Early Education Programs in the US: Background and Evaluations*

Sneha Elango
The University of Chicago

Jorge Luis García
The University of Chicago

James J. Heckman
American Bar Foundation
The University of Chicago

Andrés Hojman
The University of Chicago

November 21, 2014

*This research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, the Buffett Early Childhood Fund, NIH grants NICHD R37HD065072, NICHD R01HD54702, and NIA R24AG048081, an anonymous funder, Successful Pathways from School to Work, an initiative of the University of Chicago's Committee on Education funded by the Hymen Milgrom Supporting Organization, a grant from the Institute for New Economic Thinking (INET) supporting the Human Capital and Economic Opportunity Global Working Group (HCEO)—an initiative of the Center for the Economics of Human Development (CEHD) and Becker Friedman Institute for Research in Economics (BFI). We are very grateful to Joshua Ka Chun Shea for data managing, research assistance, and useful comments. We also thank Marianne Haramoto, Sylvi Kuperman, Ana Paula Melo, Jack Schmerold, Haonan Zhou, and Anna Ziff for research assistance. Carey Cheng, Fernanda Díaz de la Vega, Sindhu Gnanasambandan, Robert Moffitt, and Damini Sharma helped us improve this draft with useful comments. The views expressed in this paper are those of the authors and not necessarily those of the funders or persons named here or the official views of the National Institutes of Health.

Contents

1 Introduction **5**

1.1 Plan of the Rest of the Paper 14

2 The State of Childcare and Early Education Programs in the U.S. **16**

2.1 Demographic Differences in the Use of Time: Childcare Arrangements 16

2.2 Enrollment in Educational Programs 18

2.2.1 Trends in the Take-Up of Educational Programs 18

2.2.2 Demographic Differences in the Take-Up of Educational Programs . . 19

3 The Evolution of the Main Childhood Programs in the US **20**

3.1 Enrollment Levels of the Different Childhood Programs in the US 23

3.2 Funding Levels of the Different Childhood Programs in the US 24

3.3 Head Start (HS) 25

3.3.1 Description 25

3.3.2 Curricula 25

3.3.3 Program History 26

3.4 Early Head Start 27

3.4.1 Description 27

3.5 State Preschool Programs 28

4 Results from Randomized Controlled Trials of Early Childhood Interventions **30**

4.1 Perry Preschool Project (PPP) 33

4.1.1 Historical Context 33

4.1.2 Educational Content 34

4.1.3 Program Description 34

4.1.4 Evidence from the Literature 35

4.2 The Carolina Abecedarian Project (ABC) . . . 37
4.2.1 Historical Context . . . 37
4.2.2 Educational Content . . . 37
4.2.3 Program Description . . . 38
4.2.4 Evidence from the Literature . . . 38
4.3 Project CARE: The Carolina Approach to Responsive Education (CARE) . . . 39
4.3.1 Historical Context . . . 39
4.3.2 Educational Content . . . 39
4.3.3 Program Description . . . 40
4.4 Infant Health and Development Program (IHDP) . . . 40
4.4.1 Historical Context . . . 40
4.4.2 Educational Content . . . 41
4.4.3 Program Description . . . 41
4.5 Early Training Project (ETP) . . . 43
4.5.1 Historical Context . . . 43
4.5.2 Educational Content . . . 43
4.5.3 Program Description . . . 44
4.6 Comparison Across RCTs . . . 44
4.6.1 Methodology . . . 44
4.6.2 The Model . . . 44
4.6.3 Estimation Issues . . . 45
4.6.4 Results . . . 47

5 Public Early Childhood Interventions **52**
5.1 Randomized Controlled Trials . . . 52
5.1.1 Issues in the Evaluations . . . 52
5.1.2 The Tennessee Voluntary Prekinder Program (TN-VPK) . . . 54
5.1.3 The Head Start Impact Study (HSIS) . . . 55

5.1.4 Analyses of HSIS that Address Control Contamination 56
5.2 Non-randomized Evaluations of Head Start 57
5.2.1 Fixed Effects Studies 59
5.2.2 Regression Discontinuity Designs 60
5.3 Evidence on State Programs 61
5.3.1 Regression Discontinuity Designs to Evaluate State Preschool 61
5.3.2 The Chicago Child-Parent Center (CPC) 62
5.4 Impacts of Generic Center and Non-center Based Childcare 62

6 Next Steps **63**

Appendix A List of Abbreviations **74**

Appendix B Data Sources **75**

1 Introduction

Children's skill development depends critically on the investments they receive during childhood (Cunha and Heckman, 2007; Cunha et al., 2010a). Investment in children can generally be expressed in terms of (i) time spent with their parents or other adults and (ii) goods and services. In the US, a sizable proportion of households face serious constraints when deciding their investment in children. This is particularly true for disadvantaged households –which have very limited capacity investment capacity for at least three reasons: (i) they do not have access to high-quality childcare; (ii) they cannot provide human capital enrichment at home due to parenting knowledge or practice limitations (see Table 1); (iii) they cannot afford goods and services devoted to child development.

Table 1: Disadvantages Arise Early in Life in Mothers' Speech and Child Vocabulary, Hart and Risley (1995)

Average Vocabulary and Number of Words Heard by Child per Hour, Age 3

Family Socieo-Economic Status	**Actual Differences in Quantity of Words Heard**	**Actual Differences in Quantity of Words Heard**	**Child's Cumulative Vocabulary**
Welfare	616 words	5 affirmatives, 11 prohibitions	500 words
Working Class	1,251 words	12 affirmatives, 7 prohibitions	700 words
Professional	2,153 words	32 affirmatives, 5 prohibitions	1,100 words

In the last 50 years there have been two major economic changes affecting parental ability to invest in children. First, more children live in single-parent households, mainly due to sizable increases in births out of wedlock (see Figures 1a and Figure 1b). Relative to households with two parents, single-parent households are able to invest less time in producing child development. Second, mothers' labor supply has drastically increased (see Figure 1c). The consequences of this are ambiguous. While this implies an increase in the resources mothers could spend on goods and services devoted to child development, it

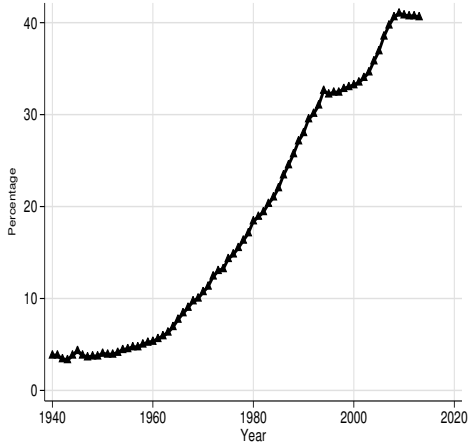
also implies mothers have less discretionary time that that *may* reduce time spend on child development. The two changes increase the relevance of childcare and early interventions as means of child development.

It is difficult to make a clear distinction between childcare and early childhood intervention. For the purpose of this paper, we consider as childcare any program in which children spent just enough time so that their parents are able to work. We consider early interventions to be programs with important educational or stimulation components. Naturally, there are examples of programs sharing childcare and early intervention components. Blau (2003) and Blau and Currie (2006) provide extensive surveys on childcare and we do not compete in aims or scope with them. Instead, we try to motivate and analyze whether there is a case for childcare and early intervention and then take a more focused approach on early interventions, as explained below.

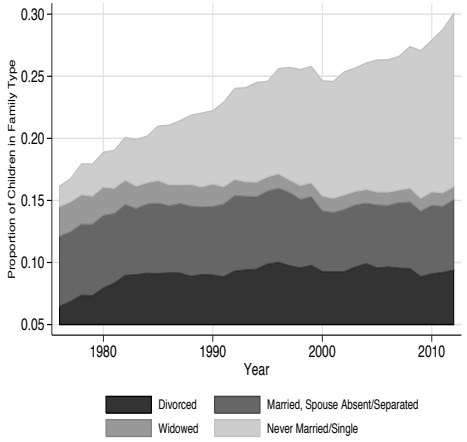
A policy relevant question in this context is whether there is a case for interventions in early childhood. A more comprehensive argument approaches the accident of birth– Heckman (2008, 2012) and Cunha et al. (2010b) formulate an argument not only based on equity but also on efficiency. Aiyagari et al. (2002) approach the accident of birth as a market failure which is efficient to fix. The market fails in that children cannot insure against poor parents and disadvantaged environments. Heckman and Mosso (2014) present a dynamic model in which parents maximize their children's human capital. If investment across time exhibits dynamic complementarity in the technology of skill formation, the socially efficient solution implies investing in relatively disadvantaged children.

Figure 1: Children and their Family Structure in the US

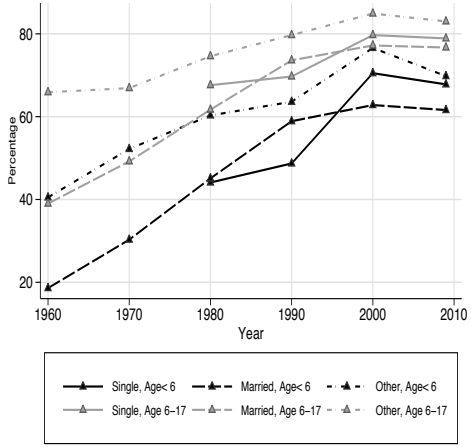
(a) Percent Birth to Unmarried Women



(b) Children under 18 Living in Single Parent Households



(c) Mother's Labor Force Participation Rates by Marital Status and Children's Age



Note: Panel (a): mother's age is 15 to 44. Source: Center for Disease Control and Prevention. Panel (b): parents are defined as the head of the household; children are defined as individuals under 18 living in the household and being the children of the head of household. Children who have been married or are not living with their parents are excluded from the calculation. Separated parents are included in "Married, Spouse Absent/Separated". Source: March IPUMS. Panel (c): other category includes widowed, divorced, or separated (including married, spouse absent). Source: Census 2012, Statistical Abstract.

Various structural models have been used to analyze policies subsidizing the disadvantaged. Examples include subsidies to investment in early childhood, e.g. Cunha (2007), restricted and unrestricted transfers, e.g. Del Boca et al. (2014), and increases in parental borrowing limits, e.g. Caucutt and Lochner (2011). In general the findings are positive and, therefore, establish a strong economic case for early intervention for the disadvantaged in order to improve child investment–suggesting theoretical and economic support for means testing (see Heckman, 2008).

The fields studying childcare and early intervention are vast. Two comprehensive surveys of childcare programs already appear in the literature: Blau (2003) and Blau and Currie (2006).[1] This paper is much more focused. It summarizes the current state of knowledge of the funding and effectiveness of early childhood programs. It also seeks to build an understanding of successful interventions to inform the mechanisms of family influence on child development (see Cunha and Heckman, 2009; Heckman and Mosso, 2014). The contrast between what families do and what external interventions do, while traditional, is artificial. Paraphrasing Dewey (1916), *good preschools do what good parents do.* Studies of household child production have strong implications for the design and interpretation of early childhood interventions and vice versa.

We review the history, evaluations, and evidence reported in the literature from different ECIs that include educational components. This task is daunting. Intervention programs differ greatly in the populations served, the nature of their curricula, and the form of child-family-childcare provider interactions. Evaluations differ in the measures taken, the length of the follow up, and the statistical method to control for selection bias. Our goals are more modest than those of Blau (2003) or Blau and Currie (2006), but achieving them is still very challenging.

We document the following 7 findings from the literature (the programs mentioned are

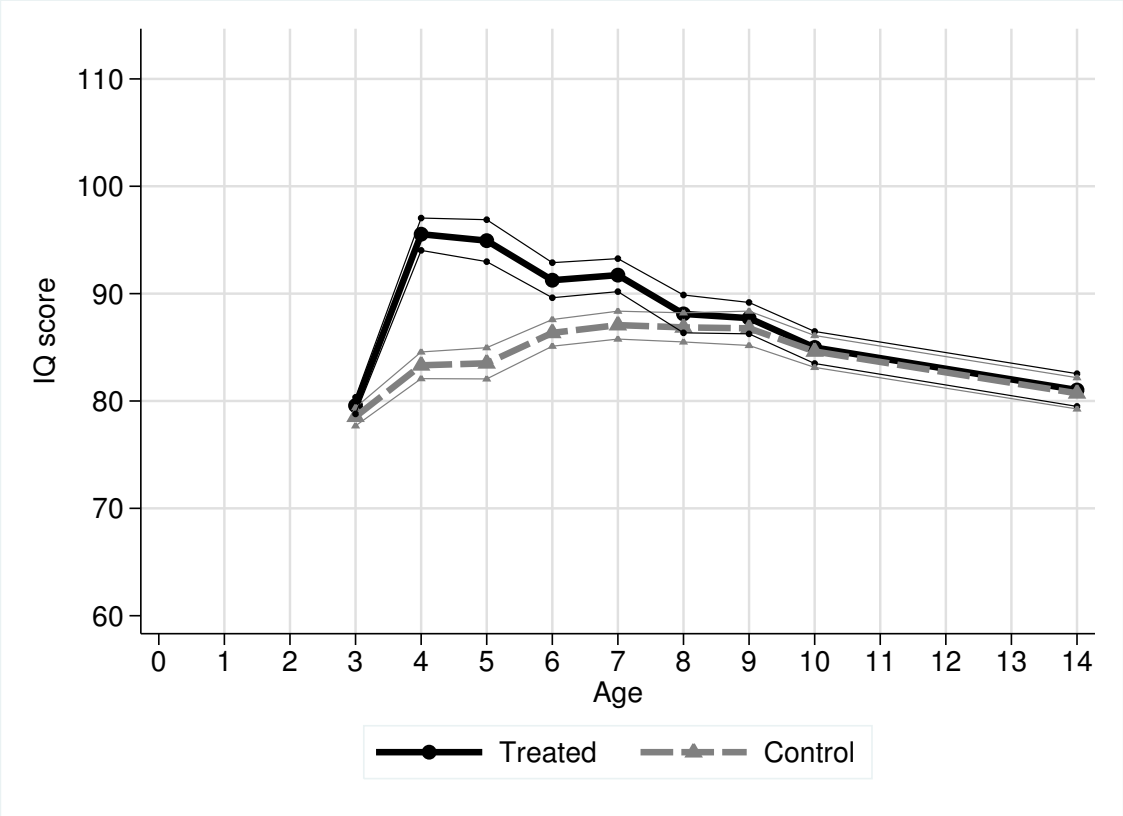
[1]The literature often separates the study of childcare from the study of early childhood programs (Blau (2003) and Blau and Currie (2006) are important exceptions). Yet the two are closely related aspects of household production and labor supply (Becker, 1991).

discussed in detail below).

F-1 **Early Childhood Interventions (ECIs) strongly boost IQ in the short-run for disadvantaged children. However, treatment effects often fade out. For very early interventions (before age 3), impacts persist into adulthood.** For the 5 US-based educational ECIs that have followed participants into adulthood, the short-term effects on cognition, as measured by different IQ tests, have an approximate magnitude of one standard deviation. The impacts fade out, especially for programs targeting children after age 3.[2] In some cases, the treatment effect persists –to a much smaller magnitude– and in other cases it completely vanishes after only a few years (see Hojman, 2014). The finding that early life cognition mediates later-life outcomes (see below) opens an important question about the importance of this early increase in cognition for later life outcomes. Figure 2 illustrates IQ dynamics for the Perry Preschool Program (PPP) as an example of the fade-out in the ECIs we analyze.

[2]For ABC, which began when the children were born, there are lasting treatment effects on IQ throughout adulthood.

Figure 2: Cognitive Skills Dynamics in the Perry Preschool Program



Note: thin lines surrounding trajectories are standard errors. Cognitive skills have been measured by the Stanford-Binet IQ test, which is constructed to have a national mean of 100 and a standard deviation of 15. Source: Hojman (2014).

F-2 **ECIs boost non-cognitive skills in the short run for disadvantaged children. Due to the lack of data and because of the difficulties associated with measuring personality traits, it is often hard to determine the real size of the impacts on these skills and whether they are affected in the long run.** For many ECIs, we are able to document treatment effects on non-cognitive skills in the short run. These effects appear to persist.[3] Non-cognitive skills are fundamental in mediating later-life outcomes (Heckman et al., 2013; Heckman and Mosso, 2014). They appear to be more important than cognitive skills as mediators of many adult treatment effects (García and Heckman, 2014). Figure 3 illustrates this for the case of the Carolina Abecedarian Project (ABC).

F-3 **ECIs are especially effective in boosting early and later life outcomes for disadvantaged children.** This is a recurring finding in the literature. We note, however, that most interventions are targeted toward the disadvantaged. When they are, the evidence supports greater effectiveness for children from disadvantaged families.[4] There are two main reasons for this finding. Parents with better socio-economic status typically provide children with an enriched home environment.[5] In addition, the substitutes for center-based childcare of less affluent children are generally of lower quality.[6] The available evidence provides a clear case for the means testing of child development programs.

[3]See Almlund et al. (2011) for discussions on relevance, dynamics, measurement, and interpretation of non-cognitive skills.

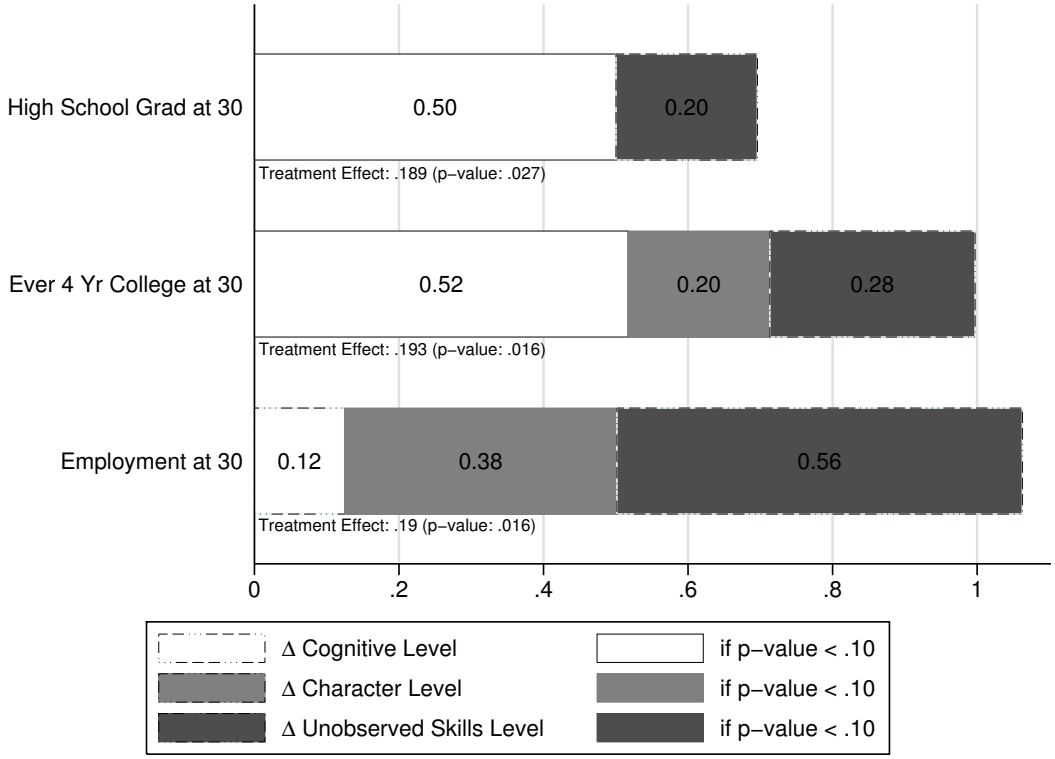
[4]An example of this comes from the Infant Health and Development Program (IHDP). This is the only ECI with long-term follow-up in which we find substantive variation across children's socioeconomic backgrounds. The effects of IHDP are much stronger for disadvantaged children (see Brooks-Gunn et al., 1992; Duncan and Sojourner, 2013; Heckman and Setzler, 2014).

[5]By design, the implementation of ECIs could crowd out parent-child interaction. It is important to grasp this and analyze the alternative to parent-child interaction. García and Hojman (2014) find that the reduction in parental-time interaction diminishes the positive total effects from a particular ECI on cognitive and non-cognitive skills.

[6]Bernal (2008), Bernal and Keane (2010), and Bernal and Keane (2011) find that when parental-time interaction is substituted for informal childcare there is a negative impact on children's cognition. However, they find no effect when the substitution is towards center-based childcare. See also Del Boca et al. (2013) and the discussion of the evidence in Heckman and Mosso (2014).

F-4 **ECIs stimulate parenting and parent-child interactions.** A main mechanism for the success of ECIs is parenting stimulation (see García and Heckman, 2014; Heckman and Mosso, 2014). Improved parental environments and parenting practices enhance the development of children in a lasting way. Importantly, lack of parenting knowledge and practice is common in disadvantaged families as Lareau (2011), Badev and Cunha (2013), and Cunha (2012) document and Table 1 illustrates.

Figure 3: Cognitive and Character Skills as Education and Employment Mediators of Treatment Effects, ABC All



Note: this figure shows how cognitive and non-cognitive (character) skills mediate treatment effects in the relevant outcome displayed in the y-axis. For the cognitive measure, we use achievement as measured by the Achenbach test at age 21. For the character measure, we use Conscientiousness at age 15, based on a factor summarizing a set of items designed to measure personality traits. We make arbitrary choices on the items based on what we understand as conscientiousness. To avoid the arbitrary scale issues that tests and factors have, we use a variable representing 25 quantiles instead of either the score test or the factor itself. The bars display the components of three-folded Laspeyres decomposition exercises on each relevant outcome. To obtain them, we estimate two models, one for the treatment group and one for the control group. We first estimate the treatment effect and display it as Δ (Predicted). Then, we construct differences in the coefficients, which we call Δ (Cognitive Tech.) and Δ (Character Tech.) and differences in mean skills, which we call Δ (Cognitive Level) and Δ (Character Level). The difference of the coefficients is scaled up by the mean difference of the control group. The difference of the mean skills is scaled up by the control coefficient. Δ (Unobserved Skills) corresponds to the difference in the intercepts in each model. We interpret this as a residual, i.e. the proportion of variance cognitive and characters skills cannot explain. The length of the bars is not identical to one because we avoid displaying the components accounting for control variables. The interval bars are displayed when the asymptotic one-sided p-value is less than .10. Source: García and Heckman (2014).

F-5 **Substitution bias is pervasive in recent program evaluations.**[7] **As the quality of alternatives to any program increases, the *measured* treatment effects of an offer to participate in an ECI decreases** We use the term *substitution bias* or *control contamination* to describe the phenomenon in which control group participants of an intervention find a worthy substitute. In this case, the estimate of the impact of an ECI against a world without any substitutes is difficult to obtain. Evaluations of ECIs in which there is control contamination follow two different strategies: (i) Estimate the treatment effects under traditional methods and estimate the treatment effect of participating in the ECI against the next best alternative of non-participants, be it participating in an alternative childcare program or not. Clearly, the better the alternatives chosen, the smaller the estimated impacts of the ECI; (ii) Apply methodologies that address the possibility of multiple choices (e.g., ECI, other childcare, or no childcare) and self-selection and, therefore, obtain the various counterfactuals of interest –ECI vs. other childcare, ECI vs. no childcare, other childcare vs. no childcare. While the second strategy delivers more information, it has important methodological challenges. In an early study, Heckman (1974) addresses this problem.[8]

F-6 **The Available Evidence Suggests that Head Start is not a failure, despite claims to the contrary.**[9] Head Start is a victim of the "gold standard" mentality, i.e. that randomization is the "gold standard" for evaluating social programs. This mentality ignores the complexity of most social experiments and the problem of substitution bias. When control contamination is addressed, the effects of Head Start on different measures of cognition and achievement appear to be substantively higher.[10]

[7] See Heckman et al. (1999, 2000) for discussions of the general problem of substitution bias.

[8] See also Burchinal et al. (1989); Heckman et al. (2000); Campbell et al. (2014a); Kline and Walters (2014); Feller et al. (2014).

[9] See, for example, Andrew J. Coulson (2010); Steve Aos and Roxanne Lieb and Jim Mayfiel and Marna Miller and Annie Pennucci (2004); Muhlhausen and Lips (2010); Zigler and Styfco (2010); Barnett (2011); Haskins (2010); Klein (2011); Burke and Muhlhausen (2013).

[10] Kline and Walters (2014); Feller et al. (2014) adjust for contamination to evaluate HSIS and find positive results.

Careful non-experimental evaluations on Head Start using very different methodologies also report positive results of the programs in dimensions as childhood achievement, labor market, behavioral and health outcomes (see Currie and Thomas, 1995; Garces et al., 2002; Ludwig and Miller, 2007a; Deming, 2009; Carneiro and Ginja, 2014). The literature generally shows that Head Start is a program with important impacts, at least for children who would otherwise stay at home.

F-7 **The available cost-benefit analyses suggest that investing in ECIs is not only socially fair but economically efficient.** A comprehensive cost-benefit analysis (CBA) includes long-run outcomes to account for life-cycle gains in income, employment, health, etc. Unfortunately, few ECIs have long-term data follow-ups. There are two ECIs with long term follow-ups, PPP and ABC. The CBA of PPP is a long-term comprehensive study accounting for later life outcomes, with enhancements in earnings and reductions in crime as its main components (see Heckman et al., 2010b). It accounts for dead-weight losses of taxation aimed to finance the program and nonetheless still finds an annual rate of return between 7% and 10%, which is above the long-term stock market rate of return on equity of 5.8% in the post World War II era before 2008 (see Heckman et al., 2010b). The CBA for ABC is much reduced in scope (Barnett and Masse, 2007). It is only based on earnings through age 21. Not surprisingly, the rate of return they find is very low, 2.5%. Recent evidence through age 35 suggests that health outcomes and crime reduction are major components of the gains of ABC (Campbell et al., 2014b). Rates of return for ABC are under preparation, and will be reported in the final draft of this paper.

1.1 Plan of the Rest of the Paper

This paper reports an extensive analysis of 5 ECIs with long-term follow-ups: PPP, ABC, the Carolina Approach to Responsive Education (CARE), IHDP, and the Early Training Program (ETP), which we are conducting as part of a larger and broader project. We

document all of their components (nature of the intervention, targeted population, years, ages of participants, and intensity of implementation, etc.), survey their main findings, and investigate the sources of treatment effects. Also, we summarize federal and state ECIs and their respective evaluations.

Blau (2003) presents a comprehensive survey of child-care subsidies. He provides a clear description of the main child-care programs in the US, as well as their eligibility requirements. He tangentially discusses the effectiveness of ECIs but his main outcomes of focus are maternal employment and welfare. This paper centers its attention on the effects of ECIs on skill formation and human development.

We build on previous studies by Currie (2001), Cunha et al. (2006), Blau and Currie (2006), and Duncan and Magnuson (2013) in several ways: (i) we update these studies using newly available data for the same programs covered by previous surveys;[11] (ii) we consider analyses based on new methodological approaches such as mediation analysis, small sample inference, and corrections for attrition and substitution bias; (iii) instead of reusing old results reported extensively in the literature (see Cunha et al., 2006), we use the primary data to do more precise comparisons among ECIs; (iv) we analyze several less-known studies that have been ignored in previous surveys and help generalize lessons from other programs; (v) we explicitly compare Head Start impacts, as obtained in several different careful evaluations, with impacts of the most successful ECIs.

Some recent studies use meta-analysis as their main analytical tool (see Duncan and Magnuson, 2013; Camilli et al., 2010; Nores and Barnett, 2010). Meta-analysis is a useful descriptive technique for presenting overviews of the literature. In our judgment, ECIs are too diverse to pool together as in standard meta-analysis which implicitly assumes a level of comparability among studies that is at odds with reality. Adjustments to create standardization are at best crude. Some of the interventions included in a meta-analysis will have flawed evaluations, which can muddle the results of the exercise. We prefer to focus on

[11] For example, in the case of ABC, we discuss unique results on adult health that had never been covered by a previous survey (Campbell et al., 2014b; Conti et al., 2014)

the few studies where it is possible to analyze and compare different populations.[12]

The structure of the remainder of the paper is as follows. Section 2 shows the basic trends and demographic differences in types of childcare and educational programs take-up. This highlights the disparities between advantaged and disadvantaged populations. Section 3 discusses the main public programs offered in the US, their levels of enrollment, and funding. It focuses on Head Start because it is the only nationwide ECI. Moreover, its history is very rich as it has been the main channel through which the U.S. government supports childcare.

Section 4 analyzes in depth 5 high-quality randomized, controlled ECIs. These are primarily demonstration or pilot projects. Section 5 surveys the literature evaluating publicly provided ECIs. Section 6 summarizes the paper and suggests avenues for future research.

2 The State of Childcare and Early Education Programs in the U.S.

In this section we review the characteristics of childcare and educational programs take-up in the US. We discuss (i) the take-up of different types of childcare in the US; (ii) the take-up of different types of educational programs in the US.

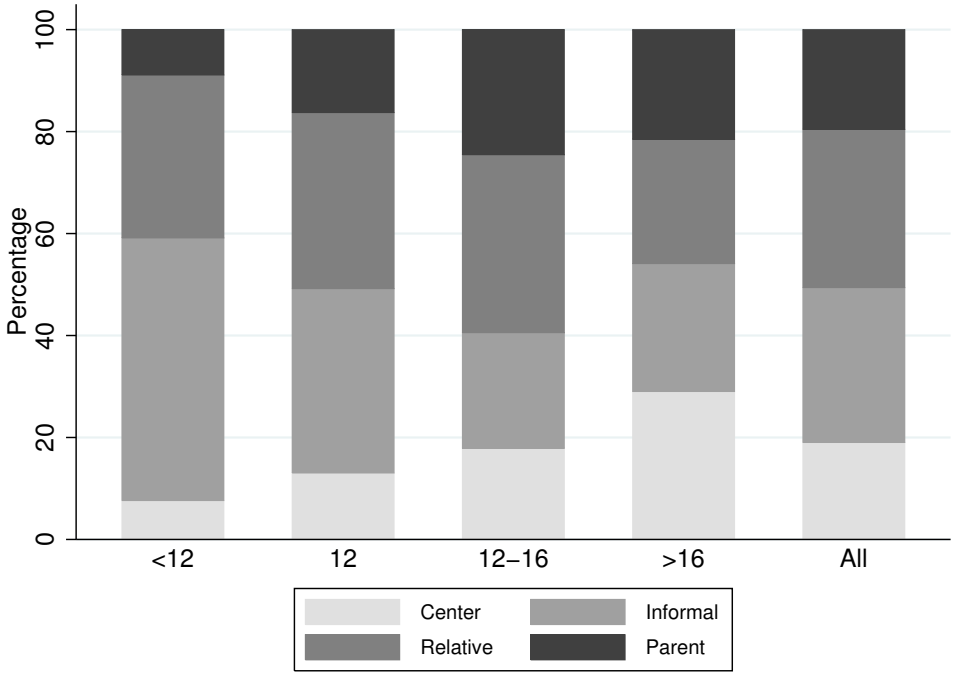
2.1 Demographic Differences in the Use of Time: Childcare Arrangements

We start by discussing where children spend most of their time. To that end, we present Figure 4. It shows the childcare arrangements used for children ages 0-5, taken from the Survey of Income and Program Participation. The trends for the aggregate population are mostly unchanged for all the years available in the data (1997-2011), so we do not show them here. Instead, we show the childcare arrangements disaggregated by mother's

[12] Paraphrasing Cary (1950, p. 98): "one word of know something is worth a million words of know nothing".

educational level. This is especially relevant for our survey because we focus on disadvantaged populations.

Figure 4: Primary Childcare Type by Mother's Years of Education



Source: Survey of Income and Program Participation 2011. Note: Center-based care includes daycare centers, nursery schools, and Head Start. Informal care includes care provided by non-relatives,family daycare homes, and other irregular arrangements.

The type of childcare selected by families varies with their background. Figure 4 shows that over 50% of mothers who did not complete high school primarily place their children in informal care[13], and less than 8% in center-based[14] care. Also, we observe that children from households with more educated mothers spend more time in organized care and less time in informal care. For instance, 30% of children with mothers who completed a bachelors degree or more are primarily cared for at care centers, and fewer than 28% of them report informal care as being the child's primary source of care.

[13]Informal care includes care provided by non-relatives,family daycare homes, and other irregular arrangements.

[14]Center-based care includes daycare centers, nursery schools, and Head Start.

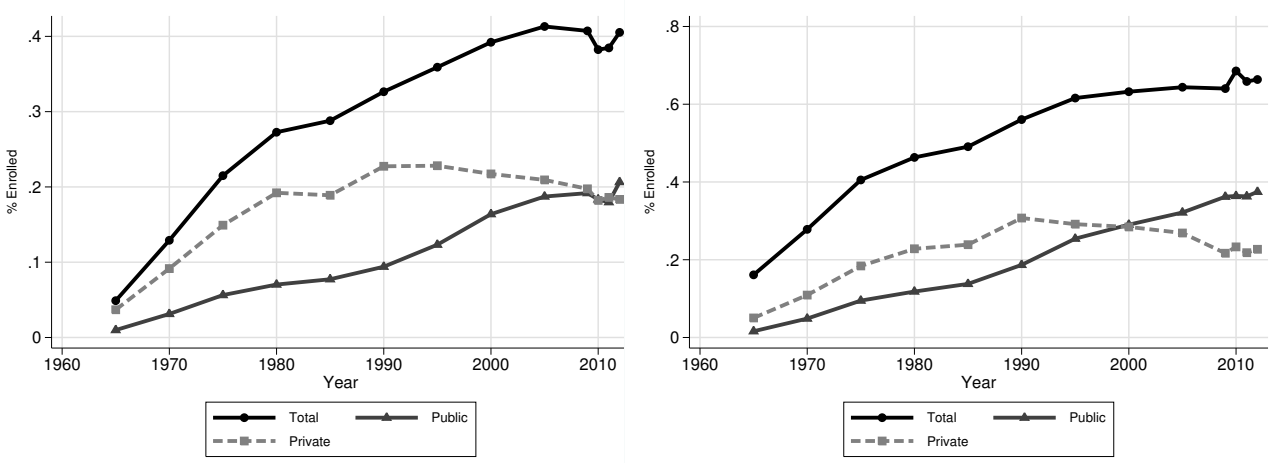
2.2 Enrollment in Educational Programs

Childcare and education do not have an obvious relationship: it is possible that children who are spending most of their time in center-based childcare are not receiving any specific cognitive or emotional stimulation. Moreover, having the mother as the main caretaker for a child does not preclude him from attending a focused preschool program. Even though the different childcare arrangements might have important consequences for the development of a child, in this survey we are mostly interested in specific educational programs. As some of the most important programs in this survey are for children aged 3 and 4, in this section we focus on participation in these programs at these ages.

2.2.1 Trends in the Take-Up of Educational Programs

Figure 5 shows the percent of 3- and 4-year olds enrolled in preschool in the US for every five years since 1965.

Figure 5: Public and Private Enrollment in Educational Programs



(a) Age 3

(b) Age 4

Source: Current Population Survey 2013.

While enrollment has been increasing, more than half of all 3-and 4-year-olds do not use early education – either because they do not have access or by parental choice. With research

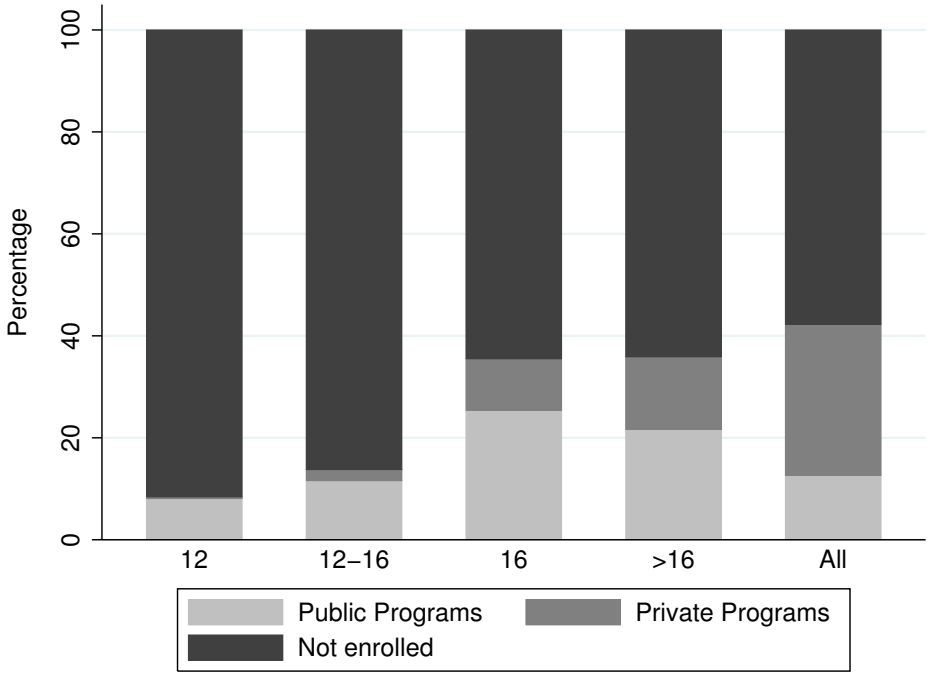
supporting the benefits of early education permeating into public and political consciousness and with the increase in the demand for early childcare brought on by working parents, the U.S. has seen a call for the expansion of ECI funding in recent years. In the last 15 years, enrollment in public programs has overtaken enrollment in private programs.

2.2.2 Demographic Differences in the Take-Up of Educational Programs

Besides the historical trends, we are interested in how participation in educational programs changes by different demographic characteristics. Enrollment has been increasing across different ethnicities, and though enrollment disparities between whites and minorities have decreased, they have persisted, particularly for Hispanics. Part of this difference may be explained by the difference in enrollment by income and mother's education. The rates of preschool participation is 55% for children of high school dropouts, 63% for high school graduates, and 87% for college graduates. The disparities by income are much more striking. The lowest rates of participation at age 4 are for families earning between $20,000 and $30,000 at 55% (lower even than for families earning $10,000 or less, which may be evidence for the success of federal programs targeting low-income families), and the highest rates are for families earning $100,000 or more, at 89%. This gap increases to 51% for participation at age 3 for the same families (Barnett and Yarosz, 2007).

As an illustration of how the take-up of programs is different for more disadvantaged groups, Figure 6 shows these statistics for ages 3 and 4 pooled but disaggregated by mother's education.

Figure 6: Percentage Enrollment of 3- and 4-year-olds by Mother's Years Education



Source: Current Population Survey 2013.

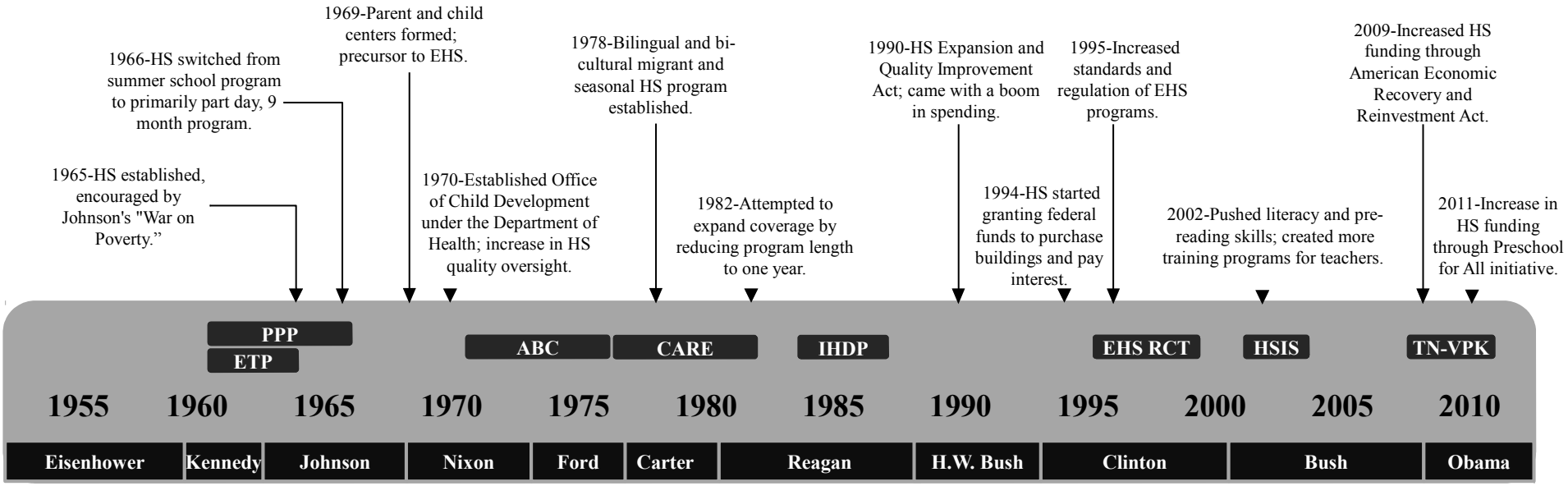
As expected, as mother's level of education increases, so does the overall level of educational program enrollment for their children. Moreover, more educated mothers are more likely to use private programs (in fact, the increase in the shares of children attending private programs gets larger for the more educated mothers). At the time of this writing we are looking for measures of quality of these programs. We expect to find a strong gradient by education status of mothers.

3 The Evolution of the Main Childhood Programs in the US

Having noted the disparities in ECI enrollment, we can now move on to discuss the federal and state governments' efforts to remedy them. We recognize that this is only part of the story as churches, cultural groups, and philanthropists play active roles in providing childcare.

The main public educational programs that 3-4 years-old children in the US attend are Head Start and State Preschool Programs. In this section, we discuss these programs. As we are interested in the very early stages of life, we also consider Early Head Start. We present a comparison of the total enrollment and funding levels of those programs. For each of them, we briefly discuss its description, history and curricula. To give a historical context, we start by presenting a timeline of the most important points of the history of the public ECIs in the US in Figure 7 based on various sources (see Haxton, 2013; Administration for Children and Families, 2008; Zigler and Styfco, 2010).

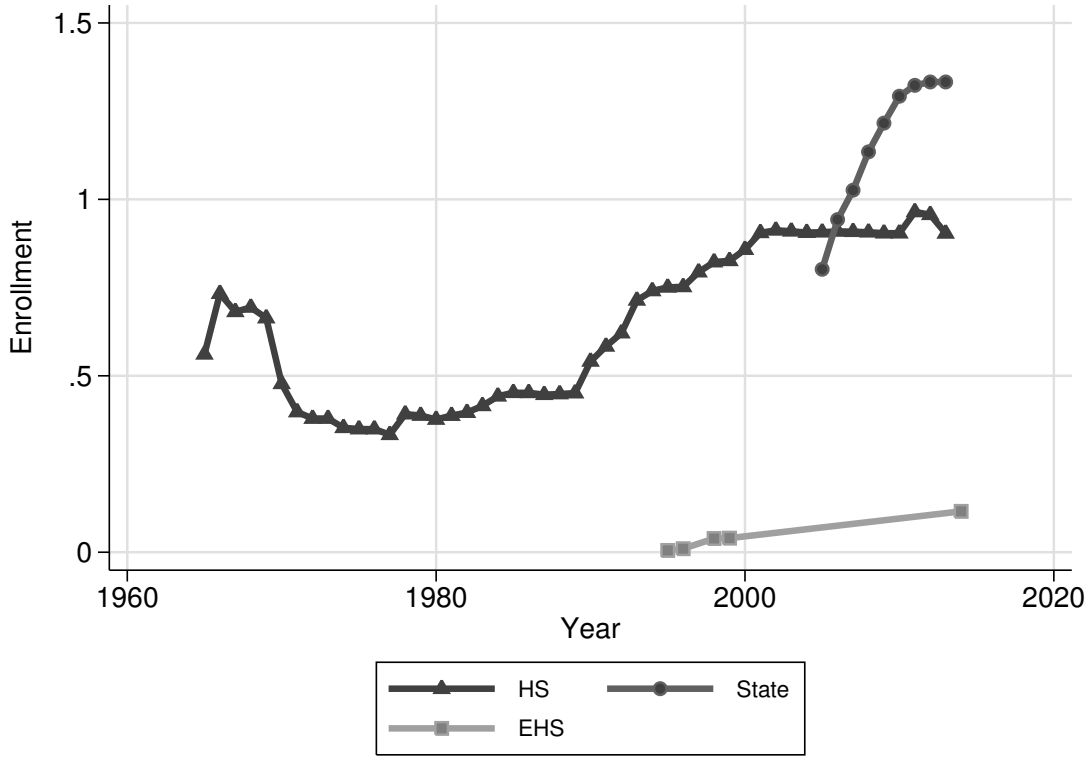
Figure 7: Head Start and Early Childhood Interventions in the US: History and Social Context



3.1 Enrollment Levels of the Different Childhood Programs in the US

The following graph presents the evolution of the enrollment in the main public preschool programs in the US. We combine data from different sources, so we have different lengths of time for our statistics.

Figure 8: Head Start, Early Head Start and State Preschool Enrollment (in Millions of Children)



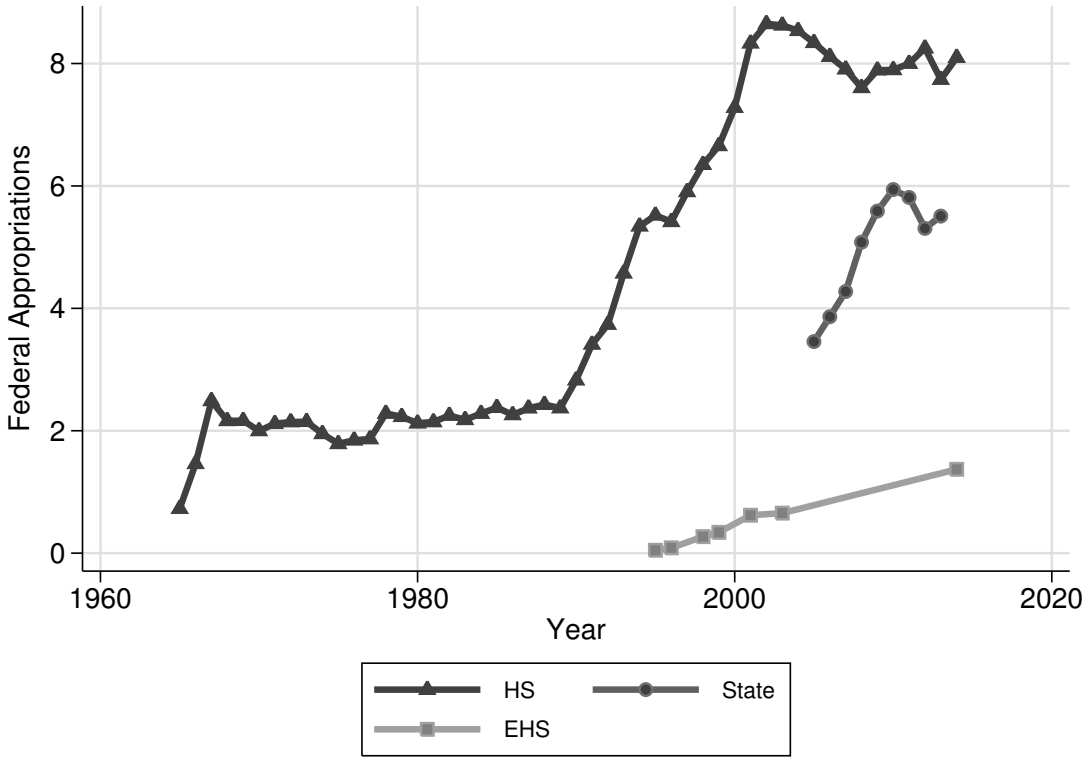
Sources: Love et al. (2004); NIEER State Yearbooks 2005-2013; Office of Head Start (2011). We display the available enrollment data in millions for Early Head Start, State Preschool Funding, and Head Start Funding since 1965.

Nationwide enrollment in Head Start has been steadily increasing over the last 30 years (as has Early Head Start, since its creation), but has recently been surpassed by enrollment in state-funded preschool programs.

3.2 Funding Levels of the Different Childhood Programs in the US

We now present the funding for the main public preschool programs in the US. Again, we combine data from different sources, so we have different lengths of time for our statistics.

Figure 9: Head Start Funding (in Billions of 2014 Dollars)



Sources: Love et al. (2004); NIEER State Yearbooks 2005-2013; Office of Head Start (2011). We display the available funding data in billions of 2014 USD for Early Head Start, State Preschool Funding, and Head Start Funding since 1965.

Federal Head Start funding is significantly higher than the sum of state funding to preschool. Early Head Start funding is approximately 10% of Head Start funding.

3.3 Head Start (HS)

3.3.1 Description

HS is the most widely known and well-funded of the ECIs we discuss. It is a federal program implemented in the United States and its territories to foster cognitive and non-cognitive development and school readiness in low- and middle-income children. It offers grants to agencies who qualify to provide HS services and maintain HS quality and performance standards. Blau (2003) notes that HS is a childcare subsidy program, and examines its structure, recipient characteristics, and potential effects. Our goal in this section is to consider its services and history, but we do so briefly, as the main discussion of HS in this survey has to do with its impacts on participants.

3.3.2 Curricula

The HS curriculum is not a well-defined set of didactic subject-oriented objectives but rather, takes on a "whole child" approach.[15] Educational and support services for parents are emphasized. Program standards explicitly address parental involvement and education about parental skills and make accessible resources that will help them better themselves and their families' outcomes (Finlay et al., 1998).

HS centers must go above and beyond the typical expectations for cognitive stimulation, like language and pre-literacy. Centers must verify their participants' health status and look out for abnormal behavior or signs of mental health problems. It is also the center's responsibility to help families receive the care they need—whether this is by educating parents about health concerns, helping them to find the right health insurance or providing mental health consultation through the program itself. Efforts must also be made to serve families with special needs, such as those dealing with substance abuse issues and home-

[15] According to HS guidelines (see Administration for Children and Families (2010)), HS *curriculum* means a written plan that includes: (i) the goals for children's development and learning; (ii) the experiences through which they will achieve these goals; (iii) what staff and parents do to help children achieve these goals; and (iv) the materials needed to support the implementation of the curriculum.

lessness, even and especially when this means helping families find and coordinate program services with entities outside of HS. There are also standards in place to ensure the quality of food served to children in the program and ensure that families are educated about their children's nutritional needs and ways to fulfill them.

3.3.3 Program History

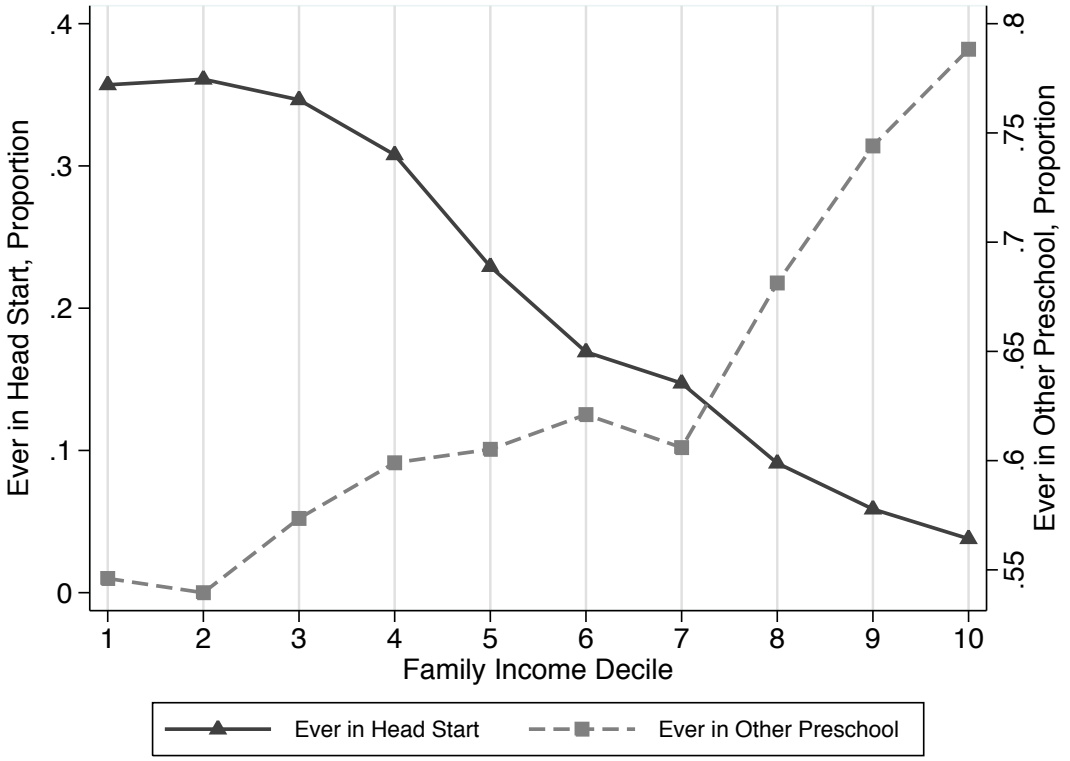
HS was first developed at a time when the outcomes of childhood intervention programs had not yet been widely studied. Thus, the Early Training Project[16] (ETP) at the Peabody College for Teachers became a valuable source of inspiration (Zigler and Styfco, 2010). ETP and other programs like it shifted attention to the importance of environment on child development and acted as an example for how programs can improve the quality of impoverished environments.

At its inception, HS was meant to end poverty in the next generation rather than merely remove individuals from impoverished circumstances.While this goal remains at the core of the program, HS implementation has changed drastically. HS in its early years is not the same as HS of the 80's, and neither of these is on par with HS after the 90's. Notable changes include the implementation of teacher and staff qualification and program quality and compliance evaluations. HS has also developed policies and services to target particular populations, like the homeless and migrant worker populations.

It is important to point out that the HS population tends to be much more disadvantaged than the population of children attending other preschools. Figure 10 shows how the enrollment in those programs changes by level of income of the family.

[16]See Section 4 for a description of ETP.

Figure 10: Enrollment in Head Start and Other Preschool Options



Note: this graph shows Head Start enrollment compared to enrollment in other preschool options. This is an updated version of the graph in Garces et al. (2002), and it is created using the data in the cNLSY survey.

3.4 Early Head Start

3.4.1 Description

Early Head Start (EHS) was created in 1994 and expanded HS program benefits to low-income families with children under age 3. EHS program services may include development services, child care, parenting education, case management, health care (including referral), and family support. Following in the HS tradition, EHS programs also partner with other community service providers to extend their reach and impact. These partnerships are expected to meet the same quality standards as EHS, including child-to-adult ratios and staff educational qualifications (see Vogel et al. (2006)). In fact, EHS programs can be thought of as HS programs that serve families with children under age 3. It has been subject

Table 2: Early Head Start Use by Type of Program

	Description	% Enrollees using as primary form of care Age 1-2	Age 2-3
Home-Based	Weekly visits and group socialization	66%	75%
Center-Based	Care at a childcare center	37%	51%
Mixed	Combination of home-based and center-based	50%	66%

Source: Love et al. (2004). Note: this table shows the percentage of Early Head Start enrollees who use each type of program as their primary childcare.

to the same policy changes as HS and receives its funding as a part of the overall HS budget.

EHS services fall under three main service delivery options—home-based (which includes weekly home visits and group socialization), center-based, and mixed-approach—that grantee organizations can mix and match to suit their participating families' needs. Surveys suggest that families would have liked to use more childcare (Love et al., 2004). Indeed, families typically needed and used care during nonstandard hours (and this seems to be the case across families enrolled in programs using any of the delivery methods). In fact, the primary child care arrangement for 34% of children using EHS included evening hours, for 21% it was during weekend hours, and for 16%, primary care was provided overnight.

EHS was established after HS increased its criteria for staff qualifications and follows its regulations. Moreover, following HS precedent, EHS supported multi-faceted services in cognitive, non-cognitive, health and parental skill development areas. Many programs sought accreditation, including through the National Association for the Education of Young Children accreditation system. Programs also initiated systems to monitor quality for both on-site childcare and for care in community settings. This included assessing individual classrooms and helping teachers take steps to improve quality.

3.5 State Preschool Programs

In addition to receiving federal funds for ECIs development, most states also have their own programs with substantial funding coming from state resources. Consistent with Figure 8, National Institute for Early Education Research (NIEER) studies show that more children are enrolled in state-funded preschool than in any other publicly-funded ECIs: 28% of 4-

year-olds are enrolled in state-funded programs, 11% in HS, 3% in other public programs, and 3% in special education, not including special education children who are also enrolled in state-funded pre-K or Head Start. (NIEER, 2013; Barnett and Yarosz, 2007)

There has been tremendous growth in state-funded programs over the last twenty-five years. In 1980, only 4 states subsidized any preschool programs, and by 1987, this number grew to 11. By the mid-nineties, fifteen states subsidized preschool. This number has steadily grown through 2014. Currently, 4% of 3-year-olds and 28% of 4-year-olds in the U.S. are enrolled in some kind of state pre-k program, with 40 states and Washington D.C. offering a total of 53 state-funded preschool programs. Twenty-six states have programs only for 4-year-olds, and ten states have no programs: (i) Hawaii, (ii) Idaho, (iii) Indiana, (iv) Mississippi, (v) Montana, (vi) New Hampshire, (vii) North Dakota, (viii) South Dakota, (ix) Utah, (x) and Wyoming (NIEER, 2013). These programs are typically evaluated on the basis of their enrollment statistics and quality of their programs.

Among the states that offer some preschool program, only nine (and Washington D.C.) had greater than 50% enrollment of 4-year-olds in 2014: (i) DC (94%) , (ii) Florida (78%), (iii) Oklahoma (74%), (iv) Vermont (71%), (v) Wisconsin (64%), (vi) West Virginia (62%), (vii) Iowa (60%), (viii) Georgia (58%), and (ix) Texas (52%). Of these, only four meet at least half of the quality requirements (based on teacher credentials, class size, staff-to-child ratio and other criteria) set forth by the NIEER. Florida and Vermont fare poorly, meeting only three and four of the criteria, respectively (NIEER, 2013). The Georgia and Oklahoma universal programs are praised for their balance of quality and cost. They score an eight and nine respectively on the NIEER's quality index.

The largest federal fund to finance state preschools is the Child Care Development Fund (CCDF). This fund does not specifically support ECIs, but it allocates funds to states to provide grants for low-income and special needs families to subsidize childcare. Much freedom is given to states to define low-income and special needs. However, they serve families whose income level does not exceed 85% of the State median income for a family of the same size,

and families may choose any childcare provider that passes state regulations (though no curriculum quality standards) for any children under age thirteen. States must also spend funds educating families about their healthcare decisions (Administration for Children and Families, 2013).

4 Results from Randomized Controlled Trials of Early Childhood Interventions

There have been several previous studies on the effects of ECIs by economists (see Currie, 2001; Cunha et al., 2006; Blau and Currie, 2006; Duncan and Magnuson, 2013). Duncan and Magnuson (2013) present a meta-analysis of the impacts on IQ (or achievement) for 84 interventions. They find that (i) the impacts of PPP and ABC are on the high end of the spectrum of effect sizes (average .35 Standard Deviations near the end of the program); (ii) programs designed by researchers had higher effect sizes; (iii) older programs had higher effect sizes, which they attribute to higher-quality home environments and increases in other forms of center-based care; and (iv) there is no clear pattern on the duration of program or starting age. They explain that the evidence on program components is still weak, but that there are hints that teaching quality might be more important than structural variables (class size, teacher education, etc.). In this survey, we dig deeper and explicitly discuss the differences in curricula between ECIs and cite new evidence about their importance.

We generally share the perspective of these previous studies, but substantially expand their analyses. In particular, we focus much of our attention on two programs: the Perry Preschool Program (PPP) and the Carolina Abecedarian Project (ABC). We obtain the following conclusions across the two programs: (i) IQ score gains fade out in PPP but persist in ABC, which began when the children were younger; (ii) both programs have strong but different impacts on adult outcomes; (iii) PPP easily passes a cost-benefit test. These conclusions are revisited in this section.

We proceed as follows. First, we describe the historical context of each program and its components. Then, we survey the evidence from the literature. After reviewing all of the programs, we propose a methodology to perform our own evaluation and provide evidence supporting the facts in Section 1.

Only a handful of studies may be classified as long-term RCTs of early educational interventions. To the best of our knowledge, the only two experimentally evaluated ECIs implemented in the United States that have been followed for enough time to present reliable estimates of at least their labor market effects are ABC (and the closely related CARE project) and PPP. Through their long-term follow-ups, we can measure their impacts on labor market outcomes, crime, and health in adulthood. A few other interventions have followed individuals after they finish school, but many of their participants were still enrolled in educational programs when last surveyed. Although we cannot learn about these programs' impacts on labor market outcomes, they allow us to study an array of relevant results such as high school completion, college enrollment, and idleness. To the best of our knowledge, the only two experimentally evaluated educational ECIs implemented in the US with follow ups up to those ages are the Infant Health and Development Program (IHDP) and the Early Training Project (ETP). In this section, we use comparable methodologies and samples, and correct for several statistical problems that are present in the earlier literature. As we understand it, all other US-based RCTs of early educational interventions with long-term follow-ups have design or attrition problems that make them unsuitable for obtaining long-term conclusions.

We begin by summarizing and comparing PPP, ABC, CARE, IHDP, and ETP in Table 3. We also summarize the baseline characteristics of the sample for each ECI in Table 4.

Table 3: High-quality Early Childhood Interventions, Summary Table

	PPP	ABC	CARE	IHDP	ETP
Program Overview[1]					
Implemented years	1962–1967	1972–1985	1978–1985	1985–1988	1962-1968
# Cohorts	5	4	2	1	2
N (Treatment:Control)	123 (58 : 65)	111 (57 : 54)	64 (41 : 23)	985 (377 :608)	88 (43 : 45)
Age of Entry	3–4	0	0	0	3–4
Duration	1–2 years	3–8 years	5 years	3 years	2–3 years
Treatment					
Home visits (per month)	4	1–2	2.5–2.7	0.5–1	4
Center care (weeks per year)	30	50	50	50	10
Center care (hours per week)	12–15	45	30	20+	20
Parent involvement		✓	✓	✓	
Nutrition		✓	✓		
Diapers/Child Care Goods		✓	✓	✓	
Health Check-ups		✓		✓	
Medical Care		✓	✓	✓	
Counseling			✓	✓	
Parenting Instruction	✓		✓	✓	✓
Job Counseling					
Control[2]					
Nutrition		✓	✓		
Diapers		✓	✓		
Health Check-ups				✓	
Medical Care				✓	
Randomization Protocol[3,4]					
Adjustment Factors	Balanced gender ratio Balanced on % working mothers		Lower HRI criteria for diversity	Birth weight & sites Gender Maternal education Maternal race Primary home language Participation in another study	
Compromises	Siblings receive same assignment Working moms switched to control	2 extremely needy switched to control 1 passed away each in control and treatment			
Counterfactual		Stay at home or childcare		Day Care (40 % low low birth weight)	No access to summer schooling
Site selection		Newly Built Center (FPGC) in NC	Samples live near FPGC	Competitive review	Segregated black schools
Program Eligibility[5]	Cultural Deprivation Scale < 11 Low IQ (< 85)	High Risk Index > 11 Biologically healthy No signs of mental retardation	High Risk Index > 11	Live within 45 min from center Birth weight < 2500g Gestational age < 37 weeks No severe illnesses or neurological defects	Home environment Education of parents Occupation of parents
Curriculum[6]					
Adult-Child Ratio	1:5–1:6	1:4–1:6	1:31:6	1:3–1:4	1:4–1:6
Staff & Certifications					
Teachers	B.A.$^{\oplus}$	HS grad–M.A.; experience with kids	HS grads$^{\circ}$	College grads	$^{\circ}$
Specialists	Research staff, Ph.D. Special Ed. Teachers$^{\oplus}$	Doctors Social Workers	College grads$^{\circ}$ M.A.$^{\circ}$	Education directors, M.A. Clinical staff	Teaching Assistants, college & PhD students Home visitors$^{\oplus\circ}$
Language Development	✓	✓	✓	✓	✓
Motor Development		✓	✓	✓	
Cognitive Development	✓	✓	✓	✓	✓
Socio-Emotional Development	✓	✓	✓	✓	✓
Task Orientation		✓			✓
High-Risk Behavior		✓			
School Preparation		✓	✓		✓
Costs (2014 USD)	$20,911	$88,737		$22,187	

Sources for this table are as follows. PPP (Weikart et al., 1964, 1978; Weikart, 1970; Schweinhart et al., 2005). ABC (Campbell and Ramey, 1994; Ramey et al., 1979; Masse and Barnett, 2002). CARE (Wasik et al., 1990; Burchinal et al., 1997; Campbell et al., 2008). IHDP (Brooks-Gunn et al., 1994a,b; McCarton et al., 1997; Ramey et al., 1992). ETP (Gray and Klaus, 1965; Klaus and Gray, 1968; Gray and Klaus, 1970). [1] In ABC, at Preschool period: Group E consists of 57 children who received a 5-year preschool intervention. A second randomization resulted in Group EE (25): who received another 3-year school-age intervention; EC (24): who did not receive the school-age intervention, but received the 5-year preschool intervention; CE (24): who did not receive the preschool intervention but underwent 3-year school-age intervention. At Preschool period: Group C consists of 54 children who received no preschool intervention. A second randomization result in Group CC: 23 children who did not receive the school-age intervention program. In the treatment group of CARE, 15 children received both child care services and family education (referred to as Ch+H); the 26 remaining children received only family education (referred to as H).The two cohorts of CARE lived in or near a Southeastern university town, and were born between the spring of 1978 and early 1980. CARE originally selected 65 families with 67 children. 64 children are included in the final sample. In IHDP, an additional 105 twins were also followed in the study, but are not analyzed in the literature. These twins were assigned to the same treatment group as their siblings. For each site, the program lasted until the youngest child turned 36 months old, correcting for prematurity. In ETP's treatment group, 22 of the children received three rounds of summer preschool, in addition to three years of weekly meetings with home visitor (these children are referred to as T1) ; the remaining 21 children in the treated group received two rounds of summer preschool, in addition to two years of weekly meetings with a home visitor (these children are referred to as T2). T1 received their first year of summer school at age 3. T2 received their first year of summer preschool at age 4. [2] The control group of the first cohort ABC received health check ups for the first year, after which this practice was discontinued. [3] The randomization protocol was the following. PPP: (1) Matched on C.D. scale and SB scores; (2) Rank by IQ, separate even and odds; (3) Randomly assign even and odd to treatment and control. ABC: (1) Identified candidates before birth; (2) Pair matched on gender, maternal IQ, number of siblings, high risk index (the High Risk Index used in ABC and CARE is calculated through a weighted average of parents' age, education levels, family income, mother's IQ, father absence, poor school performance of siblings and seven other factors); (3) Randomly assigned one to control, other to treatment. Care randomly assigns 65 families using the ABC protocol. IHDP: Randomized by the National Study Office using an adaptive randomization model. IHDP balanced the treatment groups across two birth weight strata: higher low birth weight (2000g–2500g), and lower low birth weight (< 2000g). ETP: 61 Children from Abbotfield, TN were randomized into three groups, two treatment groups and one control group [4] In PPP, home visits were intended to involve the mother in the educating the child, increase her understanding of the educative process, and to extend the curriculum beyond the classes and into the homes. Monthly group meetings for parents was also available, but is not well documented. In CARE, the reported figures are the amount of home visits: 2.5 visits per month for the H; 2.7 visits per month for Group Ch+H. The original design was to have weekly home visits. During IHDP home visits, families in treatment groups were given toys with instructions on how to play with their child with the toys. This was to extend the curriculum beyond the classroom. Home visits also sought to improve the parents ability to problem solve, cope with personal issues, and function as parents. In addition, parent groups were offered as a chance for parents to share information and concerns with each other, and to provide them with the opportunity to learn about child education and community resources. Surveys were conducted by college graduates. In ETP, T1 parents received two 9-month training sessions, while T2 parents received one 9-month training session. During these training sessions, the objective of the intervention was made clear to mothers during visits to schools. Mothers were encouraged to engage in their children's learning, as well as to expand the experiential environment of the child (e.g. trips to the library). [5] In PPP, criteria for home environment included education of parents, occupational level of father, maternal employment, and household density. [6] $^{\circ}$ signifies that staff were specially trained for the program. $^{\oplus}$ signifies that staff were state certified.

Table 4: Background Characteristics at Baseline, All the Programs

	PPP		ABC		CARE		IHDP		ETP	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Black	1.00	0.00	0.97	0.16	0.15	0.36	0.37	0.48	1.00	0.00
IQ	79.02	6.44	90.42	11.46	94.26	12.96	88.00	20.16	87.33	12.53
Mother's Age	29.10	6.57	19.89	4.82	21.68	4.73	24.87	6.00	29.71	8.49
Mother's Education	9.42	2.20	10.23	1.84	10.97	1.54	12.41	2.43	9.02	2.56
Mother Works	0.20	0.40	0.73	0.45	0.31	0.47	0.34	0.47	0.44	0.50
Father at Home	0.53	0.50	0.29	0.46	0.23	0.42	0.56	0.50	0.86	0.35
Father's Age	32.81	6.88	23.21	5.91	24.46	5.95	27.64	6.67	32.27	9.06
Father's Education	8.60	2.40	10.95	1.76	11.03	1.76	13.16	2.89	9.45	2.80
Father Works	0.86	0.35	0.87	0.34	0.82	0.39	0.57	0.50	0.98	0.15
# Siblings	4.28	2.59	0.64	1.10	0.66	0.94	1.02	1.17	3.51	2.16
Treatment	0.47	0.50	0.52	0.50	0.42	0.50	0.39	0.49	0.39	0.49

Note 1: initial sample sizes are: **PPP**: 123; **ABC**: 111; **CARE**: 64; **IHDP**: 985; **ETP**: 88.
Note 2: mother and father's years of education are counts of the number of years of schooling completed by the mother and father, respectively, at the time of program entry. We generate an indicator variable that evaluates to 1 if the mother is working at the time of program entry, and 0 otherwise. We do the same with the fathers. We also generate an indicator variable that evaluates to 1 if the father lives at home at the time of entry, and 0 otherwise. The number of siblings is reported at program entry. **PPP**. Child's IQ at age 3 is measured using the Stanford Binet Intelligence Scale. **ABC**. Child's IQ at age 2 is measured using the Stanford Binet Intelligence Scale. Mother's age is reported at the time of program entry. **CARE**. Child's IQ at age 2 is measured using the Stanford Binet Intelligence Scale. Mother's age is reported at the time of program entry. **IHDP**. Child's IQ at age 3 is measured using the Stanford Binet Intelligence Scale. **ETP**. Child's IQ at age 4 prior to entry into summer school, and we measure this using the Stanford Binet Intelligence Scale. Mother's age is reported at the time of program entry, when the child is age 4.

4.1 Perry Preschool Project (PPP)

4.1.1 Historical Context

Ypsilanti, Michigan, is a small town about 35 miles west of Detroit. Ypsilanti's proximity to Detroit's industry positively affected the economic opportunities for residents of Ypsilanti. In the 1960s and 1970s, when the subjects of the program were young, the availability of manufacturing sector work expanded, incentivizing individuals (especially males) to forgo school for more lucrative work in the factories around South East Michigan.[17] Although the Civil Rights Movement was at its apex during these years, and despite the large black population throughout South East Michigan, blacks continued to struggle in the area socially and economically.

[17] Heckman et al. (2010a, p. 36).

4.1.2 Educational Content

The educational content of PPP is described in detail in Heckman et al. (2014): four teachers state-certified in either elementary or special education planned structured learning activities for 23-26 students per year. Children experienced two years of experimental methods of instruction emphasizing active child learning for cognitive and language development within an intentionally teacher-structured and resource rich environment. The teachers offered individualized instruction to students, as needed, during the home visit. (Weikart, 1967, 1970). In the last years of PPP, the children were introduced to a more formal daily schedule that emphasized child-planned activities followed by a brief group time in which teachers guided students in evaluating their previous activities.[18] PPP teachers promoted cognitive and socio-emotional development as equally important aspects of the program.

4.1.3 Program Description

PPP might be the most well-known experimental preschool program and has been evaluated in an extremely rigorous way, as it was assigned randomly to a treatment and a control group; it has been followed for 40 years now and its rates of attrition are exceptionally low for this type of program.[19] The intervention was intensive, but not very different in terms of material inputs and human resources than some of the high quality early childhood programs that exist nowadays. Table 3 shows that the program focused on black children aged 3-5, and only ran for two years for around 3 hours every day. Although not all teachers had college degrees, they had experience and expertise in teaching. The student-teacher ratios were around 6:1, which is not unusual for high-quality preschool education programs. It also included weekly home visits by staff members. Apart from it s experimental design and low attrition, an important characteristic of PPP is that children in the control group had no access to substitute preschool programs. This could be part of the explanation to the

[18] These activities were later called Plan, Do, Review by HighScope.

[19] Small departures of the randomization were assessed in Heckman et al. (2010a); the attrition rates were less than 9% at age 40, which is much less than usual for interventions of this type.

relatively higher impacts that are observed in PPP compared with other programs in this survey (Weikart, 1970).

4.1.4 Evidence from the Literature

PPP has been evaluated many times by many analysts. Here, we rely mostly on Heckman et al. (2010a) for adult outcomes. That paper (i) accounts for compromises in the randomization of the program, such as non-random switching between groups; (ii) uses inference methods that control for size in families of variables tested, thus being robust to multiple hypothesis testing; (iii) uses permutation tests that are tailored to the randomization procedure of the program and that are exact, even in small samples.

Heckman et al. (2010a) report results by gender. For females, they find impacts of 56% on high school graduation and an impact on years of education completed of a little more than one extra year. They also find that the program reduces the total number of arrests up to age 40 by almost three. The probability of being unemployed during the last year at age 27 is reduced by 30%. No significant impacts on health or earnings are found. For males, no effects on education or health are found, but in terms of crime, they find a large reduction of almost 5 lifetime arrests. Treatment group individuals are 17% less likely to have been unemployed during the last year at age 30. Finally, the monthly earnings in the current job are increased by a little more than $1,000 in 2006 dollars at age 27 for the treated group, which is an increase of 70% compared to the control group.

Weikart (1970) gives an early analysis of the impacts on cognitive and non-cognitive skills. In terms of non-cognitive skills, that paper presents a few positive results, but lacks a coherent measurement scheme. It also describes for the first time the well-known pattern of impacts on IQ of PPP: there were very substantial impacts of the program in all types of IQ tests during the time children participated in the preschool program. In the case of PPP, the magnitude of those impacts was around a whole standard deviation. However, during the first years of elementary school, those impacts had dissipated to the point where there

was no significant difference between the groups. This is one of the clearest known examples of the fadeout phenomenon, which is widely observed in preschool interventions. This is illustrated in Figure 2, presented in the introduction to this paper.

These puzzling patterns sparked a new question in the literature: why did this program have so many substantial impacts on adult outcomes if the program did not have long-lasting impacts on IQ? This question was recently tackled by Heckman et al. (2013). In that paper, the impacts of PPP are decomposed to study the different contributions of different gains in skills to the final impacts on outcomes, using mediation analysis. They find that a substantial part of the impacts of the programs are explained by previous gains on non-cognitive skills. In particular, they find that for males, 40% of the total reduction in lifetime arrests and 20% of the reduction of unemployment at age 40 are explained by gains in externalizing behavior. On the other hand, for females, 30% of the gains in achievement tests, and 40% of the decrease in unemployment at age 27 are explained by gains in academic motivation. Consistent with males, up to 65% of the decrease in lifetime violent crimes is explained by a reduction in externalizing behavior.

The costs and benefits of PPP are evaluated in Heckman et al. (2010b). This report improved on several aspects of previous CBAs, including accounting for compromised randomization, presenting standard errors for the rate of return estimates, and accounting for deadweight loss. For comparability with other programs, we use no deadweight loss when reporting their results. We present disaggregated results for this paper in Table 7. The results are very positive, with a benefit-cost ratio of 8.6.[20] Benefits to the parents of the children (free child care, more labor market participation) are not considered, and they form a substantial part of the benefits in the case of other programs, so this evaluation might be conservative in estimating benefits.

[20] A very highly cited finding of this paper, the 7:1 benefit-cost ratio, is calculated using positive deadweight loss estimate.

4.2 The Carolina Abecedarian Project (ABC)

4.2.1 Historical Context

ABC (and CARE, in the next subsection) took place in Chapel Hill, North Carolina, a semi-rural municipality with a strong university presence. During the 1970s, proponents of African-American rights worked to build on the legal victories of the Civil Rights Movement, sometimes in the face of opposition from local government and community leaders (Hall, 2005, p. 1235). Although North Carolina was no exception, in 1969, before the youngest children of the study were born, Howard Lee was elected the first black mayor of Chapel Hill (Jerome, 1995, p. 668). This progressive occurrence among others, such as the prominence of the tolerant Frank Porter Graham (the president of the University of North Carolina), allowed the Chapel Hill area to forestall the amelioration of its racial issues (Chafe, 1981, p. 239). Children were selected based on a vulnerability score index. Most children were living with single mothers, and almost all of them were black.

4.2.2 Educational Content

The educational content of the program is described in detail in Heckman et al. (2014): teachers demonstrated a wide variation in education and included high school graduates and college graduates holding a B.A. or M.A., and an overall average of 7 years of professional experience with children. Staff training in the ABC approach featured weekly, structured supervision led by Ph.D and graduate level early childhood educators, clinical psychologists, and developmental psychologists. Instructional methods were designed and formally evaluated with each successive ABC cohort. ABC's enriched caregiving and play-based learning activities (Sparling and Lewis, 1979) were expressly designed to enhance adult-infant interactions that support children's language, motor, cognitive development and social-emotional competence, including task-orientation, and minimize infants' maladaptive, high-risk behaviors (Sparling, 2010; Ramey and Campbell, 1979; Ramey and Haskins, 1981). Organized

formal centers enabled structured and unstructured child-driven learning, and private spaces for "alone time" in the full-day program were valued. The daily schedule featured dramatic play, reading, art, music, nature, housekeeping, and science.

4.2.3 Program Description

ABC is a well-known and highly studied intervention that took place between 1972 and 1977 in Chapel Hill, NC. ABC is the most intensive of the experimental interventions we study. It provided day-long high-quality childcare and medical services from ages 0-5 and a home-visiting program until age 8. Additionally, children were transported to the program site by bus drivers and received two meals and a snack daily, along with regular health check-ups (Ramey et al., 1979; Campbell and Ramey, 1994).

4.2.4 Evidence from the Literature

The Abecedarian Project data has been used in dozens of papers. It is the only major randomized ECI that had general long-lasting impacts on IQ (Hojman, 2014). For example, Campbell et al. (2002) report a significant difference between treatment and control groups of 5 IQ points at age 21. The last general evaluation of adult outcomes is Campbell et al. (2012). They find positive impacts of 1.2 extra years of education; an increase of 17% in high school graduation; a non-significant difference in yearly earnings of $16,803 2014 dollars per year (50% more than the control group mean); and no impacts on crime.

A unique feature of the ABC data is that around age 34, a new round of data was gathered and medical tests were performed on program subjects. The results are discussed in Campbell et al. (2014a) and Campbell et al. (2014b). The sample for this study is smaller than in the normal ABC data due to an attrition rate of around 30%. Using testing procedures that are exact in small samples; controlling for size in multiple hypotheses groups; and correcting for attrition, they find significant treatment effects for males in blood pressure, hypertension and various measures of combined heart risk. For females, they find impacts

on abdominal obesity (21% difference), one combined risk measure and on one measure of prehypertension.

Barnett and Masse (2007) evaluate the costs and benefits of the ABC program. Benefits are estimated from (i) earnings from participants; (ii) earnings from future generations; (iii) maternal earnings; (iv) school education savings; (v) improved health; (vi) higher education costs; and (vii) welfare use. All of these benefits are compared with those of PPP and another intervention in Table 7. The costs of the program are calculated using a mix of reports of the implementation of the program and estimates for the costs of inputs. The cost of preschool for children in the control group is discounted (including the cost of parental care), but no deadweight loss associated with extra cost is considered. Using a 3% discount rate, their total benefits per child are $94,802 in 2002 dollars. The program benefits-cost ratio is estimated at 2.5:1. This is positive, but considerably worse than the estimate for PPP. However, the benefits considered by Barnett and Masse (2007) only include health benefits related to smoking, and do not include costs related to arrests. Moreover, their earnings are measured at age 21, so it is very likely that these benefit estimates are underestimated.

4.3 Project CARE: The Carolina Approach to Responsive Education (CARE)

4.3.1 Historical Context

See Section 4.2.

4.3.2 Educational Content

The educational content of the program is described in detail in Heckman et al. (2014): Home visitors encouraged enriched caregiving and developmentally appropriate adult-child play-for-learning interactions, and further helped parents identify suitable materials already in the home for these activities. Home visitors also identified and referred families to community

social work agencies as needed (Wasik et al., 1990; Burchinal et al., 1997).

4.3.3 Program Description

Project CARE, created and implemented by the same organization as ABC, began immediately after ABC ended. It compared the efficacy of two service delivery models: one group received a similar center-based treatment as ABC, adding weekly home visits from the child's teacher to the family. In the other group, home visitors made weekly visits but center-based child care was not offered (Wasik et al., 1990). In both approaches, the frequency of the home visits decreased as children aged. In this paper, we only compare the high intensity group with the control group in our estimations for simplicity, and because we focus on interventions with a center-based component. An issue with CARE is that the intervention group had between 12 and 25 families—it is hard to form conclusions based on such a small sample.

The literature on CARE is not nearly as extensive as the one on ABC. Maybe the two most interesting studies are Wasik et al. (1990) and Campbell et al. (2008). The first presents outcomes up to 54 months of age, focusing on IQ measurements. The patterns of the impacts are surprising: the scores of the intensive treatment group are consistently higher than the other two groups, but the control group is consistently higher than the home visit group. One possible reason is that the control group children attended other preschool centers slightly more than the home visit group. However, the differences are not significant, perhaps due to small sample size. The latter paper finds treatment impacts of the intensive intervention in college attendance when ABC and CARE samples are pooled.

4.4 Infant Health and Development Program (IHDP)

4.4.1 Historical Context

IHDP took place at eight universities in the following towns: Little Rock, Arkansas; Bronx, New York; Boston, Massachusetts; Miami, Florida; Philadelphia, Pennsylvania; Dallas,

Texas; Seattle, Washington; and New Haven, Connecticut (Brooks-Gunn et al., 1994b, p. 3035). Although the specific social and economic atmosphere varied from city to city, the national struggle for economic and social inclusion of African-Americans was prominent even after the legal victories of the Civil Rights Movement (Hall, 2005, p. 1235).

4.4.2 Educational Content

The educational content of the program is described in detail in Heckman et al. (2014): IHDP adapted ABC/CARE's treatment for families of low birth weight infants. Paralleling ABC's adult-child interactive learning activities for infants and toddlers, *Early Partners*[21] and *Partners for Learning*[22] were incorporated in home visits and in the child development centers (see Sparling and Lewis, 1979; Ramey and Ramey, 1998; Brooks-Gunn et al., 1994b; Ramey et al., 1992). As in CARE, the IHDP home visiting treatment featured the *Parent Problem Solving* approach designed to help parents cope with the unique responsibilities of caring for a vulnerable child and included information specific to the developmental challenges faced by low birth weight children (Ramey et al., 1992; Sparling et al., 1991).

4.4.3 Program Description

IHDP was a multi-site randomized intervention designed as multiple replications of ABC implemented in 1985. It was implemented by schools of medicine and hospitals in 8 different sites across the United States. Maybe its most unique characteristic is that instead of targeting children based on socioeconomic status, it targeted premature infants ($<$ 37 weeks of gestational age) with low birth weight. The creators of the program distinguished two different strata from the beginning of the program: the lighter low birth weight (LLBW) group ($\leq$ 2000 grams) and the heavier low birth weight (HLBW) group (2001-2500 grams).

The intervention was similar to the intensive treatment of ABC and CARE: it included center-based care, weekly or bi-weekly home visits, and medical services. However, it only

[21] For infants aged 24 to 40 weeks.
[22] For infants and toddlers aged 41 weeks to 36 months.

lasted for the first three years of life, not for five, as in ABC or CARE.

The literature on IHDP has usually obtained estimations separately for the two weight groups (Brooks-Gunn et al., 1994a; McCormick et al., 2006), obtaining stronger impacts for the HLBW group. As an example, the age 18 impacts of IHDP reported on McCormick et al. (2006) show negative but mild impacts of the intervention for the LLBW group (3 points in verbal IQ, 4 in reading achievement). Yet, there are positive effects of the intervention for the HLBW group (5 points in verbal IQ, and in reading achievement, and 6 in math achievement).

In this paper, we present only one general estimate for the program. Moreover, we use all the children in the sample, including those who are not part of low-income families. Having middle class families is a unique feature of IHDP that sets it apart from all of the RCTs that we consider in this survey. As opposed to the previous evaluations, the pooled results we present, evidence little effectiveness of the program.

To understand these impacts, it is also relevant to mention that different studies have pointed out that the relatively smaller impacts of IHDP may be due to its focus on non-economically vulnerable populations. In the first of those studies, Brooks-Gunn et al. (1992) discuss impacts on IQ scores at three years of age. They find that children of mothers with a college degree did not increase their scores, while children of mothers with high school degrees or less did. They also study whether the impacts were different by birth weight, and they find that in several of the demographic groups, it did not matter, and in the one group that mattered, heavier birth weight babies benefited more from the program. These two results point to the selection mechanism of IHDP, operating towards negative treatment effects. A recent paper, Duncan and Sojourner (2013), focuses on the income level of mothers and finds similar results: for the HLBW sample, the program has a significant and strong impact in IQ at ages 3 and 5 for children from disadvantaged families. Disadvantaged children gain 1.32 and 0.87 standard deviations more than higher-income children, who gain 0.32 and -0.26 respectively at ages 3 and 5. For higher-income children these gains are non-significant, but

for low-income children they are highly significant.

4.5 Early Training Project (ETP)

4.5.1 Historical Context

ETP included subjects from two cities in rural Tennessee during the early 1960s where both legal and *de facto* segregation were institutions despite previous victories on civil rights. Abbotfield, where ETP took place, was rural and poor, and although this poverty included whites, residential and school segregation existed along racial lines. In 1970, about 15% of the city's population was black. After World War II, the expansion of the industrial sector provided opportunities for blacks, but it was not until 1964 with the passage of the Civil Rights Act when black workers were included in the actual manufacturing process instead of confined to the service sector (see Gray et al., 1982).

4.5.2 Educational Content

The educational content of the program is described in detail in Heckman et al. (2014). ETP promoted cognitive, language, and social-emotional development to offset "cultural deprivation". Teachers positively reinforced children's attitudes and efforts with tangible rewards, verbal approval, and positive body language. To foster motivation, children were encouraged to persevere, compete, improve, and attempt new and challenging tasks. Teachers cultivated effective social interactions by empowering children to use spoken language to attain goals. Teachers maximized learning opportunities with the selection of racially-aligned classroom materials, such as dolls, pictures, and puzzles featuring black people (Klaus and Gray, 1968). ETP researchers intentionally hired male classroom staff to serve as role models for children whose own fathers were absent. Emphasizing the role of the parent as the child's most important teacher, home visitors hoped to change family attitudes and aspirations towards children's achievement (Gray and Klaus, 1965). Home visitors helped parents plan adult-child learning experiences that could be implemented within existing family activities, used

role-playing to demonstrate activities to parents, and emphasized the importance of positive reinforcement with school reports and homework.

4.5.3 Program Description

ETP is the oldest of the interventions we cover. It was implemented from 1962 to 1965 in Tennessee as a randomized experiment aimed for black, disadvantaged children. It was less intensive than PPP, relying strongly on the home visits, and having summer schools for children, but not traditional preschool services for the rest of the year. Table 3 shows the characteristics of the program (Gray and Klaus, 1965; Gray, 1969; Klaus and Gray, 1968). Tables 4 shows the characteristics of participants.

Gray et al. (1982) evaluates ETP: for IQ, they find the usual pattern of very strong initial impacts (approximately one standard deviation for the experimental group with higher intensity) followed by fadeout of the impacts during elementary school. The program does not seem to have had a positive effect on employment. The experimental groups members were around 15% more likely than the control group members to attend college after finishing school, but the differences were not significant. Overall, the pattern of results seems similar but weaker than the pattern in the PPP. This might be because the program was less intense, including the lack of year-round center-based activities.

4.6 Comparison Across RCTs

4.6.1 Methodology

This section presents a methodology enabling us to explore the treatment effects of the ECIs we analyze on early-life cognitive and character skills and later-life economic outcomes.

4.6.2 The Model

Let D_i denote a treatment indicator for individual i, where i is a generic individual in the set $\mathcal{I}$. Let M_i be a male indicator and X_i a vector of background characteristics. D_i takes

the value $d \in \{0, 1\}$ for individuals in the control or treatment groups, respectively. Let $Y_{i,s}^{d}$ denote outcome s for individual i where $\mathcal{S}$ denotes the set of outcomes and $s \in \mathcal{S}$ is a generic outcome, for $d \in \{0, 1\}$. Thus, the counterfactual outcome s for individual i is

$$Y_{i,s} = Y_{i,s}^{1} D_i + Y_{i,s}^{0} (1 - D_i) . \tag{1}$$

For $s \in \mathcal{S}$, our objective is to estimate the coefficients in the following linear model

$$Y_{i,s} = \beta_0^s + \beta_M^s M_i + \beta_D^s D_i + \beta_{MD}^s M_i D_i + \varepsilon_{is} \tag{2}$$

where ε_{is} is an error term. We suppress X_i to simplify notation but we control for it in all of our empirical analyses. As explained in Section 4.6.3, conditioning on certain covariates is especially important in the analysis of ABC and PPP.

Table 5: Parameters of Interest

Treatment Effect, Male	Treatment Effect, Female	Treatment Effect, Gender Difference: Male - Female
β_D^s	$\beta_D^s + \beta_{MD}^s$	β_{MD}^s

Note: this table lists the meaning of the parameters in (2) for each outcome. We estimate (2) for a collection of parameters $\mathcal{S}$, with typical element s.

Thus, our empirical objective is to test whether β_D^s is statistically and economically significant for females and likewise $\beta_D^s + \beta_{MD}^s$ for males. We also present estimates pooling females and males. This is useful in this context given the small sample sizes. However, some results may be sex specific.

4.6.3 Estimation Issues

We face four issues when estimating and performing inference on the coefficients of (2): (i) compromised randomization; (ii) small sample size; (iii) multiple hypothesis testing; (iv) item non-response.

The first arises in ABC and PPP from the initial randomization: some children were swapped between treatment and control due to their background characteristics as explained

in Section 4. The second is due to the design of the program. Thus, usual asymptotic inference methods generate imprecise estimations for the p-value of each test.[23]

The third issue arises from the nature of our research question. We will test the same hypotheses for various outcomes that seem to capture a lot of the impact of these programs. Thus, some researchers would be curious for us to correct the inference for multiple hypothesis testing. Roughly speaking, we need to correct for the fact that the probability of incurring in Type I error increases as the number of hypotheses to test increase. We think of the hypothesis we test as independent to each other, so we also present uncorrected p-values.

Finally, there are two ways in which we could observe item non-response. Either we do have data on follow up at the relevant outcome age but we miss information on a particular question or there is attrition in the sense that we do not observe data at all. Methodologically, we treat both of these as item non-response.

To solve these four issues, we use the methodology in Heckman et al. (2010a). Roughly speaking, we do the following. To solve the first issue, we control for the background characteristics altering the randomization design detailed in Section 4. In other words, we work under a "matching on observed variables" assumption. To solve the small sample size issue, we use a bootstrap based inference to avoid relying on asymptotic inference methods. In brief, our inference method is non-parametric. To solve the third issue, we correct multiple hypothesis testing through a "step-down procedure". Lastly, to solve the non-response issue, we use an inverse probability scheme.

In all of our estimations, we report two estimates of the p-value. The first corrects for all the estimation issues except for multiple hypothesis testing. The second corrects for all of the estimation issues. We label the first as "non-parametric p-value" and the second "non-parametric, step-down p-value". Table 6 makes this explicit.

[23] Heckman et al. (2010) develop a more robust procedure.

Table 6: What Estimation Issues Does Each Type of p-value Address?

$p-value$	Compromised Randomization	Small Sample	Multiple Hypotheses Testing	Item Non-response
Non-parametric	✓	✓	×	✓
Non-parametric, Step-down	✓	✓	✓	✓

Note: this table details the estimation issues each of the two p-value's we present addresses.

4.6.4 Results

Table 7 presents the results on costs and benefits from two of the studies that were discussed in this section, and for another important study that will be discussed in Section 5. This table shows PPP had outstanding benefits to cost relationships. Note that cost-benefit calculations on ABC do not account for crime; for PPP, they do not account for child care benefits; and for both ABC and PPP, they do not account for missing maternal earnings. The evidence in this table is, to the best of our knowledge, the best available on cost-benefit relationships of ECIs. It shows a high level of effectiveness for the two programs. Many other studies that do not have long-term data do cost-benefit analysis using forecasts. However, the forecasts are usually very rough, and have scarce theoretical or empirical support. We do not include them in this analysis.

We present the results for youth outcomes in Table 8. We base the interpretation of these results on non-parametric one sided p-values. We present step-down p-values to provide further information for the interested reader. However, we do not use them as the basis of our interpretation, as our objective is not to test whether the outcomes are jointly significant. Rather, our objective is to test whether each outcome of interest, particularly skills and later-life relevant economic outcomes, are significant by themselves. It is important to note that in a lot of cases pooling the data for men and women reinforces power and, therefore, increases the significance of the results.

For PPP, the first year of post-treatment results for IQ align with F-1 in Section 1. For both males and females, the results indicate a boost on IQ of more than half a standard

Table 7: Costs and Benefits of PPP and ABC

	PPP	**ABC**
Program Cost	-$20,956	-$83,788
Program Benefits		
Child care		$36,448
Child abuse and neglect		
Earnings	$92,052	$49,541
Earnings of Future Generations		$7,553
Maternal Earnings		$90,721
K-12 education	$5,104	$11,664
Smoking/Health		$23,471
Higher Education Costs		-$10,729
Welfare/AFDC	$4,364	$259
Crime	$78,800	
Total benefits	$180,319	$208,927
Net present value	$159,363	$125,139
Benefit-cost ratio	8.6:1	2.5:1
(s.e.)	(3.9)	

Note: dead-weight cost is zero, 3% rate of discount, all values are in 2014 dollars. Standard errors are obtained using bootstrapping. PPP estimates from Heckman et al. (2010b); ABC estimates from Barnett and Masse (2007).

deviation. Consistent with F-1 as well, the effect fades out for the last observation we have of this variable, age 14. The results are very similar for the rest of the interventions. The effects for ABC and IHDP are also very sizable at the beginning but do fade out. CARE has much lower impacts even at the beginning. This could merely be a consequence of the lack of precision caused by CARE's small sample size. We only contrast the control branch with the full-treatment branch. Thus, our treatment group has only 17 observations—which is further exacerbated by the fact that we estimate effects by gender. The effects for ETP do not go beyond half of a standard deviation but are sizable as well.

The evidence on non-cognitive skills is less clearly summarized, but let us qualify further this statement. Character is measured from batteries reflecting various behaviors, as does any psychometric inventory (see Heckman and Kautz, 2012). Moreover, in many studies the data available to measure character is sparse. Nonetheless, the literature reports effects on character for PPP, ABC, and IHDP. For CARE there are no statistically significant effects, and we speculate that the reasons are the same as in the case of cognition. For ETP, the measures are unreliable.

Table 8: Treatment Effects on Skills by Gender, All the Programs

	PPP			**ABC**			**CARE**			**IHDP**			**ETP**		
	Female	Male	Pooled	Female	Male	Pooled	Female	Male	Pooled	Female	Male	Pooled	Female	Male	Pooled
IQ, 1st Yr Post-Treatment	10.029	8.157	10.444	10.405	7.648	12.997	2.068	-6.172	6.840	10.677	14.265	11.560	6.546	4.132	3.598
Non-Parametric *pvalue*	**(0.000)**	**(0.008)**	**(0.000)**	**(0.000)**	**(0.014)**	**(0.000)**	(0.372)	(0.782)	**(0.056)**	**(0.000)**	**(0.000)**	**(0.000)**	(0.177)	(0.303)	(0.235)
Step-Down, Non-Parametric *pvalue*	**[0.000]**	**[0.026]**	**[0.000]**	**[0.000]**	**[0.090]**	**[0.000]**	[0.832]	[0.96]	[0.252]	**[0.000]**	**[0.000]**	**[0.000]**	[0.225]	[0.375]	[0.333]
IQ, Last Observation	-1.965	-0.134	-1.630	1.293	0.937	3.390	-2.267	-14.826	2.403	-0.542	2.793	0.304	-0.914	4.160	0.391
Non-Parametric *pvalue*	(0.812)	(0.524)	(0.778)	(0.220)	(0.360)	**(0.064)**	(0.620)	(0.920)	(0.340)	(0.674)	**(0.052)**	(0.402)	(0.590)	(0.303)	(0.440)
Step-Down, Non-Parametric *pvalue*	[1.000]	[0.820]	[0.994]	[0.692]	[0.902]	[0.230]	[0.992]	[0.998]	[0.766]	[0.680]	[0.248]	[0.386]	[0.733]	[0.468]	[0.665]
Achievement	0.299	0.250	0.342	0.435	0.223	0.674	-0.004	-1.019	0.398	0.045	0.215	0.094			
Non-Parametric *pvalue*	**(0.024)**	(0.172)	**(0.016)**	**(0.010)**	(0.194)	**(0.004)**	(0.502)	(0.932)	(0.102)	(0.260)	**(0.024)**	**(0.084)**			
Step-Down, Non-Parametric *pvalue*	**[0.096]**	[0.668]	**[0.058]**	**[0.030]**	[0.562]	**[0.002]**	[0.960]	[0.996]	[0.344]	[0.650]	[0.118]	[0.296]			
Conscientiousness	0.174	0.056	0.239	0.142	0.050	0.396	-0.363	-1.123	-0.028	-0.007	0.200	0.035			
Non-Parametric *pvalue*	**(0.056)**	(0.386)	**(0.054)**	(0.194)	(0.430)	**(0.040)**	(0.828)	(0.920)	(0.566)	(0.554)	**(0.026)**	(0.312)			
Step-Down, Non-Parametric *pvalue*	**[0.090]**	[0.774]	**[0.082]**	[0.686]	[0.964]	[0.180]	[0.998]	[0.996]	[0.918]	[0.828]	[0.158]	[0.652]			
Extraversion	0.045	-0.014	0.073	-0.014	0.072	0.215	-1.231	-2.027	-1.222	0.027	0.096	0.075			
Non-Parametric *pvalue*	(0.208)	(0.594)	(0.138)	(0.562)	(0.344)	**(0.096)**	(0.980)	(0.994)	(0.996)	(0.302)	(0.102)	**(0.096)**			
Step-Down, Non-Parametric *pvalue*	[0.600]	[0.982]	[0.358]	[0.934]	[0.872]	[0.270]	[0.992]	[1.000]	[1.000]	[0.634]	[0.312]	[0.206]			
Agreeableness	0.065	0.097	0.141	-0.193	-0.178	-0.075	-0.953	-0.051	-0.181	0.028	-0.055	0.045			
Non-Parametric *pvalue*	(0.270)	(0.298)	(0.186)	(0.818)	(0.772)	(0.632)	(0.964)	(0.604)	(0.62)	(0.298)	(0.776)	(0.188)			
Step-Down, Non-Parametric *pvalue*	[0.602]	[0.878]	[0.408]	[0.986]	[0.974]	[0.890]	[0.996]	[0.506]	[0.832]	[0.712]	[0.770]	[0.514]			
Stability	0.174	0.239	0.172	-0.230	-0.361	-0.129	-0.382	-1.417	-0.150	-0.184	-0.237	-0.139			
Non-Parametric *pvalue*	(0.128)	(0.114)	(0.118)	(0.920)	(0.912)	(0.782)	(0.866)	(0.994)	(0.686)	(1.000)	(1.000)	(0.988)			
Step-Down, Non-Parametric *pvalue*	[0.498]	[0.102]	[0.102]	[0.990]	[0.984]	[0.928]	[0.930]	[1.000]	[0.666]	[1.000]	[1.000]	[1.000]			
Openness	-0.130	-0.038	-0.072	-0.312	-0.289	-0.181	-1.113	-1.406	-0.494	0.044	0.057	0.099			
Non-Parametric *pvalue*	(0.822)	(0.570)	(0.666)	(0.982)	(0.950)	(0.862)	(0.968)	(0.936)	(0.884)	(0.188)	(0.202)	**(0.044)**			
Step-Down, Non-Parametric *pvalue*	[0.822]	[0.956]	[0.984]	[0.982]	[0.950]	[0.862]	[1.000]	[0.998]	[0.996]	[0.322]	[0.400]	**[0.078]**			

Note 1: initial sample sizes are: **PPP**: 123; **ABC**: 111; **CARE**: 64; **IHDP**: 985; **ETP**: 88.
Note 2: non-parametric $p-value$ (in parentheses) accounts for compromised randomization, small sample size, and item non-response. Step-down $p-value$ [in brackets] accounts for the same and for multiple hypotheses testing.
Note 3: when calculating each treatment effect for each program we control for three baseline characteristics: (i) mother works; (ii) father at home; (iii) number of siblings.
Note 4: in ABC we add birth-weight and gestational age as controls, because the five individuals initially assigned as control status were swapped to treatment status based on "life-at-high-risk" status as indicated by their health conditions.
Note 5: in PPP we add SES-index (as constructed by the program designers) and Stanford-Binet IQ score at age 3, because the two individuals initially assigned to control status were swapped to treatment status based on those criteria.
Note 6: **PPP** Baseline IQ is measured at age 4, and adult IQ is measured at age 14. Achievement is measured factor analyzing all available achievement test scores. Conscientiousness reflects the child's tendency to be organized, responsible, and hardworking. Extroversion is a measure of the child's preference for the outer world of people over the inner world of subjective experience. Agreeableness is a measure of the child's cooperation and altruism. Stability is a measure of the child's emotional stability. Openness refers to the child's tendency to be receptive of new aesthetic, cultural or intellectual experiences. Factors for each of the five personality traits listed are created for each age they are observed. The set of factors for each trait is then factor analyzed to generate a factor representative of all the years we observe the trait. **ABC**. Baseline IQ is measured at age 3, and adult IQ is measured at age 21. Achievement is measured factor analyzing all available achievement test scores. Conscientiousness reflects the child's tendency to be organized, responsible, and hardworking. Extroversion is a measure of the child's preference for the outer world of people over the inner world of subjective experience. Agreeableness is a measure of the child's cooperation and altruism. Stability is a measure of the child's emotional stability. Openness refers to the child's tendency to be receptive of new aesthetic, cultural or intellectual experiences. Factors for each of the five personality traits listed are created for each age they are observed. The set of factors for each trait is then factor analyzed to generate a factor representative of all the years we observe the trait. **CARE**. Baseline IQ is measured at age 3, and adult IQ is measured at age 8. Achievement is measured factor analyzing all available achievement test scores. Conscientiousness reflects the child's tendency to be organized, responsible, and hardworking. Extroversion is a measure of the child's preference for the outer world of people over the inner world of subjective experience. Agreeableness is a measure of the child's cooperation and altruism. Stability is a measure of the child's emotional stability. Openness refers to the child's tendency to be receptive of new aesthetic, cultural or intellectual experiences. Factors for each of the five personality traits listed are created for each age they are observed. The set of factors for each trait is then factor analyzed to generate a factor representative of all the years we observe the trait. **ETP**. Baseline IQ is measured at age 4, and adult IQ is measured at age 17. Achievement is measured factor analyzing all available achievement test scores. Conscientiousness reflects the child's tendency to be organized, responsible, and hardworking. Extroversion is a measure of the child's preference for the outer world of people over the inner world of subjective experience. Agreeableness is a measure of the child's cooperation and altruism. Stability is a measure of the child's emotional stability. Openness refers to the child's tendency to be receptive of new aesthetic, cultural or intellectual experiences. Factors for each of the five personality traits listed are created for each age they are observed. The set of factors for each trait is then factor analyzed to generate a factor representative of all the years we observe the trait. **IHDP**. Baseline IQ is measured at age 3, and adult IQ is measured at age 18. Achievement is measured factor analyzing all available achievement test scores. Conscientiousness reflects the child's tendency to be organized, responsible, and hardworking. Extroversion is a measure of the child's preference for the outer world of people over the inner world of subjective experience. Agreeableness is a measure of the child's cooperation and altruism. Stability is a measure of the child's emotional stability. Openness refers to the child's tendency to be receptive of new aesthetic, cultural or intellectual experiences. Factors for each of the five personality traits listed are created for each age they are observed. The set of factors for each trait is then factor analyzed to generate a factor representative of all the years we observe the trait.

We present the results for later-life outcomes in Table 11. This table is based on Heckman et al. (2010a) and Campbell et al. (2014b). The results are consistent with the overall discussion in the paper: we evaluate PPP and ABC as very successful programs in boosting later life outcomes. To make interpretation easier, we reverse negative outcomes such as obesity. While these results are selected out of a large number of results that were reported in each of the papers, they achieve significance levels after controlling for size to account for multiple hypothesis tests among selected groups of outcomes. The results presented here are some of the most important impacts found for PPP in labor outcomes and some of the most important impacts found for ABC in health. We can see in the table that the increase in HS graduation in Perry was 56% for the females, and that the increase in current employment percentage was close to 30% for both males and females. For ABC, there are important reductions in adult abdominal obesity of around 20% for females and 30% for males. The differences in health indicators shown in this table can have very strong consequences in health later in the life cycle. The takeaway from this table is that these interventions dramatically changed the lives of their beneficiaries.

Table 9: Selected Treatment Effects at Adulthood by Gender, PPP and ABC

	PPP*			**ABC****	
	Female	Male		Female	Male
HS Graduation, age 19	0.56	0.02	**Abdominal Obesity, age 30**	0.198	0.294
Partial linear *pvalue*	**(0.000)**	0.416	Block *pvalue*	**(0.080)**	(0.137)
Step-down, Partial linear*pvalue*	**[0.000]**	0.583	Step-down *pvalue*	**[0.080]**	[0.218]
Monthly Earnings, age 27	0.64	1.01	**Obesity and Hypertension, , age 30**	-0.028	0.529
Partial linear *pvalue*	**(0.000)**	**(0.011)**	Block *pvalue*	(0.501)	**(0.016)**
Step-down, Partial linear*pvalue*	**[0.000]**	**[0.037]**	Step-down *pvalue*	[0.641]	**[0.016]**
Current Employment, age 27–40	0.28	0.29	**Framingham Risk Score, age 30**	0.331	3.253
Partial linear *pvalue*	**(0.042)**	**(0.011)**	Block *pvalue*	**(0.070)**	**(0.038)**
Step-down, Partial linear*pvalue*	**[0.094]**	**[0.024]**	Step-down *pvalue*	**[0.070]**	**[0.038]**

Note 1: initial sample sizes are: **PPP**: 123; **ABC**: 111.
Note 2: * results taken from Heckman et al. (2010a). One-sided $p-value$ based on the FreedmanLane procedure, using the linear covariates maternal employment, paternal presence, and StanfordBinet IQ, and restricting permutation orbits within strata formed by Socioeconomic Status index (SES) being above or below the sample median and permuting siblings as a block (in parenthesis). Adjusted for step-down –i.e., multiple hypothesis testing– [in brackets].
Note 3: ** results taken from Campbell et al. (2014b). One-sided block permutation $p-value$ (in parenthesis). Adjusted for step-down –i.e., multiple hypothesis testing– [in brackets].
Note 4: **PPP** monthly earnings is in thousands of 2010 dollars at age 27. High school graduation (at age 19) and current employment (at age 27) are self-explanatory. **ABC** all at age 34. Abdominal obesity: waist-hip rario >0.9 for males; >0.85 for females. Obesity and hypertension: interaction of two variables. Obesity: ≥ 30 in Body-mass index. Hypertension: systolic blood pressure ≥ 140 & diastolic ≥ 90. Framingham risk score: is a sex-specific algorithm used to estimate the 10-year cardiovascular risk of an individual – it is based on basic health indicators and habits.

5 Public Early Childhood Interventions

In the previous section, we presented evidence on a set of demonstration programs that have strong long-term evaluations. However, our conclusions do not necessarily apply to large scale public programs. Arguably, the most important discussion in the ECIs literature concerns the effectiveness of government programs because of several reasons: (i) they are funded with public money, which generates deadweight losses; (ii) their scales are generally large;[24] (iii) policy-makers modify them and we wish to inform their decisions.

This section analyzes public ECIs and their evaluations. Our main objective is to inform facts F-4, F-5, and F-6 in Section 1. We survey the literature and point out advantages and pitfalls of the different evaluations.

5.1 Randomized Controlled Trials

With the objective of evaluating the effectiveness of public ECIs, the US government finances or financed a few massive RCTs. To the best of our knowledge, programs without evident design flaws are: (i) the Comprehensive Child Development Program (CCDP) ; (ii) Even Start (ES); (iii) Early Head Start (EHS), (iv) the Head Start Impact Study (HSIS); (v) The Tennessee Voluntary Pre-K Program (TN-VPK). We do not explain in depth these programs. Instead, we summarize them in Table 10. We include all of these evaluations because they are all the RCTs of public ECIs that exist and, not coincidentally, they share very similar pitfalls.

5.1.1 Issues in the Evaluations

We now discuss the main issues in the evaluations of these RCTs of public ECIs and relate them to the facts in Section 1. First, as we argue in F-5 of Section 1, low take-up and control contamination are major issues. Table 10 shows that for all of these programs the control

[24] García and Heckman (2014) studies how large scale versions of the ECIs in Section 4 would work and how might them impact social mobility.

Table 10: RCTs of Public Programs Including an Early Education Component

	Even Start	**CCDP**	**Early Head Start**	**HSIS**	**TN-VPK**
Program yrs.	1999/2000	1990/1991-1994/1995	1995/1996-1998/1999	2002/2003	2009/2010-2010/2011
Delivery mode	Family Literacy: child and adult education	Case management	Home Visits, Center-Based and mixed	Home Visits, Center-Based and mixed	Center-based
Eligible population	Parent eligible for adult educ. & child age ≤ 8 (average star child age 3.2)	Under poverty line Child age ≤ 1 (or pregnant)	90% under poverty line Child age ≤ 1 (or pregnant)	90% under poverty line Child age 3 (3yc) or 4 (4yc)	Priority to free lunch-eligible children
Sample size	18 grantees, 463 families	21 projects, 3961 families	17 grantees, 3001 families	383 grantees, 4667 children	3025 full sample, 1078 ISS
Last follow-up	After 2 years (avg. age 5)	Age 5	5th Grade (approx. age 11)	3rd Grade (approx. age 9)	1st Grade (approx. age 7)
% Response rate	76% (less in analyses)	74% Treated 78% Control	54.4%	76% Treated 71% Control	Around 96% Treated in ISS: approx. 60% Control in ISS: approx. 53%
Attrition correction	Apparently not	Apparently not	Yes	Yes	No
Parameter	Intent to Treat	Intent to Treat	LATE. Participant defined as receiving 1+ home visits or attending 2+ weeks to a center	Intent to Treat	LATE. Participant defined as found in admin. data (full) or attended 20+ days (ISS)
Experience of treated group	Year 2: 72% in child educ.; 28% in parenting educ. Avg. family participated 8 (of 12) mos.	33% enrolled 5 years Center-based child care: Age <3 avg. 36.6 hrs./mo. Age 3-5 avg. 53.9 hrs./mo.	73.3% had 20+ hrs. of center care or 1+ home visit per week at least one of the 3 years	Approx. 83% had some HS on the year of evaluation	ISS: Avg. 149 day attended
Experience of control group	Year 2: 32% in child educ.; 17% in parenting educ.; 16% in Even Start	Center-based child care: Age <3 avg. 19.2 hrs./mo. Age 3-5 avg. 36.8 hrs./mo.	13.8% had 20+ hrs. of center care or 1+ home visit per week at least one of the 3 years	Approx. 40% center care Approx. 15% Head Start	49% stayed with parents 15% private child care 11% Head Start
Multilevel estimation	No	Yes: sites inversely weighted proportional to variance	Yes: sites equally weighted		Yes
Multiple hypotheses	No correction	No correction	No correction	Benjamini-Hochberg	No correction
Impacts: Number of Positive Significant Effects / Number of Variables Measured					
Cognitive	After 2 years: 0/8	Age 3 and 5: 0/4 (PPVT-R, TVIP, Kaufman)	Age 3: 4/4. PPVT (ES 0.12) Age 10: 0/10	HS Year, 4yc: 7/12, 3yc: 8/12 PPVT ES 4yc: .09, 3yc: .18 Age 9, 4yc: 1/11 Age 9, 3yc: 0/11 (1 negative)	ISS, PreK: 7/7 Woodcock Johnson ES .31 ISS, 1st Grade: 0/10 (1 neg.) Woodcock Johnson: No impact
Personality	After 2 years: 0/5	Age 3: 0/6 (Child Behavior Checklist) Age 5: 0/3 (Child Behavior Checklist)	Age 3: 4/9. CBCL-Aggressive (ES -0.1) Age 10: 1/11 Socio-Emotional Index (ES 0.1)	HS Year, 4yc: 0/9, 3yc: 2/9 Hyperactive (ES -0.21) Age 9, 4yc 2/19 (4 negative) Age 9, 3yc 1/19	ISS, PreK: 3/6 Cooper-Farran Social (ES .34) ISS, 1st grade: 0/6 Full sample, KG: less retention
Parenting/ environment	After 2 years: 0/9	Age 3: 0/8 (NCAST) Age 5: 0/4 (NCAST)	Age 3: 8/17 HOME (ES 0.11) Age 10: 0/9	HS Year, 4yc: 1/5, 3yc: 3/5 Age 9, 4yc: 1/15, 3yc: 1/15	
Parents' self-sufficiency	NA	0/23 (employment, income, welfare)	Age 3: 2/8. Training weeks (ES .18) Age 10: 0/5	NA	NA

Note: abbreviations used in this table: CBCL: Child Behavior Checklist; CCDP: Comprehensive Child Development Program; ES: Effect Size; HOME: Home Observation for Measurement of Environment instrument; HSIS: Head Start Impact Study; HS: Head Start; ISS: Intensive Subsample (in TN-VPK); KG: Kindergarten; LATE: Local Average Treatment Effect; NCAST: Parent-Child Interaction Scales. PPVT: Peabody Picture Vocabulary Test; TVIP: Test de Vocabulario en Imagenes Peabody; PreK: Prekindergarten; TN-VPK: Tennessee Voluntary Prekindergarten Program. 3yc and 4yc: 3-years cohort and 4-years cohort (in HSIS). The row "Multilevel Estimation" indicates whether the estimators used to calculate the impacts of the program accounted for the hierarchical structure of the data. Outcomes Section: For all sets of outcomes and age, we note the number of outcomes that were significant at the 5% level (one tail), and the total number of outcomes measured (for example, a 4/9 entry in cognitive means 4 significant impacts out of 9 cognitive tests that were taken).For the TN-VPK intervention, the ISS sample are the children in the control and treatment group for whom informed consent was obtained for testing. On the other hand, "full sample" is the complete sample, for which results were obtained in administratively-recorded outcomes. See the description of the program for an analysis of how the evaluation on the ISS subsample might be considered non-experimental.

group had a substantial level of preschool enrollment. Interestingly, HSIS is the one study with the highest rate of control contamination (40%). Second, treatment heterogeneity is a major issue. For example, in HSIS, multiple curricula were used and the quality of the programs greatly varied across sites. While the HS centers were of higher quality on average than the control centers, the distribution of quality strongly overlapped between both groups (see Torcasso, 2014). Third, attrition is high. Finally, none of these studies have long-term follow-ups. The problem of control contamination is not well-solved in the literature.

5.1.2 The Tennessee Voluntary Prekinder Program (TN-VPK)

This is the only state-funded and state-implemented ECI that counts with an RCT. The characteristics of this program are in Table 10.

This intervention has one important problem: parental consent was asked for after the randomization had been done, making consents potentially affected by the program. The sample that gave consents is referred to as the Intensive Substudy (ISS). They were assessed by the research team with a battery of achievement measures. For cohort 1, 46% of the treatment groups parents and 32% of the control groups parents gave their consent to be part of the ISS –cohort 2: 74% and 68%. This might potentially generate selective attrition. Acknowledging this problem, the research team –which evaluated the program– implemented a propensity score matching strategy to ameliorate possible differences across groups. They find that both groups were very similar in terms of baseline characteristics. However, it is still possible that the differential rates of attrition have made the groups in the ISS different in unobserved ways.

There was a positive impact on school retention –an outcome that was measured both for the whole sample (4.0% retention for the treatment group compared with 8.0% for the controls) and for the ISS (4.1% and 6.2%, respectively). For the ISS, during the preschool year, positive treatment effects were found on all the direct assessments of academic skills. The effect size for the Woodcock Johnson Test (achievement) was 0.24. In contrast, at the

end of first grade, there were no effects on cognitive skills, and no differences for personality outcomes. The authors of this report were very cautious on discussing these results, stressing that more time was necessary to obtain conclusions.

5.1.3 The Head Start Impact Study (HSIS)

Head Start is by far the largest ECI in the US. Accordingly, HSIS is the largest RCT of an ECI that exists. This is important because any small sample issues raised about the RCTs in Section 4 are absent in this intervention. For ethical concerns, HSIS is a one-year randomly assigned version of Head Start and includes a 3-year-old cohort and a 4-year-old cohort. It granted treatment to 4,677 children through 383 grantees. The most current follow-up, so far, goes up to the third grade of elementary school. The attrition percentage is 25%.

Given that there was an excess of demand, the intervention was randomized at the site level. The control group was not granted treatment in the relevant Head Start site. However, nothing prevented control individuals from applying to other Head Start sites. In practice, 40% of the control group attended center-based care. Moreover, 15% of the individuals in the control group attended some other Head Start center. We discuss the methodological challenges of control contamination below.

The main report of the intervention, Puma et al. (2010), reports intent to treat effects (ITT). For the 3- and 4-year-old at baseline cohorts, 7 and 8 out of 12 cognitive outcomes, respectively, were significant and positively impacted. These treatment effects fade out strongly by age 9. The results on character measures were much smaller; only a few of them were positive. Moreover, the program had very few impacts on parenting (see Table 10). Given that an substantial number of the control group attended HS itself, the ITT estimates in this case have no meaningful economic interpretation (i.e. no counterfactual scenario). All of the remaining papers using HSIS data in this section start by estimating the Local Average Treatment Effects (LATE), which is the effect of treatment compared to the next best preferred option (see Heckman and Vytlacil, 2007). However, they do not address

the problem of control contamination, as most of the children using substitute treatments attended preschools other than HS. They find impacts on cognition of around .2 standard deviations after one year of the intervention, with a decrease to around .1 after two years of the program (see, for example Ludwig and Phillips, 2008). These results are much smaller in magnitude when compared to any of the results in Section 4. As we argue in Section 1 and explain in what follows, this is partly a consequence of (i) the availability of substitutes; (ii) children's heterogeneous backgrounds; (iii) a program duration one year.

Bitler et al. (2014) explores the heterogeneity of treatment effects in HSIS. They estimate Quantile Treatment Effects (QTE) and QTE under endogeneity, respectively analogous to ITT and LATE. They find that the impacts of HSIS are stronger in the lower percentiles of the distribution of impacts. They also find impacts by baseline scores. Consistently with the literature, they find that the impacts of the program are stronger in the bottom of the distribution (although they find an unusual pattern of U-shaped distribution of impacts).

5.1.4 Analyses of HSIS that Address Control Contamination

Two recent and preliminary studies address control contamination in the HSIS data. These are Feller et al. (2014) and Kline and Walters (2014). Both papers go beyond standard ITT and LATE using a framework for multiple treatment, multiple outcome models.[25] Both papers rely heavily on normality assumptions to secure estimates. Only Kline and Walters (2014) analyze self-selection into the various options available to participants. They model families as having three choices: (i) Head Start, h; (ii) other Preschool Program, p; and (iii) staying at Home, n. They identify 5 exhaustive and mutually exclusive groups: (i) Always-Head Start (11%); (ii) Always-Preschool (11%); (iii) Always-Home (12%); (iv) Preschool-Compliers (20%); and (v) Home-Compliers (45%).[26] They discuss the identification problems related to having a single randomization to identify the impacts of three

[25] See Heckman and Vytlacil (2007) and Heckman and Urzua (2010) for development of the general methodology.

[26] These numbers are from Feller et al. (2014). The numbers in Kline and Walters (2014) are very similar.

different choices.[27] This literature shows that, while the proportions of the different strata can be identified directly from the data, the treatment effects cannot be identified without additional assumptions. We now discuss the different methodological approaches these papers use to identify treatment effects.

Both papers share some conclusions: both find that HS has an impact that is very similar to the other available alternative center-based preschool settings, and both find that HS has significant effects on test scores for specific groups of children. These effects are moderate but this is meaningful, especially considering that the interventions last only one year and the follow-up is short. However, the magnitudes of their preferred impacts are quite different –between 0.2 standard deviations in Feller et al. (2014) and 0.47 standard deviations in Kline and Walters (2014). The magnitudes are not really comparable, as Kline and Walters (2014) obtain an Average Treatment Effects parameter, while Feller et al. (2014) obtain a LATE for the home compliers. The first paper finds a negative selection in gains in HSIS, so it is natural that after correcting for selection, the impacts are higher.

The results in this section speak to F-5 and F-6 in Section 1. First, they document that, as has been established in other applied literature, it is fundamental to account for substitution possibilities when evaluating ECIs. This is especially important in the case of large scale programs –the wider the intervention, the more substitutes available for families. Given that there are substitutes for Head Start for many controls, the intent to treat effect are much smaller when compared to the results on ECIs in Section 4.

5.2 Non-randomized Evaluations of Head Start

[27] See Heckman and Urzua (2010) for a general analysis of the problem of using one instrument to estimate treatment effects in a model with multiple competing outcomes.

Table 11: Treatment Effects at Adulthood by Gender, All the Programs

Study	Currie and Thomas (1995)	Garces et al. (2002)	Ludwig and Miller (2007b)	Deming (2009)	Carneiro and Ginja (2014)	Feller et al. (2014)	Kline and Walters (2014)	Perry Preschool (own calculations)	Abecedarian (own calculations)
Dataset	C-NLSY	PSID	Multiple	C-NLSY	C-NLSY	HSIS	HSIS	Own data	Own data
Subpopulation	Black	Black, mother educ. $\leq$ high school		Black	Males			Black, low child IQ at entry & SES	98% Black, low mother IQ
Years of birth	1979-1987	1966-1977	1960-1975	1979-1986	1977-1996	1998-1999	1998-1999	1959-1964	1972-1977
Impacts									
IQ/achievement, age 3-4	-	-	-	-	-	**0.180**	**0.473**	**0.696**	**0.866**
	-	-	-	-	-	(0.026)	(0.120)	(0.136)	(0.181)
IQ/achievement, age 5-6	**0.44**	-	-	**0.287**	-	**0.110**	-	**0.297**	**0.359**
	(0.129)	-	-	(0.095)	-	(0.23)	-	(0.122)	(0.153)
IQ/achievement, age 7-21	0.201	-	-	**0.127**	-	**0.100**	-	-0.109	**0.226**
	(NA)	-	-	(0.075)	-	(0.026)	-	(0.149)	(0.142)
Grade retention ever	0.218	-	-	**-0.107**	-	-	-	-	-
	(0.295)	-	-	(0.056)	-	-	-	-	-
High School grad. (no GED)	-	0.00	0.117	0.067	-	-	-	-	-
	-	(0.071)	(0.080)	(0.044)	-	-	-	-	-
Attended some college	-	0.031	**0.028**	**0.136**	-	-	-	-	-
	-	(0.067)	(0.019)	0.049	-	-	-	-	-
Earnings, age 23-40	-	0.051	-	-	-	-	-	$7,584	$7,249
	-	(0.357)	-	-	-	-	-	(6,299.82)	(6,480.17)
Idle (no work or study)	-	-	-	-0.030	-	-	-	**0.204**	**0.273**
	-	-	-	(0.053)	-	-	-	(0.093)	(0.100)
Ever booked/charged crime	-	**-0.126**	-	0.051	-	-	-	-	-
	-	(0.05)	-	0.050	-	-	-	-	-
Behavior Problem Index, age 12-13	-	-	-	-	-0.647	-	-	-	-
	-	-	-	-	(0.582)	-	-	-	-
CES Depression Scale, age 16-17	-	-	-	-	-0.552	-	-	-	-
	-	-	-	-	(0.489)	-	-	-	-
Head Start Related Mortality, ages 5-9	-	-	**-1.198**	-	-	-	-	-	-
	-	-	(0.796)	-	-	-	-	-	-

Note: impacts are in bold whenever they would be significant in a t-test at the 10% significance level. Impacts on IQ/achievement scores are reported in standard deviations. Currie and Thomas (1995) originally report impacts on IQ/achievement in terms of test scores. PPVT at age 8 in Currie and Thomas (1995) is calculated using their interaction of Head Start and age coefficient. The SE for the predicted impact at this age is not reported. Our calculations use bootstrapped standard errors. Grade retention is measured at age 5 in Currie and Thomas (1995) and at age 18 in all other studies. Ludwig and Miller (2007b) use census data, Vital Statistics, and the NELS. For the sake of brevity, we limit the number of estimates we present from Ludwig and Miller (2007b) to only one per data set: the impact of treatment on mortality is from the Vital Statistics, impact on high school completion is from the NELS, and impact on attending some college is from the census. Impact on high school completion and college attendance are for children roughly 18-24 years old. Earnings in Garces et al. (2002) are measured in logs. Feller et al. (2014) originally reported 95% posterior intervals of (0.13, 0.23) during the Head Start Program, (0.06, 0.15) during pre-K/Kindergarten, and (0.05, 0.15) during Kindergarten/first grade. Impacts reported in Kline and Walters (2014) are estimated from a summary index created from Peabody Picture Vocabulary Tests and Woodcock Johnson III Preacademic Skills tests taken in Spring 2003; this index is standardized to have mean 0 and a standard deviation of 1. The Center for Epidemiological Studies Depression Scale in Carneiro and Ginja (2014) measures symptoms of depression in percentile scores, where higher scores are negative.

Many researchers use diverse econometric methods and quasi-experimental designs to directly evaluate Head Start. This evidence speaks to F-6 in Section 1. It actually complements the evidence in Section 5.1.3 –which indicates that Head Start is not a failure.

5.2.1 Fixed Effects Studies

Currie and Thomas (1995, 1999); Garces et al. (2002); Deming (2009) use family fixed effects in their identification strategy. This relies on strong assumptions. They control for the effects of competing early childhood programs by conditioning on usage. Currie and Thomas (1995) find that there are gains in test scores of equal magnitudes for whites and blacks at the end of the program, the gains fade out during elementary school for blacks –which again speaks to F1 in Section 1. Similarly, Deming (2009) finds no effect for whites. He finds effects for blacks which strongly fade out. Both papers find beneficial impacts in grade repetition, the first for white children and the second for black children.

Garces et al. (2002) and Deming (2009) measure treatment effects for adult outcomes. Both find impacts on high school completion and some college attendance. The first for whites and the second for blacks. While Garces et al. (2002) finds impacts in crime for blacks, Deming (2009) finds no impacts on crime at all. The first paper does not find significant impacts in earnings –possibly because the respondents were, on average, only 23 years old – and the second does not measure it.

The most puzzling pattern is the larger impacts of the program for whites in Currie and Thomas (1995) and Garces et al. (2002). This is inconsistent with F-4 in Section 1. However, Currie and Thomas (2000) suggests that blacks went to lower-quality HS, which speaks to F-5 in Section 1.

Of the studies of Head Start, we are most skeptical about family fixed effects strategies. It is hard to understand why the strict exogeneity assumptions required hold in this context.

5.2.2 Regression Discontinuity Designs

Carneiro and Ginja (2014) study the long-term impact of Head Start on health (frequent visits to the doctor, frequent use of medicines, ever having smoked, etc.) and behavior (grade repetition, school damage, special education etc.) outcomes. They exploit exogenous variation in program eligibility rules. In sum, they exploit multiple discontinuities based on year, state, family size, and family structure to identify diverse margins enabling them to estimate causal effects of interest. The multiplicity in the eligibility thresholds distinguishes this paper from classic regression discontinuity designs. They do not focus on individuals located in a single threshold. Instead, they include various cohorts and, thus, provide effects informing on causal impacts for different population groups.

This identification strategy enables them to identify the marginal effect of relaxing the eligibility requirement of Head Start. A somewhat puzzling empirical finding of the paper is that eligibility discontinuities shift the propensity to participate of Head Start only for boys. The interpretation of this is that the marginal individuals for which the effects are found are boys. The authors do not take a stand on a reason why this may be true.

The authors combine data from the National Longitudinal Survey of Young 1979 (NLSY79) and the children of the NLSY79, the cNLSY79. The former is a nationally representative survey of the individuals aged 14 to 22 years old in 1979. The latter surveys the children (or surveys parents about their children). It is possible and standard in the literature to match these two samples. Doing this, the authors use the household level information to replicate the Head Start eligibility rules that applied to each of the individuals in the sample drawing from the NLSY79 and draw the individual specific outcomes from the cNLSY79. Importantly, it is hard to state to what extent the cNLSY79 is nationally representative, since it surveys the children of the NLSY79 without further sampling design. By sample design, the authors are able to assess impacts of Head Start for individuals receiving treatment in the 1980s and the early 1990s. In a nutshell, the findings are the following. The program decreases behavior problems, chronic diseases, and obesity at ages 12 and 13, depression and

obesity at ages 16 and 17, and crime at ages 20 and 21. These findings contrast results of the evaluations of HSIS finding no treatment effects (see Zigler, 2010) and align with evaluations of HSIS, finding relevant treatment effects.

Ludwig and Miller (2007a) exploit information on technical assistance on the implementation of Head Start to 300 poor counties offered by the Office of Economic Opportunity. The authors find an empirical discontinuity in the participation and funding rates, which is 50 to 100 percent higher in the 300 counties just above the cutoff which makes them qualify as eligible for technical assistance for Head Start implementation. The authors point out that the discontinuity in other federal social spending is small and not significant and, therefore, argue that the discontinuities on which their estimations are based are not only exogenous but solely produced by the technical assistance described above. The main estimates are based on a regression discontinuity design comparing treatment and control counties near the cutoff. That is, the "treatment group" consists of individuals in the 300 counties which received technical assistance and the "control group" are individuals in counties with characteristics very similar to the treatment counties but did not receive any treatment whatsoever. The authors find reductions in mortality and evidence on schooling improvement (it is suggestive due to data limitations).

5.3 Evidence on State Programs

5.3.1 Regression Discontinuity Designs to Evaluate State Preschool

A recent set of papers assess state preschool programs based on regression discontinuity designs (see Gormley and Gayer, 2005; Gormley Jr et al., 2005; Hill et al., 2012). These papers only measure the effect of these programs on academic achievement for a single year compared to those who will get the program, but have not yet received it. Thus the estimated impact is very short-term and limited in scope, and is not a basis for informing public policy.

5.3.2 The Chicago Child-Parent Center (CPC)

CPC began in 1967 and targets disadvantaged children. It is the second oldest federally funded preschool program after Head Start. Assignment is not random but follow-ups are available both for participants and non-participants up to age 26. Its evaluation compares children attending preschool and kindergarten to individuals who did not go to preschool but did go to kindergarten.

Gensowski et al. (2014) reanalyze CPC. They test and control for selection through a standard control function approach (see Heckman, 1979). They incorporate the information on the quality of the children's neighborhood and their access to the program. They also explore the interactions between the program and the quality of a child's neighborhood. Similar to Reynolds et al. (2011), Gensowski et al. (2014) find that the program increased men's educational attainment by over a year and significantly improved the economic outcomes for both men and women. Gensowski et al. (2014) also find that, contrary to the literature, the CPC was less effective for children from severely disadvantaged backgrounds. However, the authors do not find statistically significant impacts on criminal behavior.

5.4 Impacts of Generic Center and Non-center Based Childcare

Bernal (2008); Bernal and Keane (2010, 2011) use three different methodologies to assess the impact of local child-care on cognitive and non-cognitive development: (i) a structural approach –using a fully structural model; (ii) a quasi-structural approach –using joint estimation of the reduced form decision rules and structural production functions; (iii) instrumental variables approach. The first paper uses a sample of married women. The last two use a sample of single mothers and exploit exogeneous changes in the welfare structure as variation affecting the probability of the child being in child-care. The three papers show child-care has negative impacts on cognition, with an effect of 0.13-0.14 standard deviations. Importantly, Bernal and Keane (2011) shows this effect is mainly driven by informal child-care. We do not analyze these studies in depth. They deal with non-specific child care arrangements.

However, it is important to point out they inform F-5: the effects of child-care are limited by the extent of its quality.

6 Next Steps

This paper documents the 7 findings about Early Childhood Interventions in the US. presented in Section 1. We briefly summarize fruitful directions for further research in the next lines.

Research on ECIs has considerably progressed in the last 10 years. There are now rich data sets with long run outcomes. Follow up studies will continue to be implemented and, therefore, analysis of later-life outcomes and cost-benefits will improve. Up to now, it has only been possible to evaluate impacts at age 40 for PPP and at ages 30 to 34 for ABC. Longer term data is desirable. It is of great interest to understand the long term impacts on social mobility of the children of the participants of PPP and ABC. Moreover the long term results on health in Campbell et al. (2014b) are very promising.

New rounds of data should also inform the debate on Head Start –especially its randomized controlled trial, HSIS. The intervention was short and cognitive skills fade out. However, early life skills often mediate later-life outcomes and, therefore, the relevance of long run gains in cognition are not as important as the economic outcomes. The same is true for programs like EHS and IHDP.

A lot of questions remain in this literature. Is it better to start preschool as early as in ABC or IHDP? Or is it better to start late, as in Perry? What is the exact nature of the child-parent-teacher interaction? What are the market and non-market gains and losses for parents? What do they gain and lose from preschool?

Another set of questions is based on the nature of preschool. Is informal childcare always worse than center-based programs? Presumably, it depends on the relative quality of the two options. Evidence from CARE and ETP suggest that on-site programs work better than

home visitation programs, but we find it hard to establish this as a fact. Scenarios in which informal environments are very high quality are not unrealistic. Thus, the economically relevant way to analyze a program is against *all* alternatives: (i) staying at home; (ii) informal childcare; and (iii) other sources of formal care –including public and private programs.

Missing in the entire literature on early childhood interventions is a full integration of the market and the family. Childcare is an input to household production and, when publicly provided, it is also a subsidy for work (Aguiar and Hurst, 2007; Del Boca et al., 2013). Importantly, recent literature reveals that more work does not necessarily mean less child development –our final draft will develop these statements further. Child quality is an output of household activity. Both goods and services are inputs for household production. A full consideration of appropriate policies would recognize that subsidies, for goods, services and time of various forms, may be more be efficient than direct provision of goods and services.

Blau (2003) discusses some of these issues. More broadly, Heckman and Mosso (2014) discuss the evidence on the effectiveness of alternative policies that recognize: (i) the lack of information by agents on parenting and child quality; (ii) the time devoted to child development is not necessarily inversely related to parental employment; (iii) the parental response to public provisions of early childcare enhances child development.

References

Administration for Children and Families (2008). History of head start. http://www.acf.hhs.gov/programs/ohs/about/historyofheadstart. Accessed: 2014/09/02.

Administration for Children and Families (2010). Head start definitions. http://eclkc.ohs.acf.hhs.gov/hslc/standards/hspps/.htm. Accessed: 2014/09/06.

Administration for Children and Families (2013). CCDF Fact Sheet. http://www.acf.hhs.gov/programs/occ/factsheetocc. Accessed: 2014/09/12.

Aguiar, M. and E. Hurst (2007). Measuring trends in leisure: The allocation of time over five decades. *The Quarterly Journal of Economics 122*(3), 969–1006.

Aiyagari, S. R., J. Greenwood, and A. Seshadri (2002). Efficient investment in children. *Journal of Economic Theory 102*(2), 290–321.

Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wößmann (Eds.), *Handbook of the Economics of Education*, Volume 4, pp. 1–181. Amsterdam: Elsevier.

Andrew J. Coulson (2010). Head start: a tragic waste of money. http://www.cato.org/publications/commentary/head-start-tragic-waste-money. Accessed: 2014/11/21.

Badev, A. and F. Cunha (2013). Maternal learning about the technology of skill formation. University of Pennsylvania.

Barnett, W. S. (2011). Effectiveness of Early Educational Intervention. *Science 333*(6045), 975–978.

Barnett, W. S. and L. N. Masse (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review 26*(1), 113–125.

Barnett, W. S. and D. J. Yarosz (2007). Who goes to preschool and why does it matter? *Preschool Policy Matters 7*.

Becker, G. S. (1991). *A Treatise on the Family*. Harvard university press.

Bernal, R. (2008). The Effect of Maternal Employment and Child Care on Children's Cognitive Development. *International Economic Review 49*(4), 1173–1209.

Bernal, R. and M. P. Keane (2010). Quasi-structural Estimation of a Model of Childcare Choices and Child Cognitive Ability Production. *Journal of Econometrics 156*(1), 164–189.

Bernal, R. and M. P. Keane (2011). Child Care Choices and Children's Cognitive Achievement: the Case of Single Mothers. *Journal of Labor Economics 29*(3), 459–512.

Bitler, M., H. Hoynes, and T. Domina (2014). Experimental Evidence on Distributional Effects of Head Start.

Blau, D. (2003). Child Care Subsidy Programs. In *Means-tested transfer programs in the United States*, pp. 443–516. University of Chicago Press.

Blau, D. and J. Currie (2006). Preschool, Daycare, and Afterschool Care: Who's Minding the Kids? *Handbook of the Economics of Education 2*, 1163–1278.

Brooks-Gunn, J., R. T. Gross, H. C. Kraemer, D. Spiker, and S. Shapiro (1992). Enhancing the cognitive outcomes of low birth weight, premature infants: for whom is the intervention most effective? *Pediatrics 89*(6), 1209–1215.

Brooks-Gunn, J., C. M. McCarton, P. H. Casey, M. C. McCormick, C. R. Bauer, J. C. Bernbaum, J. Tyson, M. Swanson, F. C. Bennett, D. T. Scott, et al. (1994a). Early Intervention in Low-birth-weight Premature Infants: Results through Age 5 Years from the Infant Health and Development Program. *Jama 272*(16), 1257–1262.

Brooks-Gunn, J., C. M. McCarton, P. H. Casey, M. C. McCormick, C. R. Bauer, J. C. Bernbaum, J. Tyson, M. Swanson, F. C. Bennett, D. T. Scott, et al. (1994b). Early intervention in low-birth-weight premature infants: results through age 5 years from the infant health and development program. *Jama 272*(16), 1257–1262.

Burchinal, M., M. Lee, and C. Ramey (1989). Type of day-care and preschool intellectual development in disadvantaged children. *Child Development*, 128–137.

Burchinal, M. R., F. A. Campbell, D. M. Brayant, B. H. Wasik, and C. T. Ramey (1997). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development 68*(5), 935–954.

Burke, L. M. and D. B. Muhlhausen (2013). Head Start Impact Evaluation Report Finally Released. *The Heritage Foundation (ed.), Issue Brief 3823*.

Camilli, G., S. Vargas, S. Ryan, and W. S. Barnett (2010). Meta-analysis of the Effects of Early Education Interventions on Cognitive and Social Development. *The Teachers College Record 112*(3).

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, and R. Pinto (2014a). Early childhood program improves adult health. *Science*.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, and R. Pinto (2014b). The effects of early intervention on human development and social outcomes: Provisional evidence from abc and care. *The Economics Journal*, to appear. Unpublished manuscript, University of Chicago, Department of Economics.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014a). Early childhood investments substantially boost adult health. *Science 343*(6178), 1478–1485.

Campbell, F., G. Conti, J. J. Heckman, S. H. Moon, R. Pinto, E. Pungello, and Y. Pan (2014b). Early Childhood Investments Substantially Boost Adult Health. *Science 343*(6178), 1478–1485.

Campbell, F. A., E. P. Pungello, M. Burchinal, K. Kainz, Y. Pan, B. H. Wasik, O. A. Barbarin, J. J. Sparling, and C. T. Ramey (2012). Adult outcomes as a function of an early childhood educational program: an abecedarian project follow-up. *Developmental psychology 48*(4), 1033.

Campbell, F. A. and C. T. Ramey (1994). Effects of early intervention on intellectual and academic achievement: a follow-up study of children from low-income families. *Child development 65*(2), 684–698.

Campbell, F. A., C. T. Ramey, E. Pungello, J. Sparling, and S. Miller-Johnson (2002). Early childhood education: Young adult outcomes from the abecedarian project. *Applied Developmental Science 6*(1), 42–57.

Campbell, F. A., B. H. Wasik, E. Pungello, M. Burchinal, O. Barbarin, K. Kainz, J. J. Sparling, and C. T. Ramey (2008). Young Adult Outcomes of the Abecedarian and CARE Early Childhood Educational Interventions. *Early Childhood Research Quarterly 23*(4), 452–466.

Carneiro, P. and R. Ginja (2014). Long-term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy 6*(4), 135–173.

Cary, J. (1950). *The Horse's Mouth: a Novel*. Harper.

Caucutt, E. M. and L. J. Lochner (2011). Early and late human capital investments, borrowing constraints, and the family. Unpublished manuscript, University of Western Ontario. Retrieved from http://economics.uwo.ca/faculty/lochner/papers/earlyandlatehumancapital_Oct12.pdf.

Chafe, W. H. (1981). *Civilities and Civil Rights: Greensboro, North Carolina, and the Black Struggle for Freedom*. Oxford University Press.

Conti, G., J. Heckman, and R. Pinto (2014). The long-term health effects of early childhood interventions. Under revision, *Economic Journal*.

Cunha, F. (2007). *An Essay on Skill Formation*. Ph. D. thesis, University of Chicago.

Cunha, F. (2012). Eliciting maternal beliefs about the technology of skill formation. Presented at, Family Inequality Network: Family Economics and Human Capital in the Family, November 16, 2012.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2009, April). The economics and psychology of inequality and human development. *Journal of the European Economic Association 7*(2–3), 320–364. Presented as the Marshall Lecture, European Economics Association, Milan, Italy, August 29, 2008.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010a). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010b, May). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Currie, J. (2001). Early Childhood Education Programs. *Journal of Economic Perspectives*, 213–238.

Currie, J. and D. Thomas (1995, June). Does Head Start Make a Difference? *American Economic Review 85*(3), 341–64.

Currie, J. and D. Thomas (1999). Does Head Start Help Hispanic Children? *Journal of Public Economics 74*(2), 235–262.

Currie, J. and D. Thomas (2000). School quality and the longer-term effects of head start. *Journal of Human Resources 35*(4), 755–74.

Del Boca, D., C. Flinn, and M. Wiswall (2013). Household choices and child development. Forthcoming.

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

Del Boca, D., C. J. Flinn, and M. Wiswall (2013). Household choices and child development. Unpublished manuscript, New York University. Forthcoming *Review of Economic Studies*.

Deming, D. (2009). Early Childhood Intervention and Life-cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics 1*(3), 111–134.

Dewey, J. (1916). *Democracy and Education*. The Macmillan Company.

Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. *The 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.

Feller, A., T. Grindal, L. Miratrix, and L. Page (2014). Compared to What? Varation in the Impacts of Early Childhood Education by Alternative Care-Type Setting.

Finlay, B., H. Blank, and N. O. Poersch (1998). Head start: Helping families move from welfare to work.

Garces, E., D. Thomas, and J. Currie (2002). Longer-Term Effects of Head Start. *The American Economic Review 92*(4), 999–1012.

García, J. L. and J. J. Heckman (2014). Ability, Character, and Social Mobility.

García, J. L. and A. Hojman (2014). Parental Time Allocation and Children's Cognitive and Non-cognitive Development.

Gensowski, M., M. Griffin, and J. J. Heckman (2014). Effectiveness of chicago child parent centers in producing young-adult outcomes. *Unpublished Manuscript*.

Gormley, W. T. and T. Gayer (2005). Promoting School Readiness in Oklahoma: an Evaluation of Tulsa's Pre-K Program. *Journal of Human Resources 40*(3).

Gormley Jr, W. T., T. Gayer, D. Phillips, and B. Dawson (2005). The Effects of Universal Pre-K on Cognitive Development. *Developmental Psychology 41*(6), 872.

Gray, S. W. (1969). Selected longitudinal studies of compensatory education–a look from the inside.

Gray, S. W. and R. A. Klaus (1965). An experimental preschool program for culturally deprived children. *Child Development*, 887–898.

Gray, S. W. and R. A. Klaus (1970). The early training project: a seventh-year report. *Child Development*, 909–924.

Gray, S. W., B. K. Ramsey, and R. A. Klaus (1982). *From 3 to 20: The Early Training Project.* Baltimore: University Park Press.

Hall, J. D. (2005). The long civil rights movement and the political uses of the past. *The Journal of American History 91*(4), 1233–1263.

Hart, B. and T. R. Risley (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: P.H. Brookes.

Haskins, R. (2010, October 11). Finally, Obama Administration is Putting Head Start to the Test.

Haxton, B. (2013). A brief history of head start.

Heckman, J., S. Kuperman, and C. Cheng (2014). Comparing the mechanisms producing key outcomes in high quality early childhood programs with long-term evidence. Department of Economics, University of Chicago.

Heckman, J., S. H. Moon, R. Pinto, P. Savelyev, and A. Yavitz (2010a). 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. (1974). Effect of Childcare Programs on Women's Work Effort. *Journal of Political Economy*, S136–S163.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 153–161.

Heckman, J. J. (2008). Schools, Skills, and Synapses. *Economic Inquiry 46*(3), 289–324.

Heckman, J. J. (2012). *Giving Kids a Fair Chance (A Strategy That Works)*. Cambridge, MA: MIT P.

Heckman, J. J., N. Hohmann, J. Smith, and M. Khoo (2000, May). Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics 115*(2), 651–694.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour economics 19*(4), 451–464.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3A, Chapter 31, pp. 1865–2097. New York: North-Holland.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010b). The Rate of Return to the HighScope Perry Preschool Program. *Journal of Public Economics 94*(1), 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, and P. Savelyev (2013). Understanding the Mechanism through which and Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–86.

Heckman, J. J., R. Pinto, A. Shaikh, and A. Yavitz (2010). Inference with imperfect randomization: The case of the Perry Preschool Program. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J. and B. Setzler (2014). A Reanalysis of the Infant Health and Development Program. Forthcoming.

Heckman, J. J. and S. Urzua (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics 156*(1), 27–37.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4875–5144. Amsterdam: Elsevier.

Hill, C., W. Gormley, and S. Adelstein (2012). Do the Short-term Effects of a Strong Preschool Program Persist. *Center for Research on Children in the United States Working Paper 18*.

Hojman, A. (2014). Evidence on the fade-out of iq gains from early childhood interventions: a skill formation perspective. Working paper, University of Chicago.

Jerome, C. C. (1995). Black People in White Face: Assimilation, Culture, and the Brown Case. *William & Mary Law Review 36*.

Klaus, R. A. and S. W. Gray (1968). The early training project for disadvantaged children: A report after five years. *Monographs of the Society for Research in Child Development*, iii–66.

Klein, J. (2011, July 7). Time to Ax Public Programs That Don't Yield Results.

Kline, P. and C. Walters (2014). Evaluating Public Programs with Close Substitutes: The Case of Head Start.

Lareau, A. (2011). *Unequal Childhoods: Class, Race, and Family Life* (2 ed.). Berkeley, CA: University of California Press.

Love, J. M., J. Constantine, D. Paulsell, K. Boller, C. Ross, H. Raikes, C. BradySmith, and J. BrooksGunn (2004). The role of early head start programs in addressing the child care needs of lowincome families with infants and toddlers: Influences on child care use and quality. *US Department of Health and Human Services Head Start Bureau*.

Ludwig, J. and D. L. Miller (2007a). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *The Quarterly Journal of Economics 122*(1), 159–208.

Ludwig, J. and D. L. Miller (2007b). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly Journal of Economics 122*(1), 159–208.

Ludwig, J. and D. A. Phillips (2008). Long-term effects of head start on low-income children. *Annals of the New York Academy of Sciences 1136*(1), 257–268.

Masse, L. N. and W. S. Barnett (2002). *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention*. National Institute for Early Education Research New Brunswick, NJ.

McCarton, C. M., J. Brooks-Gunn, I. F. Wallace, C. R. Bauer, F. C. Bennett, J. C. Bernbaum, R. S. Broyles, P. H. Casey, M. C. McCormick, D. T. Scott, et al. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants: The Infant Health and Development Program. *Jama 277*(2), 126–132.

McCormick, M. C., J. Brooks-Gunn, S. L. Buka, J. Goldman, J. Yu, M. Salganik, D. T. Scott, F. C. Bennett, L. L. Kay, J. C. Bernbaum, et al. (2006). Early intervention in low birth weight premature infants: results at 18 years of age for the infant health and development program. *Pediatrics 117*(3), 771–780.

Muhlhausen, D. B. and D. Lips (2010). Head Start Earns an F: No Lasting Impact for Children by First Grade. *The Heritage Foundation*.

NIEER (2013). The state of preschool 2013.

Nores, M. and W. S. Barnett (2010). Benefits of Early Childhood Interventions across the World: (Under) Investing in the Very Young. *Economics of Education Review 29*(2), 271–282.

Office of Head Start (2011). Biennial Report to Congress: The Status of Children in Head Start Programs.

Puma, M., S. Bell, R. Cook, C. Heid, G. Shapiro, P. Broene, F. Jenkins, P. Fletcher, L. Quinn, J. Friedman, et al. (2010). Head Start Impact Study. Final Report. *Administration for Children & Families*.

Ramey, C., G. McGinness, L. Cross, A. Collier, S. Barrie-Blackley, and K. Borman (1979). The Abecedarian approach to social competence: Cognitive and linguistic intervention for disadvantaged preschoolers.

Ramey, C. T. et al. (1979). The Abecedarian Approach to Social Competence: Cognitive and Linguistic Intervention for Disadvantaged Preschoolers.

Ramey, C. T., D. M. Bryant, B. H. Wasik, J. J. Sparling, K. H. Fendt, and L. M. La Vange (1992). Infant Health and Development Program for low birth weight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics 89*(3), 454–465.

Ramey, C. T. and F. A. Campbell (1979). Compensatory education for disadvantaged children. *The School Review*, 171–189.

Ramey, C. T. and R. Haskins (1981). The modification of intelligence through early experience. *Intelligence 5*(1), 5–19.

Ramey, C. T. and S. L. Ramey (1998). Prevention of intellectual disabilities: early interventions to improve cognitive development. *Preventive medicine 27*(2), 224–232.

Reynolds, A. J., J. A. Temple, B. A. White, S.-R. Ou, and D. L. Robertson (2011). Age 26 Cost-Benefit Analysis of the Child-Parent Center Early Education Program. *Child Development 82*(1), 379–404.

Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores (2005). *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: High/Scope Press.

Sparling, J. (2010). Highlights of Research Findings from the Abecedarian Studies.

Sparling, J. and I. Lewis (1979). Learningames for the first three years: A guide to parent-child play.

Sparling, J., I. Lewis, C. T. Ramey, B. H. Wasik, D. M. Bryant, and L. M. LaVange (1991). Partners A Curriculum to Help Premature, Low Birthweight Infants Get Off to a Good Start. *Topics in Early Childhood Special Education 11*(1), 36–55.

Steve Aos and Roxanne Lieb and Jim Mayfiel and Marna Miller and Annie Pennucci (2004, September). Benefits and Costs of Prevention and Early Intervention Programs for Youth. Technical report, Washington State Institute for Public Policy.

Torcasso, J. (2014). Head start program and the head start impact study. Working paper, University of Chicago.

Vogel, C. A., N. Aikens, A. Burwick, L. Hawkinson, A. Richardson, L. Mendenko, and R. ChazanCohen (2006). Findings from the survey of early head start programs: Communities, programs, and families. final report. *US Department of Health and Human Services*.

Wasik, B. H., C. T. Ramey, D. M. Bryant, and J. J. Sparling (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child development 61*(6), 1682–1696.

Weikart, D. P. (1967). Preliminary Results From a Longitudinal Study of Disadvantaged Preschool Children. Technical report, Paper presented at the convention of the Council for Exceptional Children, St. Louis, MO.

Weikart, D. P. (1970). Longitudinal Results of the Ypsilanti Perry Preschool Project. Final Report. Volume II of 2 Volumes. Technical report.

Weikart, D. P., J. T. Bond, and J. T. McNeil (1978). *The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade*. High/Scope Educational Research Foundation.

Weikart, D. P., C. K. Kamii, and N. L. Radin (1964). Perry Preschool Project: Progress Report. Technical report, Ypsilanti Public Schools, Ypsilanti, Mich.

Zigler, E. (2010). Putting the National Head Start Impact Study into a Proper Perspective. *National Head Start Association (NHSA) 13*(1), 1–6.

Zigler, E. and S. J. Styfco (2010). *The Hidden History of Head Start*. Oxford University Press.

A List of Abbreviations

ABC	Carolina Abecedarian Project
BA	Bachelors Degree
CARE	Carolina Approach to Responsive Education
CAT	California Achievement Test
CBA	Cost Benefit Analysis
CBCL	Achenbach Child Behavior Checklist
CBI	Classroom Behavior Inventory
CCDF	Child Care Development Fund
CCDP	Comprehensive Child Development Program
cNLSY79	Children of the National Longitudinal Survey of Youth 1979
CPC	Chicago Child-Parent Center
ECI	Early Childhood Interventions
EHS	Early Head Start
ES	Even Start
ETP	Early Training Project
FPGC	Frank Porter Graham Child Development Institute
HLBW	Higher Low Birth Weight (IHDP)
HOME	Home Observation for Measurement of Environment Instrument
HS	Head Start
HSIS	Head Start Impact Study
IBR	Bayley's Infant Behavior Record
IHDP	Infant Health and Development Program
IQ	Intelligence Quotient
ISS	Intensive Substudy
ITT	Intent to Treat
K&R	Kohn and Rosman Test Behavior Instrument
KG	Kindergarten
LLBW	Lighter Low Birth Weight
LATE	Local Average Treatment Effect
LPN	Licensed Practical Nurse
MA	Masters Degree
MAT	Metropolitan Achievement Test
MSW	Master of Social Work
NCAST	Parent-Child Interaction Scales
NIEER	National Institute for Early Education Research
NLSY79	National Longitudinal Survey of Youth 1979
NSO	National Study Office (for IHDP)
OECD	Organization for Economic Co-operation and Development
OEO	Office of Economic Opportunity
PBI	Pupil Behavior Inventory
PDO	Program Development Office (for IHDP)
PFA	Preschool for All
PPP	Perry Preschool Program
PPVT	Peabody Picture Vocabulary Test
PPVT-R	Peabody Picture Vocabulary Test Revised
PreK	Prekindergarten
PSID	Panel Study of Income Dynamics
RAPS	Research Assessment Package for Schools
RCT	Randomized Controlled Trial
RN	Registered Nurse
SB	Stanford-Binet Intelligence Scale
SD	Standard Deviation
STAR	Tennessee Project Student-Teacher Achievement Ratio
TN-VPK	Tennessee Voluntary Pre-K Program
TVIP	Test de Vocabulario en Imagenes Peabody (PPVT test for Spanish-speaking and bilingual students)
WIS	Wechsler Intelligence Test
WISC	Wechsler Intelligence Scale for Children
WJ	Woodcock Johnson Test of Achievement
YRS	Ypsilanti Rating Scale

B Data Sources

In order to explore the treatment effects on early-life outcomes of the ECIs, we generate factors to act as measures of several outcomes of interest. In particular, we factor analyze cognitive and character outcomes during the early ages. We choose age 8 to be our point of comparison because this is the age at which our data is most comparable across our five programs: Perry, ABC, CARE, IHDP, and ETP. Tables A.1 and A.2 list the instruments we use to generate the early age outcomes.

As our measure of cognitive ability, we use standardized IQ and achievement tests scores. In the case of Perry and ETP, achievement tests scores are converted from their raw form to their standardized forms at the University of Chicago Test Center.

To measure the non-cognitive early childhood outcomes, we follow the Big Five paradigm and construct factors measuring openness to experience, conscientiousness, extroversion, agreeableness, and emotional stability.

- **Openness to experience** refers to the child's tendency to be receptive of new aesthetic, cultural or intellectual experiences.
- **Conscientiousness** reflects the child's tendency to be organized, responsible, and hardworking.
- **Extroversion** is a measure of the child's preference for the outer world of people over the inner world of subjective experience.
- **Agreeableness** is a measure of the child's cooperation and altruism.
- **Emotional stability** is most easily understood as the reversal of neuroticism, which is an assessment of how prone an individual is to psychological distress. Neurotic individuals are easily stressed or upset, whereas emotionally stable ones are more resolute and resilient in their emotions.

To generate the factors measuring each trait described above, question items from instruments measuring personality traits are analyzed and designated to a single trait. Each group of items is then standardized and factor analyzed using the principal factor method, retaining only one factor. Mean imputation is applied to items with missing values after standardization, and prior to factor analysis. For each program, this procedure is repeated for every age that a personality test was administered. Similar to IQ and achievement, we then factor analyze the factors measuring each trait across the ages available, generating another set of character factors that is representative of the early years of the children's lives.

Table A.1: Sources for Outcomes at age 8 in ABC and CARE

Measure	ABC	CARE
IQ	WIS	WISC
Achievement	CAT, PIAT, WJ	WJ
Openness	CBI, CBCL Teacher, Walker	CBI, CBCL Teacher
Conscientiousness	CBI, CBCL Teacher, Walker	CBI, CBCL Teacher
Extroversion	CBI, CBCL Teacher, Walker	CBI, CBCL Teacher
Agreeableness	K&R, CBI, CBCL Teacher, Walker	K&R, CBI, CBCL Teacher
Emotional Stability	K&R, CBI, CBCL Teacher, Walker	K&R, CBI, CBCL Teacher

Note: IQ Tests: SB, Stanford Binet Intelligence Scale; WIS, Weschler Intelligence Scale; WISC, Wechsler Intelligence Scale for Children. **Achievement Tests:** CAT, California Achievement Test; MAT, Metropolitan Achievement Test; PIAT, Peabody Individual Achievement Test; WJ, Woodcock-Johnson Test of Achievement. **Personality Tests:** CBCL, Achenbach Child Behavior Checklist; CBI Classroom Behavior Inventory; IBR, Bayley's Infant Behavior Record; K&R, Kohn and Rosman Test Behavior Instrument; PBI, Pupil Behavior Inventory; RAPS, Research Assessment Package for Schools; Richman, Richman Child Behavior Problems Checklist; Walker, Walker Problem Behavior Identification Checklist; YRS, Ypsilanti Rating Scale.

Table A.2: Sources for Outcomes at age 8 in ETC, IHDP, and Perry

Measure	ETP	IHDP	Perry
IQ	SB	WISC	SB
Achievement	MAT	WJ	CAT
Openness	Teaching Rating	Richman, CBCL Parent, Teacher Survey, RAPS Teacher	PBI, YRS
Conscientiousness	Teaching Rating, Reputation Test	IBR, CBCL Parent, Teacher Survey, RAPS Teacher	PBI
Extroversion	Teaching Rating, Reputation Test	IBR, CBCL Parent, Teacher Survey	PBI YRS
Agreeableness	Teaching Rating, Reputation Test	Richman, CBCL Parent, Teacher Survey, RAPS Teacher	PBI
Emotional Stability	Teaching Rating, Reputation Test	Richman, CBCL Parent, Teacher Survey, RAPS Teacher	PBI, YRS

Note: IQ Tests: SB, Stanford Binet Intelligence Scale; WIS, Weschler Intelligence Scale; WISC, Wechsler Intelligence Scale for Children. **Achievement Tests:** CAT, California Achievement Test; MAT, Metropolitan Achievement Test; PIAT, Peabody Individual Achievement Test; WJ, Woodcock-Johnson Test of Achievement. **Personality Tests:** CBCL, Achenbach Child Behavior Checklist; CBI Classroom Behavior Inventory; IBR, Bayley's Infant Behavior Record; K&R, Kohn and Rosman Test Behavior Instrument; PBI, Pupil Behavior Inventory; RAPS, Research Assessment Package for Schools; Richman, Richman Child Behavior Problems Checklist; Walker, Walker Problem Behavior Identification Checklist; YRS, Ypsilanti Rating Scale.