wedlock and cohabitate with new partners while their children grow up relative to their control-group counterparts. They are more likely to stay married while their children grow up and read more to their children. Larger average labor incomes earned by the original treatments translate into more resourced home environments. Lower incarceration rates translate into greater parental presence at home, especially for treatment-group fathers.

We analyze the Perry data analyzed by García et al. (2021) and newly available data on the children of the original participants of ABC. Panel e. of Table 1 presents fertility information on the 102 original participants surveyed in the age-54 Perry follow-up and 96 original participants surveyed in the age-45 ABC follow-up. There are no economically or statistically significant average differences across experimental groups on fertility variables, including whether participants have children or not, age at onset of fertility, number of children, and whether participants have more than five children. Overall, the data indicate that the program had minimal impact on childbearing, making differences in experimentally induced fertility a secondary consideration.[14] Information losses due to not observing children yet to be born are also a minor issue in the age-54 and age-45 follow-ups, as the vast majority of the original participants are likely to have completed childbearing.

We construct intergenerational outcomes as follows. 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

[14] The original participants are only asked about their first five children. This does not result in a major loss of information because only a small fraction of original participants of the two programs report having more than five children.

is

$$\bar{Y}^{c}_{i,j} := \frac{1}{\#\mathcal{C}_i} \sum_{c \in \mathcal{C}_i} Y^{c(i)}_{i,j} \cdot \mathbf{1}\left[\#\mathcal{C}_i > 0\right], \tag{7}$$

where "#" denotes cardinality and $\mathcal{C}_i$ indexes the children of first-generation or original participant i, $c(i) \in \mathcal{C}_i$. We analyze $\bar{Y}^{c}_{i,j}$ (outcome j for each original participant, which is the outcome for the "average child" of i). We treat average child outcomes as treatment and control outcomes for the original participants. The observed outcome is at the original-participant level and thus is defined as in Equation (5).

Table 2 presents the outcomes observed and corresponding treatment-effect estimates.[15] We use the same mean-difference estimator as in Section 3.3. There are important intergenerational spillovers.[16] As previously noted, there is heterogeneity by sex in the impacts of the first generation (e.g., García et al., 2018). We confirm that this heterogeneity also arises in the second generation. Both programs have economically and statistically sizable intergenerational impacts on employment or non-idleness rates, average health status, and non-divorced rates of the average male child of the original treatment-group participants relative to the average child of their control-group counterparts. Perry also has a sizable negative intergenerational impact on arrest rates. Both programs sizably increase high-school graduation for girls. ABC also increases college female graduation rates.

3.6 Mechanisms Producing Intergenerational Treatment Effects

Figure 5, reproduced from García et al. (2021), succinctly illustrates the mechanisms generating Perry's intergenerational treatment effects. It plots marriage rates, average earnings,

[15] Some outcomes are unavailable for ABC because the children of the original participants are younger than in Perry. We limit the sample to children in the age categories labeled in the table. These limits allow us to estimate treatment effects in comparable age ranges. For example, we evaluate impacts on employment after school age or impacts on parenthood during teenagehood or young adulthood.

[16] García et al. (2021) discuss and compare their results to other studies in the literature exploiting quasi-experimental variation in the availability of Head Start (Barr and Gibbs, 2022) and preschool programs in Denmark (Rossin-Slater and Wüst, 2020) to estimate intergenerational impacts of early childhood education programs. The results in these other studies generally align with Perry's intergenerational impact.

Table 2. Summary of Intergenerational Outcomes: Children of the Original Participants of Perry and ABC

	Male Children			Female Children		
	Control Mean	Mean Difference (MD)	MD p-value	Control Mean	Mean Difference (MD)	MD p-value
Panel a. **Perry**						
High School Graduate (Age 18 or older)	0.67	-0.01	0.582	0.74	0.13	**0.026**
College Graduate (Age 23 or older)	0.04	0.08	**0.063**	0.31	-0.09	0.846
Employed (Age 23 or older)	0.48	0.19	**0.040**	0.41	0.09	0.218
Never Arrested (Age 18 or older)	0.37	0.14	**0.089**	0.78	0.06	0.210
In Good Health (Age 18 or older)	0.82	0.12	**0.006**	0.85	0.10	**0.030**
Not a Parent (Ages 14 to 22)	1.00	0.00	1.000	0.83	0.12	0.234
Never Divorced (Age 23 or older)	0.93	0.07	**0.028**	0.86	0.11	**0.016**
Panel b. **ABC**						
High School Graduate (Age 18 or older)	0.66	-0.06	0.718	0.28	0.18	**0.067**
College Graduate (Age 23 or older)	0.55	-0.08	0.683	0.18	0.25	**0.068**
Not Idle (Age 15 or older)†	0.91	0.06	**0.083**	0.98	0.00	0.572
In Good Health (Age 18 or older)	0.83	0.18	**0.000**	0.88	0.10	0.133
Not a Parent (Ages 14 to 22)	0.63	0.17	**0.069**	0.94	-0.01	0.584

Note: Panel a. presents the control-group mean and treatment-control mean difference (MD) for the intergenerational outcome in the label for the Perry Preschool Project (Perry). Intergenerational outcomes are for the average child. We construct them by averaging within original program participants across up to their five eldest children. For each mean difference, we present the permutation *p*-value associated with the null hypothesis that the mean difference less than or equal to 0. We bold *p*-values when they are less than 0.10. Panel b. is analogous in format to Panel a. for the Carolina Abecedarian Project (ABC). †Idle: enrolled in school or working.

and average arrest outcomes for the original participants of Perry by treatment status, both as a function of their age and as a function of the average age of their children. There is greater stability of the family lives for the children of the Perry participants than for control-group members. Parents of these children were more likely to be living together during the childrearing years, and family incomes were higher relative to control-group counterparts. Fathers were less likely to be involved with the criminal justice system and were more likely to provide for their families. Data on the corresponding figures for ABC are not available. It is likely that similar mechanisms also operate in them.

3.7 Intergenerational Relationships

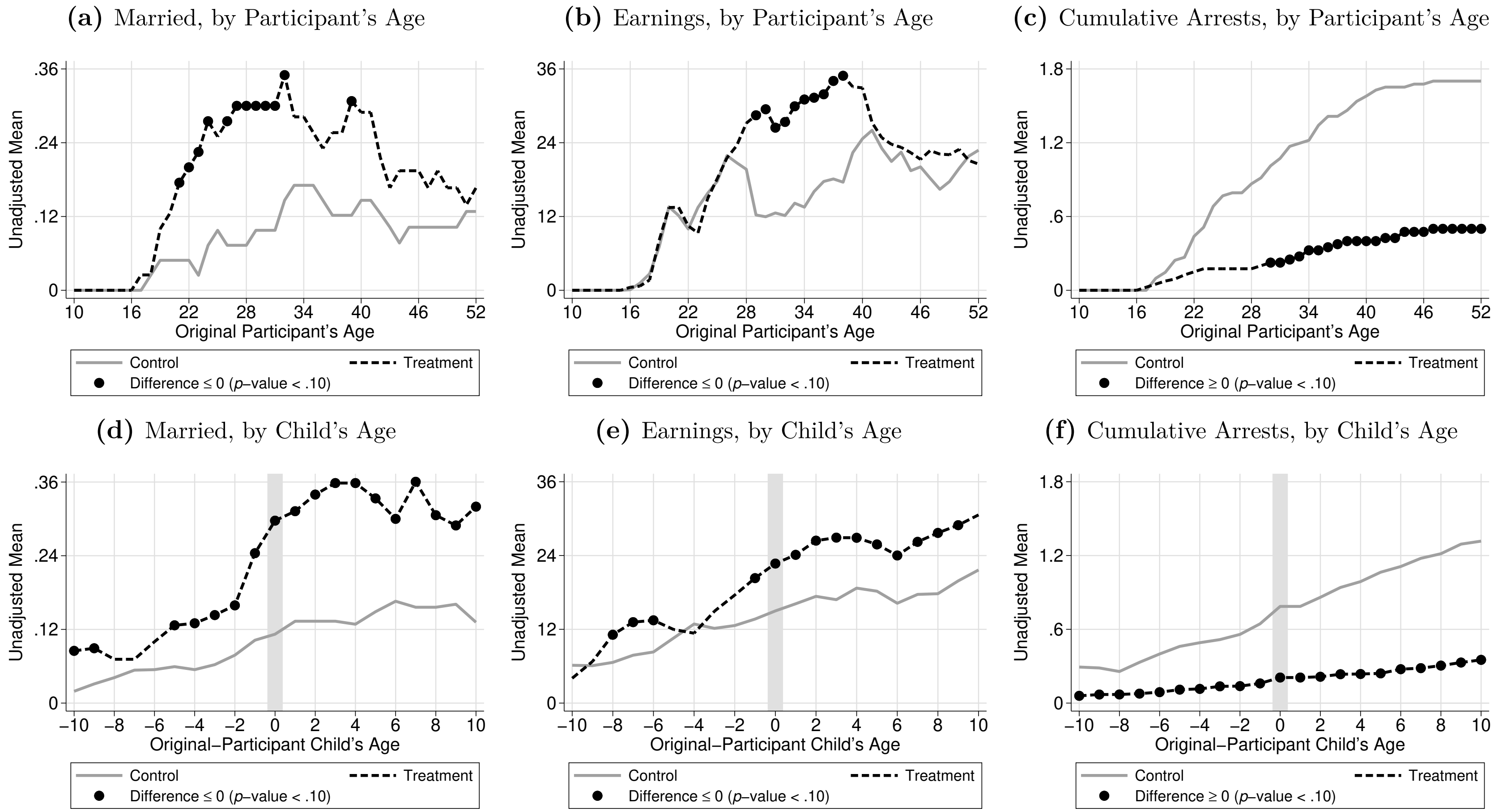
Outcomes of the original participants and their children allow us to estimate the intergenerational relationships shown in Figure 6a. These are the coefficients on parental values of child outcomes (second generation) on the parent's outcome (first generation). We estimate father-son and mother-daughter relationships for crime and education. These are key outcomes connecting the skills of the original participants, significantly boosted by the program, and midlife outcomes such as labor income and marriage stability, which in turn mediate impacts on the second generation.[17] Differences across experimental groups suggest that Perry effectively breaks intergenerational persistence in outcomes. Treatment breaks the extent to which the success of the second generation, as measured by education and crime, depends on the success of their parents. The effect is especially strong for males.

3.8 Net Social Benefits

The evidence shows that Perry and ABC are effective at boosting the life-cycle outcomes of their original participants. Do the benefits of the programs outweigh their costs? Table 3 summarizes the evidence in García et al. (2021) (Perry) and García et al. (2020) (ABC), who conduct cost-benefit analyses accounting for the full social opportunity cost of public

[17]For this exercise, we focus on Perry, for which we observe the relevant outcomes given that its second-generation participants are older.

Figure 5. Original-Participant Marriage, Earnings, and Crime by their Age and by their Children's Age, Perry

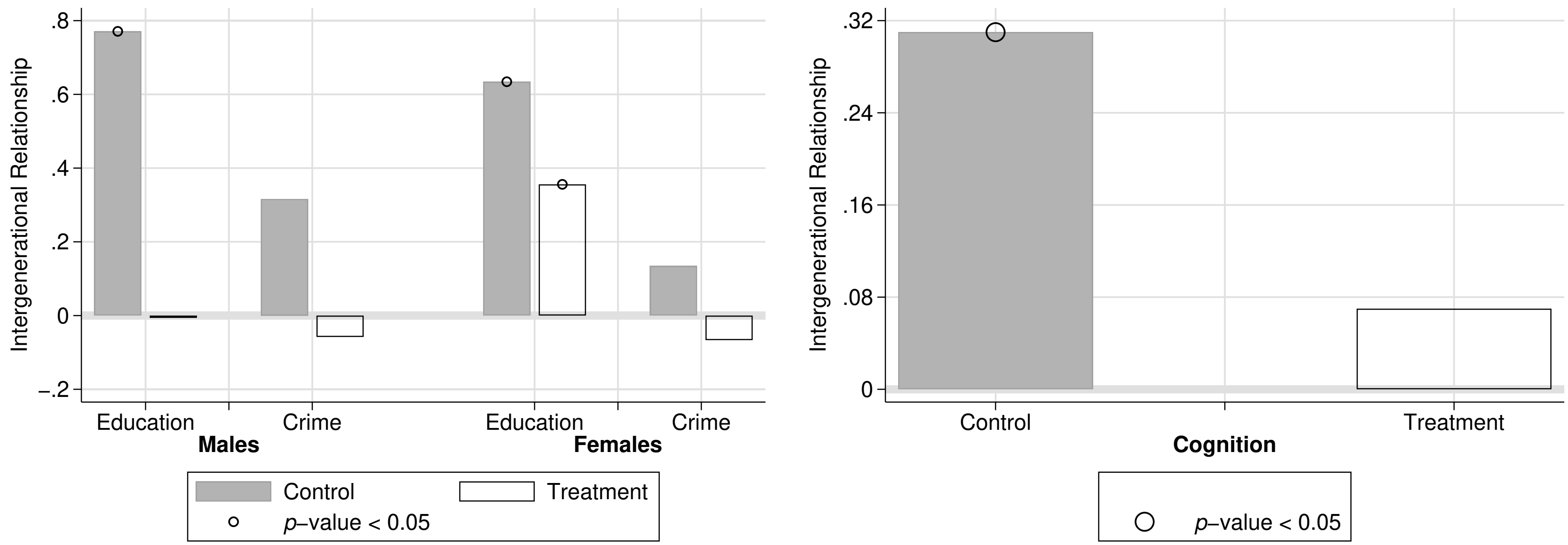


Source: García et al. (2021). **Note (directly from source):** Panel (a) displays the control-group and treatment-group unadjusted means of a married-status indicator by age of the original participants who reported having children. We mark the treatment-group mean when the unadjusted treatment-control mean difference has a permutation p-value less than 0.10. The null hypothesis for the difference is that it is less than or equal to 0. Panel (b) is analogous in format to Panel (a) for annual earnings in 1,000s of 2017 USD. Panel (c) is analogous in format to Panel (a) for cumulative violent misdemeanor and felony arrests. For Panel (c) the null hypothesis for the difference is that it is greater than or equal to 0. Panels (d) to (f) are analogous in format to Panels (a) to (c), but they are plotted by age of the children of original participants. For Panels (d) to (f) the outcomes are first averaged within original participants across up to five eldest children before constructing control and treatment means.

Figure 6. Intergenerational Outcome Relationships, Perry and Preparing for Life

(a) Participants and their Children, Perry

(b) Participants and their Parents, Preparing for Life



Note: Panel (a) displays intergenerational relationships between first-generation (original participants) and second-generation (children of original participants) participants of the Perry Preschool Project. Each relationship is the slope of a regression of the outcome of the average children of the original participants on the outcome of the original participants (i.e., β in $y = \alpha + \beta x + \varepsilon$, where standard notation applies). We estimate male-male relationships (average male children on original male participants) or female-female relationships (average female children on original female participants). We mark relationships when the permutation p-value associated with the null hypothesis that they are less than or equal to 0 is less than 0.05. Panel (b) is analogous in format to Panel (a). It displays the slope of a regression of a measure of cognition of the child participants of Preparing for Life on a measure of cognition of their mothers. For Panel (b) we use the estimates in Doyle (2020), who reports correlations (i.e., coefficients of standardized variables).

expenditure.[18]

Table 3. Summary of Benefit-Cost Analysis of the Perry and ABC

	Perry	**ABC**
Benefits		
Parental Income	N/A	133,326
Education	303	-5,151
Labor Income	68,348	146,672
Crime	88,065	513,420
Health	54,048	63,794
Other	N/A	-21,408
Costs		
Total Program Cost	23,478	105,530
Net Social Benefit (Benefits Less Costs)		
Baseline Program Cost	187,287	725,124
Subtract Deadweight Loss	175,548	672,359
Benefit-Cost Ratio		
Baseline Program Cost	9.0	7.9
Subtract Deadweight Loss	6.0	5.2

Note: Reproduced from García et al. (2021) (Perry) and García et al. (2020) (ABC), after conversion to 2021 dollars. The benefit components of the former are based on observation, except for health which is based on forecast, while the benefit components of the latter are based on forecast. The total cost is observed for both programs. For Perry, there are no monetized benefits for the parents due their potentially improved income given the *de facto* childcare component of the program. Other costs refer to savings due to less expenditure in childcare alternatives, which were not available for control-group Perry participants.

The cost per participant of Perry is 23,478 in 2021 dollars. The cost per participant of ABC is 105,530 dollars. García et al. (2021) and García et al. (2020) monetize the average treatment effects of these programs over the life cycle. The first study is based on annual longitudinal data. The second study combines observational and experimental data and forecasts outcomes over post-sample periods using causal structural models based on Equations (1) to (3). These studies report that the program has an average net social benefit

[18] We do not use the misleading "marginal value of public funds" proposed in Hendren and Sprung-Keyser (2020). This measure is conceptually problematic when used to gauge the net social benefits of programs inclusive of the social cost of public funds. García and Heckman (2022) compare social welfare analyses based on the net social benefits and social welfare analyses based on the marginal value of public funds and show that the latter ignores the social opportunity cost of raising public funds.

per participant (average total benefits less cost per participant) of 175,548 and 672,359, respectively, which differ statistically from 0 at 10% significance level. Estimates of the net social benefits account for the welfare cost of distorting taxation required to fund programs. The corresponding benefit-cost ratios, which differ statistically from 1 at 10% significance level, are 6.0 and 5.2. The source studies show that the reported estimates are robust to extensive robustness checks of the assumptions underlying their estimation.

García et al. (2021) report an additional intergenerational contribution to the net social benefits of 43 thousand dollars per average male child of original participants of Perry and 14 thousand dollars per average female child. They also report an additional intragenerational contribution to the net social benefits of 68 thousand dollars per average male sibling of the original participants and 13 thousand dollars per average female sibling.[19] Although the latter estimates are imprecise, these results indicate that the program generates additional benefits without additional costs that cover the average cost per original participant.

3.9 Evidence on Perry and ABC from Head Start

The evidence discussed thus far strongly supports the effectiveness of Perry and ABC in promoting social mobility and in eliminating gaps across socioeconomic groups. Given that in 2013, 68% of Head Start programs used either the Perry or ABC curricula, evidence on the effectiveness of Head Start is indirect evidence on its source programs. It is also evidence on whether these programs suffer from "voltage effects" (List, 2022), since the scale of Head Start is large: 900,000 children enrolled in it in 2019 (U.S. Department of Health and Human Services, 2021).

The evidence on Head Start shows that it promotes schooling, labor income, employment, health, and that it reduces crime—see the discussion in Appendix A2. Clearly, Perry and ABC can be successfully scaled at least in the form of elements embedded in Head Start.

[19]The p-values for these estimates are as follows. Male child: p-value = 0.006. Female child: p-value = 0.092. Male sibling: p-value = 0.180. Female sibling p-value = 0.359.

4. A Targeted National Implementation of Policies that Promote Social Mobility

We next apply the framework of Section 2 to use the experimentally determined increase in earnings from labor income to evaluate the likely impact of national implementations of Perry or ABC. We assess the impact of national implementations on the mean adult Black-White gap in earnings. This gap is a common metric of racial disparity. Our methodology can be applied to assess impacts of any policy on gaps in other outcomes and among other demographic groups. We abstract from "voltage effects" arising from the degradation of program gains as a consequence of going to scale.

Towards this end, it is useful to generalize the notation in Section 3.2. Let $\boldsymbol{Y}^{d,\ell}(b, \boldsymbol{x})$ be the vector of potential outcomes of interest for an individual of race b with observed pre-program or baseline characteristics $\boldsymbol{x}$ under treatment status d in program ℓ, with $\ell \in \{\text{Perry, ABC}\}$. In this case $b = 1$ if a person is Black and $b = 0$ if they are White; $d = 0$ and $d = 1$ represent the control and treatment counterfactuals, respectively. Perry targeted Blacks. ABC did not, but the area in which it was conducted was such that virtually all child participants were Black. This limits our ability to construct some interesting counterfactuals.

There are many possible comparisons of interest within and across racial groups. For instance, for the group with characteristics $(b, \boldsymbol{x})$, it is of interest to know

$$\text{ATE}^{\ell}(b, \boldsymbol{x}) := \mathbb{E}\left[\boldsymbol{Y}^{1,\ell}(b, \boldsymbol{x}) - \boldsymbol{Y}^{0,\ell}(b, \boldsymbol{x})\right] \text{ for } \boldsymbol{x} \in \mathcal{X} \text{ and } b \in \{0, 1\}, \tag{8}$$

the average gain to persons of race b with characteristics $\boldsymbol{x}$ from participation in ℓ. Defining $\tilde{b}$ as the complement of b in $\{0, 1\}$, the gain of b from ℓ relative to the untreated complement set $\tilde{b}$ for persons of characteristics $\boldsymbol{x}$:

$$\text{ATE}^{\ell}\left(b, \tilde{b}, \boldsymbol{x}\right) := \mathbb{E}\left[\boldsymbol{Y}^{1,\ell}(b, \boldsymbol{x}) - \boldsymbol{Y}^{0,\ell}\left(\tilde{b}, \boldsymbol{x}\right)\right] \text{ for } \boldsymbol{x} \in \mathcal{X}. \tag{9}$$

The gain from participation in ℓ for b for eligibles adds the additional restriction $\boldsymbol{x} \in \mathcal{B}_{\ell} \subseteq \mathcal{X}$.

To illustrate how we combine experimental and observational data to identify parameters of interest note that, for program ℓ, the experiment allows us to identify

$$\text{ATE}^{\ell}\left(b=1, \boldsymbol{x} \in \mathcal{B}_{\ell}\right) := \mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{1,\ell}\left(b=1, \boldsymbol{x}\right)\right] - \mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{0,\ell}\left(b=1, \boldsymbol{x}\right)\right], \tag{10}$$

where $\boldsymbol{Y}^{0,\ell}(b, \boldsymbol{x})$ is the outcome in the untreated state for a person with characteristics $(b, \boldsymbol{x})$ and $\boldsymbol{Y}^{0,\ell}(b, \boldsymbol{x}) = \boldsymbol{Y}^{0}(b, \boldsymbol{x})$ since the general population is not treated. $\text{ATE}^{\ell}(b=1, \boldsymbol{x} \in \mathcal{B}_{\ell})$ is the average gain for the Black population of type $\boldsymbol{x}$ from program ℓ. The average Black-White gap in the untreated state for the group with characteristics $\boldsymbol{x} \in \mathcal{B}_{\ell}$ is

$$\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{0}\left(b=1, \boldsymbol{x}\right)\right] - \mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{0}\left(b=0, \boldsymbol{x}\right)\right], \tag{11}$$

which can be estimated from observational data. The gain for Blacks with characteristics $\boldsymbol{x} \in \mathcal{B}_{\ell}$ relative to Whites with these same characteristics from participating in the program is the sum of the two previous expressions:

$$\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{1,\ell}\left(b=1, \boldsymbol{x}\right)\right] - \mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{0}\left(b=0, \boldsymbol{x}\right)\right], \tag{12}$$

which can be estimated combining experimental and observational data for eligible population, $\boldsymbol{x} \in \mathcal{B}_{\ell}$. In experimental data, we can estimate $\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{1,\ell}\left(b=1, \boldsymbol{x}\right)\right]$.[20] In observational data, we construct eligibility criteria $\mathcal{B}_{\ell}$ and estimate the fraction of eligibles $\Pr\left(b=1, \boldsymbol{x} \in \mathcal{B}_{\ell}\right)$ as well as $\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{0}\left(b=0, \boldsymbol{x}\right)\right]$. Note that we can also estimate $\mathbb{E}_{\boldsymbol{x} \notin \mathcal{B}_{\ell}}\left[Y^{0}\left(b=0, \boldsymbol{x}\right)\right]$ and $\mathbb{E}_{\boldsymbol{x} \notin \mathcal{B}_{\ell}}\left[\boldsymbol{Y}^{0}\left(b=1, \boldsymbol{x}\right)\right]$, which are useful for other comparisons. All these quantities can be integrated over the relevant support of $\boldsymbol{x}$.[21]

[20] Note that, in both Perry and ABC, those eligible to participate in the program accepted the invitation to participate. We thus do not need to condition this probability on d. Randomization determined its value in the experimental data and $d = 0$ for everyone in the observational data.

[21] García et al. (2020) verify that it is possible to construct the eligibility criteria of ABC. They also document full overlap in baseline characteristics ($\boldsymbol{x} \in \mathcal{B}_{\text{ABC}}$) between ABC and the National Longitudinal Survey of the Youth 1979 (NLSY79), which is important for valid sample use. Before conducting the exercises

We provide estimates of the following counterfactuals:

(i) Mean gain for eligible treated Blacks compared to ineligible untreated Blacks (i.e., poor Blacks compared to non-poor Blacks). That is,

$$\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_\ell} \left[\boldsymbol{Y}^{1,\ell} \left(b = 1, \boldsymbol{x} \right) \right] - \mathbb{E}_{\boldsymbol{x} \notin \mathcal{B}_\ell} \left[\boldsymbol{Y}^{0} \left(b = 1, \boldsymbol{x} \right) \right].$$

We can compare this to the observed gap between eligible and ineligible Blacks when neither group is treated:

$$\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_\ell} \left[\boldsymbol{Y}^{0} \left(b = 1, \boldsymbol{x} \right) \right] - \mathbb{E}_{\boldsymbol{x} \notin \mathcal{B}_\ell} \left[\boldsymbol{Y}^{0} \left(b = 1, \boldsymbol{x} \right) \right].$$

(ii) Mean gain for eligible treated Blacks compared to eligible untreated Whites. That is,

$$\text{ATE}^{\ell} \left(b = 1, \tilde{b} = 0, \boldsymbol{x} \in \mathcal{B}_\ell \right).$$

We can compare this to the observed gap between eligible Blacks and eligible Whites when neither group is treated:

$$\mathbb{E}_{\boldsymbol{x} \in \mathcal{B}_\ell} \left[\boldsymbol{Y}^{0} \left(b = 1, \boldsymbol{x} \right) - \boldsymbol{Y}^{0} \left(\tilde{b} = 0, \boldsymbol{x} \right) \right].$$

(iii) All Blacks, eligible or not, compared to all Whites, eligible or not, but only targeting

described in this section, we verify the same is true for the case of Perry (recall that the eligibility criteria of the two programs are similar). The oversampling of disadvantaged individuals in the NLSY79 simplifies the construction of the overlapping samples.

eligible Blacks:

$$\underbrace{\Pr\left(b=1,\boldsymbol{x}\in\mathcal{B}_\ell\right)\mathbb{E}_{\boldsymbol{x}\in\mathcal{B}_\ell}\left[\boldsymbol{Y}^{1,\ell}\left(b=1,\boldsymbol{x}\right)\right]+\Pr\left(b=1,\boldsymbol{x}\notin\mathcal{B}_\ell\right)\mathbb{E}_{\boldsymbol{x}\notin\mathcal{B}_\ell}\left[\boldsymbol{Y}^{0}\left(b=1,\boldsymbol{x}\right)\right]}_{\text{Mean of Total Black Population}}$$

$$-\underbrace{\Pr\left(b=0,\boldsymbol{x}\in\mathcal{B}_\ell\right)\mathbb{E}_{\boldsymbol{x}\in\mathcal{B}_\ell}\left[\boldsymbol{Y}^{0}\left(b=0,\boldsymbol{x}\right)\right]+\Pr\left(b=0,\boldsymbol{x}\notin\mathcal{B}_\ell\right)\mathbb{E}_{\boldsymbol{x}\notin\mathcal{B}_\ell}\left[\boldsymbol{Y}^{0}\left(b=0,\boldsymbol{x}\right)\right]}_{\text{Mean of Total White Population}}.$$

We can compare this to the observed (pre-treatment) Black-White gap (Equation (11) for $\boldsymbol{x}\in\mathcal{X}$).

An alternative to (iii) would be to treat eligible Whites as well as Blacks. However, we do not observe $\boldsymbol{Y}^{1,\ell}\left(b=0\right)$ in the experimental data. Long-term follow-ups of the Head Start Impact Study, which are not yet available, would allow us to estimate this alternative.

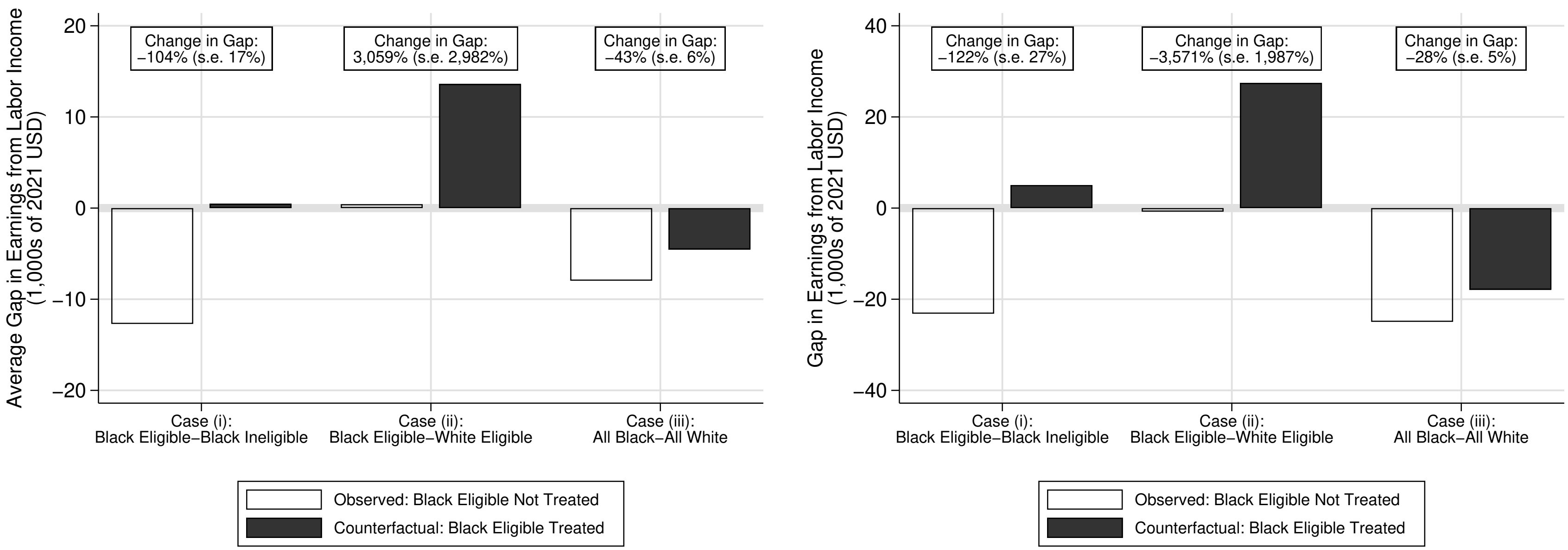
Following García et al. (2020), we use the National Longitudinal Survey of Youth 1979 (NLSY79 Bureau of Labor Statistics, 2015), a nationally representative sample whose individuals were 14 to 22 years old in 1979, to estimate observational counterparts required to form (i) to (iii). We focus on earnings from labor income between ages 20 and 40 when we study Perry and earnings from labor income at age 45 for ABC. These outcomes are summarized in Table 1 for the experimental samples. A national implementation of Perry would eliminate the gap between poor (eligible) and non-poor (ineligible) Blacks. In a no-implementation scenario, eligible Blacks and Whites have essentially the same average earnings. A national implementation targeting Blacks would thus put them at a sizable average advantage that roughly amounts to the treatment effect reported in Table 1. A national implementation of Perry would shrink the overall Black-White gap by 43%. The results for ABC are similar, which is sensible given the similarity across programs in eligibility criteria and treatment effects.[22]

[22] Duncan and Sojourner (2013) estimate that the gap in "income explained by cognitive and achievement scores" would be eliminated upon a national implementation of the Infant Health and Development Program, a program modeled after ABC. This exercise focuses on income predicted solely from cognitive test scores. As discussed here and elsewhere, cognitive scores explain a relatively small fraction of the impact of early childhood interventions. In addition, plugging experimental estimates of skill enhancements into

Figure 7. Gaps in Annual Earnings from Labor Income After National Implementations of Perry and ABC

(a) Earnings between Ages 20 and 40 after Implementing Perry

(b) Earnings at Age 45 after Implementing ABC



Note: Panel (a) first displays the observed average gap in average earnings from labor income between ages 20 and 40. It then displays a counterfactual gap when Black eligible individuals participate in the Perry Preschool Project. The panel then displays similar comparisons between eligible Blacks and Whites and the overall gap. In all three counterfactual cases, only Black eligible individuals participate in Perry. Estimates combine the experimental data from Perry and the National Longitudinal Survey of the Youth 1979. Details on the method for implementing the counterfactual scenarios are in this section. Earnings are in 2021 US dollars. The change in the gap is the percentage change in the gap between the observed and counterfactual scenarios. Unlike the rest of the paper in which permutation inference is used, this section uses the bootstrap procedure in García et al. (2020), which considers sampling variation in both the experimental and observational data. Panel (b) is analogous in format to Panel (a) for the Carolina Abecedarian Project, using earnings from labor income at age 45.

4.1 Comparing Methods for Forecasting Out-of-Sample Treatment Effects

Few early childhood education programs have long-term follow-ups. Forecasting long-run treatment effects from short-term evaluations is a major challenge. García et al. (2020) develop and implement a principled approach using economic models.[23] Building on Heckman et al. (2013), they develop a three-step process for forecasting out-of-sample future outcomes. They use Equation (2) to measure the cognitive and non-cognitive skills, $\boldsymbol{S}_a$, of treatments and controls. They plug the estimated skills into estimates of Equation (1) to recursively forecast future skills. They use Equation (3) to forecast future behaviors (including earnings) based on estimated skills. They estimate models on national observational datasets. Partially testable structural invariance assumptions ensure portability of estimated functions across samples and time, after controlling for differences in environments. When satisfied, they make empirical models fitted on non-experimental data empirically grounded vehicles for forecasting. They combine experimental and non-experimental data and account for selection bias that plagues the biometric approach (as in Athey et al., 2019). Using invariant relationships also circumvents cohort effects that arise in using current and retrospective data to predict future relationships.

Using these tools, García et al. (2020) forecast earnings, as well as other outcomes like health and criminal behavior. We compare the performance of their method to a general forecasting practice, which has multiple versions in the literature. The procedure aims to forecast the life-cycle treatment effect on earnings from a single observation of earnings during adulthood and a cognitive-test score observed during childhood.[24] Many authors estimate the relationship between earnings at an adult age a and an early-life cognitive-test

relationships fit on non-experimental observational data is problematic. The experimental are causal. The non-experimental data are not, although with some care not usually taken in the literature they can be made so. García et al. (2020) is an exception.

[23]They develop this approach to extrapolate treatment effects of ABC after age 30, when they last observe program participants.

[24]See Chetty et al. (2011) and Kline and Walters (2016) for versions of this practice. Recent studies emulating their procedures include Cascio (2021) and Ganimian et al. (2021). See García et al. (2021) for additional discussion.

score observed at age a':

$$\text{Earnings}\,(a) = \alpha + \beta \cdot \text{Test Score}\,(a') + \varepsilon_a, \tag{13}$$

where standard notation applies. The approach is sometimes more general than this and non-parametric methods are used. Relationship (13) or its non-parametric version is estimated in non-experimental data, which likely yields a biased estimate of β for usual reasons for ability bias.[25] The estimated "return" to an additional unit of Test Score $(a'\ )$ is assumed to apply in the experimental data and to hold annually throughout the life cycle. An experimentally estimated gain in Test Score (a'), $\Delta\,(a')$, is used to estimate the life-cycle discounted treatment effect attributable to the experiment, from the age in which earnings are observed to retirement, $\bar{a}$. The estimator is:

$$\text{PV} := \beta\Delta\,(a') \cdot \sum_{j=a}^{\bar{a}} \frac{1}{(1+r)^{(j-\underline{a})}}, \tag{14}$$

where r is the discount rate and $\underline{a}$ is the age of child participants when the program starts.

To evaluate this approach, we plug in an estimate of β from the NLSY79 and an estimate of $\Delta\,(a')$ from Perry to estimate PV.[26] In this example, we let $a' = 5$ and $a = 30$, thus illustrating a life-cycle prediction based on observing earnings at age 30 and a test score at age 5. We assume a retirement age of 65 and a discount rate of 0.03, as in García et al. (2021). We obtain an estimate of PV of $103,159$ (s.e. $16,203$) 2021 dollars. Relying on actual observation, García et al. (2021) obtain a counterpart estimate of the life-cycle treatment effect on earnings of 68,354 (s.e. 35,820). Relying on the method of García et al. (2020), they report an alternative estimate of 69,987 (s.e. 45,288). The prediction method based on

[25] See Heckman et al. (2008, 2006) for general discussions.

[26] Our estimate of β in the NLSY79 is 10,769.89 (s.e. 259.95). That is, a one-standard deviation in increase in the cognitive-test score at age 5 is associated with an increase of earnings from labor income at age 30 of 10,769.89 2021 dollars. Our standardized estimate of $\Delta\,(5)$ in the Perry sample is 0.85 (s.e. 0.13). That is, Perry increases a cognitive-test score (Stanford-Binet) at age 5 by 0.85 standard deviations. Standard errors for these calculations are calculated as explained in Figure 7.

solely early-life cognition overestimates the life-cycle treatment effect, likely because it suffers from ability bias in the coefficient on the test score in the earnings equation arising in non-experimental samples. The structural approach, based on well-known models of earnings, performs remarkably well when applied to Perry.

5. Isolating Components of Programs that Successfully Promote Social Mobility

A recurring feature of successful programs is the enhanced engagement of the parents in the life of the child induced by them. In an attempt to isolate this component of successful child development from the other features of omnibus programs, we study the benefits of home-visiting programs that specifically aim to improve parent-child interactions. Doing so we partially break our rules for relying on primary data sources. We have primary data for some programs but for other programs we rely on secondary sources, which we attempt to screen carefully.[27]

We start by describing two pioneering home-visiting programs: Jamaica Reach Up and Learn—henceforth *Jamaica*—and Preparing for Life (PFL), on which we have access to primary data. *Jamaica* is a prototypical and influential home-visiting program targeting disadvantaged children (Grantham-McGregor et al., 1997). It was implemented in Kingston, Jamaica and has frequent follow-ups through age 31 that indicate sustained impacts. PFL was implemented in disadvantaged neighborhoods of Dublin, Ireland (Doyle, 2020). It has follow-ups available through age nine, which indicate positive impacts on cognitive and non-cognitive skills as well as reading and mathematical achievement. We also describe recent programs patterned after *Jamaica* implemented in China (Heckman et al., 2020; Sylvia et al., 2021), Colombia (Attanasio et al., 2014), and India (Andrew et al., 2020). The content and delivery of these recent programs is based on *Jamaica* but adapted to the contexts of application.

[27]We analyze primary data for Perry, ABC, Jamaica Reach Up, and ChinaReach. Our research partner, Orla Doyle, conducted and analyzed primary data on PFL. We rely on secondary sources for other programs.

After describing these programs, we compare their impacts to those of Perry and ABC, the omnibus programs analyzed in Section 3. We use newly constructed and harmonized data from Perry and ABC and the sources cited in Tables 4 and 5 to provide a side-by-side comparison based on comparable measures across programs. We document that many impacts of home-visiting programs on home environments, cognitive and non-cognitive skills, and reading and math achievement through adulthood are close to those of omnibus programs, as are impacts on education, earnings, criminal behavior, alcohol use, and drug use.

5.1 Jamaica Reach Up and Learn

This program focuses on home visiting. In its original implementation, home visitors had similar levels of education as those of the mothers of the child participants, who were disadvantaged (roughly 76% of the mothers had nine years of education or less). They visited the mother-child participant pairs for one hour per week. They taught mothers how to play with their children, using a program that had an ascending series of tasks, in terms of developmental stages, which were created based on playbooks made from locally available materials. Mothers were encouraged to interact with their children. In addition, nutritional supplements were given to some participants.

This program was evaluated by a randomized trial with a sample of 129 stunted children who were between nine and 24 months old. The program individualized nutritional and cognitive stimulation. There were four intervention groups: (i) no intervention; (ii) nutritional intervention only; (iii) cognitive stimulation only; and (iv) both nutritional and cognitive interventions. Studies evaluating *Jamaica* find that the nutritional component had no impact. Thus, they pool groups (i) and (ii) into a control group and groups (iii) and (iv) into a treatment group. We follow this convention when presenting results for *Jamaica* below.

Table 4. Features of Omnibus and Original Home-Visiting Programs

	Omnibus Programs		**Pioneering Home-Visiting Programs**	
	Perry	**ABC**	***Jamaica***	**Preparing for Life**
***Panel a.* Features**				
Setting	Ypsilanti, Michigan	Chapel Hill, North Carolina	Poor neighborhoods in Kingston, Jamaica	Disadvantaged neighborhoods of Dublin, Ireland
Year of program start	1962	1972	1986-1987	2008
Annual cost per child participant	9,391 (2021 USD)	21,106 (2021 USD)	862 (2021 USD)	2,363 (2021 USD)
Sample at baseline	65 control; 58 treatment	56 control; 58 treatment	65 control; 64 treatment	118 control; 115 treatment
Socioeconomic characteristics of participants	Disadvantage by several measures, which determined eligibility	Disadvantage by several measures, which determined eligibility	Generally disadvantaged; all child participants were stunted at baseline	Generally disadvantaged (high unemployment, low levels of education)
Child age at start of program	3 years	0 (program started at birth)	9 to 24 months	When mothers pregnant
Program duration	2 years	5 years	2 years	5 years
Education of home visitors	College or teaching degree	No visits were implemented. Staff in childcare centers were a mix of HS graduates and education-certified staff	All had at least secondary education	All had a college degree
Experience required for home visitors	Most staff had certification or experience in education	Education-certified staff was present in childcare centers	None; received 8 weeks of mandatory training on child development	Intensive two-day initial training on program manual and child-development matters. Follow-up to this training throughout the following six months.
Education of mothers at baseline	Most mothers did not have high school completed	Most mothers did not have high school completed	Only 24% had more than 9 years of education	Relatively low; 30% had less than 16 years
Frequency of home visits	Weekly during the school year; one-hour per session	No visits were implemented	Weekly; one-hour per session	Fortnightly; one-hour per session
***Panel b.* Home-Environment Measures**				
Child age at measurement	0.5 to 4.5	3 to 5	Not available	3, 5, and 9
Measures available	Parental Attitude Research Instrument	HOME Inventory		Age 3: HOME inventory; Age 5: home learning-environment index; Age 9: parent-involvement index
***Panel c.* Very Early-Life Skill Measures**				
Child age at measurement	5	5	3 to 4	3 and 5
Measures available	Cognition: Stanford-Binet IQ Test; Non-cognitive: not available	Cognition: Stanford-Binet IQ Test; Non-cognitive: not available	Cognitive: Griffith Mental Development Scale (performance scale) Non-Cognitive: not available	Cognitive (5): BAS (spatial, pictorial, verbal sections) . Non-Cognitive (3 and 5): internalizing, externalizing, pro-social, and peer problem behavior inventories
***Panel d.* Early-Life Skill Measures**				
Child age at measurement	7 to 9	7 to 9	7 to 8 and 11 to 12	9
Measures available	Cognition: Stanford-Binet IQ Test; Non-cognitive: Problem Solving Inventory (externalizing behavior)	Cognition: McCarthy Scale of Children's Abilities; Non-cognitive: not available	Cognition: Stanford-Binet IQ Test (age 7 to 8) and Wechsler Intelligence Scale for Children (11 to 12)	Cognition: BAS Test (spatial, pictorial, verbal sections). Non-cognitive: internalizing, externalizing, pro-social, and peer problem behavior inventories

Sources: Section 3 for Perry and ABC. Doyle (2016), Doyle (2020), and Doyle (2021) for PFL. Grantham-McGregor et al. (1997), Walker et al. (2000), Walker et al. (2005), Walker et al. (2011), Gertler et al. (2014), Heckman et al. (2020), Gertler et al. (2021), and Walker et al. (2022) for *Jamaica*.

5.2 Preparing for Life

Preparing for Life, a home-visitation program similar to *Jamaica*, was implemented in disadvantaged neighborhoods of Dublin, Ireland. It was evaluated by a randomized trial that assigned 118 children to a low-dose treatment group (control) and 115 children to a high-dose treatment group (treatment). The treatment group received an average of about one home visit per month, which lasted for an average of one hour over the course of the first five years of the lives of the enrolled children. These home visits, provided by trained visitors, were the main component for the treatment group. The role of the visitors was to mentor parents in promoting child development.[28] The treatment-group parents also received baby massage classes during the first year of the program and parent training sessions during the third year. Both the treatment and control groups received developmental toy and book packs, two framed professional photographs of the target child, access to a support worker to assist in non-parenting related issues (control only), invitations to attend workshops on healthy eating and stress control, and PFL specific social events.

Panel a. of Table 4 provides additional details on *Jamaica* and PFL. Both programs targeted disadvantaged children and had much lower annual cost per child than Perry and ABC. *Jamaica* employed home visitors who had as little as secondary education while PFL employed home visitors with college degrees. However, none of the programs required previous experience in providing child-development services. Successful implementation at low cost and lack of experience requirements are relevant for scalability. Another important aspect affecting scalability is the programs' low dose. In *Jamaica*, mother-child participant pairs were treated one hour a week. In PFL, they were treated fortnightly. A general feature of *Jamaica*, PFL, and the programs patterned after *Jamaica* is that home visitors were trained and then supervised. They had supervisors who enforced implementation guidelines

[28] The mentoring was based on three core principles: (i) providing knowledge and active guidance on appropriate parenting techniques; (ii) helping parents to identify and promote children's developmental milestones; and (iii) encouraging parents to provide greater stimulation to their children, particularly during infancy.

and helped with problems that occurred during home visits.

5.3 Programs Patterned After Jamaica

Jamaica's success and plausible adaptability to other environments prompted the interest of academicians and policymakers in replicating the program in diverse settings. Examples which have been evaluated are as follows.

ChinaReach. ChinaReach is a large-scale ongoing program. Its replicates *Jamaica* in a very distinct setting. It tests *Jamaica*'s replicability at large scale. It was evaluated by a randomized trial in Huachi County of Gansu, one of the poorest areas in western China. At baseline, 852 children between zero and twenty months old were assigned to control and 715 to treatment. Follow-up data through endline, about 22 months after the initiation of the intervention, are available.

Home visitors have the same level of education as that of the mother (i.e., ten years of education on average). They were trained by supervisors with greater levels of schooling. Home visitors made weekly visits and provided a one-hour session of parenting or caregiving support based on the *Jamaica* protocols. Like *Jamaica*, ChinaReach emphasized teaching and encouraging caregivers to talk with their children through playing games, making toys, singing, reading, and storytelling to stimulate the child's cognitive, language, motor, and socio-emotional skill development. Home visitors were supervised so that they followed guidelines on a series of tasks for children organized across a progression of levels based on an influential protocol developed by Uzgiris and Hunt (1975). Supervisors accompanied home visitors once a month to each enrolled household and observed how the intervention was administered.

Other Successors. Programs patterned after *Jamaica* have been implemented in Colombia, China (other than ChinaReach), and India. All of these programs were evaluated by randomized trials. Their implementers worked with local governmental and non-governmental agencies to adapt them to their contexts, though all of them were based on the principles

Table 5. Features and Available Skill Measures for Home-Visiting Programs Patterned after *Jamaica*

	ChinaReach	**China**	**Colombia**	**India**
***Panel a.* Features**				
Setting	Villages in Huachi County of Gansu, China	Villages Shaanxi, China	Semi-urban municipalities in central Colombia	Urban disadvantaged neighborhoods in Cuttack County of Odisha, India
Year of program start	2015	2014	2010	2013
Annual cost per child participant	602 (2021 USD)	Not available	380 (2021 USD)	175 (2021 USD)
Sample at baseline	852 control; 715 treatment	296 control; 212 treatment	626 control; 635 treatment	212 control; 209 treatment
Socioeconomic characteristics of participants	Generally disadvantaged; > 70% of household participants living in cave dwelling	Generally disadvantaged; ∼26% participant households qualified for minimum-living standards social program	Eligible for a social programs targeting households belonging to the poorest 20% in the population	Generally disadvantaged; ∼50% of participant households below the poverty line
Child age at start of program	0 to 20 months	18 to 30 months	12 to 24 months	10 to 20 months
Program duration	∼22 months	6 months	18 months	18 months
Education of home visitors	10 years (average)	Most completed at least community college; 29% completed college	8.5 years (average)	26% did not have high-school; 74% had at least high-school
Experience required for home visitors	None	None	None	None
Education of mothers at baseline	10 years (average)	∼27% had less than 9 years of education; ∼73% had more than 9 years of education or more	7.5 years (average)	6.7 years control (average); 8.1 years treatment (average)
Frequency of home visits	Weekly; one-hour per session	Weekly; one-hour per session	Weekly; one-hour per session	Weekly; one-hour per session
***Panel b.* Home-Environment Measures**				
Child age at measurement	2 to 3.5	2 to 3.5	2.5 to 3.5	1.5 to 3
Measures available	HOME inventory	Items similar to those observed in the HOME inventory	Items similar to those observed in the HOME inventory, classified as either time or material resources	Items similar to those observed in the HOME inventory
***Panel c.* Very Early-Life Skill Measures**				
Child age at measurement	2 to 3.5	2 to 3.5	2.5 to 3.5	1.5 to 3
Measures available	Cognitive: Denver Developmental Screening Test (language and cognition sections). Non-cognitive: Denver Developmental Screening Test (socio-emotional section)	Cognitive: Bayley Mental Developmental Index (cognitive section) for cohort 1; Griffith Mental Development Scale (performance scale) for cohort 2. Non-cognitive: ASQ social problems inventory for both cohorts	Cognitive: Bayley Mental Developmental Index (cognitive section). Non-cognitive: IQQ and ECBQ inventories	Cognitive: Bayley Mental Developmental Index (cognitive section). Non-Cognitive: not available

Sources: Heckman et al. (2020) for ChinaReach. Sylvia et al. (2021) for China. Attanasio et al. (2020) for Colombia. Andrew et al. (2020) for India.

developed for *Jamaica*. Table 5 describes these programs and shows how they compare to ChinaReach. We do not have access to primary data on them and thus rely on secondary sources. The four programs are similar: (i) they targeted disadvantaged households; (ii) they had a low annual cost per child; and (iii) they did not require that the home visitors had experience in providing child-development services. Further, the programs in Colombia and India show that programs patterned after *Jamaica* can be implemented by employing visitors with relatively low levels of education—and not that different from the levels of education of the mother participants.[29]

Successful implementation of these programs in different contexts and with different sample sizes, including relatively large samples, indicates that scaling *Jamaica* is feasible. In addition to the programs in Table 5, other programs emulating *Jamaica* have been implemented in Bangladesh, Brazil, and Zimbabwe. Results from them are preliminary and we do not discuss them here. However, all of them show promising impacts on very early-life cognition (see, e.g., Brentani et al., 2021; Tofail et al., 2013).

5.4 Evidence on Life-Cycle Outcomes and Comparison to Omnibus Programs

Panel b. of Tables 4 and 5 describes the measures of the home environment observed for each program. They are based on either the HOME inventory described in Section 3 or very similar measures.[30] Recall that these inventories aim to capture the quality of the environment in which children grow up, inclusive of parental investment and parent-child interactions. Unless noted otherwise, impacts are measured at the end or soon after the end of the programs as in Figure 1. This limits analysis of life-cycle impacts for some programs.

[29]The programs in China employed visitors with relatively high levels of education in a context where education levels are relatively high due to national mandates. Note that even when the targeted households are disadvantaged, the mother participants in the Chinese programs had relatively high levels of education.

[30]For ABC, ChinaReach, and PFL at age 3, the HOME inventory is observed. For China, Sylvia et al. (2021) report that the inventory observed contains items very similar to those of the HOME (e.g., number of times per week family reads, sings, goes out, spends time watching television with the child). For Colombia, Attanasio et al. (2020) report observation of very similar items. For India, the items are not listed but Andrew et al. (2020) report that the measures of the home environment observed are similar to those observed for China and Colombia.

The measures and the units in which they are reported are known and allow for a comparison of impacts across programs.[31] For Perry and ABC, we use the measures summarized in Figure 2, which, as before, are standardized by subtracting control-group means and dividing by control-group standard deviations.[32] The estimates shown in Figure 8a indicate that Perry and ABC improve home environments by 0.3 standard deviations from a no-treatment counterfactual mean of 0. The measures observed for other programs are standardized in a similar fashion.

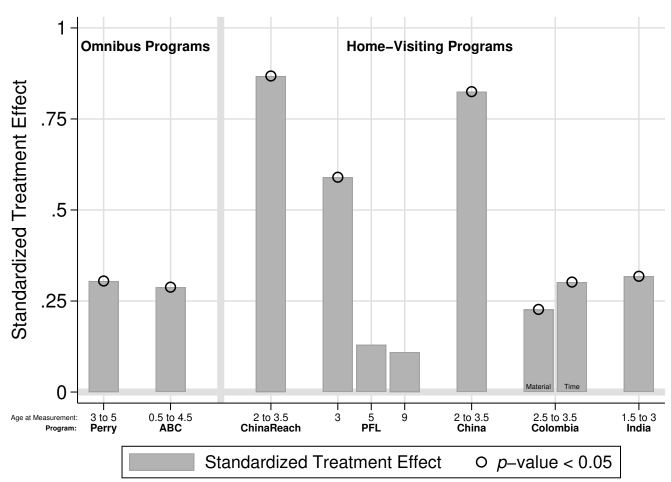
Unfortunately, we do not observe impacts on home environments for *Jamaica*. Later replications, however, show substantial impacts. The two programs emulating *Jamaica* in China generated larger impacts than Perry and ABC, improving home environments by 0.8 standard deviations. These two programs were implemented in disadvantaged rural and semi-rural areas with substantial scope for improvement. The measures for *Colombia* enable us to report separate impacts on material resources for child development and time invested by parents. Both improve by about 0.25 standard deviations after the program. The impact of the program in India is similar in magnitude. There is also a sizable initial impact by PFL, which then diminishes. Only the measure observed at age 3 for PFL is comparable to the HOME scores observed for other programs. The evidence indicates that programs emulating *Jamaica* are at least as effective as Perry and ABC at improving parenting and parent-child interactions, as measured by how the home environment shifts relative to that of no-treatment families. In many cases, the home-visiting programs have a larger impact than the omnibus programs.

[31] We either use a latent variable that summarizes multiple items, like in Section 3, or an index of multiple items. These indices differ from latent variables in that they combine items assigning weights provided by the designers of the inventories. In many cases, these weights are the same for each item in the inventory and, thus, the indices are arithmetic averages across items. When constructing latent variables, weights are estimated by exploiting covariances across items.

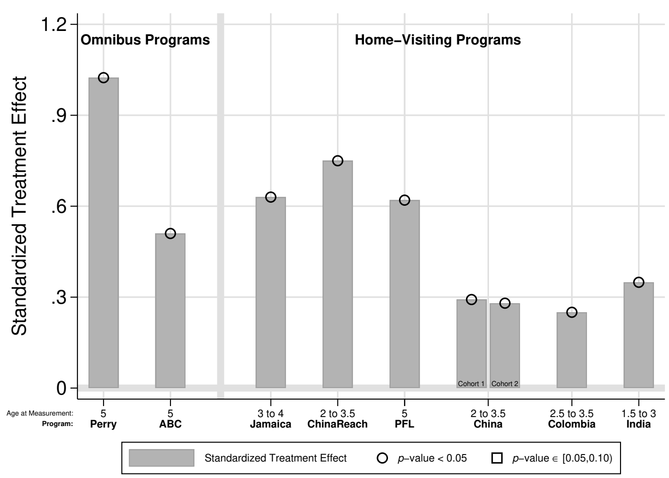
[32] When an age range for the age at measurement is reported in Figures 8 to 10, we either (i) observe one inventory for each child participant and the age of child participants belongs to an age range or (ii) we observe inventories across different ages for each child and construct or observe latent variables or indices based on them (i.e., inventories observed across ages within an age range are used to form the latent variables or indices analyzed).

Figure 8. Impacts on the Home Environment and Very Early-Life Skills, Omnibus and Home-Visiting Programs

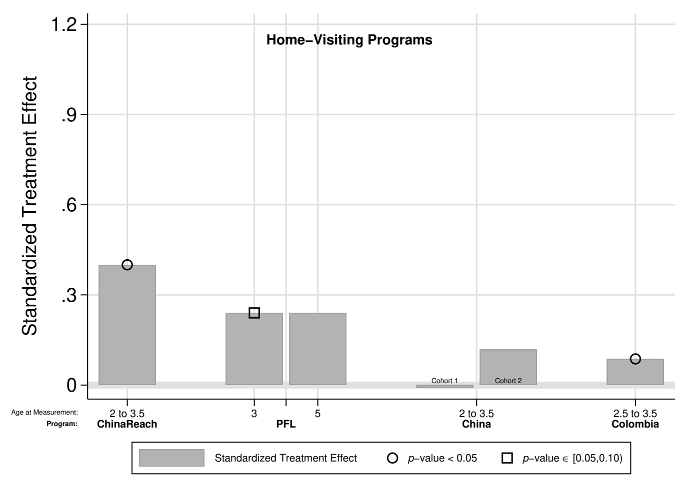
(a) Home Environment



(b) Cognitive Skills



(c) Non-Cognitive Skills



Note: Panel (a) displays program impacts on measures of the home environment. For Perry and ABC, the measures are described in Section 3. Recall that both measures are standardized by subtracting the control-group mean and dividing by the control-group standard deviation. The measures for the rest of the programs are standardized similarly. For all programs except for PFL, we report treatment effects (estimates of treatment-control mean differences). The impacts reported for PFL are effect sizes. We mark impacts when the *p*-value associated with the null hypothesis that they are less than or equal to 0 is less than 0.05. The measures are described in Tables 4 and 5. Panels (b) and (c) are analogous in format to Panel (a) for measures of cognitive and non-cognitive skills. *Cohorts* in "China:" For cognitive and non-cognitive skills, Sylvia et al. (2021) report separate results for two cohorts within their sample, while for the home environment they report results pooling the two cohorts. We display results as reported by them.

The c. panels of Tables 4 and 5 describe the measures of very early-life cognition for the omnibus and home-visiting programs. They are all based on inventories designed to assess child development. They are constructed and standardized like the home environment measures in Figure 8a. Impacts are also measured at the end of the programs. Except for the very large impact observed for Perry, impacts across programs are comparable and similar to the impacts for ABC, the other omnibus program.

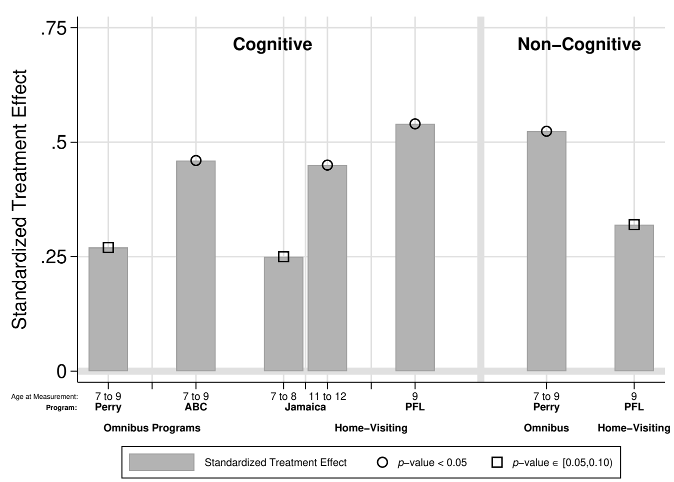
We do not observe non-cognitive skills at age five or before for ABC and Perry. We observe measures for some home-visiting programs and show comparable impacts in Figure 8c. While the measures of these impacts vary, they all aim to capture the degree of "positive personality" of children, as described in Section 3. Figure 8c indicates a consistently positive impact on non-cognitive skills. Non-cognitive skills are more difficult to measure when children are young (i.e., items and behaviors used for measurement are more difficult to observe for very young children than for older children). Reported impacts could thus underestimate true effects on non-cognitive skills. Indeed, the longer-term impacts previously reported on achievement and longer-term outcomes suggest a large impact on non-cognitive skills, which are the building blocks of the latter outcomes.

Figure 9a shows that the impact on cognitive skills persists for omnibus and home-visiting programs through age 12. It also shows that the impacts of the omnibus programs on non-cognitive skills persist through that age. We do not observe impacts on non-cognitive skills for ages 7 to 12 for the Chinese, Colombian, and Indian programs. Child participants are still too young for such follow-ups. For PFL, the impact on non-cognitive skills is not far off from the impact found for Perry. Heckman et al. (2013) show the impacts on skills found in Perry mediate the impact on long-term outcomes such as criminal behavior and employment.

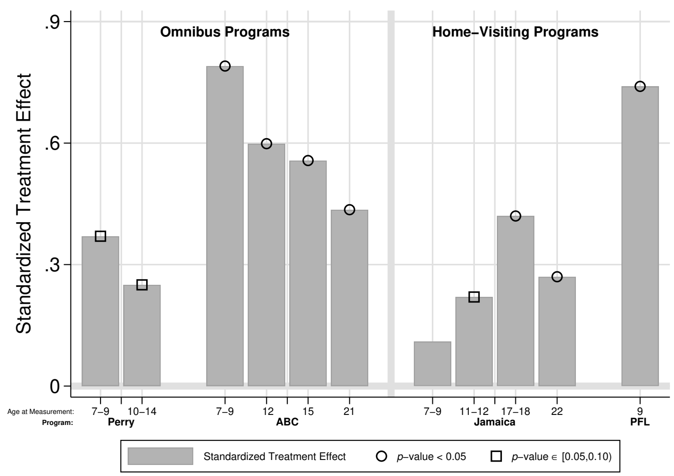
The evidence across these home-visiting interventions indicates that the mechanisms generating successful omnibus programs are also at work in home-visiting programs. They

Figure 9. Impacts on Early-Life Skills and Achievement, Omnibus and Home-Visiting Programs

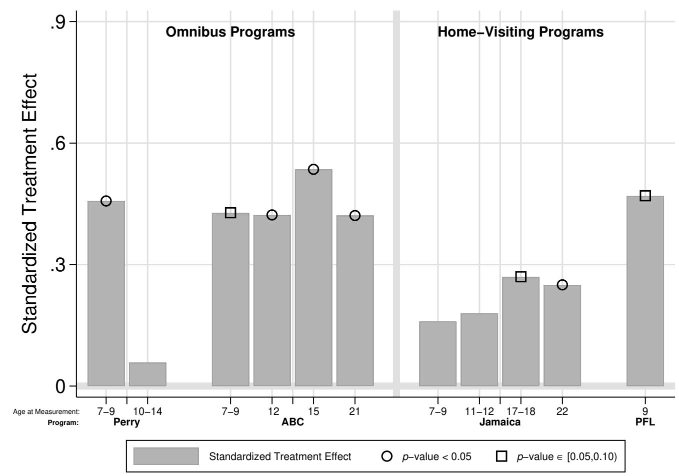
(a) Skills



(b) Reading



(c) Mathematics



Note: Panel (a) displays impacts on measures of early-life cognitive and non-cognitive skills. For Perry and ABC, the measures are standardized as described in Section 3—by subtracting the control-group mean and dividing by the control-group standard deviation. The measures for the rest of the programs are standardized similarly. For all programs except for PFL, we report treatment effects (estimates of treatment-control mean differences). The impacts reported for PFL are effect sizes. We mark impacts when the p-value associated with the null hypothesis that they are less than or equal to 0 is less than 0.05 or between 0.05 and 0.10. The measures are described in Tables 4 and 6. Panels (b) and (c) are analogous in format to Panel (a), using the reading and math achievement measures.

activate parental investment and parent-child interactions and, thereby, impact the formation of child skills. Impacts across the two types of programs are comparable and sustained through early life. A study of impacts on scores on achievement tests further corroborates this point and extends the age range of the evidence. Achievement in reading and mathematics builds on discipline, conscientiousness, and other skills beyond cognition. Measures of achievement are informative because they capture skills beyond cognition while they are obtained through tests that are comparable across and within programs throughout the life cycle.[33] Panel a. of Table 6 describes the measures of achievement observed for omnibus and home-visiting programs. Figures 9b and 9c show that impacts on reading and math achievement are sustained through age 22.

Do these impacts on skills last into adulthood and build better life outcomes? Figure 10a shows that the impacts of Perry, ABC, and *Jamaica*, for which we observe adult cognitive and non-cognitive skills, are sustained and closely aligned. More importantly, impacts on years of education, employment, violent behavior, alcohol use, and drug use, based on comparable measures, indicate similar long-term outcomes for programs that substantially vary on average total cost per child—Perry (23,478 dollars of 2021), ABC (105,530 dollars of 2021), *Jamaica* (1,724 dollars of 2021).

Figure 6b indicates that home-visiting programs replicate what is illustrated for omnibus programs in Figure 6a. For PFL, the after-program child-mother correlation of cognition in the control group is 0.31 (p-value = 0.02). Treatment decreases this correlation to 0.07 (p-value = 0.56). Home-visiting programs break the intergenerational correlation of important life outcomes, thereby interrupting cycles of family disadvantage. While formal cost-benefit analyses are not yet available, home-visiting programs emerge as cost-effective alternatives for achieving the positive impacts of Perry and ABC.

[33] See Borghans et al. (2011, 2016).

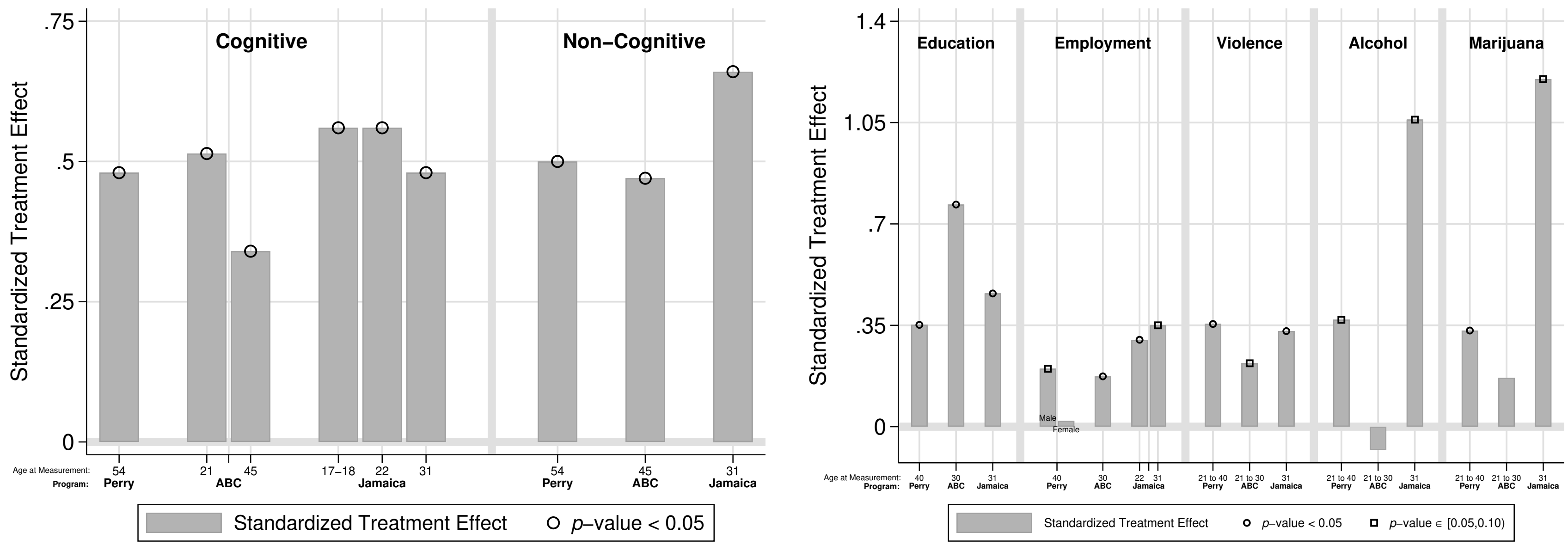
Table 6. Available Adult Skill Measures and Outcomes for Omnibus and Original Home-Visiting Programs

	Omnibus Programs		**Pioneering Home-Visiting Programs**	
	Perry	**ABC**	***Jamaica***	**Preparing for Life**
Panel a. **Achievement Measures**				
Child age at measurement	7 to 9 and 10 to 14	7 to 9, 12, 15, 21	7 to 9, 11 to 12, 17 to 18, and 22	9
Measures available	California Achievement Test (reading and math sections)	Woodcock Johnson Test (reading and math sections)	Wide Range Achievement Test (reading and math sections)	Reading and math achievement inventories
Panel b. **Adult Skill Measures**				
Child age at measurement	54	21 and 45	17, 22, and 31	Not available
Measures available	Cognition: Raven and Stroop Tests combined items. Non-cognitive: inventories of positive personality, including conscientiousness	Cognition: Wechsler Adult Intelligence Scale (21) and Raven and Stroop Tests combined items (45). Non-cognitive: inventories of positive personality, including conscientiousness (45)	Cognitive: Wechsler Adult Intelligence Scale IQ Test. Non-cognitive: Conscientiousness inventory	
Panel c. **Adult Outcomes Observed**				
Child age at measurement	21 to 40	21 to 40	22 and 31	Not available
Measures available	Education (40): years of education. Employment (40): employed in any job at age 22; employed in high-skill job at age 31. Violence (21-40): reverse-coded variables indicating engagement in fights and other violent behaviors. Alcohol/Marijuana (21-40): reverse-coded variables describing frequency and intensity of alcohol/marijuana use	Education (30): years of education. Employment (30): employed in any job at age 22; employed in high-skill job at age 31. Violence (21-30): reverse-coded variables indicating engagement in fights and other violent behaviors. Alcohol/Marijuana (21-30): reverse-coded variables describing frequency and intensity of alcohol/marijuana use	Education (31): years of education. Employment: employed in any job at age 22; employed in high-skill job at age 31. Violence (31): reverse-coded variables indicating engagement in fights and other violent behaviors. Alcohol/Marijuana (31): reverse-coded variables describing frequency and intensity of alcohol/marijuana use	

Sources: Cited in Table 4.

Figure 10. Impacts on Adult Skills and Outcomes for Omnibus Programs and Jamaica

(a) Skills **(b)** Outcomes



Note: Panel (a) presents program impacts on measures of cognitive and non-cognitive skills. For Perry and ABC, the measures are described in Section 3 for ages 54 and 45. Recall that both measures are standardized by subtracting the control-group mean and dividing by the control-group standard deviation. The other measures used for the plot are standardized similarly. We mark impacts when the p-value associated with the null hypothesis that they are less than or equal to 0 is less than 0.05. The measures are described in Table 6. Panel (b) is analogous in format to Panel (a) for adulthood outcomes. Employment is the treatment-control difference in the employment rate, except for Jamaica at age 31. For Perry, we report results for males and females. For Jamaica, employment at age 31 is the effect size for "being employed in a high-skilled job." The violence, alcohol, and marijuana outcomes are latent factor variables based on reverse-coded measures indicating violent behavior or engagement in fights (violence), alcohol use and frequency of use, and marijuana use and frequency of use.

6. Summary and Conclusions

This paper addresses the evidence on the effectiveness of early childhood education programs in promoting social mobility. Effective programs enrich home lives of disadvantaged children and promote parent-child interactions, which last throughout childhood. Enhancing parenting emerges as an essential ingredient of successful programs. Instead of presenting a barrage of undigested treatment effects like Figure 1, we focus on understanding the mechanisms underlying successful programs. Child development has a universal structure in all populations. Common mechanisms foster development. We compare programs with comparable measures for largely comparable populations. These meaningful comparisons allow us to investigate the mechanisms that successful interventions activate and enhance and look for the ingredients that activate these mechanisms.

A central lesson from the literature on child development is the crucial role of parenting—attachment, guidance, and support. All of the successful programs that we study promote parenting even though they are superficially very different. Some are omnibus programs that have many features promoting child development. Other, more focused, home-visiting programs are much less expensive and demanding to operate than omnibus programs. They are surprisingly effective at low cost. Our discovery of this common thread across successful programs is supported by the powerful evidence on the effectiveness of low-key home-visiting programs. They tap a central message of the literature on child development and filter out the persiflage promoted by "professional schools" and advocates about teacher quality, quality of facilities, need for nurses, etcetera.

Our approach to synthesizing evidence differs greatly from the common approach of finding the "best program" to implement. Instead, we search for common mechanisms and relationships that transport across environments. Using this methodology, we establish the power of parenting programs that foster the home life of children, and, thereby, promote social mobility within and across generations.

References

Agostinelli, F. and M. Wiswall (2016). Estimating the Technology of Children's Skill Formation. Technical Report w22442, National Bureau of Economic Research, Cambridge, MA.

Alfred Marshall (1890). Principles of Economics. *London, McMillan.*

Andrew, A., O. Attanasio, B. Augsburg, M. Day, S. Grantham-McGregor, C. Meghir, F. Mehrin, S. Pahwa, and M. Rubio-Codina (2020). Effects of a Scalable Home-visiting Intervention on Child Development in Slums of Urban India: Evidence from a Randomised Controlled Trial. *Journal of Child Psychology and Psychiatry 61*(6), 644–652.

Athey, S., R. Chetty, G. Imbens, and H. Kang (2019). The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely. Technical Report w26463, National Bureau of Economic Research, Cambridge, MA.

Attanasio, O., T. Boneva, and C. Rauh (2019). Parental Beliefs about Returns to Different Types of Investments in School Children. Technical Report w25513, National Bureau of Economic Research, Cambridge, MA.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia. *American Economic Review 110*(1), 48–85.

Attanasio, O., F. Cunha, and P. Jervis (2019). Subjective Parental Beliefs. Their Measurement and Role. Technical Report w26516, National Bureau of Economic Research, Cambridge, MA.

Attanasio, O., C. Meghir, and E. Nix (2020). Human Capital Development and Parental Investment in India. *Review of Economic Studies 87*(6), 2511–2541.

Attanasio, O. P., C. Fernandez, E. O. A. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014). Using the Infrastructure of a Conditional Cash Transfer Program to Deliver a Scalable Integrated Early Child Development Program in Colombia: Cluster Randomized Controlled Trial. *BMJ 349*(sep29 5), g5785–g5785.

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

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 Science in the Public Interest 21*(2), 55–97.

Barr, A., J. Eggleston, and A. A. Smith (2022). Investing in Infants: The Lasting Effects of Cash Transfers to New Families. *Quarterly Journal of Economics 137*(4), 2539–2583.

Barr, A. and C. Gibbs (2022). Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood. *Journal of Political Economy Forthcoming.*

Becker, G. S. and N. Tomes (1979). An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy 87*(6), 1153–1189.

Becker, G. S. and N. Tomes (1986, July). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics 4*(3, Part 2), S1–S39.

Borghans, L., B. H. Golsteyn, J. Heckman, and J. E. Humphries (2011). Identification Problems in Personality Psychology. *Personality and Individual Differences 51*(3), 315–320.

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and J. E. Humphries (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences 113*(47), 13354–13359.

Brentani, A., S. Walker, S. Chang-Lopez, S. Grisi, C. Powell, and G. Fink (2021). A Home Visit-based Early Childhood Stimulation Programme in Brazil—A Randomized Controlled Trial. *Health Policy and Planning 36*(3), 288–297.

Bureau of Labor Statistics (2015). National Longitudinal Surveys: The NLSY79. Website, `http://www.bls.gov/nls/nlsy79.htm`.

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), 1478–1485.

Cascio, E. U. (2021). Does Universal Preschool Hit the Target? Program Access and Preschool Impacts. *Journal of Human Resources* (Forthcoming).

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2019). On Binscatter. FRB of New York Staff Report 881, Federal Reserve Bank of New York.

Caucutt, E., L. Lochner, J. Mullins, and Y. Park (2020, September). Child Skill Production: Accounting for Parental and Market-Based Time and Goods Investments. Technical Report w27838, National Bureau of Economic Research, Cambridge, MA.

Caucutt, E. M. and L. Lochner (2020, March). Early and Late Human Capital Investments, Borrowing Constraints, and the Family. *Journal of Political Economy 128*(3), 1065–1147.

Chaparro, J., A. Sojourner, and M. Wiswall (2020, February). Early Childhood Care and Cognitive Development. Technical Report w26813, National Bureau of Economic Research, Cambridge, MA.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly journal of economics 126*(4), 1593–1660.

Chetty, R. and N. Hendren (2018a). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *Quarterly Journal of Economics 133*(3), 1163–1228.

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2020). Race and Economic Opportunity in the United States: An Intergenerational Perspective. *Quarterly Journal of Economics 135*(2), 711–783.

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

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy 126*(6), 2356–2409.

Cunha, F., I. Elo, and J. Culhane (2022). Maternal Subjective Expectations about the Technology of Skill Formation Predict Investments in Children One Year Later. *Journal of Econometrics 231*(1), 3–32.

Cunha, F. and J. Heckman (2007, April). The Technology of Skill Formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Chapter 12 Interpreting the Evidence on Life Cycle Skill Formation. In *Handbook of the Economics of Education*, Volume 1, pp. 697–812.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica 78*(3), 883–931.

Del Boca, D., C. Flinn, E. Verriest, and M. Wiswall (2019, February). Actors in the Child Development Process. Technical Report w25596, National Bureau of Economic Research, Cambridge, MA.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household Choices and Child Development. *The Review of Economic Studies 81*(1), 137–185.

Del Bono, E., J. Kinsler, and R. Pavan (2022). Identification of Dynamic Latent Factor Models of Skill Formation with Translog Production. *Journal of Applied Econometrics 37*(6), 1256–1265.

Deshpande, M. (2016, November). Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls. *American Economic Review 106*(11), 3300–3330.

Doepke, M. and F. Zilibotti (2017). Parenting With Style: Altruism and Paternalism in Intergenerational Preference Transmission. *Econometrica 85*(5), 1331–1371.

Doepke, M. and F. Zilibotti (2019). *Love, Money, and Parenting: How Economics Explains the Way We Raise Our Kids*. Princeton University Press.

Doyle, O. (2016). Preparing for Life: Testing the Effectiveness of a Five-Year Intervention Programme in India. *Unpublished Manuscript, School of Economics, University College of Dublin*.

Doyle, O. (2020). The First 2,000 Days and Child Skills. *Journal of Political Economy 128*(6), 2067–2122.

Doyle, O. (2021). Can Early Intervention Have a Sustained Effect on Human Capital. *Unpublished Manuscript, School of Economics, University College of Dublin*.

Duncan, G., A. Kalil, M. Mogstad, and M. Rege (2022). Investing in Early Childhood Development in Preschool and at Home. Technical Report w29985, National Bureau of Economic Research, Cambridge, MA.

Duncan, G. and S. Le Menestrel (2019). *A Roadmap to Reducing Child Poverty. Consensus Study Report*. ERIC.

Duncan, G. J. and K. Magnuson (2013). Investing in Preschool Programs. *Journal of Economic Perspectives 27*(2), 109–132.

Duncan, G. J. and A. J. Sojourner (2013). Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps? *Journal of Human Resources 48*(4), 945–968.

Ertem, I. O., V. Krishnamurthy, M. C. Mulaudzi, Y. Sguassero, H. Balta, O. Gulumser, B. Bilik, R. Srinivasan, B. Johnson, G. Gan, L. Calvocoressi, V. Shabanova, and B. W. C. Forsyth (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.

Ganimian, A. J., K. Muralidharan, and C. R. Walters (2021). Augmenting State Capacity for Child Development in India. NBER Working Paper w28780, National Bureau of Economic Research.

García, J. L. (2022). Social-Program Use, Savings for Retirement, Long-Term Health, and Intergenerational Impacts of Early Childhood Education. *Unpublished Manuscript, John E. Walker Department of Economics, Clemson University*.

García, J. L., F. Bennhoff, D. E. Leaf, and J. Heckman (2021). The Dynastic Benefits of Early Childhood Education. Technical Report w29004, National Bureau of Economic Research, Cambridge, MA.

García, J. L., J. Heckman, and V. Ronda (2021). The Lasting Effects of Early Childhood Education on Promoting the Skills and Social Mobility of Disadvantaged African Americans. Technical Report w29057, National Bureau of Economic Research, Cambridge, MA.

García, J. L. and J. J. Heckman (2021). Early Childhood Education and Life-cycle Health. *Health Economics 30*(S1), 119–141.

García, J. L. and J. J. Heckman (2022). On Criteria for Evaluating Social Programs. NBER Working Paper w30005, National Bureau of Economic Research.

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. *Journal of Political Economy 128*(7), 2502–2541.

García, J. L., J. J. Heckman, and A. L. Ziff (2018). Gender Differences in the Benefits of an Influential Early Childhood Program. *European Economic Review 109*, 9–22.

García, J. L., J. J. Heckman, and A. L. Ziff (2019). Early Childhood Education and Crime. *Infant Mental Health Journal 40*(1), 141–151.

Gertler, P., J. Heckman, R. Pinto, S. Chang, S. Grantham-McGregor, C. Vermeersch, S. Walker, and A. Wright (2021). Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31. Technical Report w29292, National Bureau of Economic Research, Cambridge, MA.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. *Science 344*(6187), 998–1001.

Grantham-McGregor, S. M., S. P. Walker, S. M. Chang, and C. A. Powell (1997). Effects of Early Childhood Supplementation with and without Stimulation on Later Development in Stunted Jamaican Children. *The American Journal of Clinical Nutrition 66*(2), 247–253.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2021). The Long-Term Effects of Universal Preschool in Boston. NBER Working Paper w28756, National Bureau of Economic Research.

Havnes, T. and M. Mogstad (2015). Is Universal Child Care Leveling the Playing Field? *Journal of Public Economics 127*, 100–114.

Heckman, J., B. Liu, M. Lu, and J. Zhou (2020). The Impacts of a Prototypical Home Visiting Program on Child Skills. Technical Report w27356, National Bureau of Economic Research, Cambridge, MA.

Heckman, J., L. Lochner, and P. Todd (2008). Earnings Functions and Rates of Return. *Journal of Human Capital 2*(1), 1–31.

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), 2052–2086.

Heckman, J. J. and G. Karapakula (2021). Using a Satisficing Model of Experimenter Decision-Making to Guide Finite-Sample Inference for Compromised Experiments. *Econometrics Journal 24*(2), C1–C39.

Heckman, J. J. and R. Landersø (2022). How the Welfare State Affects Inequality and Social Mobility: A Comparison of the U.S. and Denmark. *Unpublished Manuscript, Department of Economics, The University of Chicago*.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*, 307–458.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010). The Rate of Return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1-2), 114–128.

Heckman, J. J. and S. Mosso (2014). The Economics of Human Development and Social Mobility. *Annual Review of Economics 6*(1), 689–733.

Heckman, J. J., R. Pinto, S. H. Moon, A. Yavitz, and P. Savelyev (2010). Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program. *Quantitative Economics 1*(1), 1–46.

Heckman, J. J., R. Pinto, and A. M. Shaikh (2022). Inference with Imperfect Randomization: The Case of the Perry Preschool Program. *Journal of Econometrics Forthcoming*.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments. *Handbook of Econometrics 6*, 4875–5143.

Hendren, N. and B. Sprung-Keyser (2020). A Unified Welfare Analysis of Government Policies. *Quarterly Journal of Economics 135*(3), 1209–1318.

Hermes, H., P. Lergetporer, F. Peter, and S. Wiederhold (2021). Behavioral Barriers and the Socioeconomic Gap in Child Care Enrollment. IZA Working Paper IZA DP No. 14698, IZA Institute for Labor Economics.

Hertzman, C. and J. Bertrand (2007). Children in Poverty and the Use of Early Development Instrument Mapping to Improve their Worlds. *Paediatrics and Child Health*.

Hojman, A. P. C. (2016). *Three Essays on the Economics of Early Childhood Education Programs*. Ph. D. thesis, The University of Chicago.

Imbens, G. and J. Angrist (1994). Estimation and Identification of Local Average Treatment Effects. *Econometrica 62*, 467–475.

Inhelder, B. and J. Piaget (1972). *The Psychology Of The Child*. Basic Books.

Kline, P. and C. R. Walters (2016). Evaluating Public Programs with Close Substitutes: The Case of HeadStart. *Quarterly Journal of Economics 131*(4), 1795–1848.

Leibowitz, A. (1974). Home Investments in Children. *Journal of Political Economy 82*(2, Part 2), S111–S131.

Leventhal, T., A. Martin, and J. Brooks-Gunn (2004). The EC-HOME across Five National Data Sets in the 3rd to 5th Year of Life. *Parenting 4*(2-3), 161–188.

Linver, M. R., A. Martin, and J. Brooks-Gunn (2004). Measuring Infants' Home Environment: The IT-HOME for Infants between Birth and 12 Months in Four National Data Sets. *Parenting 4*(2-3), 115–137.

List, J. A. (2022). *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale.* Currency.

Loewenstein, S. (1973). Conclusion of a study of pari as a research tool. *Child Welfare 52*(6), 400–402.

Mullins, J. (2019). Designing Cash Transfers in the Presence of Children's Human. *University of Minnesota, Minneapolis and Saint Paul, MN*.

Mullins, J. (2020). A Structural Meta-Analysis of Welfare Reform Experiments and their Impacts on Children. *University of Minnesota, Minneapolis and Saint Paul, MN*.

Pavan, R. (2016). On the Production of Skills and the Birth-order Effect. *Journal of Human Resources 51*(3), 699–726.

Protzko, J. (2015). The Environment in Raising Early Intelligence: A Meta-Analysis of the Fadeout Effect. *Intelligence 53*, 202–210.

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association 53*(284), 873–880.

Quandt, R. E. (1972). A New Approach to Estimating Switching Regressions. *Journal of the American Statistical Association 67*(338), 306–310.

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 Economic Journal: Applied Economics 12*(3), 255–86.

Seror, A. (2022, September). Child Development in Parent-Child Interactions. *Journal of Political Economy 130*(9), 2462–2499.

Steinberg, L. D. (2014). *Age of Opportunity: Lessons from the New Science of Adolescence.* Houghton Mifflin Harcourt.

Sylvia, S., N. Warrinnier, R. Luo, A. Yue, O. Attanasio, A. Medina, and S. Rozelle (2021). From Quantity to Quality: Delivering a Home-Based Parenting Intervention Through China's Family Planning Cadres. *The Economic Journal 131*(635), 1365–1400.

Tofail, F., J. D. Hamadani, F. Mehrin, D. A. Ridout, S. N. Huda, and S. M. Grantham-McGregor (2013). Psychosocial Stimulation Benefits Development in Nonanemic Children but Not in Anemic, Iron-Deficient Children. *The Journal of Nutrition 143*(6), 885–893.

U.S. Department of Health and Human Services (2021). Head Start Program Facts: Fiscal Year 2019. Website, https://eclkc.ohs.acf.hhs.gov/about-us/article/head-start-program-facts-fiscal-year-2019.

Uzgiris, I. C. and J. Hunt (1975). Assessment in Infancy: Ordinal scales of Psychological Development.

Voltaire (1785). *Zaïre in Oeuvres Complètes de Voltaire*.

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-retarded Jamaican Children: Prospective Cohort Study. pp. 8.

Walker, S. P., S. M. Chang, M. Vera-Hernández, and S. Grantham-McGregor (2011). Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *Pediatrics 127*(5), 849–857.

Walker, S. P., S. M. Chang, A. S. Wright, R. Pinto, J. J. Heckman, and S. M. Grantham-McGregor (2022). Cognitive, Psychosocial, and Behaviour Gains at Age 31 Years from the Jamaica Early Childhood Stimulation Trial. *Journal of Child Psychology and Psychiatry 63*(6), 626–635.

Walker, S. P., S. M. Grantham-McGregor, C. A. Powell, and S. M. Chang (2000). Effects of Growth Restriction in Early Childhood on Growth, IQ, and Cognition at Age 11 to 12 Years and the Benefits of Nutritional Supplementation and Psychosocial Stimulation. *The Journal of Pediatrics 137*(1), 36–41.

Walters, C. R. (2015). Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start. *American Economic Journal: Applied Economics 7*(4), 76–102.

Washington State Institute for Public Policy (2015). What Works and What Does Not? Benefit-Cost Findings from WSIPP. Website, https://www.wsipp.wa.gov/ReportFile/1602/Wsipp_What-Works-and-What-Does-Not-Benefit-Cost-Findings-from-WSIPP_Report.pdf.

Weikart, D. P. et al. (1978). The Ypsilanti Perry Preschool Project: Preschool Years and Longitudinal Results Through Fourth Grade. Monograph No. 3.

WHO and de Onís (2006). Assessment of Sex Differences and Heterogeneity in Motor Milestone Attainment among Populations in the WHO Multicentre Growth Reference Study. *Acta Paediatrica 95*, 66–75.

Yazejian, N., D. Bryant, K. Freel, and M. Burchinal (2015). High-quality Early Education: Age of Entry and Time in Care Differences in Student Outcomes for English-only and Dual Language Learners. *Early Childhood Research Quarterly 32*, 23–39.

Yazejian, N., D. M. Bryant, S. Hans, D. Horm, L. St. Clair, N. File, and M. Burchinal (2017). Child and Parenting Outcomes After 1 Year of Educare. *Child Development 88*(5), 1671–1688.