

Early Learning Across Decades: Advances in Measuring Head Start Effectiveness*

Abstract

The Head Start program, launched in 1965 and targeted to children from disadvantaged backgrounds, remains the largest early childhood care and education (ECE) program in the United States and the only one deployed at the federal level. As such, the Head Start literature spans several decades and now allows for synthesis of findings from different contexts, time periods, and research designs. This paper provides a comprehensive assessment of the rigorous evidence measuring Head Start's effects on children, their families, and society. The focus is on how (1) the contrast between program and counterfactual conditions, (2) takeup of the program among eligible populations, and (3) treatment-effect heterogeneity inform interpretation and applicability of key findings. The paper presents implications of the evidence for the modern-day Head Start program and an adjacent, policy-relevant research agenda.

JEL classification

I24, I38, J13

Keywords

early childhood, Head Start, child development

Corresponding author

Chloe R. Gibbs

chloe.gibbs@nd.edu

* *Acknowledgements:* I thank David Romer and four anonymous reviewers for their valuable feedback. I am grateful to Martha Bailey, Steven Durlauf, Hilary Hoynes, Kathleen McGarry, and participants at the "Lessons from the 1960s for Inequality" workshop at UCLA in January 2024 for insightful comments on a related presentation, and to Victoria Barone, Jennifer Doleac, Lisa Gennetian, Trevon Logan, Jens Ludwig, Doug Miller, and attendees at talks at Duke's Center for Child and Family Policy and UCL's Centre for Education Policy and Equalising Opportunities for helpful conversations. Remaining errors and omissions are my own.

1 Introduction

The Head Start program, launched as a foundational component of the War on Poverty, began in the summer of 1965. The program provides comprehensive supports to young children from low-income families to prepare them for formal school entry. Throughout its 60-year history, measuring Head Start’s impact on children and their communities has been the subject of much inquiry. And across the decades, there has been contention about the program’s objectives and its effectiveness as well as the applicability of lessons learned to current policy contexts. The literature’s maturity now facilitates extraction of key findings and insights from across studies, empirical strategies, and cohorts of participating children. As social science methods have improved, researchers have more carefully addressed the endogeneity of Head Start exposure with study designs that remove from estimates of program impact underlying differences between Head Start participants and their non-participating peers.

Previous summaries of the extant literature synthesize findings on preschool programs generally (Cascio and Schanzenbach, 2014; Currie, 2001; Duncan and Magnuson, 2013; Elango et al., 2015), and Head Start specifically (Gibbs et al., 2011; Ludwig and Phillips, 2007, 2008). In a recent *Science* “Policy Forum,” Burchinal and colleagues (2024) review evidence on preschool effectiveness with a focus on experimental studies. These summaries suggest that early childhood programs can have important short- and long-term effects, but not in every instance, and these effects may be declining over time (i.e., more recent deployments of Head Start and other preschool programs may be less effective) (Burchinal et al., 2024; Duncan and Magnuson, 2013).

Existing work also points to the pattern of “fadeout” of initial program effects in the years following the program, particularly on measures of cognitive skills, highlighting that the evaporation of early test score advantages is both a common finding and can still be consistent with realization of improvements on long-term outcomes (Duncan and Magnuson, 2013; Gibbs et al., 2011, 2016). Scholars have pointed to the types of skills that early childhood care and education (ECE) interventions affect and the subsequent environments that children experience as potential explanations for these patterns (Bailey et al., 2017; List and Uchida, 2024).

This literature evolved as much applied empirical work in economics has, beginning with study designs employing matching, attempting to control for—often a very limited set of—observable characteristics of children. Researchers refined their approaches to studying Head Start over time with an increasing focus on accounting for unobservable factors that influence program exposure or participation. Early attempts to address selection employed

family fixed effects, using within-family variation in Head Start participation, but these approaches may be limited in their ability to isolate program effects when sibling-specific shocks or time-varying differences remain and additionally raise important questions about selection into identification.

As causal methods improved, so too has our collective understanding of Head Start program impact, through quasi-experimental studies leveraging features of program availability and eligibility, and a large, nationally representative randomized control trial commissioned by the federal government and conducted in the early 2000s. The latter study facilitated documentation and exploration of treatment-effect heterogeneity, providing better understanding of who benefits in what contexts. Importantly, existing syntheses of early childhood program, and Head Start, effectiveness pre-date or exclude several of these important developments in the literature. While evidence suggests that Head Start *can improve* children’s short- and long-term outcomes, there is no single Head Start program and effectiveness varies considerably across centers.

This analytic review of the Head Start literature first briefly presents the economic perspective on investment in disadvantaged children’s early childhood years, a conceptual foundation that mirrors the motivation of Head Start’s creators in the 1960s. The paper then discusses the importance of program–no program contrast when understanding Head Start impact, and documents how both Head Start and counterfactual conditions have changed over time in important ways.

With this context, this review considers the nature of the comparisons drawn in rigorous studies of Head Start impact, the treatment effects estimated under different assumptions and the heterogeneity thereof, and the challenges in measuring takeup. The paper synthesizes what we know—and still need to know—about Head Start effectiveness, and discusses the relevance of existing findings, particularly from early deployments of the Head Start program, for current policy. The evidence coalesces around several key themes. Head Start can be effective at improving children’s outcomes but there is considerable variation in program effectiveness by context, children served, and the nature of the program delivered. Children who enter the program with greater disadvantage and those who would otherwise be in informal care are the most likely to benefit. This evidence points to the importance of program quality, and in particular, identifying children, families, and places for whom and for which Head Start programming would generate the largest contrast, relative to counterfactual experiences, in inputs to children’s early development.

In addition, the literature suggests that cognitive test scores over the short-term may be uninformative or loosely informative about the likely long-term effects of the program. A greater understanding of the pathways through which ECE investments alter long-run life

chances could facilitate better measurement and assessment of program effectiveness more immediately. Focusing exclusively on measuring the direct effects of Head Start on participants misses the dual role ECE programming plays in supporting children’s development while allowing parents to engage in gainful activities. Though the evidence is more limited, Head Start does increase maternal employment and may generate important spillovers to peers, siblings, and communities. Finally, and relatedly, our current tools for measuring program impact are ill-suited for assessing Head Start’s effectiveness in the changing landscape of ECE in the United States.

The paper proceeds as follows: Section 2 discusses the economic arguments for investments in early childhood, the impetus of Head Start, and describes the characteristics of children served by the program over time. Section 3 presents a conceptual framework for understanding program—no program contrast in inputs to human capital production, and evidence about how this contrast has evolved over the past several decades. Section 4 provides an overview of approaches to measuring Head Start effectiveness and the treatment effects estimated in different research designs, and Section 5 outlines implications for a Head Start research agenda building on what we have learned.

2 Early Childhood Investments

Early childhood as a foundational, and particularly impactful, period for investment in children’s development has garnered much interest among scholars, policymakers, and practitioners alike. Economists come to the study of early childhood through the lens of human capital investment and the returns to such investment across the lifecycle. Many of those same arguments undergirded the creation of the Head Start program and its inclusion in the broader War on Poverty effort.

2.1 Early Years as a Critical Period

Early childhood interventions have generated attention as an efficient, and potentially equity-enhancing, means to support children’s early skill development and later-life wellbeing (Currie, 2001; Heckman, 2000; Heckman and Masterov, 2007). The underlying conceptual foundation suggests early childhood investments have the potential to generate substantial returns to individuals, their families and communities, and society as a whole. First, researchers have extensively documented the early emergence and persistence of gaps in skills and outcomes by family socioeconomic status (Duncan and Magnuson, 2013; Lee and Burkam, 2002; Waldfogel and Washbrook, 2011). In addition to these documented gaps in both cognitive and

non-cognitive skills, researchers have established both the importance of early skill development for the manifestation of subsequent skills, and the role that that early gaps play in the formation and persistence of later gaps in educational attainment and labor market success (Currie and Thomas, 1999; Duncan et al., 2007; Heckman, 2000; Waldfogel and Washbrook, 2011).

The productivity of early childhood investments stems from complementarities between early skill development and subsequent human capital formation as well as the long time horizon over which they can realize their return. Researchers also point to the malleability of the young brain, fostering acute susceptibility to developmental intervention at young ages. The developmental psychology literature asserts that this plasticity is declining in age supporting the notion that earlier is better (Knudsen et al., 2006; Shonkoff and Phillips, 2000). Bailey and coauthors (2017) describe two types of pathways through which early childhood investments may generate durable skills: through the development of malleable, fundamental skills that would not have emerged otherwise, and through the production of right-time skills and capacities that allow a child to avoid bad outcomes or capitalize on opportunities at a particular moment in time (“foot-in-the-door”). These two conceptualizations of skill-building illuminate ways in which skill development in the early years could persist, or not, beyond the timing of intervention.

Against this theoretical backdrop, there is a growing body of empirical research documenting that early childhood programs *can generate* meaningful and lasting improvements in children’s life chances, including evidence from the Abecedarian Project, Boston pre-kindergarten programs, the Perry Preschool Project, the Project STAR class-size reduction intervention in kindergarten, and Head Start itself. This burgeoning literature documents effects on educational attainment, labor market success, and other measures of health and well-being into adulthood (Campbell et al., 2014; Dynarski et al., 2013; Chetty et al., 2011; Deming, 2009; Gray-Lobe et al., 2022; Schweinhart et al., 2005). Because of the impact on important long-run outcomes, with societal implications beyond the privately realized benefits, programs targeting the earliest years of life can produce sizable social benefits when assessed in a marginal value of public funds (MVPF) framework (Hendren and Sprung-Keyser, 2020).

2.2 Head Start Origins

Built on these arguments and designed by developmental psychologists and social scientists, the Head Start program began in 1965 as a component of Lyndon B. Johnson’s War on Poverty, initially funding 8-week summer preschool programs across the United States and serving over 500,000 children. President Johnson declared in May 1964—as Head Start and

the suite of War on Poverty programs were being designed—that, “poverty must not be a bar to learning and learning must offer an escape from poverty.”

Head Start’s creators intended for it to support child development and to prepare children from disadvantaged households for school entry, but it varied considerably in how (and how intensely) it was deployed at the local level, by design. The White House Office of Economic Opportunity (OEO) issued a grant application that emphasized “medical, cultural, and self-esteem or self-discovery activities,” notably omitting any explicit mention of cognitive or academic skill-building as a critical component. Additionally, there was an emphasis—across all OEO activities—on community, and especially parent, involvement. Local control, which is emphasized to this day, was facilitated by the flow of funding directly from the federal government to local grantees. In the early years of Head Start rollout, program administrators emphasized its nationwide reach and the diversity of communities receiving funds.

Consistent with a “whole child” focus of the program, Head Start provided many services that resembled those of a health care center and community center, including medical examinations and dental care, parent education and community organizing and development, and direct employment of community members and parents in center operations. Many children received vaccinations and their first checkups with a doctor and dentist through the program, and Head Start centers operated as hubs for families in accessing other social services, including nutrition and health care. While there was variation, the program was low duration, on average, in those early years, with children only participating in a few hours of programming per week.

2.3 Children Served, Over Time

At its inception, the Head Start program served very disadvantaged participants, and children living in households below the poverty line in 1965 lived in very resource-deprived circumstances. According to U.S. Census Bureau program reports on the early years of Head Start: 50 percent of program participants were Black; less than one-third of participating children’s mothers had high school diploma or more schooling; nearly 70 percent of mothers were unemployed; and approximately 10 percent of families reported no running water at home.

The average Head Start participant has changed over time, due in part to demographic change, but also changes to the material conditions children in low-income families experience. As displayed in Figure 1, since 1988, Head Start has increasingly served younger participants (i.e., more infants and toddlers and relatively fewer preschool-aged children).

The proportion of Head Start participants that are Black has declined and nearly 40 percent of present-day participants are Hispanic. About a quarter of Head Start children speak Spanish as their primary home language. Characterizing the context in which Head Start centers operate, participating children are less likely to have no health insurance or to have no consistent dental care at the time of Head Start enrollment (since 2003 when programs began collecting these measures, and certainly since program inception when children and families had far less access to health insurance and care).

Today, Head Start serves over 800,000 young children with over \$11 billion in annual federal appropriations, and the program continues to serve children from families earning income at or below the federal poverty level. As such, Head Start evidence speaks to the effects of a targeted, means-tested program, which could be quite different from the effects of a universal or near-universal program like many state public preschool programs currently in operation or under consideration. In fact, Cascio (2023) documents larger gains for children from low-income families from universal programs relative to targeted programs, justifying their relatively higher costs per disadvantaged child served.

3 Understanding Counterfactuals

In the 60 years since Head Start began, the program has evolved. Alongside those changes, the early childhood experiences of U.S. children, including children from low-income households, have changed considerably. Interpreting the trends in Head Start effects — and the relative magnitude of changes in counterfactual conditions and the nature of Head Start inputs to child development over time — has been the subject of interest among researchers, though to a limited extent. Researchers have pointed to this evolution of the relative experiences, particularly in the gap in Head Start and counterfactual inputs to early childhood development for children from low-income families, as potentially contributing to declining Head Start effectiveness over time. Policymakers, however, are vitally interested in this question of whether, in fact, Head Start has become relatively “worse” over time. This line of inquiry rests, of course, on an empirical question, but one that is difficult to answer with available data.

3.1 Assessing Contrast

Because differences in outcomes for program participants and non-participants are likely a function of differences in inputs to the production of human capital, it is important to understand *both* what the Head Start program provides and what eligible, exposed, or par-

participating children would otherwise experience in the early childhood years in the absence of the program. Both of these pursuits are challenging in terms of data availability and our capacity to measure — and our knowledge of how to best do so — key aspects of the early childhood environment.

Any gap in developmental inputs likely corresponds to the expected change, or lack thereof, in outcomes. Figure 2 presents a simple depiction of comparing inputs to the production of children’s human capital (1) in the absence of Head Start (“home” and other early childhood experiences) and (2) with Head Start across the distribution of family socioeconomic status. The quantity and quality of counterfactual experiences are increasing in family resources, both historically and today, and since Head Start serves children from disadvantaged families, the change in human capital inputs for Head Start-participating children is only realized at the low-end of the family resources distribution (up to \bar{s} , the SES-eligibility threshold for Head Start participation). The Appendix presents a simple corresponding model to this framework in which outcomes for Head Start children and non-Head Start children are a function of the human capital inputs.

The figure depicts improvements in both the Head Start experiences and the counterfactual “home” or non-Head Start experiences of young children over time. There are likely two important *changes over time* to the relationship between family socioeconomic status and human capital inputs, in that counterfactual experiences have improved across the family resources distribution (an upward shift in the curve) and perhaps, although less well documented, more pronouncedly for high-income families (a steepening of its slope). The Head Start program has also changed over time, likely constituting an improvement in both the quantity and quality of inputs that the program provides.

Two predictions derive from the simple conceptual model: (1) that for any individual child or group of children in Head Start, their outcomes are a function of the gap in inputs generated from the program–no program contrast in their circumstances and settings, and (2) that changes over time in the effectiveness of Head Start are a function too of whether those differences have grown or shrunk over the same period. Straightforwardly, the open question about Head Start effectiveness over time — whether the program has gotten better or worse — corresponds to the relative size of the gaps in program–no program experiences historically and today.

While changes for the entire family SES distribution would be of interest when thinking about universal or quasi-universal preschool programs, we are focused on the left tail. If the change in the quantity and quality of Head Start inputs from 1965 to 2025 is larger than the change in the quantity and quality of counterfactual inputs for children from low-SES families over that same timeframe, Head Start would be more effective today than in early

iterations of the program. Conversely, if counterfactual conditions have improved more from 1965 to 2025, Head Start would be less effective. And finally, what may be empirically true in specific contexts in the U.S., if both Head Start and counterfactual experiences have improved to a similar degree over time, Head Start could be similarly effective today, relative to a much improved set of outside options.

3.2 Changes in the Program Over Time

While legacies of the original Head Start program continue to shape its modern-day implementation, much has changed as well. Figure 3 shows the growth in federal appropriations and in program enrollment, depicting the quick pivot in the early years from a summer program serving more children to a more intensive school-year program serving fewer children. The Head Start program grew in the 1990s, concurrent with welfare reform and expansions of the program, and largely leveled off in the early 2000s. The program contracted in the COVID-19 pandemic and remains smaller in total enrollment and funding relative to its pre-pandemic size.

Changes to programming accompanied the program's growth. As depicted in Figure 4, the 1990s and 2010s constituted periods of increasing program intensity as measured by per child spending in the program. To put Head Start's per pupil spending in context of K-12 expenditures, current levels of annual spending per enrolled child are similar, but growth in Head Start spending has been much more dramatic, increasing 140 percent since 1990 as K-12 expenditures have increased 49 percent in that same timeframe. Over this time period, the service hours and mix for the average Head Start child are also changing. While Head Start continues to provide a range of programming, including health and nutrition services and social and emotional skill development, with a focus on parent involvement and community partnerships, there is now an explicit focus on cognitive skill development, including literacy and numeracy. The Head Start program has intensified in some ways alongside its emphasis on academic school readiness, including heightened expectations for teacher credentials, increased duration, and smaller child-caregiver ratios.

Figure 5 depicts some of these programmatic changes, as supplied by Head Start grantees in annual reports. In particular, Head Start children are much more likely today to be in a full-day, full-week spot relative to the 1990s version of the program. Children are less likely to have access to transportation from their program. The proportion of Head Start teachers who exit in the course of a year ("teacher turnover"), which we can measure starting in the early 2000s, typically hovers around 15 percent. In recent COVID-affected years, it jumped to as high as 40 percent and remains elevated. Turnover for other types of staff in

Head Start centers has been relatively constant since 1988, at about 10 percent. According to the Program Information Reports data, child-caregiver ratios reached their series low of 10:1 just prior to the pandemic and have reverted to 15:1, where they hovered from 2003 to 2013.¹ Child-home visitor ratios are high at the center level (100:1) because home visitors only typically provide intensive services to a subset of Head Start-participating children and conduct the one-on-one services on a rotating basis. These ratios have been stable over the series with small decreases in the pandemic-affected years.

We know little about measuring “quality” in Head Start, and in ECE settings more generally. Several tools exist to capture quality and states build information and reporting systems around measures of quality in ECE, but there is scant causal evidence on whether specific features or characteristics of programs lead to improved child outcomes. Suggestive evidence points to the importance of attached, stable relationships between child and caregiver as a critical component (Pianta, 1997; Sabol and Pianta, 2012). Other work investigating Head Start quality documents correspondence between higher Head Start spending and more funding directed to child-specific uses and better child outcomes (Currie and Neidell, 2007), and between home visiting and full-day programming and center-level effectiveness (Walters, 2015). The latter finding has been confirmed in experimental studies of full-day settings in other early childhood contexts (Atteberry et al., 2019; Gibbs, 2023).

3.3 Counterfactual Conditions Over Time

There have been substantial changes to children’s early childhood experiences, apart from Head Start availability and participation, fundamentally altering the counterfactual to Head Start for likely-eligible children. Parents spend more time with children and invest in their development in other ways, including more reading time, learning materials, and educational activities (Bassok et al., 2016). While these increases have been realized for all families regardless of socioeconomic status, some argue that the growth has been most pronounced for higher-income families (Kalil et al., 2016).

The formal early childhood care and education landscape has also changed considerably alongside the growth in Head Start. Despite some declines during the COVID-19 pandemic, far more children participate in formal “preschool” in public and private settings before entering K-12 schooling in the U.S., growing from about 10 percent in 1965 to 60 percent today, though there are pronounced disparities by race/ethnicity and family socioeconomic status (Cui and Natzke, 2021; Flood et al., 2022). Forty-four states and DC now have public

¹It is worth noting that some reduction in ratios would be mechanically related to the changing age composition of Head Start participants, with the shift in the last two decades from preschool-aged children towards serving more toddlers, as displayed in Figure 1.

pre-K programs, and 37% of four year olds participate in state-funded pre-K (Friedman-Krauss et al., 2024). Participation rates in public pre-K are far lower when focusing on three year olds, reaching eight percent in the 2023–2024 school year.

As mentioned, the average Head Start participant today is less likely to enter the program without health insurance or having never received dental care. These trends point to important changes to the context in which Head Start operates, particularly with respect to social services and supports for low-income families. Many of these programs, including nutrition assistance and health care coverage, developed and expanded alongside the Head Start program.

4 Measuring Head Start Effectiveness

Researchers and those interested in evidence-based policymaking know that measuring the impact of the Head Start program is complicated by evaluation challenges that have existed since the program began and are common across social policy impact evaluation. Early studies of Head Start were limited in their ability to overcome these challenges, but recent methodological innovations have allowed for advances in this area.

4.1 Challenges in Research Design

Researchers have over time sought to identify comparable groups of non-exposed children, families, or communities to better estimate and understand Head Start impact. Selection into Head Start exposure manifests in both the availability of Head Start across communities and in participation in Head Start by children and families, hindering researchers’ ability to isolate the effects of the program from other potential confounding factors. Because these underlying selection mechanisms could lead to countervailing predictions about the interaction between participant characteristics and program effectiveness, it is difficult to sign the direction of the likely bias in comparing those exposed to Head Start and not.

In particular, families eligible for Head Start are likely negatively selected because of their disadvantaged circumstances, but among eligible families, those that seek out and secure Head Start for their children may be positively selected on unobservable dimensions. Similarly, communities that operate Head Start programs have otherwise unmet need for early childhood programming for children from disadvantaged families and may lack other resources; however, communities that have established partnerships to provide Head Start and apply for and receive Head Start funding may have infrastructure and social capital advantages relative to their non-receiving counterparts.

Identifying the causal effects of Head Start requires research designs that can plausibly rule out alternative explanations for differences in outcomes, including the underlying selection mechanisms — selection into provision and selection into participation. Early attempts to estimate program simply compared Head Start participants to non-participants (Westinghouse Learning Corporation, 1969), and then evolved to attempts to control for some (limited) observable characteristics of participants and non-participants (Barnow and Cain, 1977). Research has evolved to identifying plausibly exogenous variation in program exposure or participation, attempting to address unobservable determinants of exposure or participation, but those applications continue to be relatively rare.

Finally, measurement of Head Start effectiveness requires data that can facilitate such analyses, and the data needs can be quite intense when trying to understand the impact of the program across contexts and over the long run. Studying Head Start requires *data on the program*, either its existence at a particular time and in a particular place, intensity, level of funding, enrollment, or specific aspects of program delivery, and *data on the outcomes of interest*, which could be at the individual child, family, or community level. In the case of Head Start, long-term effects may be of particular interest as the program was conceived of as an anti-poverty program and an investment in children’s future productivity — and that long time horizon for such outcomes to be realized compounds these data challenges.

4.2 Drawing Comparisons

Given these challenges, there are a limited number of instances in which the three critical conditions are met: (1) the researcher can observe Head Start participation, availability, or exposure in data, (2) the researcher can credibly isolate the effect of Head Start with plausibly exogenous variation in participation, availability, or exposure, and (3) there are available measures of the relevant outcomes, including potentially over the long run. The modern literature deploys quasi-experimental methods, leveraging sibling comparisons, eligibility rules, or program rollout or expansion, as well as the nationally representative Head Start Impact Study (HSIS), which employed a multi-site randomized control trial, to explore Head Start effects, and variation therein. In particular, recent advancements in the quasi-experimental literature provide evidence from earlier iterations of the program (cohorts participating from the late 1960s through the 1990s), and reanalyses of the HSIS allow for exploration of short-run effects for more recent cohorts.

4.2.1 Family Fixed Effects

The earliest studies of Head Start that took seriously unobservable factors influencing selection into program exposure or participation employed comparisons within family, drawing on cohorts of Head Start-participating children from the late 1960s through the 1990s (Currie and Thomas, 1995; Deming, 2009; Garces et al., 2001). This literature relies on family fixed effects (FFE) to control for common aspects of the family environment and upbringing and leverage comparisons among siblings who differ in their exposure to or participation in Head Start. Earlier evidence points to positive medium- and long-term effects of program participation in the forms of increased educational achievement and attainment and reduced grade repetition and engagement in crime and risky behaviors (Deming, 2009), with some work documenting concentrated effects among white children (Currie and Thomas, 1995; Garces et al., 2001). More recent reanalyses, however, present a mixed picture of long-term results (Bauer and Schanzenbach, 2016; Pages et al., 2020).

Even if FFE designs convincingly address selection concerns in isolating the Head Start treatment effect — which may be more plausible for earlier cohorts of Head Start exposed families due to program introduction — recent evidence highlights differential selection *into identification* in these settings. This finding has implications for the *external* validity of such estimates, and their comparability to other estimates in the literature. Perhaps most importantly for the interpretation of FFE designs, Miller, Shenhav, and Grosz (2021) demonstrate that families that contribute to the identifying variation in these studies differ in important ways from the average Head Start family, and Head Start participating child. In particular, selection into identification of Head Start treatment effects, conditional on child/family characteristics, results in over-representation of larger families (Miller et al., 2021). These differences between the identifying sample and broader sample of interest, coupled with treatment effect heterogeneity, imply that the family fixed effects estimates overstate the average treatment effect (ATE), i.e., children from large families likely experience different (larger) effects than the average Head Start participant. In two longitudinal datasets, the authors find that reweighted-on-observables fixed effects estimates, intended to capture the Head Start participants' ATE, are substantially smaller (e.g., educational attainment effect for white Head Start participants in the Panel Study of Income Dynamics is 83% smaller when reweighted relative to the FFE estimate) (Miller et al., 2021).

Further complicating the interpretation of FFE estimates is recent evidence documenting sibling spillovers of ECE investments, in the context of the Perry Preschool Project (Bennhoff et al., 2024; García et al., 2021). To the extent that one child's exposure to programming alters parents' behavior, resources in the home, or sibling interactions with one another, the program could have indirect effects on other children in the family. Evidence from across

the Head Start literature, and discussed in more detail in subsequent sections, does point to important effects on parents which could manifest in better outcomes for siblings. In the case of the Perry Preschool Project, the authors suggest that these documented within-family spillovers contribute to an even more favorable for benefit-cost ratio for the intervention (Bennhoff et al., 2024; García et al., 2021).

4.2.2 Discontinuities in Program Exposure

Literature leveraging differences in Head Start exposure via discontinuities in receipt of grant-writing assistance and families' program eligibility also focuses on earlier time periods, analyzing cohorts from the late 1960s through the mid 1970s (Ludwig and Miller, 2007) and participants from the early 1980s through the early 2000s (Carneiro and Ginja, 2014) respectively. Ludwig and Miller (2007) use a discontinuity in grant-writing assistance, provided to the 300 poorest counties to bolster their capacity to apply for the federal funding in the early years of the program. The grant-writing assistance discontinuity produced large increases in Head Start funding levels, and the authors found that Head Start reduced child mortality due to "program-susceptible" causes and may have also increased educational attainment (Ludwig and Miller, 2007).

Looking at more recent cohorts and focusing on boys, Carneiro and Ginja (2014) use discontinuities in program eligibility rules to explore adolescent and young adult health and behavioral outcomes for Head Start participants in the 1980s through the 1990s. Discontinuities in program eligibility predict Head Start participation for male children only, so they focus on this sample. Head Start participation leads to reductions in health and behavioral problems in adolescents, and lower likelihood of criminal engagement and idleness for young adults (Carneiro and Ginja, 2014).

4.2.3 Geographic Variation in Program Access

The most recent evidence on Head Start effectiveness builds on an approach deployed in the literature on the impact of in-kind transfers, including nutrition programs for pregnant women and families (Almond et al., 2011; Hoynes and Schanzenbach, 2009; Hoynes et al., 2011, 2016) and the War on Poverty's Community Health Centers (Bailey and Goodman-Bacon, 2015). These studies capitalize on variation in Head Start access or exposure across geography and over time (Bailey et al., 2021; Barr and Gibbs, 2022; Johnson and Jackson, 2019; Kose, 2023; Thompson, 2018).

Head Start has the additional features, relative to other government transfer and safety net programs, of narrow age targeting and family income-based eligibility. Researchers have

capitalized on those program attributes alongside cross-county and over-time variation in program introduction, lending further credence to the application of this design in isolating the causal effect of Head Start exposure. Bailey and coauthors (2021), Johnson and Jackson (2015), and Thompson (2018) all use geographic variation in program exposure, particularly as Head Start initially rolled out across the country, to explore impact. They generally find effects on human capital and economic self-sufficiency in adulthood (Bailey et al., 2021), positive effects on educational attainment and earnings and reductions in poverty and incarceration that were enhanced when followed by increased K-12 spending (Johnson and Jackson, 2019), and improved health, educational attainment, and labor market outcomes through age 48 (Thompson, 2018).

Given these documented effects on the first generation, Barr and Gibbs (2022) investigate intergenerational transmission of Head Start effects using similar archival records of the early introduction of the program, and fortuitously timed survey data, finding substantial spillovers to the children of those exposed to Head Start in its rollout in the late 1960s, including reduced criminal engagement and teen parenthood and increased educational attainment — effects corresponding to \$35,000 improvement in expected earnings and large enough to effectively close the 2nd-generation gap in long-term outcomes between individuals with more and less advantaged grandmothers (Barr and Gibbs, 2022). The quasi-experimental literature is well positioned to explore long-term effects given the nature of the designs. The evidence suggests that those exposed to Head Start as children engage in different parenting behaviors and approaches later in life, which could serve as one relevant pathway to intergenerational spillovers (Bauer and Schanzenbach, 2016; Barr and Gibbs, 2022). Indeed, Barr and Gibbs (2022) document improved home environments in early life for the children of Head Start-exposed mothers.

Using more recent variation via funding expansions in the 1990s, Kose (2023) uses federal funding changes across counties in Texas in that time period to explore impact on children’s subsequent academic achievement, and documents that Head Start funding expansions improved test scores, particularly for Hispanic students from low-income households, who experienced increased language proficiency and lower likelihood of identification for special education services (Kose, 2023).

4.2.4 Interpreting Quasi-Experimental Impact Estimates

Taken together, this literature coalesces around the central empirical finding of pronounced positive effects for those exposed to Head Start from the late 1960s through the 1990s, particularly in the form of increased educational attainment. Figure 6 presents the estimates of program impact on likelihood of high school graduation from across the quasi-experimental

literature, shown as estimates of the implied treatment on the treated (TOT) effect, by timing of program cohort.² TOT estimates (as presented in Figure 6) inflate intent-to-treat estimates by the respective authors’ preferred measure of program takeup or participation — with the exception of FFE estimates that already have a TOT interpretation — in accordance with the nature of their research design. Importantly, authors typically encourage caution when interpreting scaled estimates because of difficulty in measuring the first stage of Head Start participation due to recall bias in self-reported participation and misclassification error (Barr and Gibbs, 2022). First-stage estimates range from 0.24 (Thompson, 2018), to 0.53 (Barr and Gibbs, 2022), to 0.74 for male respondents in the eligible age group (Carneiro and Ginja, 2014). Measurement of the first stage, which looks different across applications, has important implications for interpreting the magnitudes as well as for comparability of estimates across settings. Notably, the confidence intervals of the underlying intent-to-treat estimates of program impact on high school completion overlap, providing suggestive evidence that the estimates may not differ statistically.

While previous work (Duncan and Magnuson, 2013) has suggested declining Head Start impact over time when looking at test score measures immediately following the program, there is no clear pattern of time trends in effects when looking at the longer-term outcome of high school completion. Importantly, work in this vein that looks at both test score effects and longer-term effects affirms that long-run benefits can be realized even in the absence of intervening test score improvements (Barr and Gibbs, 2022; Deming, 2009; Ludwig and Miller, 2007). This pattern of either no short-term cognitive gains with emerging longer-run effects or evaporating test score effects is also realized in studies of the Perry Preschool Project, deployed with disadvantaged preschool-aged children in Ypsilanti, Michigan in a similar timeframe as the introduction of Head Start, and Sure Start, modeled after aspects of Head Start and introduced in 1999 for children under age five in England (Carneiro et al., 2025; Schweinhart et al., 2005). Researchers point to the forms of skill development that ECE programs support and the interaction of those skills with features of subsequent schooling environments as potential explanations for skill “fadeout” (Bailey et al., 2017; Heckman et al., 2013; List and Uchida, 2024).

To the extent there are spillovers from program exposure within a cohort or community, estimates of direct effects on participants should be interpreted with some caution. In fact, comparisons of participants and non-participants in the same places or schools may not be the correct way to think about impact estimates of Head Start, which has features of a place-based policy and is intended to generate family- and community-wide effects. For example,

²While high school completion is the most common long-run outcome across studies, years of schooling and college-going show similar patterns when included in the set of investigated outcomes.

one type of spillover relevant for fully measuring Head Start effectiveness is the impact of Head Start-exposed students on their classroom peers in subsequent school environments. Because of non-random sorting to schools and classrooms, these "peer effects" or classroom compositional effects are difficult to isolate. In a study that capitalizes on the within-school formation of peer groups for identification, the authors document spillovers of peer preschool (including but not limited to Head Start) participation to their classmates' reading and math performance, but little evidence of effects on non-cognitive skills (Neidell and Waldfogel, 2010).

4.3 Reanalyses of the Head Start Impact Study

The national HSIS employed a multi-site randomized control trial in which Head Start applicants to oversubscribed centers in 2002-2003 were randomly assigned to the offer of a slot or not. This approach produced a nationally representative sample of 4,667 newly entering three- and four-year old children who were followed from their Head Start participation through the primary grades of elementary school.

The much-discussed main findings from the HSIS showed initial skill advantages, at the end of the Head Start year and kindergarten entry, that faded by first and third grades (Puma et al., 2005, 2010, 2012). These lackluster results for Head Start *on average* were widely presented as evidence that the program does not work. There has been no further follow-up of the HSIS participants, so these estimates are limited to relatively short-term outcomes, immediately following the program year and in the primary grades of schooling. As discussed, this pattern of "fadeout" in test score effects could relate to the type of skills that are and are not impacted by Head Start participation and could also be shaped by the subsequent schooling environments into which children transition. Re-analyses of HSIS data suggest a more nuanced picture that relates back to the importance of understanding the counterfactual to Head Start in different contexts (Montialoux, 2016).

While the quasi-experimental literature has attempted to unpack treatment-effect heterogeneity, these pursuits are generally underpowered to detect variation in treatment effectiveness across subgroups. Researchers have used the HSIS to demonstrate that average affects mask substantial heterogeneity in the effectiveness of Head Start across contexts and for different children and families. In particular, authors have documented considerable variation in impact by Head Start center (Morris et al., 2018; Walters, 2015). Walters (2015) characterizes the variation across centers, nearly 20 percent of a standard deviation in test scores, as larger than standard estimates in the literature of the variation in teacher- or school-level effects. Morris and coauthors (2018) document similarly sized variation across

centers using different methods. Centers with full-day programming and those providing home visiting services generate larger effects on test scores though those estimates are not causal (Walters, 2015).

Importantly, many children who did not get Head Start participated in other forms of formal, nonparental ECE. Effects are more pronounced among children who would otherwise be in parental or relative care (Bloom and Weiland, 2015; Feller et al., 2016; Kline and Walters, 2016), and Head Start centers that draw young students from other forms of preschool produce very little or no improvement on measures of skill development (Feller et al., 2016; Walters, 2015). There is also suggestive evidence that effects are larger and more persistent for the three year old cohort in the HSIS, perhaps pointing to the less robust outside options for this age group (Gibbs et al., 2011; Kline and Walters, 2016). Additionally, across methodological approaches, evidence shows that children with low skills at program entry experienced the greatest benefit and effects for Spanish speakers persisted (Bitler et al., 2014; Bloom and Weiland, 2015; Feller et al., 2016).

4.4 Defining Takeup

In the existing literature, authors have estimated program takeup from historical records and program eligibility rules, but participation rates among a study population vary over time and geography, resulting in first stages that are quite different across studies. These magnitudes have implications for the interpretation of treatment on the treated estimates.

In Figure 6, estimates of Head Start impact are scaled by the authors' preferred first stage. Often, particularly when assessing treatment effects for the historical Head Start program, these rest on a number of assumptions and are not well established in existing data. Over time, as documented in this paper, alternative options arose which alter the interpretation of the Local Average Treatment Effect (LATE) estimated in many studies. As Kline and Walters (2016) show, with multiple counterfactual conditions, estimating standard just-identified instrumental variables (IV) recovers a weighted average of sub-LATEs which are treatment effects relative to specific counterfactual conditions. They consider specifically alternative center-based preschool and no preschool (or home care), and partition Head Start applicants into always takers, compliers who switch from other center-based care, compliers who switch from home care, never takers in other center-based care, and never takers in home care. As such, the group of children who enroll in Head Start is comprised of the first three categories.

In the case of Head Start, and particularly present-day Head Start, the IV coefficient takes the form of:

$$\beta_{hs}^{IV} = \omega LATE_{0 \rightarrow hs} + (1 - \omega) LATE_{ec \rightarrow hs}$$

where $LATE_{0 \rightarrow hs}$ is the LATE for Head Start participants who would have alternatively been in no formal early childhood setting, likely parental or relative care, and $LATE_{ec \rightarrow hs}$ is the LATE for children who would have otherwise been in an alternative early childhood program.³ Consideration of the component parts of the LATE estimate becomes especially relevant when assessing evidence on the modern-day implementation of the Head Start program, against a backdrop of multiple care settings and providers. Depending on the nature of the policy regime, the social planner may be more or less interested in particular sub-LATEs. For example, a regime focused on expanding slots where preschool participation is otherwise low would be interested in the comparison to home care while a regime focused on improving the quality or intensity of Head Start slots may want the deployment of those resources to outperform available alternatives.

Kline and Walters (2016) document substantial variation within the HSIS in the complier share across study years and age cohorts. That is, their estimates of the share of Head Start compliers shifted from preschool alternatives ranges from 0.28 to as high as 0.72. They show that the alternative setting for compliers moving from other center-based care is quite similar — publicly funded with comparable programming. The most common outside option is a state-funded preschool program. Considering the alternative ECE settings has important implications for interpretation of effects and benefit-cost analyses. Berne (2024) and Zepa (2025) discuss these considerations in scaling ITT estimates of pre-K expansions, highlighting critical assumptions about both substitution across programs, but also in determining who is affected by expansions that may alter the mix of children in particular settings and could affect resources and quality outside the focal setting.

4.5 Documenting Broader Effects

Evidence from both strands of the literature points to important effects beyond the direct effects on Head Start participants, suggesting other important pathways to longer-run and broader Head Start impact through program effects on families. In particular, quasi-experimental evidence suggests that certain subgroups of mothers worked more and sought more schooling as a result of their child’s Head Start access (Pihl, 2022; Wikle and Wilson, 2023). In the more recent HSIS, authors similarly document effects on some mothers’ employment and educational attainment (Sabol and Chase-Lansdale, 2015; Wikle and Wilson,

³The weight is calculated as follows: $\omega = \left(-\frac{E[D=0|Z_{hs}(1)] - E[D=0|Z_{hs}(0)]}{E[D=hs|Z_{hs}(1)] - E[D=hs|Z_{hs}(0)]} \right)$.

2023). These effects are concentrated among single mothers with some suggestive evidence of more pronounced effects for Black mothers.

Also in the context of the HSIS, Gelber and Isen (2013) estimate program effects on parenting practices, including time spent with children engaged in reading and on math activities and time non-resident fathers spent with their children. These effects persisted beyond the program year (Gelber and Isen, 2013). Effects on parents point to potential pathways through which Head Start could affect children’s own longer-run outcomes and through which the program could generate intragenerational spillovers to siblings and other peers.

5 What We Know and Need to Know

Measuring the longer-term and broader effects of Head Start can inform how we think about public investments in early childhood, and the return on those investments. Because of the scale and maturity of the Head Start program, researchers have built a robust evidence base on its effects. The Head Start literature speaks specifically to the impact and effectiveness of a targeted ECE program, and can help us move beyond the debate over, “Does Head Start work?” to a deeper understanding of when such investments work, relative to what, in what contexts, with what features, and for which children. Recent rigorous contributions to the literature can facilitate portability of retrospective findings to consideration of the present-day program and its current participants.

While Head Start has been the subject of impact research since its early years, the last decade has moved our collective understanding of the program’s effectiveness substantially forward, providing insights for the broader literature on early childhood investments as well as for policy. We have evidence that Head Start *can alter* children’s trajectories in meaningful ways, with implications for their own wellbeing, their families, and society, but more recent evidence is muddied by a lack of long-term followup—which would facilitate linkage of short-run patterns with long-run outcomes of interest—and thus more complex to interpret. It highlights, in particular, the importance of understanding relevant comparisons.

In particular, advances in the quasi-experimental literature, and in approaches to analyzing experimental data, have allowed for greater exploration of how Head Start effects vary across settings, over time, and for children with particular characteristics — dimensions along which we would expect, *ex ante*, program effectiveness to vary. Questions remain, however, about whether the long-term effects realized by Head Start cohorts in the late 1960s through the 1990s persist for more recent participants in the program.

This synthesis of existing work focuses on (1) children served in Head Start, (2) program–

no program contrast for these children over time, and (3) the comparisons drawn in the literature to extract key insights and lessons learned for the current Head Start program and a forward-looking research agenda. Several key themes emerge:

- The literature suggests Head Start can be effective at improving important life outcomes, and is likely most effective for the most disadvantaged at program entry and those for whom the outside option is parental or relative care.
- Patterns of short- and medium-term test scores effects are not particularly informative in projecting long-run outcomes, pointing to the likely important role of (less commonly and consistently measured) social, emotional, and behavioral skills.
- Though more limited, evidence documents important effects on the parents of Head Start participants as well as the potential for spillovers to siblings, peers, and the children of participants which suggests our typical approaches to cost-benefit analysis of the program may understate benefits.
- Our focus on rigorous causal designs comparing Head Start-participating or Head Start-exposed children to those who are not is limited in providing evidence-based lessons for policymaking in a fragmented, mixed delivery system of which Head Start is one part.

5.1 Social Welfare Implications

When assessing the broader welfare implications of Head Start investments, most analyses focus on the later-life improvements in outcomes with ramifications for society, including educational attainment, criminal justice system involvement, and health. Research also documents cost-benefit assessments of other ECE programs, generating a range of estimates of three to seven dollars in savings for every dollar spent providing programs such as the Perry Preschool Project, the Carolina Abecedarian Project (ABC), and the Carolina Approach to Responsive Education (CARE) (García et al., 2020; Heckman et al., 2010). These are likely upper bounds on the returns to ECE investments as the programs were intensive, comprehensive programs deployed with disadvantaged populations when few alternatives existed. As presented in the theoretical foundation for investments in early childhood, ECE programs have a long time horizon over which to realize returns. Such investments may also exhibit complementarities with subsequent investments. And, ECE investments may produce positive externalities in excess of their privately realized benefits. As a result, some ECE programs can generate large or even infinite MVPF estimates, as is the case with Head Start (Hendren and Sprung-Keyser, 2020).

Existing benefit-cost estimates of Head Start suggest that the program likely pays for itself, and even modest effects on standardized test scores in the short-run can be consistent with positive net benefits because of the multiple pathways through which Head Start affects later-life wellbeing (Gibbs et al., 2011; Kline and Walters, 2016; Ludwig and Phillips, 2007, 2008). Importantly, in the presence of general equilibrium adjustments, substitution across publicly-provided or funded care types, and peer effects, participant outcomes are an incomplete capture of program effectiveness. Moreover, measures of the benefit-cost ratio or the MVPF focus on one of the dual roles that Head Start plays, that of facilitating children’s skill development. They do not account for effects on parental employment. Finally, they do not incorporate within-family spillovers to siblings or the children of Head Start participants (i.e., intra- and intergenerational spillovers) which estimates imply could be quite substantial (Barr and Gibbs, 2022; Bennhoff et al., 2024).

5.2 Avenues for Future Research

Over time, the Head Start program has evolved alongside changes in the quantity and quality of developmental inputs that young children experience in their early years in the U.S. In addition, there is considerable variation in Head Start program effectiveness at the center level, likely driven by both program quality differences and differences in counterfactual experiences. These key facts provide the foundation for a research, and evidence-based policy, agenda going forward.

First, unpacking variation in program effectiveness can have important implications for policy. Leveraging what we know, and continuing to expand what we know, about treatment–effect heterogeneity could support better program targeting. In particular, policymakers and practitioners can use knowledge about gaps in access to resource-rich environments in early childhood to (1) better understand where and for whom expected outcomes of program introduction or expansion would be greatest, and (2) direct resources toward identifying the children and families for whom a shift to Head Start may constitute the largest contrast relative to what they would otherwise do (e.g., parental or relative care). Kline and Walters (2016) document a reverse Roy model in that the children least likely to enroll in Head Start are those most likely to benefit. Identifying the best policy levers to address those disparities and fill in the gaps in access and participation could generate more bang for the marginal Head Start buck.

Relatedly, researchers can support deeper understanding of what constitutes quality in Head Start settings by identifying the “key ingredients” that generate improved child outcomes. Our current knowledge base on what constitutes high-quality programming in Head

Start centers (i.e., measures of quality that lead to improved child outcomes) is quite limited which hampers our ability to draw direct lessons for how programs should run and on what dimensions they should improve. There is scant rigorous research on the impact of investing in particular program components, structural features of the ECE setting, curriculum and instruction, teacher and caregiver characteristics, and training and professional development. While relevant to unpacking variation in Head Start centers' effectiveness, evidence on quality in ECE could also help maximize benefits of public dollars devoted to early childhood investments more broadly.

The existing evidence has been informative about retrospective iterations of the Head Start program, but the findings' applicability to current iterations of the Head Start program, or our deployment of public early childhood resources more broadly, is hampered by important limitations. In particular, our lack of understanding about the pathways through which long-term benefits are realized makes it difficult to assess the effectiveness of programs before those long-run outcomes manifest.

Unfortunately, researchers and policymakers will always be chasing the long-term effects of modern-day programs. Beyond retrospective evaluations and benefit-cost analyses of already-deployed programs, it would be immensely useful to those interested in evidence-based policymaking to adequately and concretely measure the intermediary outcomes that correspond to improved long-run outcomes. Given what we know about patterns of test score effects, these efforts should focus on a broader range of cognitive and social/emotional/behavioral skills. This pursuit would also benefit from measuring long-term effects for more recent cohorts of Head Start participants via the HSIS, a missing data point in understanding the evolution of Head Start effectiveness over time.⁴ In the context of Head Start as well as other ECE investments, greater investigation of the channels through which long-term effects are generated would support more immediate appraisal of program effectiveness.

Finally, our causal research designs rely on comparisons of Head Start participants to non-participants, making it difficult to draw conclusions about programs at scale or in general equilibrium. The non-participant experience has changed significantly over time, which importantly alters the nature and interpretation of these program–no program comparisons. It may be reasonable to conclude that our current frontier tools and approaches for measur-

⁴Using study records to follow up on HSIS participants, who are now in their 20s, would likely be both feasible and cost-effective. Researchers have developed agreements and deployed new methods of linking individuals to post-secondary enrollment outcomes in the National Student Clearinghouse and earnings, employment, and educational attainment outcomes in Internal Revenue Service tax records in influential long-term studies of both class-size reduction in Tennessee Project STAR and Boston public preschool (Chetty et al., 2011; Gray-Lobe et al., 2022).

ing the causal effects of policies and programs are not up to the task of understanding the full effects of Head Start, or any ECE investment, in current contexts. Given the fragmented system of ECE provided in homes, centers, schools, and community-based settings, and funded by both private and public sources, designs that focus on isolating participant-level effects of one type of program will be hampered by the general equilibrium considerations and spillovers mentioned herein. With higher levels of ECE participation among young children and frequent changes to the landscape of ECE offerings, perhaps — rather than understanding the impact of Head Start relative to all of those outside options, or as compared to any one of them — a new knowledge-building research agenda can focus on illuminating the optimal deployment of public resources across the patchwork of ECE funding sources, providers, and settings.

References

- Almond, Douglas, Hilary W. Hoynes, and Diane Whitmore Schanzenbach.** 2011. “Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes.” *Review of Economics and Statistics* 93 (2): 387–403.
- Atteberry, Allison, Daphna Bassok, and Vivian C. Wong.** 2019. “The Effects of Full-day Prekindergarten: Experimental Evidence of Impacts on Children’s School Readiness.” *Educational Evaluation and Policy Analysis* 41 (4): 537–562.
- Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu.** 2017. “Persistence and Fadeout in the Impacts of Child and Adolescent Interventions.” *Journal of Research on Educational Effectiveness* 10 7–39.
- Bailey, Martha J., and Andrew Goodman-Bacon.** 2015. “The War on Poverty’s Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans.” *American Economic Review* 105 (3): 1067–1104. 10.1257/aer.20120070.
- Bailey, Martha J., Shuqiao Sun, and Brenden Timpe.** 2021. “Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency.” *American Economic Review* 111 (12): 3963–4001. 10.1257/aer.20181801.
- Barnow, Burt S., and Glen G. Cain.** 1977. “A Reanalysis of the Effect of Head Start on Cognitive Development: Methodology and Empirical Findings.” *Journal of Human Resources* 12 (2): 177–197. 10.2307/145384.
- Barr, Andrew, and Chloe R. Gibbs.** 2022. “Breaking the Cycle? Intergenerational Effects of an Antipoverty Program in Early Childhood.” *Journal of Political Economy* 130 (12): 3253–3285. 10.1086/720764.
- Bassok, Daphna, Jenna E. Finch, RaeHyuck Lee, Sean F. Reardon, and Jane Waldfogel.** 2016. “Socioeconomic Gaps in Early Childhood Experiences: 1998 to 2010.” *AERA Open* 2 (3): 10.1177/2332858416653924.
- Bauer, Lauren, and Diane Whitmore Schanzenbach.** 2016. “The Long-Term Impact of the Head Start Program.” technical report, The Hamilton Project, Brookings Institution.
- Bennhoff, Frederik H., Jorge Luis García, and Duncan Ermini Leaf.** 2024. “The Dynastic Benefits of Early-Childhood Education: Participant Benefits and Family Spillovers.” *Journal of Human Capital* 18 44–73.
- Bitler, Marianne P., Hilary W. Hoynes, and Thurston Domina.** 2014. “Experimental Evidence on Distributional Effects of Head Start.” Technical Report20434, National Bureau of Economic Research, Cambridge, MA. 10.3386/w20434.
- Bloom, Howard S., and Christina Weiland.** 2015. “Quantifying Variation in Head Start Effects on Young Children’s Cognitive and Socio-Emotional Skills Using Data from the National Head Start Impact Study.” *SSRN Electronic Journal*. 10.2139/ssrn.2594430.
- Burchinal, Margaret, Anamarie Whitaker, Jade Jenkins, Drew Bailey, Tyler Watts, Greg Duncan, and Emma Hart.** 2024. “Unsettled science on longer-run effects of early education.” *Science* 384 506–508.
- Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan.** 2014. “Early Childhood Investments Substantially Boost Adult Health.” *Science* 343 1478–1485.
- Carneiro, Pedro, Sarah Cattan, Gabriella Conti, Claire Crawford, Christine Farquharson, and Nick Ridpath.** 2025. “The short- and medium-term effects of Sure Start on children’s outcomes.” technical report, Institute for Fiscal Studies, London, UK.

- Carneiro, Pedro, and Rita 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. 10.1257/po1.6.4.135.
- Cascio, Elizabeth U.** 2023. “Does Universal Preschool Hit the Target? Program Access and Preschool Impacts.” *Journal of Human Resources* 58 (1): 1–42.
- Cascio, Elizabeth U., and Diane Whitmore Schanzenbach.** 2014. “Proposal 1: Expanding Preschool Access for Disadvantaged Children.” policy Proposal, Brookings Institution.
- Chetty, Raj, John Friedman, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan.** 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *Quarterly Journal of Economics* 126 (4): 1593–1660.
- Cui, Jiashan, and Luke Natzke.** 2021. “Early Childhood Program Participation: 2019.” Technical Report NCES 2020-075REV, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC.
- Currie, Janet.** 2001. “Early Childhood Education Programs.” *Journal of Economic Perspectives* 15 (2): 213–238. 10.1257/jep.15.2.213.
- Currie, Janet, and Matthew Neidell.** 2007. “Getting inside the “Black Box” of Head Start quality: What matters and what doesn’t.” *Economics of Education Review* 26 83–99.
- Currie, Janet, and Duncan Thomas.** 1995. “Does Head Start Make a Difference?” *American Economic Review* 85 (3): 341–364.
- Currie, Janet, and Duncan Thomas.** 1999. “Early test scores, socioeconomic status and future outcomes.”
- Deming, David.** 2009. “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start.” *American Economic Journal: Applied Economics* 1 (3): 111–134.
- Duncan, Greg J., Chantelle J. Dowsett, Amy Claessens et al.** 2007. “School Readiness and Later Achievement.” *Developmental Psychology* 43 (6): 1428–1446. 10.1037/0012-1649.43.6.1428.
- Duncan, Greg J., and Katherine Magnuson.** 2013. “Investing in Preschool Programs.” *Journal of Economic Perspectives* 27 (2): 109–132. 10.1257/jep.27.2.109.
- Dynarski, Susan, Joshua Hyman, and Diane Whitmore Schanzenbach.** 2013. “Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion.” *Journal of Policy Analysis and Management* 32 (4): 692–717.
- Elango, Sneha, Jorge Luis García, James J. Heckman, and Andrés Hojman.** 2015. “Early Childhood Education.” In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, 235–297, University of Chicago Press.
- Feller, Avi, Todd Grindal, Luke Miratrix, and Lindsay C. Page.** 2016. “Compared to what? Variation in the impacts of early childhood education by alternative care type.” *The Annals of Applied Statistics* 10 (3): 1245–1285. 10.1214/16-AOAS910.
- Flood, Sarah, Joel McMurry, Aaron Sojourner, and Matthew Wiswall.** 2022. “Inequality in early care experienced by US children.” *Journal of Economic Perspectives* 36 (2): 199–222.
- Friedman-Krauss, Allison H., W. Steven Barnett, Katherine S. Hodges, Karin A. Garver, Jennifer K. Duer, G. G. Weisenfeld, and Jessica Siegel.** 2024. “The State of Preschool Yearbook 2023.” technical report, National Institute for Early Education Research, New Brunswick, NJ.
- Garces, Eliana, Duncan Thomas, and Janet Currie.** 2001. “Longer-Term Effects of Head Start.” *American Economic Review* 92 (4): 999–1012. 10.1257/00028280260344560.
- García, Jorge Luis, Frederik H. Bennhoff, Duncan Ermini Leaf, and James J. Heckman.**

2021. “The Dynastic Benefits of Early Childhood Education.” Working Paper29004, National Bureau of Economic Research. 10.3386/w29004.
- García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and María José Prados.** 2020. “Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program.” *Journal of Political Economy* 128 2502–2541.
- Gelber, Alexander, and Adam Isen.** 2013. “Children’s schooling and parents’ behavior: Evidence from the Head Start Impact Study.” *Journal of Public Economics* 101 25–38. 10.1016/j.jpubeco.2013.02.005.
- Gibbs, Chloe, Jens Ludwig, and Douglas L. Miller.** 2011. “Does Head Start Do Any Lasting Good?”, Technical Reportw17452, National Bureau of Economic Research, Cambridge, MA. 10.3386/w17452.
- Gibbs, Chloe, Jens Ludwig, Douglas L. Miller, and Na’ama Shenhav.** 2016. “Short-run Fade-out in Head Start and Implications for Long-run Effectiveness.” technical report, University of California, Davis, Center for Poverty Research, Davis, CA.
- Gibbs, Chloe R.** 2023. “Experimental Evidence on Early Childhood Investment: The Impact of a Longer School Day.” technical report, University of Notre Dame.
- Gray-Lobe, Guthrie, Parag A Pathak, and Christopher R Walters.** 2022. “The Long-Term Effects of Universal Preschool in Boston.” *Quarterly Journal of Economics* 138 (1): 363–411.
- Heckman, James J.** 2000. “Policies to foster human capital.” *Research in Economics* 54 (1): 3–56.
- Heckman, James J., and Dimitriy V. Masterov.** 2007. “The productivity argument for investing in young children.” *NBER Working Paper* Paper No. 13016.
- Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz.** 2010. “The rate of return to the HighScope Perry Preschool Program.” *Journal of Public Economics* 94 114–128.
- Heckman, James, Rodrigo Pinto, and Peter Savelyev.** 2013. “Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review* 103 (6): 2052–86.
- Hendren, Nathaniel, and Ben Sprung-Keyser.** 2020. “A Unified Welfare Analysis of Government Policies.” *The Quarterly Journal of Economics* 135 (3): 1209–1318. 10.1093/qje/qjaa006.
- Hoynes, Hilary, Marianne Page, and Ann Huff Stevens.** 2011. “Can targeted transfers improve birth outcomes?: Evidence from the introduction of the WIC program.” *Journal of Public Economics* 95 (7): 813–827. <https://doi.org/10.1016/j.jpubeco.2010.12.006>.
- Hoynes, Hilary, and Diane Schanzenbach.** 2009. “Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program.” *American Economic Journal: Applied Economics* 1 (4): 109–139.
- Hoynes, Hilary, Diane Schanzenbach, and Douglas Almond.** 2016. “Long-Run Impacts of Childhood Access to the Safety Net.” *American Economic Review* 106 (4): 905–954.
- Johnson, Rucker C., and C. Kirabo Jackson.** 2019. “Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending.” *American Economic Journal: Economic Policy* 11 (4): 310–349. 10.1257/pol.20180510.
- Kalil, Ariel, Kathleen M. Ziol-Guest, Rebecca M. Ryan, and Anna J. Markowitz.** 2016. “Changes in Income-Based Gaps in Parent Activities With Young Children From 1988 to 2012.” *AERA Open* 2 (3): 10.1177/2332858416653732.
- Kline, Patrick, and Christopher R. Walters.** 2016. “Evaluating Public Programs with Close Substitutes: The Case of Head Start.” *The Quarterly Journal of Economics* 131 (4): 1795–1848.

10.1093/qje/qjw027.

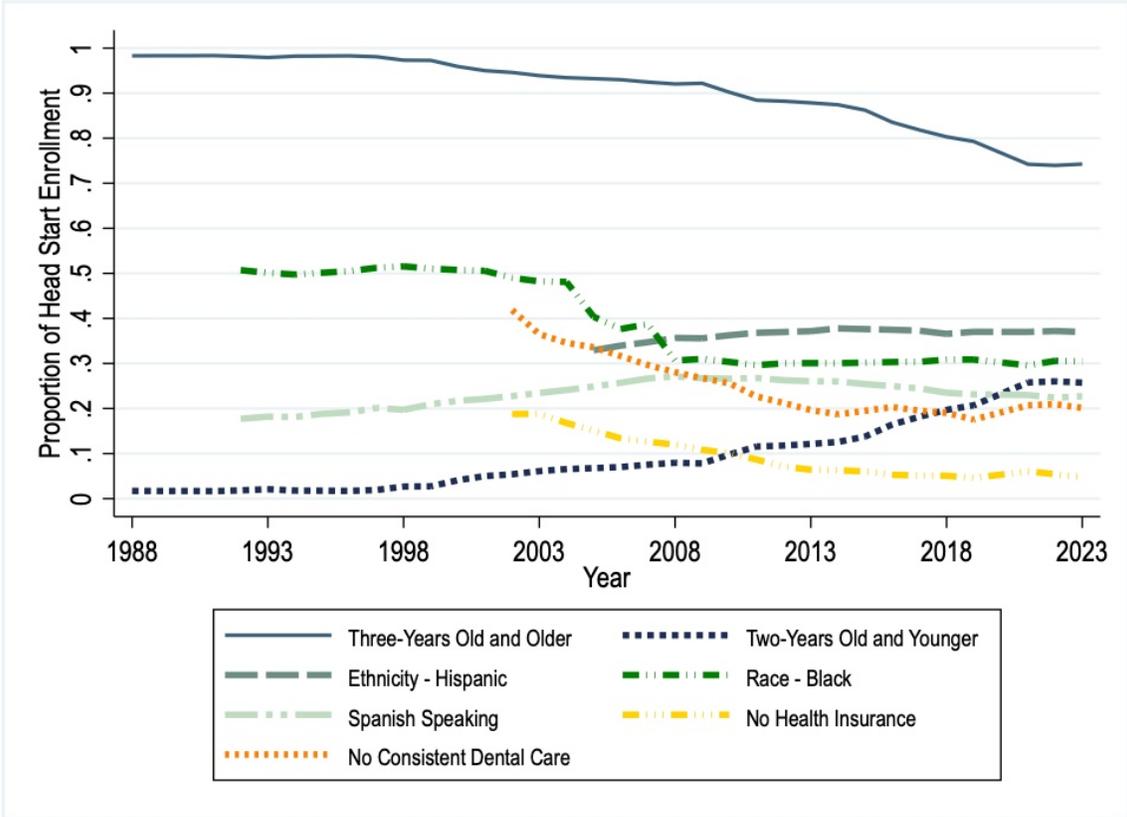
- Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff.** 2006. “Economic, neurobiological, and behavioral perspectives on building America’s future workforce.” *PNAS* 103 (27): 10155–10162.
- Kose, Esra.** 2023. “Public Investments in Early Childhood Education and Academic Performance: Evidence from Head Start in Texas.” *Journal of Human Resources* 58 (6): 2042–2069. 10.3368/jhr.0419-10147R2.
- Lee, Valerie E., and David T. Burkam.** 2002. *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*. Washington, DC: Economic Policy Institute.
- List, John A., and Haruka Uchida.** 2024. “Here Today, Gone Tomorrow? Toward an Understanding of Fade-out in Early Childhood Education Programs.” Technical Report33027, National Bureau of Economic Research. 10.3386/w33027.
- Ludwig, Jens, and Douglas L. Miller.** 2007. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” *The Quarterly Journal of Economics* 122 (1): 159–208. 10.1162/qjec.122.1.159.
- Ludwig, Jens, and Deborah Phillips.** 2007. “The Benefits and Costs of Head Start.” Technical Reportw12973, National Bureau of Economic Research, Cambridge, MA. 10.3386/w12973.
- Ludwig, Jens, and Deborah 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. 10.1196/annals.1425.005.
- Miller, Douglas L., Na’ama Shenhav, and Michel Grosz.** 2021. “Selection into Identification in Fixed Effects Models, with Application to Head Start.” *Journal of Human Resources* 56 (1): 1–31.
- Montialoux, Claire.** 2016. “Revisiting the impact of Head Start.” technical report, Institute for Research on Labor and Employment.
- Morris, Pamela A., Maia Connors, Allison Friedman-Krauss et al.** 2018. “New Findings on Impact Variation From the Head Start Impact Study: Informing the Scale-Up of Early Childhood Programs.” *AERA Open* 4 (2): 10.1177/2332858418769287.
- Neidell, Matthew, and Jane Waldfogel.** 2010. “Cognitive and Noncognitive Peer Effects in Early Education.” *The Review of Economics and Statistics* 92 562–576.
- Pages, Remy, Dylan J. Lukes, Drew H. Bailey, and Greg J. Duncan.** 2020. “Elusive Longer-Run Impacts of Head Start: Replications Within and Across Cohorts.” *Educational Evaluation and Policy Analysis* 42 (4): 471–492, Publisher: American Educational Research Association.
- Pianta, Robert C.** 1997. “Adult–Child Relationship Processes and Early Schooling.” *Early Education and Development* 8 11–26.
- Pihl, Ariel Marek.** 2022. “Head Start and mothers’ work: Free child care or something more?” *Labour Economics* 75.
- Puma, Michael, Stephen Bell, Ronna Cook, and Camilla Heid.** 2010. “Head Start Impact Study: Final Report.” technical report, U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC.
- Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer.** 2012. “Third Grade Follow-up to the Head Start Impact Study Final Report.” Technical ReportOPRE 2012-45, U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and

Evaluation, Washington, DC.

- Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, and Michael Lopez.** 2005. “Head Start Impact Study: First Year Findings.” technical report, U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC.
- Sabol, Terri J., and P. Lindsay Chase-Lansdale.** 2015. “The Influence of Low-Income Children’s Participation in Head Start on Their Parents’ Education and Employment.” *Journal of Policy Analysis and Management* 34 136–161.
- Sabol, Terri J., and Robert C. Pianta.** 2012. “Recent Trends in Research on Teacher–Child Relationships.” *Attachment Human Development* 14 213–231.
- Schweinhart, Lawrence J., Jeanne Mortie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield, and Milagros Mores.** 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, MI: High/Scope Educational Research Foundation.
- Shonkoff, Jack P., and Deborah A. Phillips.** 2000. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Research Council (US) and Institute of Medicine (US) Committee on Integrating the Science of Early Childhood Development, National Academies Press (US).
- Thompson, Owen.** 2018. “Head Start’s Long-Run Impact: Evidence from the Program’s Introduction.” *Journal of Human Resources* 53 (4): 1100–1139. 10.3368/jhr.53.4.0216.7735R1.
- Waldfogel, Jane, and Elizabeth Washbrook.** 2011. “Early Years Policy.” *Child Development Research*. 10.1155/2011/343016.
- Walters, Christopher R.** 2015. “Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start.” *American Economic Journal: Applied Economics* 7 (4): 76–102. 10.1257/app.20140184.
- Westinghouse Learning Corporation.** 1969. “The Impact of Head Start: An Evaluation of the Effects of Head Start on Children’s Cognitive and Affective Development.” technical report, Ohio University, New York, NY.
- Wikle, Jocelyn, and Riley Wilson.** 2023. “Access to Head Start and Maternal Labor Supply: Experimental and Quasi-Experimental Evidence.” *Journal of Labor Economics* 41 (4): 1081–1127. 10.1086/720980.
- Zerpa, Mariana.** 2025. “Short and Medium-Run Impacts of Preschool Education: Evidence from State Pre-K Programs.” working paper, University of Chile.

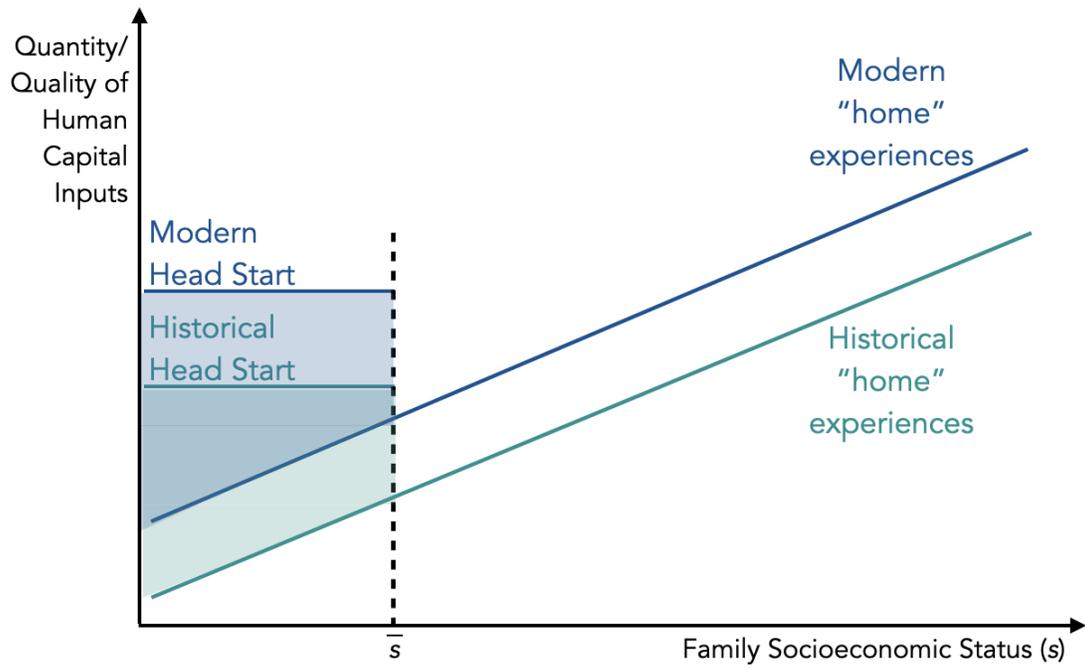
Figures

Figure 1: Head Start Participants, 1988-2023



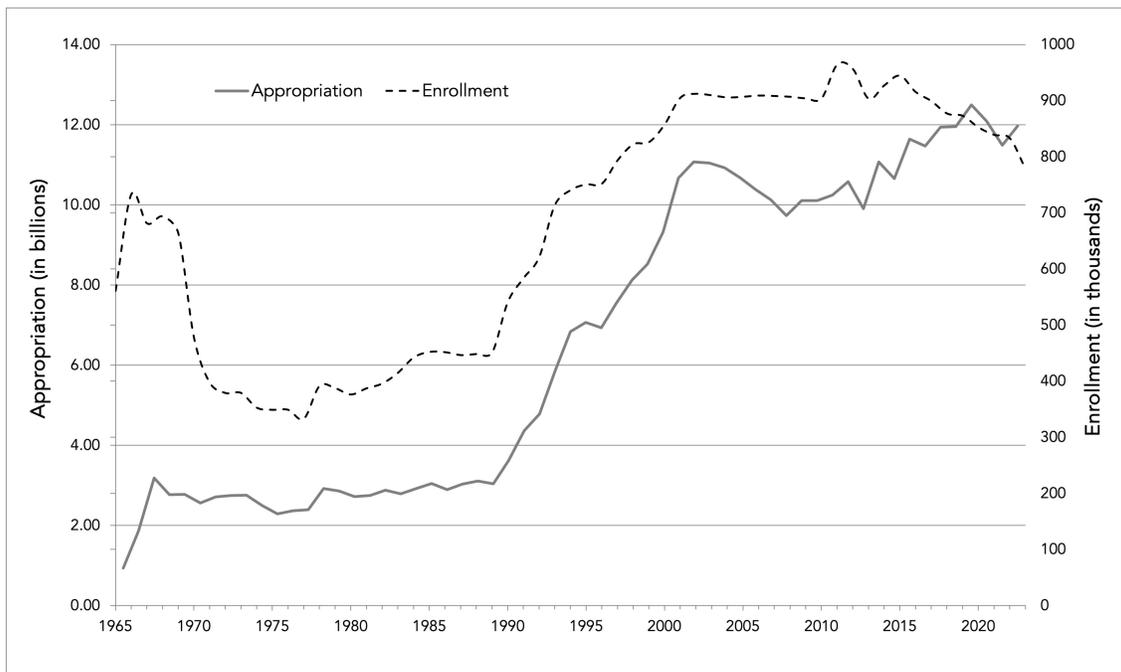
Source: Author's calculations from Head Start Program Information Reports, 1988-2023

Figure 2: Understanding Program–No Program Contrast, Then and Now



Source: Author's adaptation of conceptual model in Cascio and Schanzenbach (2014).

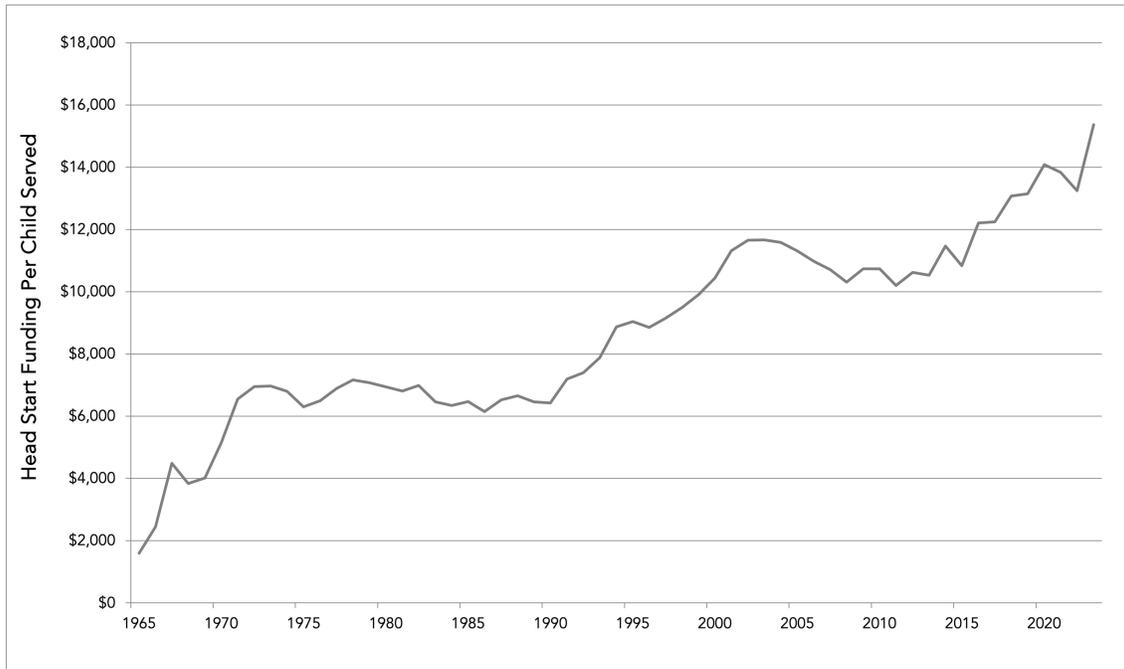
Figure 3: Head Start Enrollment and Funding, 1965-2023



Source: U.S. Department of Health and Human Services, Office of Head Start, "Head Start Program Annual Fact Sheets."

Note: Appropriations in 2023 dollars.

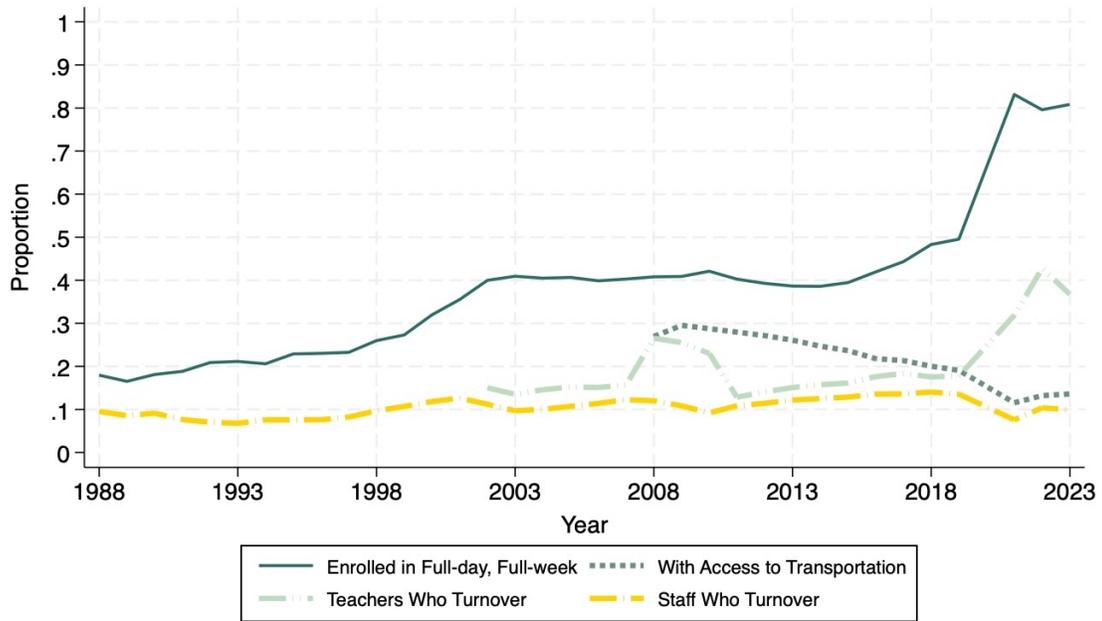
Figure 4: Head Start Funding per Child Served, 1965-2023



Source: Author's calculations from U.S. Department of Health and Human Services, Office of Head Start, "Head Start Program Annual Fact Sheets."

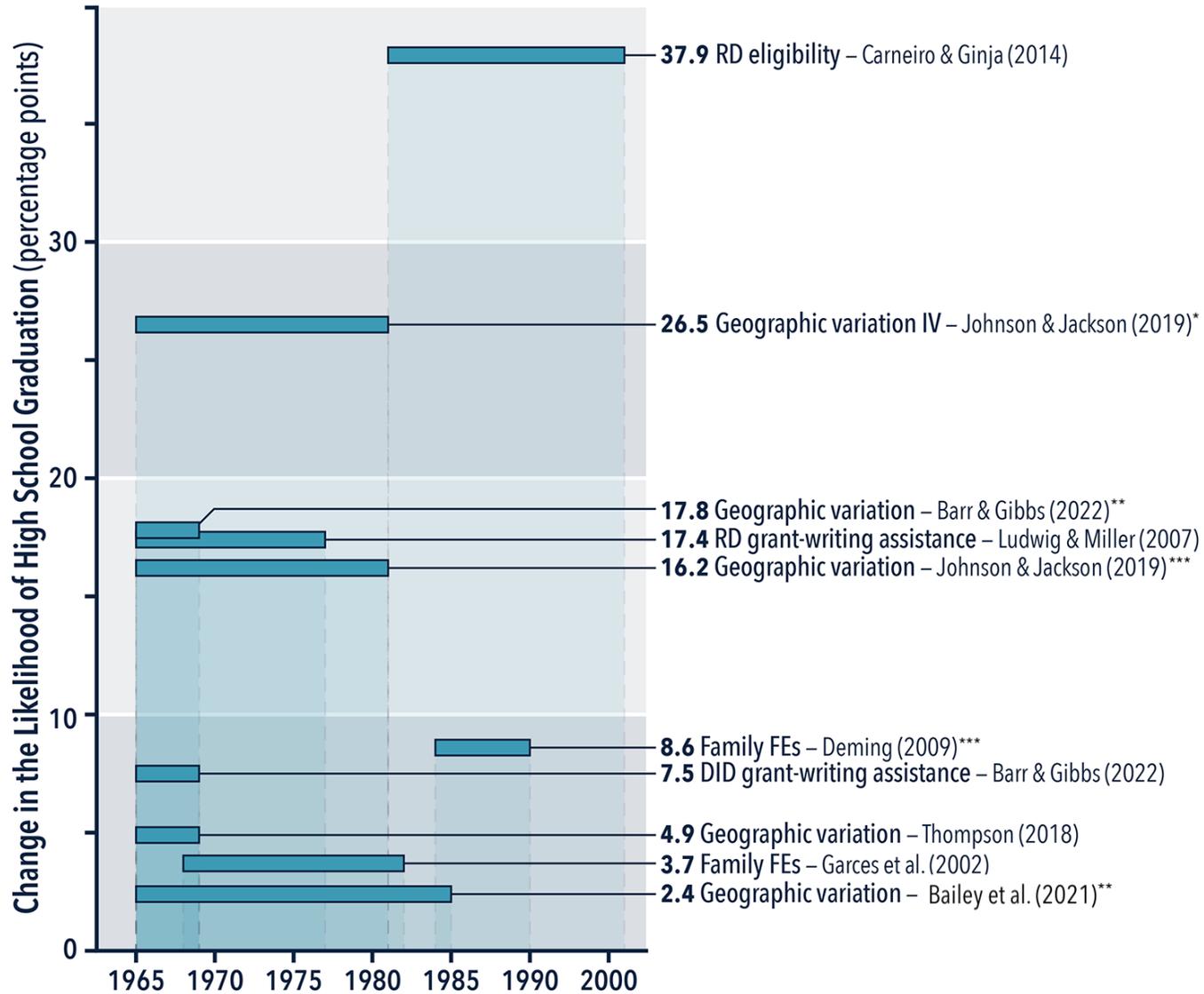
Note: Funding in 2023 dollars.

Figure 5: Head Start Programming, 1988-2023



Source: Author's calculations from Head Start Program Information Reports, 1988-2023

Figure 6: Quasi-Experimental Study Estimates of Head Start Impact



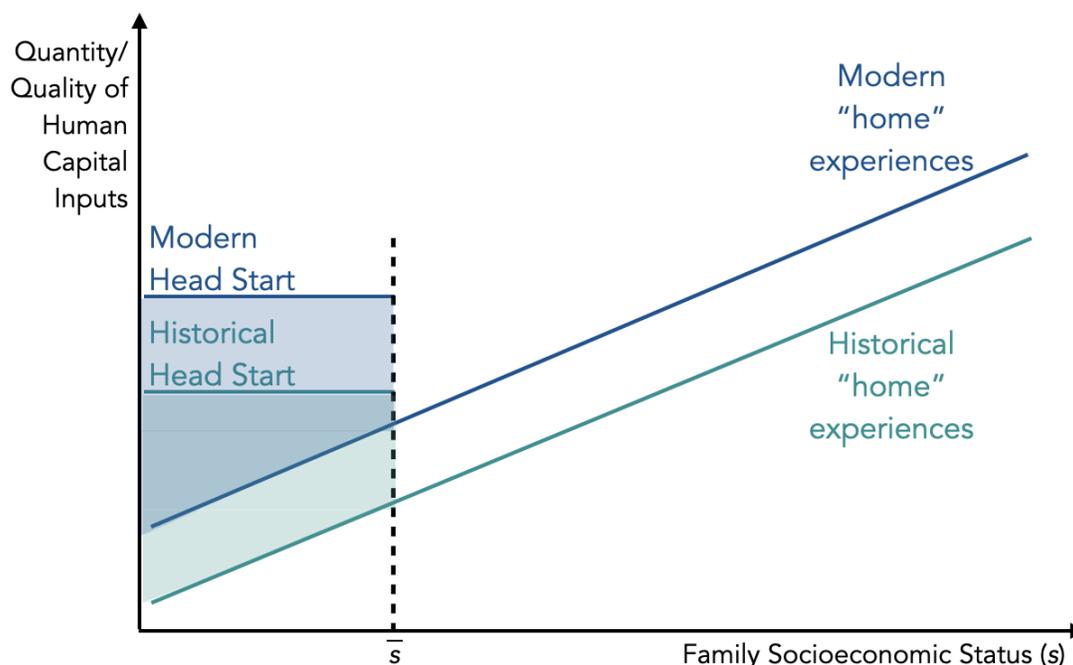
Note: Family fixed effects studies report an estimate of the treatment on the treated. For other studies, intent-to-treat estimates are scaled by the authors' reported first stage though these should be interpreted with some caution. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ denote statistical significance of the authors' corresponding intent-to-treat estimates.

Source: Estimates compiled from Figure 5 in Bailey et al. (2021) and Table H1 in the online appendix of Barr and Gibbs (2022).

Appendix: Modeling Head Start Effectiveness

A simple model, reflecting the framework captured in Figure 2 in the paper (reproduced as Figure A1), can concisely capture the key components of understanding Head Start effectiveness over time.

Figure A1: Understanding Program–No Program Contrast, Then and Now



Source: Author's adaptation of conceptual model in Cascio and Schanzenbach (2014).

Let $s \in [0, \bar{s}]$ denote family socioeconomic status (SES), where \bar{s} is the SES-eligibility threshold for Head Start participation. Additionally, let:

- H_t denote the quantity/quality of Head Start inputs in year $t \in \{1965, 2025\}$,
- $C_t(s) = \alpha_t + \beta s$ denote counterfactual (non-Head Start) inputs at SES s in year t ,
- $Y_t^{HS}(s)$ denote the outcome in the Head Start-participating condition at SES s in year t , and
- $Y_t^{CF}(s)$ denote the outcome in the non-participating condition at SES s in year t .

To link inputs to outcomes, we assume child outcomes are generated via a production function, $f(\cdot)$, mapping early years human capital inputs to outcomes:

$$\begin{aligned} Y_t^{HS}(s) &= f(H_t) \text{ if } s < \bar{s}, \\ Y_t^{CF}(s) &= f(C_t(s)) = f(\alpha_t + \beta s). \end{aligned}$$

applying, for simplicity, the same slope of “home” experiences in SES across years (β), though the level varies over time (α_t). The production of Head Start outcomes is independent of family SES while counterfactual outcomes vary with s , an intentional simplification to isolate how changes in program inputs and outside options drive outcomes (i.e., the complete-substitution assumption).

The Head Start treatment effect, as a function of the quantity and quality of inputs to human capital development, at SES level s and time t , is:

$$\delta_t(s) = Y_t^{HS}(s) - Y_t^{CF}(s) = f(H_t) - f(\alpha_t + \beta s).$$

Let $g(s)$ denote the probability density function of SES on $[0, \bar{s}]$, with $\int_0^{\bar{s}} g(s) ds = 1$.

Next, we aggregate over the range of eligible SES levels $s \in [0, \bar{s}]$, such that we generate averages over the eligible population at each time point:

$$\bar{\delta}_t = \int_0^{\bar{s}} \delta_t(s) g(s) ds = \int_0^{\bar{s}} [f(H_t) - f(\alpha_t + \beta s)] g(s) ds.$$

We assume a linear production function in which γ captures the productivity of human capital inputs, $f(x) = \gamma x$. Then:

$$\begin{aligned} \delta_t(s) &= \gamma(H_t - \alpha_t - \beta s), \\ \bar{\delta}_t &= \gamma \int_0^{\bar{s}} (H_t - \alpha_t - \beta s) g(s) ds \\ &= \gamma \left[(H_t - \alpha_t) \int_0^{\bar{s}} g(s) ds - \beta \int_0^{\bar{s}} s g(s) ds \right] \\ &= \gamma(H_t - \alpha_t - \beta \cdot \mathbb{E}[s]), \end{aligned}$$

where $\mathbb{E}[s]$ is the expected SES level among Head Start participants.

To characterize the change in Head Start effectiveness over time, we focus on two time points,

1965 and 2025, and let $\Delta H = H_{2025} - H_{1965}$ and $\Delta\alpha = \alpha_{2025} - \alpha_{1965}$. If we assume $\mathbb{E}[s]$ is stable over time, then the change in average effectiveness is captured by:

$$\begin{aligned}\delta^{\text{change}} &= \bar{\delta}_{2025} - \bar{\delta}_{1965} \\ &= \gamma(H_{2025} - \alpha_{2025} - \beta \cdot \mathbb{E}[s]) - \gamma(H_{1965} - \alpha_{1965} - \beta \cdot \mathbb{E}[s]) \\ &= \gamma(\Delta H - \Delta\alpha).\end{aligned}$$

Empirical evidence suggests that both the average Head Start experience and the average counterfactual early childhood experience have improved since 1965. That is, $\Delta H > 0$, denoted by the shift from the green to blue Head Start lines in the figure, and $\Delta\alpha > 0$, as shown by the difference in the green and blue “home” experiences lines. Straightforwardly,

- If $\Delta H > \Delta\alpha$, Head Start has become more effective relative to counterfactual conditions.
- If $\Delta H < \Delta\alpha$, Head Start has become less effective alongside larger improvements in counterfactual conditions.
- If $\Delta H = \Delta\alpha$, Head Start effectiveness remained constant relative to counterfactual conditions.

Importantly, if $\mathbb{E}[s]$ changes over time (i.e., Head Start serves children from relatively higher-SES families, within the range of eligibility, in 2025 than in 1965), that would be an additional consideration in assessing average effectiveness over time. We have insufficient data on the treated population at either time point to draw conclusions.

Online Appendix: Evidence on Head Start Effectiveness

Table A1: Summary of Key Studies on Head Start Effectiveness

<i>Panel A. Quasi-Experimental Studies</i>			
Study	Research Design	Context & Time Frame	Findings
Bailey, Sun, & Timpe (2021), “Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency,” <i>AER</i>	Difference-in-differences	Geographic rollout, 1965–1980	Greater educational attainment and economic self-sufficiency in adulthood
Barr & Gibbs (2022), “Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood,” <i>JPE</i>	Difference-in-differences	Geographic rollout, 1965–1970	Improved long-run outcomes for children of those exposed to Head Start, capturing intergenerational benefits
Bauer & Schanzenbach (2016), “The Long-Term Impact of the Head Start Program,” <i>Hamilton Project economic analysis</i>	Family fixed effects	CNLSY data, 1974–1994	Improvements in educational attainment and greater use of positive parenting practices in adulthood
Carneiro & Ginja (2014), “Long-Term Impacts of Compensatory Preschool on Health and Behavior,” <i>AEJ: Policy</i>	Regression discontinuity	CNLSY data, 1981–2000	Reductions in behavioral and health problems in adolescence and lower likelihood of criminal engagement and idleness in young adulthood, for males

Continued on next page

Study	Research Design	Context & Time Frame	Findings
Deming (2009), “Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start,” <i>AEJ: Applied</i>	Family fixed effects	CNLSY data, 1984–1990	Long-term impact of 0.23 SD on a summary index of outcomes in adolescence and early adulthood despite “fadeout” of test score effects
Garces, Thomas, & Currie (2002), “Longer-Term Effects of Head Start,” <i>AER</i>	Family fixed effects	PSID data, 1969–1981	Increased educational attainment and suggestive evidence of higher earnings; reductions in criminal engagement among Black participants; positive spillovers from older to younger siblings
Johnson & Jackson (2019), “Reducing Inequality through Dynamic Complementarity,” <i>AEJ: Policy</i>	Difference-in-differences & IV	PSID data, 1965–1980	Greater educational attainment, higher earnings, and reductions in poverty and incarceration; Head Start benefits were larger when followed by better-funded K-12 schooling
Kose (2023), “Public Investments in Early Childhood Education and Academic Performance: Evidence from Head Start in Texas,” <i>JHR</i>	Difference-in-differences	Federal funding expansions, TX administrative data, 1988–1994	Improved academic achievement, particularly for low-income Hispanic students, and greater language proficiency and lower special education identification among Hispanic students
Ludwig & Miller (2007), “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design,” <i>QJE</i>	Regression discontinuity	Grant-writing assistance at program introduction, 1965–1980	Reductions in child mortality and suggestive evidence of increased educational attainment
Miller, Shenhav, & Grosz (2023), “Selection into Identification in Fixed Effects Models,” <i>JHR</i>	Family fixed effects	PSID and CNLSY data, 1965–1990	Selection into the identifying variation in family fixed effects designs generates estimates of Head Start impact that are larger than “reweighted-on-observables” estimates intended to recover average treatment effects

Continued on next page

Study	Research Design	Context & Time Frame	Findings
Pages et al. (2020), “Elusive Longer-Run Impacts of Head Start: Replications Within and Across Cohorts,” <i>EEPA</i>	Family fixed effects	CNLSY data, 1984–2000	Extending timeframe of Deming (2009) yields mixed evidence on long-term impact with null effects overall and mostly negative effects in more recent birth cohorts
Pihl (2022), “Head Start and Mothers’ Work: Free Child Care or Something More?,” <i>Labour Econ</i>	Regression discontinuity	Census data, 1970	Increased employment in the short run for non-white mothers and decreased employment for single mothers, with persistent declines (10 yrs. later) in employment for non-white and single mothers.
Thompson (2018), “Head Start’s Long-Run Impact,” <i>JHR</i>	Difference-in-differences	NLSY79, 1966–1970	Increased earnings, additional years of schooling, and lower likelihood of a health limitation in adulthood.
Wikle & Wilson (2023), “Access to Head Start and Maternal Labor Supply: Experimental and Quasi-Experimental Evidence,” <i>JOLE</i>	Difference-in-differences	CPS, 1990s funding expansions	Led to increased short-run employment and earnings among single mothers.
<i>Panel B. Head Start Impact Study (HSIS)</i>			
Bitler, Hoynes, & Domina (2014), “Experimental Evidence on Distributional Effects of Head Start,” <i>NBER working paper</i>	RCT re-analysis	HSIS, 2000s	Gains concentrated at the bottom of the skills distribution, and cognitive gains persist for Spanish speakers while fading out for the full population.

Continued on next page

Study	Research Design	Context & Time Frame	Findings
Bloom & Weiland (2015), “Quantifying Variation in Head Start Effects on Young Children’s Cognitive and Socio-Emotional Skills Using Data from the National Head Start Impact Study,” <i>MDRC working paper</i>	RCT re-analysis	HSIS, 2000s	Considerable cross-site variation in program effectiveness, relative to local alternatives, and pronounced positive effects for certain subgroups, including dual language learners and Spanish-speaking children with low pretest scores.
Feller et al. (2016), “Compared to What? Variation in the Impacts of Early Childhood Education by Alternative Care Type,” <i>Annals Appl Stat</i>	RCT re-analysis	HSIS, 2000s	Positive effects on vocabulary for those who would otherwise be in home-based care and no discernable effects on vocabulary for those who would be in other center-based care.
Gelber & Isen (2013), “Children’s Schooling and Parents’ Behavior: Evidence from the Head Start Impact Study,” <i>JPubE</i>	Multi-site RCT	HSIS, 2000s	Sizable increases in parents’ involvement, including time spent reading, math activities, and days children spent with non-resident fathers, persisting after the Head Start period.
Kline & Walters (2016), “Evaluating Public Programs with Close Substitutes: The Case of Head Start,” <i>QJE</i>	RCT re-analysis	HSIS, 2000s	Program produces larger skill benefits for children who would otherwise not participate in preschool and for those who are less likely to enroll in Head Start.
Morris et al. (2018), “New Findings on Impact Variation from the HSIS: Informing the Scale-Up of Early Childhood Programs,” <i>AERA Open</i>	RCT re-analysis	HSIS, 2000s	Average effects conceal important variation in effectiveness across sites and by children’s characteristics and counterfactual experiences.
Puma et al. (2010), “Head Start Impact Study Final Report,” <i>US Dept of HHS report</i>	Multi-site RCT	HSIS, 2000s	Positive cognitive, social-emotional, and health effects during the program year largely evaporated by kindergarten and end of 1st grade.

Continued on next page

Study	Research Design	Context & Time Frame	Findings
Puma et al. (2012), “Third Grade Follow-up to the Head Start Impact Study,” <i>US Dept of HHS report</i>	Multi-site RCT	HSIS, 2000s	Little to no difference on a variety of elementary school outcomes measured in third grade between children with program access and those in the control group .
Sabol & Chase-Lansdale (2015), “The Influence of Low-Income Children’s Participation in Head Start on Their Parents’ Education and Employment,” <i>JPAM</i>	Multi-site RCT	HSIS, 2000s	Increases in parental education, particularly for African American parents and those with at least some college at program outset.
Schiman (2021), “Experimental Evidence of the Effect of Head Start on Mothers’ Labor Supply and Human Capital Investments,” <i>REHO</i>	Multi-site RCT	HSIS, 2000s	Increases in use of non-parental care, but no overall effect on mothers’ likelihood of employment; mothers of younger children at program entry increase labor supply and enrollment at school.
Walters (2015), “Inputs in the Production of Early Childhood Human Capital: Evidence from Head Start,” <i>AEJ: Applied</i>	RCT re-analysis	HSIS, 2000s	More cross-center variation than typical estimates of teacher and school effectiveness, with those offering full-day programming and home visitation more effective and those drawing children from other center-based care less effective.
Wikle & Wilson (2023), “Access to Head Start and Maternal Labor Supply: Experimental and Quasi-Experimental Evidence,” <i>JOLE</i>	Multi-site RCT	HSIS, 2000s	Greater maternal labor force participation, with largest effects among never-married mothers, mothers with no younger children, and for those exposed to centers with full-day programs.