EdWorkingPaper No. 26-1591

Heterogeneous Fadeout of Cognitive Gains: Evidence from Head Start Preschool

Haoxuan Liu
University of Chicago

Fadeout is defined as the decline in an intervention's initial treatment effect in subsequent follow-up periods. This paper uses the Head Start Impact Study to examine heterogeneity in the fadeout of cognitive gains from access to Head Start preschool. I find that Black children experience substantially less fadeout than non-Black children and retain a 0.23 standard deviation gain at the end of kindergarten ($p<0.10$). A selection-corrected decomposition attributes 74% of the Black--non-Black fadeout differential to differences in cognitive growth between Black and non-Black children under Head Start nonparticipation. Mediation analyses provide suggestive evidence that kindergarten school environments, particularly school achievement levels and time spent in adult-directed individual activities, partially contribute to these divergent growth trajectories. Counterfactual simulations show that reallocating existing Head Start offers toward centers serving predominantly Black applicants reduces aggregate fadeout only modestly because Black children constitute a relatively small share of applicants nationally. These findings suggest that meaningfully improving the persistence of Head Start's cognitive effects may require expanding the program and directing recruitment toward children for whom initial gains are more likely to persist.

VERSION: September 2026

Suggested citation: Liu, Haoxuan. (2026). Heterogeneous Fadeout of Cognitive Gains: Evidence from Head Start Preschool. (EdWorkingPaper: 26-1591). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/z77m-rk76

Heterogeneous Fadeout of Cognitive Gains: Evidence from Head Start Preschool

Haoxuan Liu*

September 16, 2026

Abstract

Fadeout is defined as the decline in an intervention's initial treatment effect in subsequent follow-up periods. This paper uses the Head Start Impact Study to examine heterogeneity in the fadeout of cognitive gains from access to Head Start preschool. I find that Black children experience substantially less fadeout than non-Black children and retain a 0.23 standard deviation gain at the end of kindergarten ($p < 0.10$). A selection-corrected decomposition attributes 74% of the Black–non-Black fadeout differential to differences in cognitive growth between Black and non-Black children under Head Start nonparticipation. Mediation analyses provide suggestive evidence that kindergarten school environments, particularly school achievement levels and time spent in adult-directed individual activities, partially contribute to these divergent growth trajectories. Counterfactual simulations show that reallocating existing Head Start offers toward centers serving predominantly Black applicants reduces aggregate fadeout only modestly because Black children constitute a relatively small share of applicants nationally. These findings suggest that meaningfully improving the persistence of Head Start's cognitive effects may require expanding the program and directing recruitment toward children for whom initial gains are more likely to persist.

*Harris School of Public Policy, University of Chicago, 1307 E. 60th Street, Chicago, IL 60637. Email: noahliuhx@uchicago.edu. I thank Marianne Bertrand, Robert Kaestner, Ariel Kalil, Stephen Raudenbush, Lesley Turner, and Christopher Walters, as well as seminar participants in the Harris PhD Workshop and the Committee on Education at the University of Chicago, for helpful comments and suggestions. All errors are mine.

1 Introduction

Early investments in human capital play an important role in shaping children's developmental trajectories and lifelong well-being. A large interdisciplinary literature spanning psychology, education, economics, and neuroscience has extensively documented the benefits of early childhood interventions (Knudsen et al., 2006; Cunha and Heckman, 2007; Duncan and Magnuson, 2013; Almond et al., 2018).

However, a puzzling finding has accompanied this literature from the outset: the short-term impacts of early interventions often do not persist, particularly for cognitive skills (Currie and Thomas, 1995). More recent evidence from randomized controlled trials reaches similar conclusions (Weiland and Yoshikawa, 2013; Lipsey et al., 2018; Weiland et al., 2020; Castillo et al., 2020; Watts et al., 2025). Understanding why early cognitive gains fade is important both for characterizing the dynamics of skill formation in contemporary education systems and for designing effective educational policies that sustain these gains.

Against this backdrop, I study the fadeout of cognitive skill gains in the Head Start Impact Study (HSIS), a large-scale, multisite randomized controlled trial that provided lottery-based access to Head Start preschool for low-income children in the United States. Building on the well-established finding that the average cognitive impacts of Head Start in the HSIS fade out by the end of kindergarten (Puma et al., 2010a), I examine whether the extent of fadeout differs across children facing different forms of socioeconomic and structural disadvantage. More disadvantaged children may enter more resource-constrained educational environments after preschool and consequently experience different skill-growth trajectories. Studying this heterogeneity can therefore deepen our understanding of the conditions associated with the persistence or erosion of Head Start gains and inform policies aimed at improving the persistence of program impacts and the allocation of Head Start resources.

Focusing on the four-year-old cohort, who experienced one year of preschool before entering kindergarten, I find that the average intent-to-treat (ITT) effect on a cognitive composite of reading and mathematics skills declines from 0.13 SD at the end of Head Start ($p < 0.01$) to a statistically insignificant -0.08 SD at the end of kindergarten, one year later, indicating complete fadeout of the initial cognitive gains.

Using the same cognitive outcome and time window, I then examine whether fadeout differs across six dimensions of baseline disadvantage: race, ethnicity, baseline cognitive skills, maternal education, family structure, and home risk. I find that Black children experience substantially less fadeout during the kindergarten year than observably similar non-Black children. Across specifications that account for the other dimensions of disadvantage and indicators of Head Start center quality, Black children experience between 0.36 SD ($p < 0.05$) and 0.43 SD ($p < 0.01$) less fadeout. Neither the remaining dimensions of disadvantage nor the center-quality measures significantly predict differential fadeout.

More importantly, this differential fadeout translates into substantial differences in cognitive skill effects by the end of kindergarten. At the end of Head Start, the estimated ITT effects are nearly identical for Black and non-Black children, at 0.15 SD ($p < 0.10$) and 0.16 SD ($p < 0.01$), respectively. Their subsequent trajectories, however, diverge sharply. By the end of kindergarten, the estimated effect for non-Black children declines to -0.14 SD ($p < 0.05$), whereas Black children retain a positive effect of 0.23 SD ($p < 0.10$). Estimates in first and third grade are directionally consistent with this pattern, although their smaller samples reduce their precision. I also find a similar set of results among compliers, for whom changes in the treatment effect more directly capture the attenuation of cognitive gains induced by Head Start participation.

To further understand the source of the Black–non-Black fadeout differential, I decompose it into between-group differences in cognitive growth among treatment- and control-assigned children. Distinguishing between these two components helps identify where the heterogeneity in fadeout originates and provides an important step toward understanding its underlying mechanisms. I find that differences in control-group growth account for approximately 79–84% of the ITT fadeout differential between Black and non-Black children. In particular, non-Black control-assigned children experience 0.29–0.36 SD more cognitive growth during kindergarten than otherwise similar Black control-assigned children. By contrast, the Black–non-Black growth difference among treatment-assigned children is only 0.07–0.08 SD and is statistically insignificant. Thus, the Black–non-Black fadeout differential appears to arise primarily from substantially faster kindergarten growth among non-Black children in the control group, rather than from differential growth among treatment-assigned children.

Although the ITT decomposition attributes most of the Black–non-Black fadeout differential to differences in control-group growth, this reduced-form pattern may be confounded by differential selection into Head Start participation. Black children have a substantially weaker first stage than non-Black children: the share of always-takers is more than twice as large among Black children (38% versus 16%), while the share of compliers is 18 percentage points smaller (53% versus 71%) (Angrist and Imbens, 1995). As a result, the observed control groups differ substantially in their composition: Black control-assigned children are much more likely than non-Black control-assigned children to participate in Head Start despite not receiving an offer. This compositional difference can generate an apparent Black–non-Black difference in control-group growth, and hence a large attribution of the fadeout differential to the control side, even in the absence of any true racial difference in potential cognitive growth under either Head Start participation or nonparticipation. The reduced-form decomposition therefore cannot distinguish differential selection from differential potential growth, motivating a more structural analysis that separates these two sources of the observed fadeout differential.

I therefore develop a two-period selection model that jointly models Head Start participation and cognitive outcomes at the end of Head Start and the end of kindergarten. The model allows both selection into Head Start and cognitive skill levels in each period, and therefore subsequent fadeout, to depend on observed and unobserved individual characteristics. I also model Head Start site-specific parameters as random effects to capture unobserved heterogeneity across local environments that may jointly influence Head Start participation and children's cognitive development.

Accounting for differential selection reduces the estimated contribution of control-side growth by about 10 percentage points. Conditional on the other measures of baseline disadvantage and Head Start center characteristics, Black children experience 0.63 SD less fadeout during kindergarten than otherwise similar non-Black children ($p < 0.01$). The selection-corrected decomposition attributes approximately 74% of this differential to differences in cognitive growth under nonparticipation, compared with 84% in the corresponding ITT decomposition.

Predicted growth across the four combinations of race and participation status further clarifies that the differential is driven by unusually rapid kindergarten growth among non-Black children under nonparticipation. Taking this group as the benchmark, predicted growth among Black nonparticipants, Black participants, and non-Black participants is 0.47, 0.44, and 0.60 SD lower, respectively. This pattern shows that non-Black children do not generally experience faster kindergarten growth: under Head Start participation, their predicted growth is relatively similar to that of Black children under either participation status. Non-Black children stand apart only when they do not attend Head Start, under which condition they experience substantially more rapid kindergarten growth.

This pattern raises a natural follow-up question: what explains the unusually rapid kindergarten growth of non-Black children under Head Start nonparticipation? To investigate this question, I use reduced-form specifications to examine two hypotheses that may account for the divergent kindergarten growth trajectories of Black and non-Black children: differential family compensation and differential school compensation. Focusing on a smaller complete-case sample, I find little evidence that non-Black control children receive systematically greater family investments during kindergarten that could account for their unusually rapid catch-up. By contrast, Black and non-Black control children are exposed to meaningfully different kindergarten environments. In particular, Black control children attend schools with significantly lower reading and mathematics proficiency (0.32 SD, $p < 0.01$) and spend approximately half as much class time in adult-directed individual activities as otherwise similar non-Black control children ($p < 0.05$).

A Gelbach decomposition provides suggestive evidence that these kindergarten school environments partially account for the divergent growth trajectories. Observed school inputs account for approximately 32% of the Black–non-Black control-group growth differential and 17% of the resulting fadeout differential. The largest contributions come from school reading and mathematics proficiency and time spent in adult-directed individual activities. These patterns are consistent with a differential school-compensation mechanism in which non-Black control children enter kindergarten environments that are better positioned to support rapid catch-up.

Finally, I ask whether the racial differences in the persistence of Head Start effects can be leveraged to reduce aggregate fadeout. I use the model to estimate a series of partial-equilibrium policy counterfactuals that reallocate a fixed number of Head Start offers toward centers where a majority of applicants are Black. These counterfactuals capture the direct effect of changing the allocation of existing offers while holding children's potential outcomes, local environments, and other general-equilibrium responses fixed. Under the most extensive reallocation counterfactual, the Head Start offer rate among Black applicants rises from 0.58 to 0.96, yet aggregate fadeout is attenuated by only 9.6%. The limited aggregate effect arises in part because Black children constitute a relatively small share of Head Start applicants nationally, so even a large increase in their offer rate changes access for only a modest share of the overall applicant population.

These policy simulation results suggest that policies seeking to exploit subgroup differences in fadeout through the reallocation of existing offers are unlikely to meaningfully reduce aggregate attenuation. At the same time, the greater persistence of gains among Black children provides a potential rationale for directing additional Head Start resources toward centers serving predominantly Black applicants. Resources could also be geographically targeted toward highly disadvantaged neighborhoods, which may provide an administratively feasible way to reach a larger share of Black applicants.[1]

This paper makes several contributions. First, I contribute to the literature on the effects of Head Start. Head Start has been studied intensively for decades, including through observational and quasi-experimental analyses based on sibling comparisons and cohorts exposed to the program from the 1960s through the late 1990s, as well as through the nationally representative randomized HSIS launched in the early 2000s (Currie and Thomas, 1995; Garces et al., 2002; Ludwig and Miller, 2007; Ludwig and Phillips, 2007; Carneiro and Ginja, 2014; Puma et al., 2010a, 2012; Gelber and Isen, 2013; Walters, 2015; Kline and Walters, 2016). Studies of earlier cohorts generally found lasting effects on educational attainment and later-life outcomes, whereas the HSIS documented that average cognitive gains had largely disappeared by the end of kindergarten. This finding created some tension with the earlier literature and raised concerns that the benefits observed among earlier cohorts might not be replicable within the contemporary education system, although several earlier studies also documented fadeout of initial cognitive gains alongside persistent long-term benefits.

I show that cognitive gains in the HSIS did not fade universally. Black and non-Black children experience nearly identical cognitive gains at the end of Head Start, but the gains among Black children persist at least through the end of kindergarten, whereas those among non-Black children fade completely. This finding contrasts with evidence from earlier cohorts, which found that Head Start's cognitive gains faded more rapidly among Black children and attributed this pattern partly to differences in the quality of the schools children subsequently attended (Currie and Thomas, 1995, 2000; Deming, 2009). At the same time, it is consistent with Whitaker et al. (2026), who argue that subsequent school quality alone cannot explain the weakening effects of contemporary preschool programs. My findings therefore show that the racial pattern of Head Start fadeout is not stable across cohorts and suggest that it may depend on broader changes in the educational environments in which children develop.

Second, I contribute to the literature on treatment-effect fadeout. A growing body of work

[1] Related socioeconomic and place-based targeting strategies have been used in other educational settings. After a federal court ended Chicago Public Schools' desegregation consent decree, the district replaced race-conscious admissions rules with a system that allocates selective-enrollment seats across socioeconomic tiers defined by applicants' census tracts. Similar approaches may be particularly relevant in large U.S. cities with high levels of residential segregation.

documents the attenuation of early-childhood intervention effects across settings (Hart et al., 2024; Rosengarten et al., 2024; Watts et al., 2026), including in the HSIS (Hart et al., 2026), and investigates the mechanisms underlying this pattern (Bailey et al., 2020; Li et al., 2020; Watts et al., 2023; List and Uchida, 2024; Watts et al., 2025; Kaestner and Schiman, 2026). At the same time, a small but influential body of evidence shows that early-childhood interventions can generate meaningful long-run benefits even when their short-run cognitive effects attenuate, highlighting the potential importance of skills and developmental processes not fully captured by cognitive assessments (Campbell and Ramey, 1994; Nores et al., 2005; Campbell et al., 2012; Heckman et al., 2013; García et al., 2020; García and Heckman, 2023; Gray-Lobe et al., 2023). Taken together, these two lines of work sharpen the value of studying cognitive fadeout. The fadeout or persistence of cognitive gains need not be the sole criterion for judging the value of preschool, but it remains an important dimension of program effectiveness and a useful lens for understanding the process of skill formation (Burchinal et al., 2024). The persistence of cognitive skills is also directly policy-relevant in the United States, where measures of academic achievement play a central role in school accountability systems and shape the school-choice decisions of families across elementary and secondary education (Jacob, 2005; Neal and Schanzenbach, 2010; Figlio and Loeb, 2011; Abdulkadiroğlu et al., 2020; Campos and Kearns, 2024; Corradini, 2024; Abdulkadiroglu et al., 2026).

I contribute to this literature by decomposing heterogeneous fadeout into differences in skill growth under participation and nonparticipation, thereby shifting attention from the evolution of treatment effects alone to the underlying skill-growth trajectories that generate them (Becker, 1962; Ben-Porath, 1967; Cunha and Heckman, 2008; Attanasio et al., 2020). Because subgroup differences in Head Start take-up can distort the observed treatment- and control-group growth comparisons, I further develop a two-period selection model that separates genuine differences in potential growth from differences induced by selection into participation. Applying this framework, I show that the Black–non-Black fadeout differential is driven primarily by unusually rapid cognitive growth among non-Black children under Head Start nonparticipation. I then provide suggestive evidence that differences in kindergarten school environments partially account for this growth disparity, whereas observed family investments provide little support for a differential family-compensation explanation. More broadly, these findings connect to a neighboring literature on charter-school impacts showing that treatment-effect heterogeneity can arise primarily from differences in counterfactual growth under nonattendance (Walters, 2018; Angrist et al., 2023).

Finally, this paper directly extends the heterogeneous selection model developed by Walters (2015) to study heterogeneity in the initial effects of Head Start, with particular attention to variation in alternative care environments. I extend this framework to two outcome periods by jointly modeling Head Start participation and treatment-specific potential outcomes at the end of Head Start and the end of kindergarten, while placing a parametric empirical Bayes distribution over the site-specific parameters to capture heterogeneity arising from unobserved site-level factors. More broadly, the model is related to the literature on endogenous treatment selection, particularly in settings with longitudinal outcomes (Heckman, 1979; Heckman and Vytlacil, 2001; Yau and Little, 2001; Chib and Jacobi, 2007; Heckman, 2010; Jacobi et al., 2016; Mogstad et al., 2018; Wagner et al., 2023), as well as to the empirical Bayes and random-effects literatures (Morris, 1983; Raudenbush et al., 2012; Efron, 2012; Chetty et al., 2014; Gu and Koenker, 2023; Kline et al., 2024). The resulting framework combines elements of these approaches to study heterogeneity in how treatment effects evolve across individuals and local Head Start environments.

The remainder of the paper is organized as follows. Section 2 describes the Head Start Impact Study, the analysis sample, and the cognitive outcomes. Section 3 documents average fadeout and examines heterogeneity by baseline disadvantage. Section 4 studies the greater persistence of Head Start gains among Black children and decomposes the Black–non-Black fadeout differential into treatment- and control-side growth. Section 5 develops and estimates the two-period selection model and presents the selection-corrected results. Section 6 examines the potential role of post-Head-Start environments and considers the policy implications of reallocating Head Start offers. Section 7 concludes.

2 Head Start Impact Study

2.1 Head Start Background

Head Start is the largest early childhood education program in the United States. Established in 1965 under President Lyndon B. Johnson as part of the War on Poverty, the program began as an eight-week demonstration project and has since served more than 40 million children and their families (Administration for Children and Families, 2024). Head Start aims to promote school readiness among children ages 3 to 5 from low-income families by providing comprehensive services that support early cognitive development, health, and family well-being, primarily through center-based programs. Eligibility is primarily determined by family income relative to federal poverty guidelines, though children experiencing homelessness or living in families receiving public assistance, such as Temporary Assistance for Needy Families or the Supplemental Nutrition Assistance Program, also qualify (Office of Head Start, 2024c, 2025).

Head Start also constitutes a major public expenditure in early childhood education. In fiscal year 2024 (FY2024), the federal government allocated approximately \$7.08 billion in operating funds to the Head Start Preschool program.[2] Because federal grants typically cover about 80% of total program costs, total operating resources are estimated at approximately \$8.85 billion.[3] This level of spending is substantial: it amounts to roughly 72% of the City of Chicago's FY2024 net municipal operating budget (City of Chicago, Office of Budget and Management, 2023), placing Head Start on a fiscal scale approaching that of the municipal operating budget of the third-largest city in the United States.

Head Start's per-child financial investment is comparable to, and often exceeds, that of state-level preschool initiatives. In FY2024, Head Start Preschool supported 488,792 funded enrollment slots (Office of Head Start, 2024a), implying a total resource cost of approximately \$18,115 per child. By comparison, national estimates suggest that the per-child cost of a high-quality state public pre-K program is approximately \$15,621 in 2024 dollars.[4] Estimated costs for specific state programs range from about \$12,054 in Oklahoma to \$18,942 for Boston's universal pre-K program, while Tennessee's Voluntary Pre-K program costs approximately \$13,653 (Karoly et al., 2021). While higher spending does not necessarily imply higher quality, the larger per-child investment in Head Start is consistent with its broader service model, which integrates early education with health screenings, disability services, and family support components.

[2]This figure excludes Early Head Start, American Indian and Alaska Native Head Start, and Migrant and Seasonal Head Start programs. The exact federal operations appropriation for Head Start Preschool in FY2024 was $\$7,082,932,155$.

[3]Head Start is federally funded but locally administered. The federal government awards annual grants directly to community-based organizations, including school districts and nonprofit and for-profit providers. Grantees are generally required to provide a non-federal match equal to 20% of total program costs, which may be satisfied through cash contributions or in-kind support from state and local governments or private sources. Awarded funds are typically required to be obligated within the designated budget year (Office of Head Start, 2024b).

[4]Dollar amounts originally reported in 2019 dollars are converted to 2024 dollars using the Consumer Price Index for All Urban Consumers inflation calculator from the U.S. Bureau of Labor Statistics, with an inflation adjustment factor of 1.23: `https://www.bls.gov/data/inflation_calculator.htm`.

2.2 HSIS Background and Design

The Head Start Impact Study (HSIS) is a nationally representative randomized controlled trial designed to evaluate the effects of access to Head Start on children's development. Following the 1998 reauthorization of Head Start, Congress mandated a national evaluation to assess the program's effects on children's school readiness, parental practices, and outcomes across subgroups. In fall 2002, the HSIS was launched with a nationally representative sample of newly entering, eligible children from Head Start programs where demand exceeded available capacity.[5] The restricted-use HSIS sample used in this study contains 4,442 newly entering, eligible children, including 2,449 three-year-olds and 1,993 four-year-olds, drawn from 353 Head Start centers operated by 83 grantees.[6] Children were randomly assigned either to a treatment group, which received access to Head Start services, or to a control group, which did not receive access through the study lottery but could enroll in other early childhood programs, including Head Start centers not participating in the HSIS, available in their communities (Puma et al., 2010a).

HSIS followed both treatment and control children longitudinally. Beginning in fall 2002, children in the three-year-old cohort entered their first year of Head Start and were eligible to attend a second year of preschool before entering kindergarten at age five. Consequently, during the second preschool year, children originally assigned to the control group could enroll in Head Start centers that had previously rejected their applications or participate in other preschool programs available in their communities. In contrast, children in the four-year-old cohort assigned to the treatment group were eligible for only one year of Head Start prior to kindergarten entry. Accordingly, Puma et al. (2010a) emphasize that the estimated HSIS impacts "represent the effects of one year of Head Start." After the preschool year(s), both treatment and control children entered kindergarten (see Table 1 for the data collection timeline by cohort).

[5] Consequently, the study was conducted exclusively in communities where demand for Head Start exceeded capacity. Although the HSIS is broadly nationally representative, it does not represent approximately 15 percent of Head Start programs operating in communities where available slots exceeded the number of applicants (Puma et al., 2010a).

[6] Puma et al. (2010a) reports that the original HSIS recruited 4,667 children, including 2,559 three-year-olds and 2,108 four-year-olds, from 383 randomly selected oversubscribed Head Start centers operated by 84 grantees across 23 states. The smaller restricted-use sample used here reflects the exclusion of Puerto Rico, where differences in the language of instruction and interpretation of outcomes in a predominantly Spanish-speaking context limit comparability with the mainland United States.

The HSIS collected rich data on child assessments, parent interviews, and teacher reports at baseline, during the Head Start year(s), and in follow-up waves in kindergarten, first grade, and third grade.

2.3 Sample

I focus on the four-year-old cohort in the HSIS data. Among the 1,993 children who entered the study at age four, 1,182 were assigned to the treatment group and 811 to the control group. Although the three-year-old cohort also experienced complete fadeout of its cognitive gains during the second preschool year (Walters, 2015), I exclude this cohort for two reasons. First, the three-year-old cohort presents a more complicated setting for studying how subsequent environments shape fadeout. Among three-year-olds, some treatment-assigned children leave Head Start in the second year, while many control-assigned children enter Head Start, so subsequent catch-up can partly reflect changes in program participation arising from families' choices of preschool environments. While substantively important, this makes it difficult to isolate a different question: how variation in the quality and characteristics of subsequent environments shapes fadeout once children have entered a common schooling stage. This latter margin is both theoretically less obvious and particularly relevant for policy because it speaks to which features of subsequent schools may reinforce or erode earlier preschool gains, and therefore which features schools may potentially modify. By contrast, four-year-old children generally transition to kindergarten after the focal Head Start year, providing a more uniform institutional setting in which to study how differences in subsequent school environments shape the persistence or fadeout of earlier cognitive gains. Second, many contemporary preschool programs, particularly state-funded universal prekindergarten programs, primarily serve four-year-old children. The cognitive trajectories of the HSIS four-year-old cohort therefore have greater relevance for understanding the persistence of effects in contemporary preschool settings (Stanford, 2023).

2.4 Outcomes

My analysis focuses on children's language, reading, and mathematics skills, measured using the Peabody Picture Vocabulary Test (PPVT), the Woodcock-Johnson III Letter-Word Identification assessment (WJ Letter-Word), and the Woodcock-Johnson III Applied Problems assessment (WJ Applied).[7]

I reviewed all cognitive outcomes collected in the HSIS and selected these three measures based on four considerations. First, I focus on cognitive achievement because it is the primary domain in which early childhood interventions generate measurable impacts and the domain in which fadeout has been most extensively documented (Hart et al., 2024). Evidence from HSIS is consistent with this pattern: access to Head Start increased cognitive skills by approximately 0.1 standard deviation during preschool, but most of these gains had dissipated by the end of kindergarten (Kline and Walters, 2016). The three outcomes capture distinct but related dimensions of early skill formation. Language and reading measure verbal and literacy development, while mathematics measures quantitative reasoning. Together, they provide a broad characterization of cognitive development in early childhood. In contrast, HSIS found little evidence of initial impacts on social-emotional or health outcomes, providing limited motivation to examine fadeout in these alternative developmental domains (Puma et al., 2010a).

Second, these outcomes are measured using vertically scaled standardized assessments. Each test measures ability along a common underlying continuum over a broad age range, allowing me to compare the same underlying skill across survey waves. This feature addresses a central concern in the fadeout literature: the "change of the goalpost," whereby apparent attenuation may reflect changes in the skills measured by an assessment rather than genuine changes in children's underlying abilities (García and Heckman, 2023; Heckman and Zhou, 2026).

Third, these measures are consistently observed across survey waves for the four-year-old cohort (Puma et al., 2012). This permits a coherent analysis of children's skill trajectories from the end of preschool through the end of third grade and allows me to trace when fadeout occurs. Repeated measurement also helps distinguish persistent fadeout from a "sleeper effect," in which treatment effects temporarily diminish and subsequently re-emerge.

Fourth, these three measures exhibit moderate or no concavity in skill growth in the short term, limiting the amount of fadeout that can arise mechanically from curvature. Appendix B

[7] The Spanish counterparts of these assessments are the Test de Vocabulario en Imágenes Peabody, the Batería III Woodcock-Muñoz Identificación de Letras y Palabras, and the Batería III Woodcock-Muñoz Problemas Aplicados. For Spanish-speaking children, some assessments were administered in both English and Spanish. When both versions were available, I use the assessment that best reflects the child's proficiency at that wave, so that test scores more accurately capture underlying cognitive skills rather than differences in language proficiency.

presents a stylized model that quantifies the mechanical attenuation implied by concave skill growth. I show that, over k periods, the proportional mechanical attenuation of an initial treatment effect is approximately

$$\frac{\Delta_{t+k} - \Delta_t}{\Delta_t} \approx \frac{f''(a_t)}{f'(a_t)}k.$$

Applying this framework to estimates from the National Longitudinal Survey of Youth 1979 (Farkas and Beron, 2004) and the 2006 Head Start Family and Child Experiences Survey cohort (Choi et al., 2016), I find that an effect on PPVT measured at age 5 would decline mechanically by approximately 11–13 percent over one year and 33–39 percent over three years. By contrast, the estimated growth profiles for WJ Letter-Word and WJ Applied exhibit no concavity and therefore imply no mechanically generated fadeout from preschool to kindergarten, which corresponds to the time window examined in my subsequent analysis.

3 Treatment Effects and Heterogeneous Fadeout

3.1 Balance and Retention

Table 2 presents descriptive statistics and baseline balance tests for the 1,993 children in the four-year-old cohort of the HSIS.[8] I examine three categories of covariates: child characteristics, maternal characteristics, and home environment measures. Although Head Start serves low-income families overall, participants are heterogeneous in their demographic characteristics and degree of disadvantage. Approximately 18% of children are Black and half are Hispanic. More than 40% of mothers have less than a high school education, 16% were teenage mothers, and more than 20% of children live in home environments classified as posing a moderate or high risk to their development. Families at different points along this spectrum of disadvantage may expose their children to different educational environments after preschool, creating scope for differential fadeout.

Overall, baseline covariates are well balanced across the treatment and control groups. The only exception is home risk status. Children in the treatment group are 4 percentage points more likely to come from moderate- or high-risk home environments. However, this difference is small in magnitude and only marginally statistically significant.

[8] All analyses apply the HSIS baseline sampling weights to ensure representativeness of the national population of newly entering four-year-old Head Start participants in 2002 (Puma et al., 2010b).

Attrition is examined at each follow-up period, including the end of preschool, kindergarten, first grade, and third grade. Attrition is defined as the absence of valid assessments for any of the three cognitive outcomes (language, reading, and math) at a given wave. Restricting the sample to children with valid measures for all three outcomes yields 1,977 children at baseline (99% of the initial sample), 1,612 children with both baseline and end-of-preschool measures (81%), 1,439 children observed through the end of kindergarten (72%), 1,337 children observed through the end of first grade (67%), and 1,193 children observed through the end of third grade (60%).

Table A1 shows that children assigned to the treatment group have moderately higher retention rates than those assigned to the control group at each follow-up wave, by approximately 7–10 percentage points. This pattern is consistent with prior findings in the HSIS literature (e.g., Walters, 2015). However, the magnitude of this difference is modest and remains stable across waves. Combined with the strong baseline balance between treatment and control groups, this limited differential attrition is unlikely to meaningfully bias the estimated treatment effects.[9]

3.2 Average Treatment Effects

To examine whether the selected outcomes exhibit fadeout over the follow-up period from the end of preschool through the end of third grade, I estimate intent-to-treat (ITT) effects separately at each follow-up wave using:

$$Y_{it} = \beta_{0t} + \beta_{1t} Z_i + \alpha_t' X_i + \gamma_{g(i),t} + \varepsilon_{it},$$

where Y_{it} denotes one of the three cognitive outcomes (language, reading, or math) for child i at follow-up wave t, and Z_i is an indicator for random assignment to receive an offer of Head Start. The vector X_i includes the baseline covariates listed in Table 2, with arbitrary categories omitted to avoid collinearity. The term $\gamma_{g(i),t}$ denotes grantee fixed effects, where $g(i)$ identifies the Head Start grantee associated with child i. I report specifications both

[9]Because the subsequent heterogeneity analysis focuses in particular on differences between Black and non-Black children, I additionally examine whether treatment-related retention differs by race. At the end of kindergarten, regressing an indicator for inclusion in the analysis sample on treatment assignment, a Black indicator, their interaction, the baseline covariates, and grantee fixed effects yields an assignment-by-Black interaction of 0.08 (SE = 0.07, $p = 0.24$). Thus, although treatment assignment increases retention overall, I find no statistically significant evidence that the treatment-related retention difference varies between Black and non-Black children.

with and without baseline covariates with robust standard errors; all specifications include grantee fixed effects to alleviate concerns that the estimated treatment effects are driven by cross-grantee differences in treatment-assignment probabilities and average outcomes.[10]

Within the HSIS sample, compliers constitute a particularly informative subgroup for studying fadeout. Compliers are children who enroll in Head Start when offered a seat but would not enroll otherwise (Angrist and Imbens, 1995). The local average treatment effect (LATE) therefore identifies the causal effect of actual Head Start participation for this group. Attenuation of the LATE after preschool provides direct evidence that the cognitive gains induced by Head Start participation fade out over time, rather than reflecting selection into participation.

Accordingly, I estimate the LATE using two-stage least squares (2SLS), instrumenting actual Head Start participation with random assignment. All specifications are estimated separately by follow-up wave:

$$\begin{aligned} \text{First stage:} \quad & D_i = \pi_0 + \pi_1 Z_i + \rho' X_i + \lambda_{g(i)} + u_i, \\ \text{Second stage:} \quad & Y_{it} = \beta_{0t} + \beta_{1t} \widehat{D}_i + \alpha_t' X_i + \gamma_{g(i),t} + \varepsilon_{it}, \end{aligned}$$

where D_i is an indicator equal to one if child i enrolled in Head Start during the preschool year, either at the originally assigned center or at another Head Start center outside the HSIS experimental sample, and zero otherwise. The terms $\lambda_{g(i)}$ and $\gamma_{g(i),t}$ denote grantee fixed effects in the first- and second-stage equations, respectively.[11]

Tables A2, A3, and A4 report ITT and LATE estimates for the three outcomes across four follow-up periods. I define an outcome as exhibiting fadeout when an initially significant treatment effect attenuates toward zero, becomes statistically insignificant, and remains so through the final follow-up wave. Below, I report estimates from specifications that include baseline covariates and grantee fixed effects.

Language skills do not satisfy this definition. At the end of preschool, the ITT effect is approximately 0.12 baseline control-group SD and the LATE is 0.17 SD. Although both estimates fall close to zero at the end of kindergarten, they recover by the end of first grade

[10] Although center fixed effects provide a more local comparison within Head Start centers, they leave limited within-center variation for estimating heterogeneous treatment effects across dimensions of family disadvantage. I therefore use grantee fixed effects throughout to maintain a common specification across the average and heterogeneous treatment-effect analyses. Reassuringly, specifications using center fixed effects yield estimates of the average treatment effect that are similar to those obtained with grantee fixed effects.

[11] These 2SLS estimates with covariates raise interpretive complications, which I discuss in Section 4.2.

and remain positive through third grade, with some later estimates statistically significant at the 10% or 5% level. The temporary attenuation at the end of kindergarten therefore does not constitute sustained fadeout. These results also underscore the importance of longer-term follow-up: restricting attention to kindergarten would suggest substantial attenuation, whereas the subsequent recovery indicates that the kindergarten decline was not persistent and may instead reflect transitory fluctuations in measured language performance.

Reading skills exhibit clear fadeout. The end-of-preschool ITT and LATE effects are approximately 0.10 and 0.15 SD, respectively, but both become negative and statistically insignificant by the end of kindergarten. The estimates remain close to zero and statistically insignificant through third grade.

Math skills display a similar pattern. The end-of-preschool ITT and LATE effects are approximately 0.12 and 0.19 SD, respectively. By the end of kindergarten, both estimates have attenuated to near zero and become statistically insignificant, and they remain close to zero throughout the subsequent follow-up waves.

Accordingly, the remainder of the analysis focuses on reading and math skills. To summarize fadeout across these domains and reduce outcome-specific noise, I construct a cognitive composite following prior work (Kling et al., 2007; Deming, 2009; Walters, 2015). I first normalize each outcome using the baseline control-group mean and standard deviation, average the normalized reading and math scores, and then re-standardize the resulting composite using its baseline control-group mean and standard deviation. Expressing effects in units of the baseline control-group standard deviation places all follow-up estimates on a common scale and avoids confounding changes in treatment effects with changes in test-score dispersion as children age (Cascio and Staiger, 2012). Figure 1 presents the ITT and LATE estimates for the composite, and Table A5 reports the corresponding regression results.

The composite exhibits a clear pattern of fadeout. Treatment effects are positive and precisely estimated at the end of preschool, with an ITT effect of 0.13 SD and a LATE effect of 0.20 SD, both statistically significant at the 1% level. By the end of kindergarten, these effects decline sharply to -0.08 SD and -0.12 SD, respectively, and remain statistically indistinguishable from zero through the third-grade follow-up. I therefore focus the subsequent analysis on this window from the end of preschool to the end of kindergarten. This window captures essentially all of the observed fadeout and, because it immediately follows preschool, preserves the largest possible sample for identifying heterogeneity and studying the mechanisms underlying fadeout relative to analyses using more distant follow-up periods.

3.3 Heterogeneous Fadeout During Kindergarten

Understanding heterogeneity in fadeout is important for identifying the mechanisms that drive the erosion of early gains and for informing the allocation of Head Start resources. I focus on heterogeneity by baseline disadvantage among Head Start applicants, comparing low-income children with their even more disadvantaged peers. More disadvantaged children often transition into more resource-constrained educational environments after preschool and may therefore follow different subsequent growth trajectories. Examining whether fadeout varies along this dimension can be informative about the potential role of post-preschool environments in preserving or eroding early gains. This analysis also has important equity implications. If the more disadvantaged children experience greater fadeout, the appropriate policy response may be to complement Head Start with additional investments that support their subsequent development, rather than concluding that Head Start resources are ineffective or should be redirected away from them.

I construct six indicators, drawn from the baseline measures in Table 2, to capture distinct dimensions of socioeconomic and structural disadvantage: (1) whether the child is Black, (2) whether the child is Hispanic, (3) whether baseline cognitive skills fall in the bottom quartile, (4) whether the mother has less than a high school education, (5) whether the child does not live with both biological parents, and (6) whether the home environment is classified as posing a moderate or high developmental risk. I include indicators for race and ethnicity because Black and Hispanic children may face structural barriers and forms of disadvantage that are not fully captured by the observed socioeconomic and family-background measures.

To examine heterogeneity in fadeout across the disadvantage spectrum, I estimate the following specification separately for each of the six disadvantage indicators:

$$G_i = \alpha + \beta_1 Z_i + \beta_2 \text{disadv}_i + \beta_3 \left(Z_i \times \text{disadv}_i\right) + X_i'\gamma + \lambda_{g(i)} + \varepsilon_i,$$

where G_i denotes cognitive skill growth between the end of preschool and the end of kindergarten, Z_i indicates random assignment to receive a Head Start offer, and disadv_i denotes one of the six disadvantage indicators described above. The vector X_i includes the full set of baseline covariates listed in Table 2, with any variable directly represented by disadv_i omitted to avoid duplication, and $\lambda_{g(i)}$ denotes grantee fixed effects. The coefficient β_3 captures the treatment-effect differential associated with the corresponding dimension of disadvantage. Because a negative treatment effect on kindergarten skill growth indicates fadeout, a positive estimate of β_3 indicates that children with that characteristic experience less fadeout than otherwise similar children without that characteristic.

Table 3 reports estimated fadeout differentials across six dimensions of baseline disadvantage. I find robust evidence that Black children experience less fadeout than otherwise similar non-Black children. Column (1) estimates each interaction separately. Black children experience 0.38 SD less fadeout during kindergarten than non-Black children, and this difference is statistically significant at the 1% level. In contrast, the remaining indicators provide little evidence of differential fadeout. The interaction for Hispanic children is negative and marginally significant, suggesting greater fadeout, but this pattern appears to be driven by the inclusion of Black children in the comparison group: the estimate becomes small and statistically insignificant once the Black and Hispanic interactions are included jointly in Column (2).

To account for overlap across the different dimensions of disadvantage and examine whether Black children's fadeout pattern can be explained by observed disadvantage indicators, Column (2) includes all six indicators and their interactions with treatment assignment jointly. The estimate for Black children remains large and statistically significant: conditional on the other observed dimensions of disadvantage, Black children experience 0.36 SD less fadeout than otherwise similar non-Black children. Thus, the Black–non-Black fadeout differential is not explained by the observed socioeconomic and family-background indicators available in the HSIS and may instead reflect other unobserved dimensions of disadvantage or structural differences associated with race.

Children facing different forms of disadvantage can attend Head Start centers of differing quality, which may influence the persistence of their initial treatment gains. Column (3) therefore adds interactions between treatment assignment and Head Start center characteristics, including the share of teachers with a bachelor's degree or teaching license, full-day service provision, the breadth of program services, director experience, and class size.[12] None of the center-quality indicators significantly predicts differential fadeout. At the same time, the estimate for Black children remains robust and increases slightly to 0.43 SD.

I also estimate the LATE analog of the heterogeneous fadeout specification, instrumenting Head Start participation and each participation-by-characteristic interaction with random assignment and the corresponding assignment-by-characteristic interaction. Table A7 reports the resulting fadeout differentials. Columns (1), (2), and (3) follow the same sequence of specifications as the ITT analysis. Across all three columns, the result for Black children

[12]See Table A6 for detailed variable definitions.

remains large and statistically significant: Black compliers experience between 0.52 and 0.64 SD less fadeout during kindergarten than otherwise similar non-Black compliers. The remaining disadvantage and center characteristics do not consistently predict differential fadeout.

Taken together, these results provide robust evidence that Head Start's cognitive gains fade less for Black children than for non-Black children during kindergarten. The differential remains large after accounting for observed dimensions of baseline disadvantage and measured Head Start center characteristics, suggesting that these observed factors do not account for the pattern. The following sections investigate potential explanations for this divergence. I first examine whether differential selection into Head Start participation can generate part of the observed fadeout differential, and then consider whether differences in family investments and kindergarten school environments help account for the divergence in cognitive growth.

4 Understanding Less Fadeout among Black Children

4.1 Persistence of Head Start Gains through Kindergarten

To clarify what less fadeout among Black children means in terms of cognitive skill levels, Table 4 reports the ITT effects at the end of Head Start and the end of kindergarten in Panels A and B, respectively. To hold the analysis sample fixed across the two periods, both panels restrict the sample to children observed through the end of kindergarten. The table uses the same three interaction specifications as Table 3, with cognitive skill levels at each wave as the outcomes.

Panel A shows no evidence of a differential initial treatment effect between Black and non-Black children. Column (1), which estimates the interaction with Black separately, shows that the end-of-Head-Start cognitive effects are positive and nearly identical for Black and non-Black children, at 0.15 SD ($p < 0.1$) and 0.16 SD ($p < 0.01$), respectively. Adding interactions with the other disadvantage indicators and Head Start center characteristics attenuates the subgroup-specific estimates and reduces their precision, suggesting that these characteristics capture some of the same treatment-effect heterogeneity as the Black indicator. Nevertheless, the Black–non-Black difference remains small and statistically insignificant in every specification. The greater subsequent fadeout among non-Black children therefore does not originate from a larger initial reduced-form gain. This pattern contrasts with evidence in the recent fadeout literature that larger initial effects tend to undergo greater subsequent attenuation (Watts et al., 2023).

More importantly, Panel B provides suggestive evidence that the initial gains among Black children persist through the end of kindergarten, whereas those among non-Black children have disappeared or even reversed. Column (1) shows that Black children retain a 0.23 SD gain from Head Start assignment at the end of kindergarten ($p < 0.1$). By contrast, the estimated effect for non-Black children is -0.14 SD ($p < 0.05$), indicating that their control-assigned peers have surpassed their treatment-assigned peers. I return to this negative estimate below when examining the growth trajectories that generate it and the potential subsequent compensation in facilitating catch-up. Although adding the other interaction terms reduces the precision of the subgroup-specific estimates, their signs and magnitudes remain broadly stable across specifications. The resulting Black–non-Black treatment-effect difference ranges from 0.37 to 0.44 SD and is statistically significant at the 1% level in every specification.

Table A8 extends the separate-interaction specification in Column (1) through third grade to examine the longer-run pattern of Head Start effects. I find that the negative effect for non-Black children gradually attenuates toward zero, whereas the estimated effect for Black children remains positive at 0.18 SD by third grade, although it is imprecisely estimated ($p = 0.145$), in part because of the smaller follow-up sample. The corresponding LATE estimate for Black children is 0.30 SD ($p = 0.159$).[13]

These results establish that Black and non-Black children experience similar initial gains from assignment to Head Start but substantially different levels of persistence by the end of kindergarten. To better understand the source of this divergence, I next shift attention from differences in treatment effects to differences in the underlying post-preschool skill-growth trajectories.

[13] Studying the dynamics of cognitive growth beyond kindergarten is outside the scope of this paper. One possible interpretation of the temporary negative effect among non-Black children, however, is a dynamic compensatory process. If schools disproportionately direct instructional support toward children who enter kindergarten behind, such responses could generate rapid catch-up and potentially temporary overshooting among non-Black control children during kindergarten. Once these children surpass their treatment-assigned peers, similar compensatory responses in subsequent grades could operate in the opposite direction, causing the negative treatment effect to attenuate toward zero. The available data are not sufficiently detailed to allow me to directly test this possibility.

4.2 Differential Catch-Up in the Control Group

To further understand the source of the Black–non-Black fadeout differential, I decompose it into between-group differences in cognitive growth among treatment- and control-assigned children. This decomposition identifies whether fadeout heterogeneity arises from differences in post-intervention growth among treated children or from differences in the counterfactual trajectories of untreated children. Distinguishing between these sources is theoretically meaningful because a treated-side growth differential is more naturally associated with mechanisms such as dynamic complementarity, whereas a dominant control-side growth differential is more consistent with mechanisms involving compensatory investment.

Let τ_{gt} denote the treatment effect for group g at time t, where $t \in \{P, K\}$ indexes the end of preschool and the end of kindergarten:

$$\tau_{gt} = \mathbb{E}[Y_t(1) - Y_t(0) \mid G = g] \,.$$

The treatment effect on skill growth during kindergarten is

$$\Delta_g = \mathbb{E}[Y_K(1) - Y_P(1) \mid G = g] - \mathbb{E}[Y_K(0) - Y_P(0) \mid G = g] \,.$$

Thus, the fadeout differential between groups $g = 1$ and $g = 0$ can be written as

$$\begin{aligned} \Delta_1 - \Delta_0 = & \underbrace{\{\mathbb{E}[Y_K(1) - Y_P(1) \mid G = 1] - \mathbb{E}[Y_K(1) - Y_P(1) \mid G = 0]\}}_{\text{Treated-group growth differential}} \\ & - \underbrace{\{\mathbb{E}[Y_K(0) - Y_P(0) \mid G = 1] - \mathbb{E}[Y_K(0) - Y_P(0) \mid G = 0]\}}_{\text{Control-group growth differential}} \,. \end{aligned}$$

A group may therefore experience less fadeout because its treated children grow faster, because its control children catch up less rapidly, or through both channels.

Empirically separating these channels, however, presents a complication. Although ITT estimates allow direct comparisons across assignment groups, they may conflate differences in potential growth with differential selection into Head Start participation. To illustrate, suppose there are only always-takers and compliers, so that all treatment-assigned children attend Head Start. Let a_g denote the share of always-takers in group g. Observed growth in the control-assigned group is a mixture of growth under participation and nonparticipation,

$$a_g \, \mathbb{E}[Y_K(1) - Y_P(1) \mid G = g] + (1 - a_g) \, \mathbb{E}[Y_K(0) - Y_P(0) \mid G = g] \,.$$

If group $g = 1$ has a weaker first stage because it contains a larger share of always-takers, its control-assigned group places greater weight on growth under participation. Suppose further that the two groups have identical potential growth under each participation state, but that children grow more slowly after Head Start participation than under nonparticipation, perhaps because kindergarten teachers devote greater attention to children who enter with lower skills. This selection-induced compositional difference can make the observed control group for $g = 1$ grow more slowly than the observed control group for $g = 0$, generating a spurious between-group difference in control-group growth even when the groups have identical potential growth under both participation states.

Table 5 confirms that this compositional concern is empirically relevant. I estimate first-stage effects separately for Black and non-Black children using the same specifications as in Table 3, replacing cognitive growth with Head Start participation as the outcome. Across specifications, assignment increases Head Start participation by 18–23 percentage points more for non-Black children than for observably similar Black children, documenting substantial differential selection into Head Start.

The weaker first stage for Black children is concentrated almost entirely on the control side. In Column (1), Black children assigned to the control group are 22 percentage points more likely to participate in Head Start than non-Black children, whereas the corresponding participation difference among treatment-assigned children is only 4 percentage points and statistically insignificant. Using this specification to predict take-up rates, Head Start participation is 16% for non-Black and 38% for Black children in the control group, compared with 87% and 91%, respectively, in the treatment group. Under monotonicity (no defiers) (Angrist and Imbens, 1995), these take-up rates imply that approximately 16% of non-Black children are always-takers (vs. 38% for Black children), 71% are compliers (vs. 53%), and 13% are never-takers (vs. 9%). Thus, Black children have a substantially larger always-taker share and a correspondingly smaller complier share, directly corroborating the compositional concern described above. Because this differential selection can distort the reduced-form decomposition of fadeout into treatment- and control-side growth, a more structured analysis is needed to separate selection from potential growth.

I therefore treat the ITT fadeout decomposition as a reduced-form description and baseline comparison point, and introduce a structural selection model in the next section to recover selection-corrected cognitive growth under Head Start participation and nonparticipation.[14]

[14] Although LATE accounts for imperfect take-up by identifying treatment effects for compliers and can, in principle, be decomposed into complier growth under participation and nonparticipation (Abadie, 2003), the group-specific 2SLS estimates with covariates in my setting do not necessarily correspond to average effects for naturally defined complier populations. Appendix C of Walters (2018) shows that, with a saturated

Table 6 presents the Black–non-Black ITT decomposition using the same growth outcome and specifications as Table 3. Across specifications, Black children experience 0.36–0.43 SD less fadeout than otherwise similar non-Black children, with differences in control-group growth accounting for approximately 79–84% of this differential. During kindergarten, non-Black children assigned to the control group experience 0.29–0.36 SD more cognitive growth than otherwise similar Black children assigned to the control group, and this difference is statistically significant across specifications. By contrast, the corresponding Black–non-Black growth difference among treatment-assigned children is only 0.07–0.08 SD and is statistically insignificant throughout.

The lower rows show a related pattern in terms of changes in treatment effects. The initial treatment gains for non-Black children decline significantly by 0.19–0.30 SD during kindergarten, whereas those for Black children increase by 0.08–0.18 SD, although the latter estimates are statistically imprecise. Taken together, the reduced-form evidence suggests that the Black–non-Black difference in fadeout arises primarily from divergent cognitive growth among control-assigned children, without yet accounting for differential selection into Head Start participation.

5 Two-Period Selection Model

This section develops a two-period selection model that jointly models Head Start participation and potential outcomes at the end of Head Start and the end of kindergarten. I use the model to account for differential selection into Head Start, reassess heterogeneity in fadeout by baseline disadvantage and center characteristics, and decompose fadeout differentials into growth under treatment and control regimes.

set of covariate-cell indicators, 2SLS identifies a weighted average of covariate-cell-specific LATEs, where the weights depend on cell prevalence, within-cell assignment variance, and the cell-specific first stage. The resulting pooled LATE therefore does not simply average effects according to the prevalence of compliers. Decomposing it into growth under participation and nonparticipation inherits the same weighting complication and does not provide a straightforward account of how each potential-growth trajectory contributes to the between-group differential.

5.1 Setup

Consider child i in site j. Head Start participation follows the threshold-crossing rule

$$D_{ij} = \mathbf{1}\left[\eta_{ij} < \lambda_j + P_{ij}'\psi_\lambda + \exp\left(\kappa_j + P_{ij}'\psi_\kappa\right) Z_{ij}\right],$$

where Z_{ij} is the randomized Head Start offer, P_{ij} contains the baseline disadvantage measures and center characteristics listed in Table A6, and η_{ij} captures unobserved resistance to participation. The coefficients ψ_λ and ψ_κ allow the baseline participation propensity and the offer effect, respectively, to vary systematically with observed child and center characteristics. The exponential parameterization preserves monotonicity by ensuring that an offer weakly increases participation, while λ_j and κ_j capture residual cross-site heterogeneity in the baseline participation propensity and the offer effect.

For treatment status $d \in \{0, 1\}$ and period $t \in \{P, K\}$, potential cognitive outcomes are

$$Y_{tij}(d) = \alpha_{dtj} + P_{ij}'\psi_{dt} + \varepsilon_{dtij},$$

where P and K denote the end of preschool and the end of kindergarten, respectively. The coefficients ψ_{dt} allow potential outcomes under each treatment state and in each period to vary systematically with the observed characteristics in P_{ij}, while α_{dtj} captures residual cross-site heterogeneity.

I assume that the individual-level participation and potential-outcome shocks are jointly normally distributed:

$$(\eta_{ij}, \varepsilon_{1Pij}, \varepsilon_{1Kij}, \varepsilon_{0Pij}, \varepsilon_{0Kij})' \sim \mathcal{N}(0, \Sigma),$$

with $\text{Var}(\eta_{ij}) = 1$ to normalize the scale of the participation equation. The covariance matrix Σ allows unobserved resistance to participation to be correlated with potential outcomes under both treatment states and in both periods, thereby accounting for selection on unobservables. It also allows potential-outcome shocks to be correlated across periods within each treatment state. Because cross-treatment-state covariances are not separately identified, I set them to zero.

I further assume that the site-specific parameters

$$\theta_j = (\alpha_{1Pj}, \alpha_{0Pj}, \alpha_{1Kj}, \alpha_{0Kj}, \lambda_j, \kappa_j)'$$

are jointly normally distributed: $\theta_j \sim \mathcal{N}(\theta_0, V_0)$. This specification allows for residual cross-site variation in participation and potential outcomes that is not explained by observed characteristics, while permitting these dimensions to covary across sites.

The two-period specification directly determines potential cognitive growth during kindergarten. For treatment status $d \in \{0, 1\}$, define

$$\begin{aligned} G_{ij}(d) &= Y_{Kij}(d) - Y_{Pij}(d) \\ &= (\alpha_{dKj} - \alpha_{dPj}) + P_{ij}' (\psi_{dK} - \psi_{dP}) + (\varepsilon_{dKij} - \varepsilon_{dPij}) . \end{aligned}$$

The treatment effect on kindergarten cognitive growth is therefore

$$\begin{aligned} \tau_{G,ij} &= G_{ij}(1) - G_{ij}(0) \\ &= P_{ij}' [(\psi_{1K} - \psi_{1P}) - (\psi_{0K} - \psi_{0P})] \\ &\quad + [(\alpha_{1Kj} - \alpha_{1Pj}) - (\alpha_{0Kj} - \alpha_{0Pj})] \\ &\quad + [(\varepsilon_{1Kij} - \varepsilon_{1Pij}) - (\varepsilon_{0Kij} - \varepsilon_{0Pij})] . \end{aligned}$$

Because fadeout corresponds to a decline in the treatment effect between the end of preschool and the end of kindergarten, a larger value of $\tau_{G,ij}$ indicates less fadeout.

For a given child or center characteristic p, let $\psi_{dt,p}$ denote its loading in the potential-outcome equation under treatment status d in period t. The selection-corrected fadeout differential associated with characteristic p is

$$\Delta_p^G = \underbrace{(\psi_{1K,p} - \psi_{1P,p})}_{\text{Treated-growth differential}} - \underbrace{(\psi_{0K,p} - \psi_{0P,p})}_{\text{Control-growth differential}} .$$

The first component captures how the characteristic predicts cognitive growth during kindergarten under Head Start participation, while the second captures the corresponding relationship under nonparticipation. A positive value of Δ_p^G indicates less fadeout for children or centers with that characteristic. These two components are the structural counterparts to the treated- and control-group growth differential in the reduced-form fadeout decomposition.

5.2 Estimation

I estimate the model by simulated maximum likelihood. Integrating over the distribution of the unobserved site-specific parameters, the weighted marginal log-likelihood is

$$\mathcal{L}(\Omega) = \sum_{j=1}^{J} \log \left[\int_{\mathbb{R}^6} \left\{ \prod_{i=1}^{N_j} L_{ij}(\theta_j)^{w_{ij}} \right\} \varphi_6(\theta_j; \theta_0, V_0) \; d\theta_j \right] ,$$

where $\Omega = (\theta_0, V_0, \Psi, \Sigma)$ collects the population-level parameters, Ψ collects the coefficients on observed characteristics, and the analysis weights w_{ij} are normalized to have mean one in the estimation sample. Each child's likelihood contribution combines the joint density of the two observed outcomes with the probability of the observed participation decision:

$$
\begin{aligned}
L_{ij}(\theta_j) = \Bigg[& \phi_2\left(\begin{bmatrix} Y_{Pij} - \alpha_{1Pj} - P_{ij}'\psi_{1P} \\ Y_{Kij} - \alpha_{1Kj} - P_{ij}'\psi_{1K} \end{bmatrix} ; 0, \Sigma_{11} \right) \\
& \times \Phi\left(\frac{\lambda_j + P_{ij}'\psi_\lambda + \exp\left(\kappa_j + P_{ij}'\psi_\kappa\right) Z_{ij} - \Sigma_{\eta 1}\Sigma_{11}^{-1} \begin{bmatrix} Y_{Pij} - \alpha_{1Pj} - P_{ij}'\psi_{1P} \\ Y_{Kij} - \alpha_{1Kj} - P_{ij}'\psi_{1K} \end{bmatrix}}{\sqrt{1 - \Sigma_{\eta 1}\Sigma_{11}^{-1}\Sigma_{1\eta}}} \right) \Bigg]^{D_{ij}} \\
\times \Bigg[& \phi_2\left(\begin{bmatrix} Y_{Pij} - \alpha_{0Pj} - P_{ij}'\psi_{0P} \\ Y_{Kij} - \alpha_{0Kj} - P_{ij}'\psi_{0K} \end{bmatrix} ; 0, \Sigma_{00} \right) \\
& \times \Phi\left(\frac{\Sigma_{\eta 0}\Sigma_{00}^{-1} \begin{bmatrix} Y_{Pij} - \alpha_{0Pj} - P_{ij}'\psi_{0P} \\ Y_{Kij} - \alpha_{0Kj} - P_{ij}'\psi_{0K} \end{bmatrix} - \lambda_j - P_{ij}'\psi_\lambda - \exp\left(\kappa_j + P_{ij}'\psi_\kappa\right) Z_{ij}}{\sqrt{1 - \Sigma_{\eta 0}\Sigma_{00}^{-1}\Sigma_{0\eta}}} \right) \Bigg]^{1-D_{ij}}.
\end{aligned}
$$

Here, Σ_{dd} denotes the covariance matrix of the two potential-outcome shocks under treatment state d, and $\Sigma_{\eta d}$ denotes their covariance with the participation shock. The function $\phi_2(\cdot; 0, \Sigma_{dd})$ is the corresponding bivariate normal density, while $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Because the six-dimensional site-level integral has no closed-form solution, I draw

$$
\theta_{js} = \theta_0 + C'\xi_{js}, \qquad \xi_{js} \sim \mathcal{N}(0, I_6), \qquad C'C = V_0,
$$

and maximize the simulated log-likelihood using 2,000 draws of the six-dimensional site-specific parameter vector for each site

$$
\widehat{\mathcal{L}}(\Omega) = \sum_{j=1}^{J} \log \left[\frac{1}{S} \sum_{s=1}^{S} \prod_{i=1}^{N_j} L_{ij}(\theta_{js})^{w_{ij}} \right].
$$

I compute standard errors using a sandwich covariance estimator constructed from site-level likelihood scores. This follows the structure of the marginal likelihood, which integrates over a common site-specific parameter vector and treats the resulting site-level likelihood contributions as independent across sites. Appendix C provides the full likelihood derivation and simulation procedure.

Because the model allows both potential outcomes and first-stage take-up to vary across sites, I require the corresponding site-specific quantities to be estimable. I therefore restrict the estimation sample to sites for which both the first stage and the treatment effect on kindergarten cognitive growth can be estimated.[15] After imposing this site restriction and requiring nonmissing cognitive outcomes in both periods, baseline disadvantage measures, and center characteristics, the resulting sample contains $1,150$ children across 149 sites, including individual centers and aggregated grantee-level sites. This sample represents 89% of the reduced-form analytic sample with valid two-period cognitive outcomes and complete baseline and center characteristics.

5.3 Model Fit and Estimates

Table A9 assesses model fit by comparing reduced-form moments with their model-implied counterparts in the structural-model sample. The first column reports the estimated first stage; the ITT effects on cognitive skills at the end of Head Start, at the end of kindergarten, and on growth during kindergarten; and the corresponding LATE estimates. The second column reports the average model-implied estimate across 1,000 simulated datasets, while the final column reports the difference between the model-implied and reduced-form estimates.[16] The small differences across all moments indicate that the model fits the reduced-form patterns well. Beyond matching these reduced-form moments, the model implies that observed child and center characteristics jointly explain approximately 68% of the cross-site variation in treatment-effect fadeout.

Table 7 presents the main results from the structural model.[17] I estimate how six base-

[15]I first estimate center-specific 2SLS regressions of cognitive growth on Head Start participation, instrumenting participation with randomized assignment. Centers are retained when both the first-stage and treatment-effect coefficients can be estimated. When a center is too small to identify these quantities separately, I aggregate small centers within the same grantee and re-estimate both coefficients, treating each retained grantee-level aggregate as a single site in the structural model. Because many centers have small samples, these screening regressions exclude baseline covariates to preserve degrees of freedom. A center or grantee-level aggregate is excluded only when either coefficient cannot be estimated or when its standard error is missing or zero.

[16]To construct the model-implied moments, I simulate 1,000 datasets from the estimated structural model while holding fixed the empirical site structure, randomized assignments, observed characteristics, and analysis weights. In each replication, I draw site-specific parameters from the estimated cross-site distribution and individual-level shocks from the estimated joint error distribution. I then generate Head Start participation and cognitive outcomes in both periods and re-estimate the same first-stage, ITT, and LATE specifications used in the observed data.

[17]Tables A10 and A11 report the model's core parameter estimates and the underlying covariate loadings.

line disadvantage indicators and five Head Start center characteristics are associated with selection-corrected Head Start effects on cognitive skill levels at the end of Head Start and the end of kindergarten, as well as with cognitive growth during kindergarten.

The main finding is that Black children experience significantly less fadeout. Conditional on the other observed baseline disadvantage measures and center characteristics, the treatment effect on cognitive growth during kindergarten is approximately 0.63 SD larger for Black children than for non-Black children, indicating 0.63 SD less fadeout. This difference is statistically significant at the 1% level and is consistent with the reduced-form result. In addition, the treatment effect on cognitive skills at the end of kindergarten is 0.45 SD larger for Black children relative to non-Black children. This estimate is statistically significant at the 5% level and also consistent with the reduced-form finding.

One additional pattern emerges from the structural analysis. Controlling for the other baseline disadvantage measures and Head Start features, children who are not living with two biological parents experience 0.24 SD less fadeout than children living with two biological parents ($p < 0.05$). This difference reflects a change in the relative Head Start effect from -0.10 SD at the end of Head Start to 0.13 SD at the end of kindergarten, although neither level difference is individually statistically significant. By contrast, the corresponding reduced-form fadeout differential is -0.10 SD, while the corresponding LATE estimate is -0.22 SD. Both estimates are statistically insignificant and point in the opposite direction from the structural estimate. This discrepancy likely reflects both the model's correction for differential selection into Head Start participation and its accompanying distributional assumptions. I consequently treat this additional finding as more suggestive than the robust Black–non-Black fadeout differential.

None of the remaining disadvantage indicators or center quality indicators significantly predicts differential fadeout. Smaller class size is associated with more favorable Head Start effects on cognitive skill levels at both time points relative to larger class size, a pattern consistent with a large causal literature showing that smaller classes can improve children's academic achievement and produce lasting effects on later outcomes (Krueger, 1999; Angrist and Lavy, 1999; Chetty et al., 2011). However, smaller class size is not significantly associated with less treatment-effect fadeout during kindergarten.[18]

[18]Following Walters (2015), I also test whether fadeout varies with the site-level share of compliers induced to attend Head Start rather than remain in home-based care. I find no significant heterogeneity along this dimension. This null result is perhaps unsurprising because the composition of preschool counterfactual care need not predict children's post-preschool growth trajectories. I do not report these estimates because this

Table 8 decomposes the fadeout differential associated with Black children into growth differentials under Head Start participation and nonparticipation. Conditional on the other baseline disadvantage measures and center characteristics, Black children experience an imprecisely estimated 0.16 SD more cognitive growth under Head Start participation than non-Black children. By contrast, Black children experience a precisely estimated 0.47 SD less cognitive growth under nonparticipation than non-Black children ($p < 0.01$). Given that the overall fadeout differential is 0.63 SD, approximately $\frac{0.47}{0.63} \approx 74\%$ of the selection-corrected Black–non-Black fadeout differential is attributable to differential cognitive growth under nonparticipation.

This 74% share represents a 10 percentage point reduction from the corresponding 84% implied by the ITT decomposition, which also jointly accounts for the other dimensions of baseline disadvantage and Head Start center characteristics. This reduction suggests that differential selection modestly overstates the contribution of control-side growth in the reduced-form decomposition. Nevertheless, the substantive conclusion remains largely unchanged: both approaches attribute the large majority of the Black–non-Black fadeout differential to differences in cognitive growth under Head Start nonparticipation rather than participation. This finding also aligns with an adjacent literature on charter-school impacts emphasizing that heterogeneity in counterfactual growth under nonattendance is central to understanding treatment-effect heterogeneity (Walters, 2018; Angrist et al., 2023).

Finally, I ask whether the growth disparity under Head Start nonparticipation reflects unusually slow growth among Black children or unusually rapid growth among non-Black children by comparing predicted growth across participation status and Black–non-Black groups. Estimates in Tables A10 and A11 predict that, relative to non-Black children under nonparticipation, Black children under nonparticipation experience 0.47 SD less predicted cognitive growth, Black participants experience 0.44 SD less predicted growth, and non-Black participants experience 0.60 SD less predicted growth. Thus, the other three groups exhibit relatively similar growth, whereas non-Black children under nonparticipation stand apart. This comparison points to unusually rapid cognitive growth among non-Black children under nonparticipation as the primary source of differential fadeout.

source of heterogeneity is less central to the paper, although it remains relevant to the literature on the determinants of Head Start's initial effectiveness in the HSIS.

6 Discussion

6.1 What Explains Differential Growth in the Control Group?

The structural analysis above shows that Black and non-Black children follow genuinely different cognitive-growth trajectories during kindergarten even after accounting for differential selection into Head Start participation. I next examine two hypotheses that may explain why both Black and non-Black control children enter kindergarten trailing their treatment counterparts, but only non-Black control children subsequently grow rapidly and catch up, whereas Black control children maintain a relatively stable skill gap with their treatment counterparts through the end of kindergarten.

The first hypothesis is differential family compensation. Families may respond to relatively low skills at kindergarten entry by increasing investments in their children, and this response may be stronger among non-Black families. This hypothesis generates three testable predictions. First, non-Black control families should make greater investments during kindergarten than non-Black treatment families, because their children enter kindergarten behind but subsequently catch up. Second, family investments should differ much less between Black control and treatment families, because Black control children do not exhibit comparable catch-up. Third, if differential family compensation contributes to these divergent growth trajectories, differences in family investments should account for part of both the Black–non-Black control-group growth differential and the resulting fadeout differential.

The second hypothesis is differential school compensation. Non-Black control children may enter kindergarten environments that are better able to provide individualized or compensatory instruction to children who enter school behind, whereas Black control children may enter environments with fewer resources or less capacity to provide such support. Unlike the family-compensation hypothesis, for which I observe investments directed toward the individual child, the school data primarily describe the kindergarten environment rather than child-specific compensatory inputs.

This distinction makes symmetric treatment–control predictions, as in the family-compensation hypothesis, less appropriate for the school-compensation hypothesis. In particular, the same school environment may generate greater gains for children who enter kindergarten behind. Non-Black control children therefore need not attend more compensatory school environments than non-Black treatment children in order to catch up. The key prediction is instead that non-Black control children should be exposed to kindergarten environments with greater capacity for compensatory instruction than Black control children. Further, if differential school compensation contributes to their divergent growth trajectories, differences in school environments should account for part of both the Black–non-Black control-group growth differential and the resulting fadeout differential.

To test both sets of hypotheses, I first compare a set of family and school inputs measured during kindergarten to assess whether the level differences implied by the two hypotheses are present in the data. I then conduct a linear mediation analysis to examine whether these observed inputs account for the Black–non-Black control-group growth differential and the resulting fadeout differential.[19] Specifically, I estimate

$$G_i = \beta_0 + \beta_1 \text{Black}_i + \beta_2 Z_i + \beta_3 \left(\text{Black}_i \times Z_i\right) + X_i' \gamma + M_i' \delta + \left(M_i \times Z_i\right)' \kappa + \lambda_{g(i)} + \varepsilon_i,$$

where G_i denotes cognitive skill growth during kindergarten, Z_i indicates assignment to Head Start, X_i contains the baseline covariates used in the ITT analysis, and $\lambda_{g(i)}$ denotes grantee fixed effects. The vector M_i contains the potential explanatory variables. I interact each explanatory variable with Head Start assignment, allowing its association with cognitive growth to differ between treatment- and control-assigned children. The coefficient β_1 captures the adjusted Black–non-Black growth differential among control-assigned children, while β_3 captures the corresponding Black–non-Black fadeout differential.[20]

Tables A12 and A13 describe the family and school inputs included in the mediation analysis. The family measures primarily capture home investments, parenting practices and styles, and constraints on parental time and stability, while the school measures capture the school-level academic environment, peer disadvantage and behavior, teacher quality and resources, instructional practices and responsiveness to children's developmental needs, and teachers' beliefs about child-centered teaching.

[19] I return to the reduced-form linear specification for the mechanism analysis for two reasons. First, the selection model is not naturally suited to incorporating post-treatment family and school measures as mediators, and doing so would be computationally costly. Second, the reduced-form and selection-corrected fadeout analyses yield very similar qualitative conclusions. I therefore use the more transparent linear specification for the descriptive mediation analysis.

[20] Although Head Start center fixed effects would permit a more localized comparison of Black and non-Black children within the same center, there is limited within-center racial overlap in the HSIS. Among Black children, the median leave-one-out share of Black peers in the center is approximately 0.92, compared with 0.00 among non-Black children. Consequently, specifications based exclusively on within-center racial variation have limited common support and substantially less statistical power.

Table 9 restricts the sample to 656 children with nonmissing family and school measures. Panel A compares family inputs between treatment- and control-assigned children separately for Black and non-Black children. The results provide little support for the differential family-compensation hypothesis. Non-Black control families do not report systematically greater investments than non-Black treatment families. If anything, non-Black treatment families report marginally greater cultural enrichment than non-Black control families. In addition, Black treatment families are significantly more likely than Black control families to report reading to their children every day in the past week ($p < 0.05$). Both findings run counter to the hypothesis that differential family investment is the key driver of the divergent cognitive growth trajectories during kindergarten.

Panel B compares kindergarten school inputs between Black and non-Black children within the control group and provides some support for differential school compensation. Among control-assigned children, Black children attend schools with substantially lower reading and mathematics proficiency (0.32 SD, $p < 0.01$) and spend about 10 percentage points less class time in adult-directed individual activities than non-Black children ($p < 0.05$). This difference is large relative to the non-Black control-group mean, for whom approximately 20 percent of class time is devoted to such activities. Because activity time is compositional, the lower share of adult-directed individual activities implies that Black control children spend correspondingly more time in other instructional formats, including child-chosen activities without adult direction and group-based activities.

Table 10 then uses a Gelbach decomposition to quantify how much the observed family and school mediators account for the Black–non-Black control-group growth differential and the resulting fadeout differential (Gelbach, 2016). Results in Panel B show that kindergarten school inputs together account for approximately 32% of the control-group growth differential and 17% of the overall Black–non-Black fadeout differential. Consistent with the level differences reported above, school reading and mathematics proficiency and time spent in adult-directed individual activities generate the largest attenuation of the control-group growth differential, accounting for approximately 14% and 10% of the differential, respectively.

Panel A, by contrast, shows that conditioning on the observed family inputs enlarges the Black–non-Black control-group growth differential by approximately 33% and the fadeout differential by approximately 5%. This pattern is difficult to interpret substantively given the observed family measures available in the data. As shown in Panel A of Table 9, Black and non-Black control families do not exhibit a systematic pattern of differences across these inputs, and several measures may themselves reflect endogenous responses to child behavior or proxy for other unobserved family characteristics.

Taken together, the mediation analysis provides suggestive evidence that differential school compensation may account for part of the divergent cognitive growth of Black and non-Black control children and the resulting fadeout differential. In particular, non-Black control children enter kindergarten environments with higher school achievement levels and spend more time in adult-directed individual activities, patterns consistent with environments that may be more conducive to catch-up. In contrast, I find little evidence in support of differential family compensation. Two limitations of this analysis warrant caution. First, substantial missingness in the family and kindergarten school measures limits the generalizability of the results. Second, the observed kindergarten school inputs account for only a portion of the control-group growth differential and the resulting fadeout differential.[21] More detailed data on children's within-classroom experiences would be needed to account for the remaining differential and to more directly identify the mechanisms underlying divergent skill growth during kindergarten.

6.2 Reallocating Head Start Offers to Reduce Fadeout

The main finding that Black children experience less fadeout and retain more persistent cognitive gains raises a natural policy question: holding the total number of Head Start slots fixed, how much could aggregate fadeout be reduced by reallocating offers toward children whose initial treatment gains are more persistent?

Given that Black applicants are concentrated in centers serving larger shares of Black children and that centers are the unit through which recruitment occurs, I consider a series of counterfactual policies that redirect Head Start offers toward these centers to expand access for Black children. I classify centers that received no applications from Black children as source centers and centers in which at least half of applicants are Black as recipient centers. I then simulate policies that transfer offers from source centers to recipient centers while

[21] Consistent with this limitation, the estimated effect of Head Start assignment on kindergarten cognitive growth among non-Black children in the common mechanism-analysis sample is -0.29 SD ($p < 0.01$). Conditioning on the observed family and kindergarten school measures attenuates this estimate only modestly, to -0.25 SD ($p < 0.01$), indicating that these observed inputs account for only a limited share of the unusually rapid catch-up among non-Black control children.

holding the total number of offers fixed. Although these counterfactuals are intended to reallocate resources without increasing the overall scale of the Head Start program, they do not hold total expenditure exactly fixed because the federal cost of funding a Head Start offer varies across locations and I do not observe center-specific costs.

Under each policy, I randomly select offers from the pooled set of offers across all source centers, so that the number removed from a particular source center is proportional to its initial number of offers in expectation. I then distribute the transferred offers across recipient centers in proportion to the number of applicants who did not initially receive an offer. Given the resulting counterfactual number of offers in each source and recipient center, I rerandomize all offers within each center among its applicants.

Under the observed HSIS assignment, the 94 source centers contain 709 applicants and 448 offers, while the 31 recipient centers contain 219 applicants and 139 offers. I consider three counterfactual reallocations that transfer 10% of source-center offers (45 offers), 15% of source-center offers (67 offers), or enough offers to allow every applicant in recipient centers to receive an offer (80 offers, or approximately 18% of all source-center offers). I evaluate each policy using the posterior distribution implied by the estimated structural model. For each center, I construct 2,000 candidate draws of the center-specific parameter vector and assign posterior weights according to how well each draw fits the center's observed data.[22] I then conduct 1,000 outer simulations, drawing new participation and outcome shocks for every child and redrawing the offer allocation under each policy. Within each simulation, I average outcomes across the 2,000 parameter draws using center-specific posterior weights and then aggregate across centers using the HSIS sampling weights. The reported estimates are the means of the resulting 1,000 policy moments, with the 5th and 95th percentiles summarizing their distribution.

Table 11 shows that the counterfactual policies substantially increase the share of Black applicants receiving Head Start offers but reduce aggregate fadeout only modestly. As more offers are reallocated from source centers to recipient centers, the Black offer rate rises monotonically from 0.58 to 0.96. However, even the most aggressive policy, which provides an offer to every applicant in recipient centers, reduces aggregate fadeout only from -0.28 SD to -0.25 SD, an attenuation of approximately 9.64%. The limited aggregate effect in part reflects the fact that Black children constitute only a minority of the national Head Start applicant population, so even large increases in their offer rate translate into relatively

[22] See Appendix C.3 for details on the construction of the posterior weights.

modest changes in the composition of children receiving offers overall. Thus, reallocating access toward centers serving predominantly Black applicants can modestly improve the persistence of Head Start impacts, but cannot eliminate most aggregate fadeout without broader changes in the population recruited.

These counterfactual exercises warrant a few qualifications. First, they do not account for capacity constraints at recipient centers, such as limited classroom space or teacher availability. If these constraints are already binding, recipient centers may be unable to accommodate additional children without expanding capacity or reducing program quality. Second, the model does not capture general equilibrium effects, including changes in children's subsequent learning environments. For example, as source centers lose offers, local kindergartens may receive more children without prior Head Start participation, requiring teachers to divide their fixed attention among more students entering behind and potentially slowing children's catch-up. Third, explicitly race-based resource reallocation may be politically difficult to implement. A more feasible place-based policy could instead target centers located in socioeconomically disadvantaged neighborhoods, which may approximate the concentration of Black applicants without directly conditioning resource allocation on race.

7 Conclusion

Fadeout in the average treatment effect can mask substantial heterogeneity across participants. In the context of early childhood development, understanding this heterogeneity is important not only for characterizing different trajectories of human capital formation following an early childhood intervention, but also for informing policies designed to strengthen the durability of initial gains.

This paper studies heterogeneity in the fadeout of the initial cognitive gains generated by access to Head Start preschool. I find that Black children's initial gains fade out significantly less than those of non-Black children by the end of kindergarten. A selection-corrected model attributes this differential fadeout primarily to differential cognitive growth under Head Start nonparticipation. I further show that Black participants, Black nonparticipants, and non-Black participants experience relatively similar growth during kindergarten, whereas non-Black nonparticipants stand apart and grow more rapidly during kindergarten. As a result, the initial Head Start effect reverses sign among non-Black children by the end of kindergarten, while Black children retain a positive and marginally significant cognitive gain.

I then test whether differential family and school compensation may explain the divergent growth trajectories experienced by Black and non-Black children. Using a smaller complete-case sample, I find that observed kindergarten school environments account for part of the Black–non-Black control-side growth differential and the resulting fadeout differential. The largest contributions come from differences in school reading and mathematics proficiency and time spent in adult-directed individual activities, patterns consistent with a differential school-compensation mechanism in which non-Black control children enter kindergarten environments with greater capacity to support catch-up than Black control children. In contrast, I find little evidence that non-Black control children receive greater family investments during kindergarten that could account for their more rapid catch-up.

Finally, given that Black children experience more persistent Head Start gains, I conduct a series of partial-equilibrium policy simulations that reallocate existing Head Start offers toward centers serving predominantly Black applicants and examine whether such reallocation attenuates aggregate fadeout. The simulations substantially increase Head Start access among Black children but reduce aggregate fadeout only modestly. Because Black children constitute a relatively small share of Head Start applicants nationally, even large increases in their offer rate affect too small a share of the overall applicant population to substantially improve average persistence. I therefore conclude that reallocating a fixed number of existing offers has limited capacity to reduce aggregate fadeout; larger improvements would likely require expanding program capacity and extending access to populations for whom Head Start gains are more persistent.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. Journal of econometrics 113(2), 231–263.

Abdulkadiroğlu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020). Do parents value school effectiveness? American Economic Review 110(5), 1502–1539.

Abdulkadiroglu, A., P. A. Pathak, and C. R. Walters (2026). Who gets what in education: Can school matching improve student achievement? Technical report, National Bureau of Economic Research.

Administration for Children and Families (2024). History of head start. https://acf.gov/ohs/about/history-head-start. U.S. Department of Health & Human Services. Accessed February 23, 2026.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. Journal of Economic Literature 56(4), 1360–1446.

Angrist, J., P. Hull, and C. Walters (2023). Methods for measuring school effectiveness. Handbook of the Economics of Education 7, 1–60.

Angrist, J. and G. Imbens (1995). Identification and estimation of local average treatment effects.

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. The Quarterly journal of economics 114(2), 533–575.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. American Economic Review 110(1), 48–85.

Bailey, D. H., J. M. Jenkins, and D. Alvarez-Vargas (2020). Complementarities between early educational intervention and later educational quality? a systematic review of the sustaining environments hypothesis. Developmental Review 56, 100910.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of political economy 70(5, Part 2), 9–49.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. Journal of political economy 75(4, Part 1), 352–365.

Burchinal, M., A. Whitaker, J. Jenkins, D. Bailey, T. Watts, G. Duncan, and E. Hart (2024). Unsettled science on longer-run effects of early education. Science 384(6695), 506–508.

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

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

Campos, C. and C. Kearns (2024). The impact of public school choice: Evidence from los angeles's zones of choice. The Quarterly Journal of Economics 139(2), 1051–1093.

Carneiro, P. and R. Ginja (2014). Long-term impacts of compensatory preschool on health and behavior: Evidence from head start. American Economic Journal: Economic Policy 6(4), 135–173.

Cascio, E. U. and D. O. Staiger (2012). Knowledge, tests, and fadeout in educational interventions. Technical report, National Bureau of Economic Research.

Castillo, M., J. A. List, R. Petrie, and A. Samek (2020). Detecting drivers of behavior at an early age: Evidence from a longitudinal field experiment. Technical report, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? evidence from project star. The Quarterly journal of economics 126(4), 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. American economic review 104(9), 2633–2679.

Chib, S. and L. Jacobi (2007). Modeling and calculating the effect of treatment at baseline from panel outcomes. Journal of Econometrics 140(2), 781–801.

Choi, J. Y., J. Elicker, S. L. Christ, and J. Dobbs-Oates (2016). Predicting growth trajectories in early academic learning: Evidence from growth curve modeling with head start children. Early Childhood Research Quarterly 36, 244–258.

City of Chicago, Office of Budget and Management (2023). 2024 budget overview. https://www.chicago.gov/content/dam/city/depts/obm/supp_info/2024Budget/2024-Budget-Overview_CityofChicago.pdf. City of Chicago FY2024 Adopted Budget. Accessed February 24, 2026.

Corradini, V. (2024). Information and access in school choice systems: Evidence from new york city. Unpublished working paper, University of Wisconsin.

Cunha, F. and J. Heckman (2007). The technology of skill formation. American economic review 97(2), 31–47.

Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. Journal of human resources 43(4), 738–782.

Currie, J. and D. Thomas (1995). Does head start make a difference? The American Economic Review 85(3), 341–364.

Currie, J. and D. Thomas (2000). School quality and the longer-term effects of head start. Journal of Human Resources 35(4), 755–774.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. American Economic Journal: Applied Economics 1(3), 111–134.

Duncan, G. J. and K. Magnuson (2013). Investing in preschool programs. Journal of economic perspectives 27(2), 109–132.

Efron, B. (2012). Large-scale inference: empirical Bayes methods for estimation, testing, and prediction, Volume 1. Cambridge University Press.

Farkas, G. and K. Beron (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. Social Science Research 33(3), 464–497.

Figlio, D. and S. Loeb (2011). School accountability. Handbook of the Economics of Education 3, 383–421.

Garces, E., D. Thomas, and J. Currie (2002). Longer-term effects of head start. American economic review 92(4), 999–1012.

García, J. L. and J. J. Heckman (2023). Parenting promotes social mobility within and across generations. Annual Review of Economics 15, 349–388.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the life-cycle benefits of an influential early-childhood program. Journal of Political Economy 128(7), 2502–2541.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? Journal of labor economics 34(2), 509–543.

Gelber, A. and A. Isen (2013). Children's schooling and parents' behavior: Evidence from the head start impact study. Journal of public Economics 101, 25–38.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2023). The long-term effects of universal preschool in boston. The Quarterly Journal of Economics 138(1), 363–411.

Gu, J. and R. Koenker (2023). Invidious comparisons: Ranking and selection as compound decisions. Econometrica 91(1), 1–41.

Hart, E. R., D. H. Bailey, S. Luo, P. Sengupta, and T. W. Watts (2024). Fadeout and persistence of intervention impacts on social–emotional and cognitive skills in children and adolescents: A meta-analytic review of randomized controlled trials. Psychological Bulletin 150(10), 1207.

Hart, E. R., C. M. Botvin, D. H. Bailey, and T. W. Watts (2026). Using experimental variation to examine the (co-) development of cognitive and social-emotional skills in early childhood. edworkingpaper no. 26-1369. Annenberg Institute for School Reform at Brown University.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. American economic review 103(6), 2052–2086.

Heckman, J. J. (1979). Sample selection bias as a specification error. Econometrica: Journal of the econometric society, 153–161.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. Journal of Economic literature 48(2), 356–398.

Heckman, J. J. and E. Vytlacil (2001). Policy-relevant treatment effects. American Economic Review 91(2), 107–111.

Heckman, J. J. and J. Zhou (2026). A study of the microdynamics of early-childhood learning. Journal of Political Economy 134(1), 49–85.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. Journal of public Economics 89(5-6), 761–796.

Jacobi, L., H. Wagner, and S. Frühwirth-Schnatter (2016). Bayesian treatment effects models with variable selection for panel outcomes with an application to earnings effects of maternity leave. Journal of Econometrics 193(1), 234–250.

Kaestner, R. and C. Schiman (2026). Do preschool investments depreciate and fadeout or result in dynamic complementarities? an assessment using the head start impact study. Review of Economics of the Household 24(1), 61–97.

Karoly, L. A., J. S. Cannon, C. J. Gomez, and A. A. Whitaker (2021). High-quality publicly funded pre-kindergarten programs: How much do they cost? research brief. rb-a252-1. RAND Corporation.

Kline, P., E. K. Rose, and C. R. Walters (2024). A discrimination report card. American Economic Review 114(8), 2472–2525.

Kline, P. and C. R. Walters (2016). Evaluating public programs with close substitutes: The case of head start. The Quarterly Journal of Economics 131(4), 1795–1848.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. Econometrica 75(1), 83–119.

Knudsen, E. I., J. J. Heckman, J. L. Cameron, and J. P. Shonkoff (2006). Economic, neurobiological, and behavioral perspectives on building america's future workforce. Proceedings of the national Academy of Sciences 103(27), 10155–10162.

Krueger, A. B. (1999). Experimental estimates of education production functions. The quarterly journal of economics 114(2), 497–532.

Li, W., G. J. Duncan, K. Magnuson, H. S. Schindler, H. Yoshikawa, and J. Leak (2020). Timing in early childhood education: How cognitive and achievement program impacts vary by starting age, program duration, and time since the end of the program. edworkingpaper no. 20-201. Annenberg Institute for School Reform at Brown University.

Lipsey, M. W., D. C. Farran, and K. Durkin (2018). Effects of the tennessee prekindergarten program on children's achievement and behavior through third grade. Early Childhood Research Quarterly 45, 155–176.

List, J. A. and H. Uchida (2024). Here today, gone tomorrow? toward an understanding of fade-out in early childhood education programs. NBER Working Paper 33027, National Bureau of Economic Research.

Ludwig, J. and D. 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.

Ludwig, J. and D. Phillips (2007). The benefits and costs of head start and commentaries. Child Policy Nexus 21(3), 1–20.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. Econometrica 86(5), 1589–1619.

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. Journal of the American statistical Association 78(381), 47–55.

Neal, D. and D. W. Schanzenbach (2010). Left behind by design: Proficiency counts and test-based accountability. The Review of Economics and Statistics 92(2), 263–283.

Nores, M., C. R. Belfield, W. S. Barnett, and L. Schweinhart (2005). Updating the economic impacts of the high/scope perry preschool program. Educational Evaluation and Policy Analysis 27(3), 245–261.

Office of Head Start (2024a). Head start program facts: Fiscal year 2024. https://headstart.gov/program-data/article/head-start-program-facts-fiscal-year-2024. Administration for Children and Families, U.S. Department of Health & Human Services. Accessed February 23, 2026.

Office of Head Start (2024b). Sec. 640 allotment of funds limitations on assistance. https://headstart.gov/policy/sec-640-allotment-funds-limitations-assistance. Head Start Act, 42 U.S.C. 9835. Accessed February 24, 2026.

Office of Head Start (2024c). Sec. 645 participation in head start programs. https://headstart.gov/policy/head-start-act/

sec-645-participation-head-start-programs. Head Start Act, 42 U.S.C. 9836. Accessed February 24, 2026.

Office of Head Start (2025). Head start. https://acf.gov/ohs/about/head-start. Administration for Children and Families, U.S. Department of Health & Human Services. Accessed February 24, 2026.

Puma, M., S. Bell, R. Cook, C. Heid, P. Broene, F. Jenkins, A. Mashburn, and J. Downer (2012). Third grade follow-up to the head start impact study: Final report. opre report 2012-45. Administration for Children & Families.

Puma, M., S. Bell, R. Cook, C. Heid, G. Shapiro, P. Broene, F. Jenkins, P. Fletcher, L. Quinn, J. Friedman, et al. (2010a). Head start impact study. final report. Administration for children & families.

Puma, M., S. Bell, R. Cook, C. Heid, G. Shapiro, P. Broene, F. Jenkins, P. Fletcher, L. Quinn, J. Friedman, et al. (2010b). Head start impact study. technical report. Administration for Children & Families.

Raudenbush, S. W., S. F. Reardon, and T. Nomi (2012). Statistical analysis for multisite trials using instrumental variables with random coefficients. Journal of research on Educational Effectiveness 5(3), 303–332.

Rosengarten, M. L., E. R. Hart, D. H. Bailey, M. P. McCormick, B. J. Lovett, and T. W. Watts (2024). Using meta-analytic data to examine fadeout and persistence of intervention impacts on constrained and unconstrained skills. edworkingpaper no. 24-1069. Annenberg Institute for School Reform at Brown University.

Stanford, L. (2023, January). Which states offer universal pre-k? it's more complicated than you might think. *Education Week*. Published January 25, 2023; accessed July 13, 2026.

Wagner, H., S. Frühwirth-Schnatter, and L. Jacobi (2023). Factor-augmented bayesian treatment effects models for panel outcomes. Econometrics and Statistics 28, 63–80.

Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from head start. American Economic Journal: Applied Economics 7(4), 76–102.

Walters, C. R. (2018). The demand for effective charter schools. Journal of Political Economy 126(6), 2179–2223.

Watts, T. W., C. M. Botvin, D. H. Bailey, E. R. Hart, S. Mattera, D. H. Clements, J. Sarama, D. C. Farran, and M. W. Lipsey (2026). Predicting persistence and fadeout across multi-site rcts of an early childhood mathematics curriculum intervention. Educational Evaluation and Policy Analysis, 01623737251405258.

Watts, T. W., E. R. Hart, and D. H. Bailey (2025). How general is educational intervention fadeout? a meta-analysis of educational rcts with follow-up. edworkingpaper no. 25-1366. Annenberg Institute for School Reform at Brown University.

Watts, T. W., J. M. Jenkins, K. A. Dodge, R. C. Carr, M. Sauval, Y. Bai, M. Escueta, J. Duer, H. Ladd, C. Muschkin, et al. (2023). Understanding heterogeneity in the impact of public preschool programs. Monographs of the Society for Research in Child Development 88(1), 7–182.

Weiland, C., R. Unterman, A. Shapiro, S. Staszak, S. Rochester, and E. Martin (2020). The effects of enrolling in oversubscribed prekindergarten programs through third grade. Child Development 91(5), 1401–1422.

Weiland, C. and H. Yoshikawa (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. Child development 84(6), 2112–2130.

Whitaker, A. A., M. Burchinal, J. M. Jenkins, D. H. Bailey, T. W. Watts, G. J. Duncan, E. R. Hart, and E. Peisner-Feinberg (2026). Why are preschool programs becoming less effective? Journal of Policy Analysis and Management 45(1), e70031.

Yau, L. H. and R. J. Little (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. Journal of the American Statistical Association 96(456), 1232–1244.

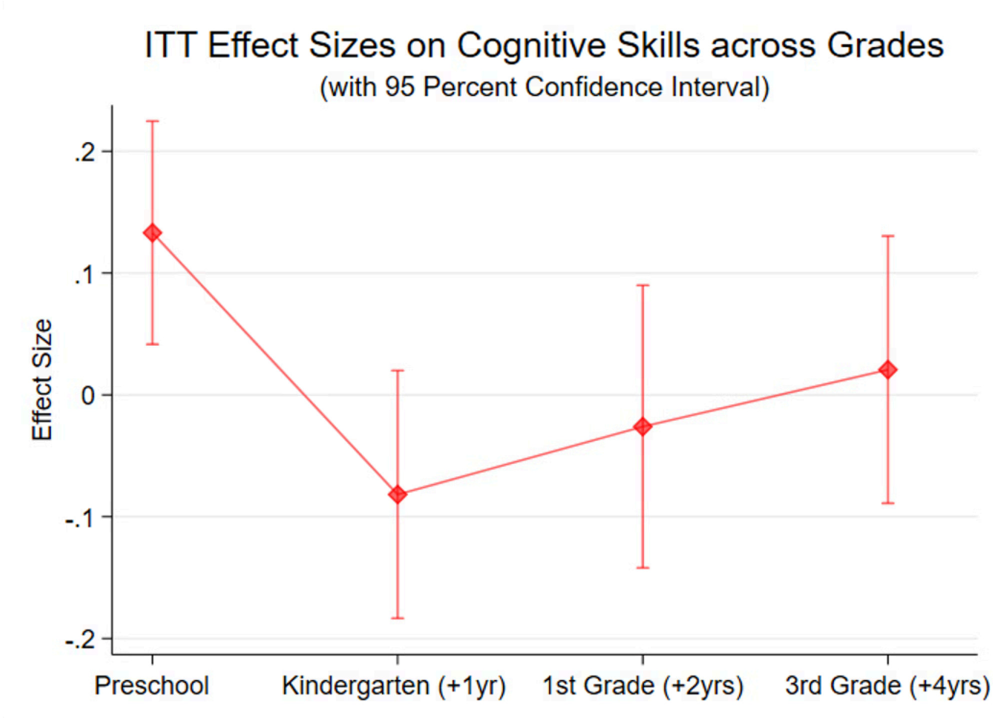
Table 1: HSIS Data Collection Timeline by Age Cohort

	3-Year-Old Cohort	4-Year-Old Cohort
Fall 2002	Baseline	Baseline
Spring 2003	End of Preschool Year 1	End of Preschool
Spring 2004	End of Preschool	End of Kindergarten
Spring 2005	End of Kindergarten	End of Grade 1
Spring 2006	End of Grade 1	—
Spring 2007	—	End of Grade 3
Spring 2008	End of Grade 3	—

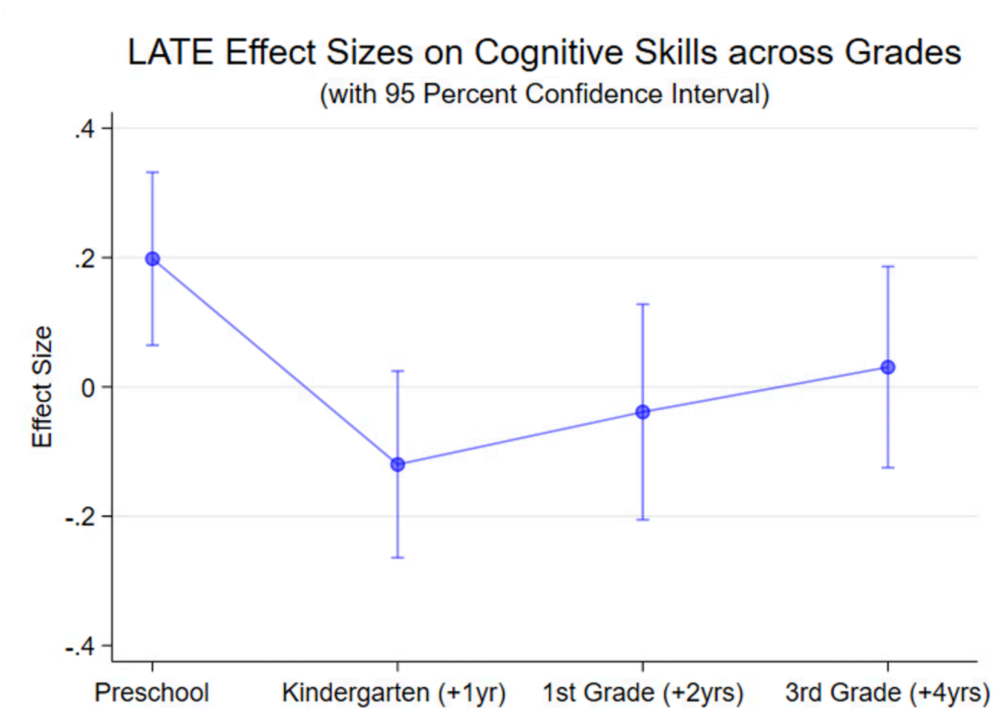
Table 2: Baseline Characteristics and Balance Tests

	Total (N = 1,993)	Control (N = 811)	Treatment (N = 1,182)	Difference
Child Characteristics				
Female	0.49 (0.50)	0.50 (0.50)	0.49 (0.50)	0.01
White	0.32 (0.47)	0.31 (0.46)	0.33 (0.47)	-0.01
Black	0.18 (0.38)	0.18 (0.38)	0.19 (0.39)	-0.01
Hispanic	0.50 (0.50)	0.51 (0.50)	0.49 (0.50)	0.02
Special needs	0.13 (0.33)	0.12 (0.33)	0.13 (0.34)	-0.01
PPVT	279.23 (37.46)	280.63 (35.83)	277.86 (38.97)	2.77
WJ Letter–Word	327.52 (30.03)	326.59 (31.98)	328.43 (27.98)	-1.84
WJ Applied	393.62 (23.66)	393.46 (24.12)	393.78 (23.21)	-0.32
Mother Characteristics				
Age	29.03 (6.75)	28.92 (6.58)	29.15 (6.91)	-0.23
Less than high school	0.43 (0.49)	0.43 (0.50)	0.42 (0.49)	0.01
High school graduate	0.32 (0.47)	0.33 (0.47)	0.30 (0.46)	0.03
More than high school	0.26 (0.44)	0.23 (0.42)	0.28 (0.45)	-0.04
Married	0.49 (0.50)	0.50 (0.50)	0.49 (0.50)	0.01
Immigrant	0.27 (0.45)	0.27 (0.45)	0.27 (0.45)	-0.00
Teen mother	0.16 (0.36)	0.16 (0.37)	0.16 (0.36)	0.00
Home Environment				
Two biological parents	0.55 (0.50)	0.56 (0.50)	0.54 (0.50)	0.02
Spanish spoken at home	0.40 (0.49)	0.40 (0.49)	0.39 (0.49)	0.01
Moderate/high home risk	0.21 (0.41)	0.19 (0.39)	0.23 (0.42)	-0.04^{*}

Notes: Entries report weighted baseline means, with weighted standard deviations in parentheses, for the full four-year-old cohort and separately for the control and treatment groups. The difference is calculated as the control-group mean minus the treatment-group mean. The full baseline sample contains 1,993 children, including 811 control-group children and 1,182 treatment-group children. Baseline cognitive assessments are available for 1,977 children, including 803 control-group children and 1,174 treatment-group children. Baseline measures are drawn from child assessments and parent interviews. HSIS baseline sampling weights are applied. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.



(a) ITT



(b) LATE

Figure 1: ITT and LATE of Head Start on the cognitive composite across grades

Table 3: Heterogeneity in Kindergarten Fadeout by Disadvantage

	(1) Separate	(2) Joint	(3) Joint w/ center features
Baseline Disadvantage			
Child Black	0.38***	0.36**	0.43***
	(0.13)	(0.15)	(0.16)
Child Hispanic	−0.19*	−0.04	−0.09
	(0.11)	(0.13)	(0.13)
Low baseline cognition	−0.03	−0.08	−0.09
	(0.13)	(0.13)	(0.14)
Mother below high school	−0.18	−0.13	−0.08
	(0.11)	(0.12)	(0.13)
Not with two biological parents	0.07	−0.02	−0.10
	(0.12)	(0.12)	(0.12)
Moderate/high home risk	0.03	−0.01	0.00
	(0.14)	(0.15)	(0.16)
Head Start Center Characteristics			
High share BA/licensed teachers			0.05
			(0.13)
Full-day services			0.11
			(0.12)
High program services			−0.03
			(0.13)
Experienced director			−0.09
			(0.14)
Small class size			0.02
			(0.13)
Disadvantage interactions estimated jointly	No	Yes	Yes
Center-feature interactions included	No	No	Yes
Observations	1,439	1,439	1,292

Notes: Entries report treatment-by-characteristic interactions from regressions of cognitive skill growth between the end of Head Start and the end of kindergarten. Column (1) estimates each interaction separately, Column (2) includes all disadvantage interactions jointly, and Column (3) additionally includes center-characteristic interactions. All specifications include baseline covariates, grantee fixed effects, and sampling weights. Robust standard errors are reported in parentheses. Positive coefficients indicate less fadeout. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: ITT Effects for Black and Non-Black Children

	(1) Separate	(2) Joint	(3) Joint w/ center features
Panel A: End of Head Start			
Non-Black effect	0.16***	0.05	−0.01
	(0.06)	(0.09)	(0.11)
Black effect	0.15*	0.12	−0.01
	(0.09)	(0.13)	(0.17)
Black–non-Black difference	−0.01	0.07	0.00
	(0.11)	(0.11)	(0.12)
Panel B: End of Kindergarten			
Non-Black effect	−0.14**	−0.14	−0.26*
	(0.06)	(0.10)	(0.13)
Black effect	0.23*	0.30*	0.18
	(0.13)	(0.16)	(0.19)
Black–non-Black difference	0.37***	0.44***	0.43***
	(0.14)	(0.15)	(0.16)
Observations	1,439	1,439	1,292

Notes: The table reports ITT effects on cognitive skills for non-Black and Black children and the corresponding Black–non-Black difference. Panels A and B use the same sample of children observed at both the end of Head Start and the end of kindergarten. Outcomes are measured in units of the baseline control-group standard deviation. Column (1) estimates the interaction with Black separately, Column (2) jointly includes interactions with all disadvantage indicators, and Column (3) additionally includes interactions with Head Start center characteristics. All specifications include baseline covariates, grantee fixed effects, and sampling weights. Robust standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: Decomposing Differential Head Start Participation for Black Children

	(1) Separate	(2) Joint	(3) Joint w/ center features
Black_i	0.22^{***}	0.25^{***}	0.23^{***}
(Control-group participation gap)	(0.07)	(0.07)	(0.08)
$\text{Black}_i + Z_i \times \text{Black}_i$	0.04	0.02	0.05
(Treatment-group participation gap)	(0.05)	(0.05)	(0.06)
$Z_i \times \text{Black}_i$	-0.18^{**}	-0.23^{***}	-0.18^{**}
(Black–non-Black first-stage differential)	(0.07)	(0.08)	(0.09)
Z_i	0.71^{***}	0.75^{***}	0.82^{***}
(Non-Black first stage)	(0.03)	(0.05)	(0.06)
$Z_i + Z_i \times \text{Black}_i$	0.53^{***}	0.52^{***}	0.64^{***}
(Black first stage)	(0.07)	(0.08)	(0.10)
Disadvantage interactions estimated jointly	No	Yes	Yes
Center-feature interactions included	No	No	Yes
Observations	1,439	1,439	1,292

Notes: The table decomposes the Black–non-Black difference in the effect of Head Start assignment on Head Start participation into participation gaps among control- and treatment-assigned children. Column (1) estimates the Black interaction separately, Column (2) includes all disadvantage interactions jointly, and Column (3) additionally includes center-feature interactions. All specifications include baseline covariates, grantee fixed effects, and sampling weights. Robust standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Decomposing Differential Kindergarten Fadeout for Black Children

	(1) Separate	(2) Joint	(3) Joint w/ center features
Black_i	-0.30^{**}	-0.29^{*}	-0.36^{**}
(Control-group growth gap)	(0.14)	(0.15)	(0.16)
$\text{Black}_i + Z_i \times \text{Black}_i$	0.08	0.07	0.08
(Treatment-group growth gap)	(0.11)	(0.11)	(0.12)
$Z_i \times \text{Black}_i$	0.38^{***}	0.36^{**}	0.43^{***}
(Black–non-Black fadeout differential)	(0.13)	(0.15)	(0.16)
Z_i	-0.30^{***}	-0.19^{*}	-0.25^{**}
(Non-Black fadeout)	(0.06)	(0.10)	(0.13)
$Z_i + Z_i \times \text{Black}_i$	0.08	0.17	0.18
(Black fadeout)	(0.12)	(0.15)	(0.18)
Disadvantage interactions estimated jointly	No	Yes	Yes
Center-feature interactions included	No	No	Yes
Observations	1,439	1,439	1,292

Notes: The table decomposes the Black–non-Black differential in cognitive skill growth between the end of Head Start and the end of kindergarten into control- and treatment-group growth gaps. Column (1) estimates the Black interaction separately, Column (2) includes all disadvantage interactions jointly, and Column (3) additionally includes center-feature interactions. All specifications include baseline covariates, grantee fixed effects, and sampling weights. Robust standard errors are in parentheses. Positive fadeout differentials indicate less fadeout for Black children. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7: Selection-Corrected Heterogeneity in Head Start Effects on Cognitive Skills

	End of Head Start $\psi_{1P} - \psi_{0P}$	End of K $\psi_{1K} - \psi_{0K}$	Growth during K $(\psi_{1K} - \psi_{0K}) - (\psi_{1P} - \psi_{0P})$
Baseline Disadvantage			
Child Black	-0.18	0.45^{**}	0.63^{***}
	(0.15)	(0.23)	(0.17)
Child Hispanic	-0.06	0.16	0.22
	(0.13)	(0.17)	(0.16)
Low baseline cognition	0.10	0.06	-0.04
	(0.13)	(0.16)	(0.14)
Mother below high school	0.05	0.05	0.00
	(0.17)	(0.15)	(0.15)
Not with two biological parents	-0.10	0.13	0.24^{**}
	(0.16)	(0.16)	(0.12)
Moderate/high home risk	0.07	-0.09	-0.16
	(0.15)	(0.15)	(0.13)
Head Start Center Characteristics			
High share BA/licensed teachers	0.13	0.18	0.05
	(0.14)	(0.14)	(0.11)
Full-day services	0.09	0.12	0.02
	(0.13)	(0.12)	(0.11)
High program services	0.09	0.14	0.05
	(0.17)	(0.13)	(0.13)
Experienced director	0.07	-0.04	-0.11
	(0.13)	(0.13)	(0.12)
Small class size	0.29^{**}	0.23^{*}	-0.06
	(0.12)	(0.13)	(0.11)

Notes: Entries report estimated loadings of the listed baseline disadvantage measures and center characteristics on the selection-corrected Head Start treatment effect. The first two columns report heterogeneity in treatment effects at the end of Head Start and the end of kindergarten, respectively. The final column reports heterogeneity in the treatment effect on cognitive skill growth during kindergarten. Positive coefficients in the final column indicate less fadeout. Cognitive outcomes and growth are measured in units of the baseline control-group standard deviation. Standard errors are based on a sandwich covariance matrix constructed from site-level likelihood scores and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 8: Selection-Corrected Decomposition of the Black–Non-Black Fadeout Differential

	Treated growth $\psi_{1K,B} - \psi_{1P,B}$	Control growth $\psi_{0K,B} - \psi_{0P,B}$	Fadeout differential $(\psi_{1K,B} - \psi_{1P,B}) - (\psi_{0K,B} - \psi_{0P,B})$
Child Black	0.16	-0.47^{***}	0.63^{***}
	(0.12)	(0.15)	(0.17)
Share attributable to control-side growth		$\lvert -0.47 \rvert / 0.63 \approx 74\%$	

Notes: The first column reports the Black–non-Black differential in potential cognitive growth during kindergarten under Head Start participation. The second column reports the corresponding differential under non-participation. The final column reports the treated-growth differential minus the control-growth differential; positive values indicate less fadeout among Black children. Cognitive growth is measured in units of the baseline control-group standard deviation. Standard errors are based on a sandwich covariance matrix constructed from site-level likelihood scores and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 9: Differences in Family and Kindergarten School Inputs

	Treatment − Control		Black − non-Black
	Non-Black	**Black**	**among controls**
Panel A. Family inputs			
Home inputs			
Parent reads to child every day	−0.06 (0.06)	0.26** (0.13)	−0.21 (0.15)
Family cultural enrichment index	0.25* (0.15)	0.48 (0.43)	0.32 (0.39)
Parenting			
Parent spanked child	−0.02 (0.05)	−0.21 (0.14)	0.16 (0.17)
Parent used time-out	0.00 (0.05)	0.15 (0.16)	−0.36** (0.17)
Authoritarian parenting	−0.01 (0.03)	0.03 (0.12)	0.03 (0.13)
Authoritative parenting	0.10 (0.06)	0.18 (0.13)	0.15 (0.15)
Permissive parenting	0.05 (0.05)	−0.10 (0.08)	−0.08 (0.12)
Family constraints			
Number of residential moves	−0.09 (0.06)	−0.21 (0.16)	0.06 (0.19)
Mother currently employed	−0.06 (0.04)	0.01 (0.13)	0.22 (0.14)
Panel B. Kindergarten school inputs			
School environment			
School reading and mathematics proficiency	−0.13*** (0.05)	−0.10 (0.10)	−0.32*** (0.11)
Share eligible for free or reduced-price lunch	−0.01 (0.02)	0.06 (0.05)	−0.01 (0.05)
Peer environment	−0.11 (0.07)	0.07 (0.20)	−0.13 (0.20)
Teacher resources			
Years of teaching experience	0.72 (0.90)	1.75 (2.44)	2.64 (2.56)
College courses related to teaching and child development	0.07 (0.93)	3.23 (2.17)	−2.72 (2.34)
Weekly hours of classroom aide and volunteer assistance	−1.98 (3.82)	1.86 (5.42)	3.08 (6.58)
Instructional practices			
Share of activity time in adult-directed individual activities	−0.01 (0.02)	0.03 (0.04)	−0.10** (0.05)
Number of daily literacy and mathematics activities	0.01 (0.28)	0.26 (1.01)	0.27 (1.10)
Teacher beliefs and responsiveness			
Teacher belief score	−0.06 (0.04)	0.03 (0.13)	−0.01 (0.13)
Teacher adaptive-response measure	0.10* (0.06)	−0.15 (0.16)	0.05 (0.18)
Observations		656	

Notes: Each row is estimated using a linear regression of the indicated family or kindergarten school input on Black, Head Start assignment, their interaction, the baseline covariates reported in Table 2, and grantee fixed effects, with robust standard errors in parentheses. The first two columns report treatment–control differences separately for non-Black and Black children. The final column reports the Black–non-Black difference among control-assigned children. All specifications use the common mechanism-analysis sample and are weighted using the HSIS child-level baseline sampling weights. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 10: Gelbach Decomposition of Growth and Fadeout Differential

	Black–non-Black control-group growth gap		Black–non-Black fadeout differential	
	Contribution	Share (%)	Contribution	Share (%)
Panel A. Family inputs				
Home inputs	−0.02 (0.03)	4.02	0.03 (0.03)	5.30
Parenting	0.10 (0.07)	−26.39	−0.02 (0.06)	−4.04
Family constraints	0.04 (0.05)	−10.54	−0.04 (0.05)	−6.13
All family inputs	**0.13** (0.09)	**−32.91**	**−0.03** (0.08)	**−4.87**
Panel B. Kindergarten school inputs				
School environment	**−0.09*** (0.05)	**23.00**	**0.08** (0.06)	**13.22**
School reading and mathematics proficiency	−0.06 (0.03)	13.97	0.01 (0.02)	1.95
Share eligible for free or reduced-price lunch	−0.02 (0.03)	5.22	0.05 (0.05)	8.10
Peer environment	−0.02 (0.03)	3.81	0.02 (0.04)	3.17
Teacher resources	0.01 (0.05)	−1.38	−0.01 (0.04)	−2.00
Instructional practices	**−0.05** (0.04)	**11.76**	**0.02** (0.04)	**3.92**
Adult-directed individual activities	−0.04 (0.04)	10.36	0.02 (0.02)	2.96
Daily literacy and mathematics activities	−0.01 (0.02)	1.40	0.01 (0.03)	0.96
Teacher beliefs and responsiveness	0.01 (0.03)	−1.61	0.01 (0.03)	2.04
All kindergarten school inputs	**−0.13** (0.09)	**31.78**	**0.10** (0.08)	**17.18**
All mediators	**0.00** (0.12)	**−1.13**	**0.07** (0.11)	**12.31**
Restricted coefficient	−0.40 (0.27)		0.58** (0.24)	
Full coefficient	−0.40* (0.24)		0.51** (0.23)	
Observations	656			

Notes: The table reports Gelbach decompositions of the Black–non-Black cognitive-growth differential among control-assigned children and the Black–non-Black fadeout differential. Contributions report the portion of the difference between the restricted- and full-specification coefficients attributable to each mediator block. Robust standard errors for directly estimated decomposition components are reported in parentheses. The share column expresses each contribution relative to the corresponding restricted coefficient, with positive values indicating attenuation of the corresponding differential and negative values indicating amplification. Indented rows decompose the school-environment and instructional-practice blocks into their component measures. The "All family inputs" and "All kindergarten school inputs" rows are sums of the corresponding decomposition components. All specifications use the common mechanism-analysis sample, include the baseline covariates reported in Table 2 and grantee fixed effects, and are weighted using the HSIS child-level baseline sampling weights. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 11: Counterfactual Reallocation of Head Start Offers and Aggregate Fadeout

	Observed HSIS assignment	Reallocate 10% of source offers	Reallocate 15% of source offers	Fill recipient centers
Black offer rate	0.58	0.83 [0.80, 0.86]	0.90 [0.87, 0.92]	0.96 [0.93, 0.98]
End-of-Head-Start ITT	0.18 [0.11, 0.25]	0.17 [0.11, 0.24]	0.17 [0.10, 0.24]	0.17 [0.10, 0.24]
End-of-kindergarten ITT	−0.10 [−0.18, −0.03]	−0.09 [−0.17, −0.01]	−0.09 [−0.17, −0.01]	−0.08 [−0.15, −0.01]
Fadeout	−0.28 [−0.36, −0.21]	−0.26 [−0.34, −0.19]	−0.26 [−0.33, −0.18]	−0.25 [−0.33, −0.18]
Fadeout attenuation ratio	—	6.07%	8.21%	9.64%

Notes: Fadeout is defined as the end-of-kindergarten ITT effect minus the end-of-Head-Start ITT effect, so more negative values indicate greater fadeout. The fadeout attenuation ratio is calculated relative to the observed HSIS assignment as $1 - \text{Fadeout}_p/\text{Fadeout}_{\text{Observed}}$. Brackets report the 5th and 95th percentiles of the simulated policy moments. Each estimate is averaged across 1,000 outer simulations. Within each simulation, outcomes are averaged across 2,000 candidate draws of each center's parameter vector using center-specific posterior weights and are then aggregated across centers using the HSIS sampling weights.

A Tables and Figures

Table A1: Retention by Treatment Status Across Follow-Up Waves

	Treatment	Control mean	N
End of Head Start	0.10***	0.75	1,993
	(0.03)		
End of Kindergarten	0.10***	0.67	1,993
	(0.03)		
End of 1st Grade	0.09***	0.61	1,993
	(0.03)		
End of 3rd Grade	0.07**	0.55	1,993
	(0.03)		

Notes: Entries report the effect of treatment assignment on the probability that a child remains observed at each follow-up wave. The control mean reports the weighted retention rate among control-group children. Coefficients are estimated from regressions of the retention indicator on treatment assignment. Sampling weights are applied. Robust standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A2: Intent-to-Treat and Local Average Treatment Effects on Language Skills

	ITT		LATE	
	No controls	Controls	No controls	Controls
End of Head Start	2.97	4.17^{**}	4.40	6.21^{**}
$N_{\text{ITT}} = 1{,}612$				
$N_{\text{LATE}} = 1{,}608$	(2.53)	(1.84)	(3.69)	(2.70)
End of Kindergarten	-0.48	1.00	-0.70	1.47
$N_{\text{ITT}} = 1{,}439$				
$N_{\text{LATE}} = 1{,}434$	(2.68)	(2.28)	(3.80)	(3.23)
End of 1st Grade	2.93	4.08^{*}	4.35	6.07^{**}
$N_{\text{ITT}} = 1{,}337$				
$N_{\text{LATE}} = 1{,}334$	(2.42)	(2.12)	(3.50)	(3.07)
End of 3rd Grade	1.05	2.70	1.54	3.98^{*}
$N_{\text{ITT}} = 1{,}193$				
$N_{\text{LATE}} = 1{,}189$	(2.06)	(1.68)	(2.93)	(2.41)

Notes: Entries report ITT and LATE estimates of Head Start on language test scores. Outcomes are measured in original test-score units. LATE estimates instrument Head Start participation with random assignment to receive a Head Start offer. Specifications with controls include the baseline covariates listed in Table 2. The 2SLS specifications exclude singleton observations in absorbed grantee fixed-effect groups, which do not contribute to within-grantee identification. Consequently, LATE sample sizes are slightly smaller than the corresponding ITT sample sizes. All specifications include grantee fixed effects and apply sampling weights. Heteroskedasticity-robust standard errors are reported in parentheses. Control-group means are 302 at the end of Head Start, 343 at the end of kindergarten, 373 at the end of first grade, and 407 at the end of third grade. The baseline control-group standard deviation is 35.83. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A3: Intent-to-Treat and Local Average Treatment Effects on Reading Skills

	ITT		LATE	
	No controls	Controls	No controls	Controls
End of Head Start $N_{\text{ITT}} = 1{,}612$	4.08**	3.25***	6.05**	4.84***
$N_{\text{LATE}} = 1{,}608$	(1.78)	(1.19)	(2.59)	(1.73)
End of Kindergarten $N_{\text{ITT}} = 1{,}439$	-2.79	-2.78	-4.06	-4.08
$N_{\text{LATE}} = 1{,}434$	(2.03)	(1.76)	(2.86)	(2.49)
End of 1st Grade $N_{\text{ITT}} = 1{,}337$	-0.86	-0.99	-1.28	-1.47
$N_{\text{LATE}} = 1{,}334$	(2.38)	(2.16)	(3.43)	(3.11)
End of 3rd Grade $N_{\text{ITT}} = 1{,}193$	0.39	1.06	0.57	1.57
$N_{\text{LATE}} = 1{,}189$	(2.12)	(1.86)	(3.01)	(2.64)

Notes: Entries report ITT and LATE estimates of Head Start on reading test scores. Outcomes are measured in original test-score units. LATE estimates instrument Head Start participation with random assignment to receive a Head Start offer. Specifications with controls include the baseline covariates listed in Table 2. The 2SLS specifications exclude singleton observations in absorbed grantee fixed-effect groups, which do not contribute to within-grantee identification. Consequently, LATE sample sizes are slightly smaller than the corresponding ITT sample sizes. All specifications include grantee fixed effects and apply sampling weights. Heteroskedasticity-robust standard errors are reported in parentheses. Control-group means are 336 at the end of Head Start, 387 at the end of kindergarten, 436 at the end of first grade, and 483 at the end of third grade. The baseline control-group standard deviation is 31.98. ${}^{*}p < 0.10$, ${}^{**}p < 0.05$, ${}^{***}p < 0.01$.

Table A4: Intent-to-Treat and Local Average Treatment Effects on Math Skills

	ITT		LATE	
	No controls	Controls	No controls	Controls
End of Head Start $N_{\text{ITT}} = 1{,}612$	3.32*	3.01*	4.93*	4.48*
$N_{\text{LATE}} = 1{,}608$	(1.94)	(1.58)	(2.83)	(2.30)
End of Kindergarten $N_{\text{ITT}} = 1{,}439$	-1.31	-1.24	-1.91	-1.81
$N_{\text{LATE}} = 1{,}434$	(1.46)	(1.34)	(2.07)	(1.90)
End of 1st Grade $N_{\text{ITT}} = 1{,}337$	-0.22	-0.33	-0.33	-0.49
$N_{\text{LATE}} = 1{,}334$	(1.32)	(1.17)	(1.90)	(1.68)
End of 3rd Grade $N_{\text{ITT}} = 1{,}193$	-0.87	0.06	-1.27	0.09
$N_{\text{LATE}} = 1{,}189$	(1.47)	(1.25)	(2.08)	(1.77)

Notes: Entries report ITT and LATE estimates of Head Start on math test scores. Outcomes are measured in original test-score units. LATE estimates instrument Head Start participation with random assignment to receive a Head Start offer. Specifications with controls include the baseline covariates listed in Table 2. The 2SLS specifications exclude singleton observations in absorbed grantee fixed-effect groups, which do not contribute to within-grantee identification. Consequently, LATE sample sizes are slightly smaller than the corresponding ITT sample sizes. All specifications include grantee fixed effects and apply sampling weights. Heteroskedasticity-robust standard errors are reported in parentheses. Control-group means are 393 at the end of Head Start, 428 at the end of kindergarten, 456 at the end of first grade, and 489 at the end of third grade. The baseline control-group standard deviation is 24.12. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A5: Intent-to-Treat and Local Average Treatment Effects on Cognitive Skills

	ITT		LATE	
	No controls	Controls	No controls	Controls
End of Head Start	0.16^{***}	0.13^{***}	0.23^{***}	0.20^{***}
$N_{\text{ITT}} = 1{,}612$				
$N_{\text{LATE}} = 1{,}608$	(0.06)	(0.05)	(0.09)	(0.07)
End of Kindergarten	-0.08	-0.08	-0.12	-0.12
$N_{\text{ITT}} = 1{,}439$				
$N_{\text{LATE}} = 1{,}434$	(0.06)	(0.05)	(0.08)	(0.07)
End of 1st Grade	-0.02	-0.03	-0.03	-0.04
$N_{\text{ITT}} = 1{,}337$				
$N_{\text{LATE}} = 1{,}334$	(0.07)	(0.06)	(0.10)	(0.09)
End of 3rd Grade	-0.01	0.02	-0.02	0.03
$N_{\text{ITT}} = 1{,}193$				
$N_{\text{LATE}} = 1{,}189$	(0.07)	(0.06)	(0.10)	(0.08)

Notes: Entries report ITT and LATE estimates of Head Start on the cognitive composite. The outcome is the average of normalized reading and math scores, expressed in units of the baseline control-group standard deviation. LATE estimates instrument Head Start participation with random assignment to receive a Head Start offer. Specifications with controls include the baseline covariates listed in Table 2. The 2SLS specifications exclude singleton observations in absorbed grantee fixed-effect groups, which do not contribute to within-grantee identification. Consequently, LATE sample sizes are slightly smaller than the corresponding ITT sample sizes. All specifications include grantee fixed effects and apply sampling weights. Heteroskedasticity-robust standard errors are reported in parentheses. Control-group means are 0.16 at the end of Head Start, 1.95 at the end of kindergarten, 3.52 at the end of first grade, and 5.20 at the end of third grade. The baseline control-group standard deviation is 1.00. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A6: Baseline Disadvantage and Center Features

Category	Variable	Detail
Disadvantage	Child Black	Indicator
	Child Hispanic	Indicator
	Low baseline cognition	Bottom quartile
	Mother below high school	Indicator
	Not living with two biological parents	Indicator
	Moderate/high home risk	Indicator
Head Start center	High share BA/licensed teachers	Top quartile
	Full-day services	Indicator
	High program services	Top quartile of service index
	Experienced director	Top quartile
	Small class size	Bottom quartile

Notes: Low baseline cognition is defined as cognitive skills in the bottom quartile of the weighted child-level distribution. For Head Start center characteristics, "high" and "experienced" denote the top quartile, while "small" denotes the bottom quartile, of the weighted child-level distribution of center exposures. Full-day services is an indicator for whether the center offers a full-day program. The program-services index summarizes whether centers provide child health services, hearing and vision screening or referrals, mental health services, nutrition services, alcohol or drug abuse treatment or counseling, family counseling or mental health services, assistance with family violence, food and nutrition assistance, foster-care payments, housing assistance, income assistance, job-training or employment assistance, adult education or literacy services, help with medical care, and help with utilities. Variables are constructed from the HSIS child assessment, center director interview, and Covariates and Subgroup Variables files.

Table A7: Complier Heterogeneity in Kindergarten Fadeout by Disadvantage

	(1) Separate	(2) Joint	(3) Joint w/ center features
Baseline Disadvantage			
Child Black	0.55**	0.52**	0.64**
	(0.22)	(0.24)	(0.27)
Child Hispanic	−0.28*	−0.06	−0.16
	(0.16)	(0.19)	(0.19)
Low baseline cognition	−0.01	−0.05	−0.03
	(0.18)	(0.18)	(0.19)
Mother below high school	−0.27*	−0.25	−0.18
	(0.16)	(0.20)	(0.21)
Not with two biological parents	0.08	−0.07	−0.22
	(0.17)	(0.20)	(0.23)
Moderate/high home risk	0.05	0.06	0.11
	(0.21)	(0.22)	(0.25)
Head Start Center Characteristics			
High share BA/licensed teachers			0.11
			(0.23)
Full-day services			0.08
			(0.19)
High program services			0.05
			(0.17)
Experienced director			−0.24
			(0.25)
Small class size			0.00
			(0.20)
Disadvantage interactions estimated jointly	No	Yes	Yes
Center-feature interactions included	No	No	Yes
Observations	1,434	1,434	1,287

Notes: Entries report treatment-by-characteristic interactions from 2SLS regressions of cognitive skill growth between the end of Head Start and the end of kindergarten. Head Start participation and its interactions are instrumented by assignment and the corresponding assignment interactions. Column (1) estimates each interaction separately, Column (2) includes all disadvantage interactions jointly, and Column (3) additionally includes center-characteristic interactions. The 2SLS specifications exclude singleton observations in absorbed grantee fixed-effect groups, which do not contribute to within-grantee identification. All specifications include baseline covariates, grantee fixed effects, and sampling weights. Robust standard errors are reported in parentheses. Positive coefficients indicate less fadeout among compliers. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A8: ITT and LATE Effects for Black and Non-Black Children over Time

	End of Head Start	End of Kindergarten	End of First Grade	End of Third Grade
Panel A: Intent-to-treat effects				
Non-Black effect	0.16***	−0.14**	−0.06	−0.01
	(0.06)	(0.06)	(0.07)	(0.06)
	[$p = 0.005$]	[$p = 0.011$]	[$p = 0.371$]	[$p = 0.895$]
Black effect	0.15*	0.23*	0.14	0.18
	(0.09)	(0.13)	(0.14)	(0.12)
	[$p = 0.092$]	[$p = 0.080$]	[$p = 0.339$]	[$p = 0.145$]
Black–non-Black difference	−0.01	0.37***	0.20	0.19
	(0.11)	(0.14)	(0.16)	(0.14)
	[$p = 0.892$]	[$p = 0.009$]	[$p = 0.216$]	[$p = 0.171$]
Observations	1,439	1,439	1,337	1,193
Panel B: Local average treatment effects				
Non-Black effect	0.23***	−0.20**	−0.08	−0.01
	(0.08)	(0.08)	(0.09)	(0.09)
	[$p = 0.006$]	[$p = 0.011$]	[$p = 0.373$]	[$p = 0.893$]
Black effect	0.27*	0.40*	0.24	0.30
	(0.16)	(0.22)	(0.25)	(0.22)
	[$p = 0.090$]	[$p = 0.070$]	[$p = 0.345$]	[$p = 0.159$]
Black–non-Black difference	0.05	0.60**	0.32	0.32
	(0.18)	(0.23)	(0.27)	(0.23)
	[$p = 0.798$]	[$p = 0.010$]	[$p = 0.230$]	[$p = 0.173$]
Observations	1,434	1,434	1,334	1,189

Notes: The table reports ITT and LATE effects on cognitive skills for non-Black and Black children and the corresponding Black–non-Black difference at each follow-up wave. The end-of-Head-Start and end-of-kindergarten estimates use the common sample observed through kindergarten; the first- and third-grade estimates use the corresponding cumulative follow-up samples observed through each respective wave. Each column estimates the interaction between the Black indicator and Head Start assignment in Panel A or Head Start participation in Panel B, without including interactions with the other disadvantage indicators. In Panel B, Head Start participation and its interaction with the Black indicator are instrumented by random assignment and the interaction between assignment and the Black indicator, respectively. Outcomes are measured in units of the baseline control-group standard deviation. The 2SLS specifications exclude singleton observations in absorbed grantee fixed-effect groups, which do not contribute to within-grantee identification. All specifications include the baseline covariates listed in Table 2, grantee fixed effects, and sampling weights. Heteroskedasticity-robust standard errors are reported in parentheses, and two-sided p-values are reported in brackets. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A9: Experimental and Model-Implied Reduced-Form Moments

	Experimental estimate	Model-implied moment	Difference
Panel A: First-stage and ITT moments			
First stage	0.75	0.78	0.03
End-of-Head-Start ITT	0.20	0.20	0.00
End-of-kindergarten ITT	−0.12	−0.10	0.01
ITT on kindergarten growth	−0.31	−0.30	0.01
Panel B: LATE moments			
End-of-Head-Start LATE	0.26	0.25	−0.01
End-of-kindergarten LATE	−0.16	−0.13	0.02
LATE on kindergarten growth	−0.42	−0.39	0.03

Notes: The first column reports reduced-form moments estimated directly in the structural-model estimation sample. The first stage is estimated from a weighted regression of Head Start participation on randomized assignment. The ITT estimates are obtained from weighted regressions of each cognitive outcome, or kindergarten cognitive growth, on randomized assignment. The LATE estimates are obtained from weighted 2SLS regressions of each outcome on Head Start participation, instrumenting participation with randomized assignment. These specifications do not include baseline covariates or fixed effects. The second column reports the mean of the corresponding estimates across 1,000 datasets simulated from the estimated selection model. The final column reports the model-implied moment minus the corresponding experimental estimate. Kindergarten growth is defined as the end-of-kindergarten outcome minus the end-of-Head-Start outcome. Cognitive outcomes and growth are measured in units of the baseline control-group standard deviation.

Table A10: Core Structural Parameter Estimates

Panel A: Intercept Parameters

	$\alpha_{1P,0}$	$\alpha_{0P,0}$	$\alpha_{1K,0}$	$\alpha_{0K,0}$	λ_0	$\log \pi_0$
Estimate	0.60***	0.56***	1.82***	2.38***	−2.32***	1.23***
	(0.09)	(0.11)	(0.11)	(0.12)	(0.37)	(0.14)

Panel B: Individual-Level Outcome-Shock Standard Deviations

	σ_{1P}	σ_{0P}	σ_{1K}	σ_{0K}
Estimate	0.65***	0.69***	0.74***	0.74***
	(0.03)	(0.03)	(0.03)	(0.05)

Panel C: Individual-Level Shock Correlations

	$\rho_{1,PK}$	$\rho_{0,PK}$	$\rho_{n,1P}$	$\rho_{n,0P}$	$\rho_{n,1K}$	$\rho_{n,0K}$
Estimate	0.49***	0.54***	0.00	0.18*	−0.42**	0.02
	(0.05)	(0.04)	(0.18)	(0.11)	(0.17)	(0.14)

Notes: Entries report natural-scale maximum-likelihood estimates. Standard errors, reported in parentheses, are based on a sandwich covariance matrix constructed from site-level likelihood scores and are transformed to the natural parameter scale using the delta method. P and K denote the end of Head Start and the end of kindergarten, respectively. Subscripts 1 and 0 denote potential outcomes under Head Start participation and nonparticipation. The parameters λ_0 and $\log \pi_0$ govern the baseline participation propensity and the logarithm of the first-stage shift induced by random assignment. Cholesky-factor parameters used to construct the cross-site covariance matrix are omitted because they do not have a direct economic interpretation. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A11: Loadings of Observed Characteristics on Structural Parameters

	$\psi_{\alpha_{1P}}$	$\psi_{\alpha_{0P}}$	$\psi_{\alpha_{1K}}$	$\psi_{\alpha_{0K}}$	ψ_{λ}	$\psi_{\log \pi}$
Baseline Disadvantage						
Child Black	−0.15	0.04	0.01	−0.43**	1.08***	−0.43**
	(0.11)	(0.15)	(0.12)	(0.19)	(0.39)	(0.18)
Child Hispanic	−0.22**	−0.16	0.00	−0.15	0.85**	−0.19
	(0.09)	(0.10)	(0.10)	(0.12)	(0.36)	(0.13)
Low baseline cognition	−0.83***	−0.93***	−0.59***	−0.65***	−0.35	0.16
	(0.09)	(0.09)	(0.11)	(0.13)	(0.29)	(0.12)
Mother below high school	−0.16	−0.21*	−0.09	−0.14	−0.12	0.12
	(0.10)	(0.11)	(0.07)	(0.12)	(0.28)	(0.13)
Not living with two biological parents	−0.02	0.09	0.00	−0.13	0.36	−0.11
	(0.09)	(0.12)	(0.08)	(0.12)	(0.25)	(0.14)
Moderate/high home risk	−0.15*	−0.22*	−0.12	−0.03	−0.15	0.07
	(0.08)	(0.13)	(0.10)	(0.12)	(0.31)	(0.14)
Head Start Center Characteristics						
High share BA/licensed teachers	0.10	−0.03	0.18**	0.00	0.39	−0.21
	(0.10)	(0.11)	(0.07)	(0.12)	(0.36)	(0.17)
Full-day services	0.19*	0.09	0.10	−0.01	0.12	−0.14
	(0.10)	(0.09)	(0.09)	(0.10)	(0.24)	(0.12)
High program services	−0.01	−0.10	0.11	−0.03	−0.75***	0.42***
	(0.14)	(0.11)	(0.08)	(0.13)	(0.27)	(0.12)
Experienced director	0.13*	0.06	0.06	0.10	0.12	−0.13
	(0.08)	(0.11)	(0.09)	(0.12)	(0.40)	(0.17)
Small class size	0.11	−0.19**	0.12	−0.12	0.34	−0.13
	(0.08)	(0.10)	(0.10)	(0.12)	(0.26)	(0.13)

Notes: Entries report the estimated loadings of observed child and center characteristics on each structural parameter. Standard errors, reported in parentheses, are based on a sandwich covariance matrix constructed from site-level likelihood scores and are transformed to the natural parameter scale using the delta method. P and K denote the end of Head Start and the end of kindergarten, respectively. Subscripts 1 and 0 denote potential outcomes under Head Start participation and nonparticipation. The parameters λ and $\log \pi$ govern the baseline participation propensity and the logarithm of the first-stage shift induced by random assignment. All listed characteristics are binary. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A12: Family Variables in the Mechanism Analysis

Block	Variable
Home inputs	Parent reads to the child every day in the past week
	Family cultural enrichment index for the past month
Parenting	Parent spanked the child in the past week
	Parent used time-out in the past week
	Authoritarian parenting style
	Authoritative parenting style
	Permissive parenting style
Family constraints	Number of residential moves during the preceding 12 months
	Mother currently employed

Notes: The table lists the family variables included in the mechanism analysis. Variables are drawn from the HSIS kindergarten parent interview file. The family cultural enrichment index is a 0–7 count of activities including going to a movie, attending a play or concert, visiting an art gallery or museum, visiting a playground or park, discussing the child's heritage, attending community events, and taking the child on errands.

Table A13: Kindergarten School Variables in the Mechanism Analysis

Block	Variable
School environment	Standardized composite of school-level reading and mathematics proficiency
	Share of students eligible for free or reduced-price lunch
	Standardized composite of peer attendance and behavior
Teacher resources	Years of teaching experience
	Number of college courses related to teaching and child development
	Weekly hours of assistance from classroom aides and volunteers
Instructional practices	Share of activity time devoted to adult-directed individual activities
	Number of literacy and mathematics activities conducted daily
Teacher beliefs and responsiveness	Teacher belief score, with higher values indicating more child-centered instruction
	Teacher adaptive-response measure

Notes: The table lists the kindergarten school variables included in the mechanism analysis. Variables are drawn from the HSIS kindergarten teacher and school-environment files and are averaged within the child's kindergarten setting where appropriate. The school-proficiency measure combines school-level reading and mathematics proficiency, while the peer-environment measure combines peer attendance and behavior, with higher values indicating a more favorable peer environment. The teacher adaptive-response measure captures the extent to which teachers use developmental assessments and respond with adjustments or additional supports tailored to children's developmental needs.

B Mechanical Fadeout Under Concave Skill Growth

This appendix derives a benchmark "mechanical fadeout" rate implied purely by concavity in a smooth skill-growth function, holding the post-treatment environment fixed.

B.1 Setup

Suppose skill evolves as a smooth concave function of developmental time,

$$Y = f(a) \quad \text{with} \quad f'(a) > 0 \quad \text{and} \quad f''(a) < 0$$

Let the control group's skill at age a_t be

$$C_t = f(a_t)$$

And let the treatment group have a skill advantage Δ_t at age a_t,

$$T_t = C_t + \Delta_t$$

Since $f(\cdot)$ is strictly increasing, there exists an "effective age shift" h_t such that the treatment group skill is

$$f(a_t + h_t) = f(a_t) + \Delta_t$$

B.2 One-period Mechanical Fadeout

Assume that after the intervention both groups face the same environment, so subsequent skill growth is governed by the same function $f(\cdot)$. Then

$$C_{t+1} = f(a_t + 1), \qquad T_{t+1} = f(a_t + 1 + h_t)$$

and the next-period gap is

$$\Delta_{t+1} = f(a_t + 1 + h_t) - f(a_t + 1)$$

Using a first-order Taylor expansion for small h_t,

$$f(a + h) \approx f(a) + f'(a)h$$

so that

$$\Delta_t \approx f'(a_t)h_t, \qquad h_t \approx \frac{\Delta_t}{f'(a_t)}$$

and

$$\Delta_{t+1} \approx f'(a_t+1)h_t$$

Substituting for h_t yields

$$\Delta_{t+1} \approx \Delta_t \frac{f'(a_t+1)}{f'(a_t)}$$

Therefore, the one-period mechanical change in the treatment-control gap is

$$\Delta_{t+1} - \Delta_t \approx \Delta_t \left(\frac{f'(a_t+1)}{f'(a_t)} - 1 \right)$$

Using $f'(a_t+1) \approx f'(a_t) + f''(a_t)$ yields the approximation

$$\Delta_{t+1} - \Delta_t \approx \frac{f''(a_t)}{f'(a_t)} \Delta_t$$

B.3 Multi-period extension

Iterating the one-period expression forward yields

$$\begin{aligned} \Delta_{t+k} &\approx \Delta_{t+k-1} \frac{f'(a_t+k)}{f'(a_t+k-1)} \\ &\approx \Delta_t \prod_{\ell=1}^{k} \frac{f'(a_t+\ell)}{f'(a_t+\ell-1)} = \Delta_t \frac{f'(a_t+k)}{f'(a_t)} \end{aligned}$$

Applying a Taylor expansion around a_t gives

$$f'(a_t+k) \approx f'(a_t) + f''(a_t)k,$$

so that

$$\Delta_{t+k} \approx \Delta_t \frac{f'(a_t) + f''(a_t)k}{f'(a_t)}, \quad \Delta_{t+k} - \Delta_t \approx \Delta_t \frac{f''(a_t)}{f'(a_t)} k$$

B.4 Quadratic growth as a benchmark

If $f(a)$ is approximated by a quadratic growth curve,

$$f(a) = \beta_0 + \beta_1 a + \beta_2 a^2 + \epsilon$$

then

$$f'(a) = \beta_1 + 2\beta_2 a \quad f''(a) = 2\beta_2$$

and the implied multi-period mechanical fadeout benchmark is

$$\Delta_{t+k} \approx \Delta_t \frac{\beta_1 + 2\beta_2(a_t + k)}{\beta_1 + 2\beta_2 a_t}$$
$$\Delta_{t+k} - \Delta_t \approx \Delta_t \frac{2\beta_2 k}{\beta_1 + 2\beta_2 a_t}$$

This benchmark isolates attenuation driven solely by concavity in skill accumulation, providing a reference point for evaluating whether observed fadeout plausibly reflects only mechanical skill-growth curvature or instead additional policy-relevant mechanisms.

C Two-Period Selection Model

C.1 Model

Selection

Consider child i in site j. I define the latent utility from attending Head Start as

$$u_{1ij} = \big(\lambda_j + P_{ij}'\psi_\lambda\big) + \exp\big(\kappa_j + P_{ij}'\psi_\kappa\big)\, Z_{ij} - \eta_{ij}, \tag{1}$$

and normalize the utility from not attending Head Start to zero:

$$u_{0ij} = 0, \tag{2}$$

where Z_{ij} is the randomized Head Start offer indicator. The vector $P_{ij} = (P_{1ij}, \ldots, P_{Mij})'$ contains observed child- and site-level inputs measured prior to treatment assignment. The parameter λ_j captures the residual baseline propensity to attend Head Start in site j, while $P_{ij}'\psi_\lambda$ captures systematic variation in baseline participation associated with observed inputs. The term η_{ij} captures unobserved individual-specific resistance to participation, such as transportation costs, scheduling constraints, or the costs associated with separating from the child.

The offer effect is parameterized on the log scale. Specifically,

$$\log \pi_{ij} = \kappa_j + P_{ij}'\psi_\kappa, \tag{3}$$

so that

$$\pi_{ij} = \exp\big(\kappa_j + P_{ij}'\psi_\kappa\big) > 0. \tag{4}$$

Here, κ_j captures residual cross-site heterogeneity in the log offer effect, while $P_{ij}'\psi_\kappa$ allows the strength of the offer effect to vary systematically with observed inputs. The exponential parameterization preserves monotonicity by ensuring that receiving a Head Start offer weakly increases the incentive to participate.

A child enrolls in Head Start when the latent utility from participation exceeds the normalized utility from non-participation:

$$\begin{aligned} D_{ij} &= \mathbf{1}\left[u_{1ij} > u_{0ij}\right] && (5)\\ &= \mathbf{1}\left[\lambda_j + P_{ij}'\psi_\lambda + \exp\big(\kappa_j + P_{ij}'\psi_\kappa\big)\, Z_{ij} - \eta_{ij} > 0\right] && (6)\\ &= \mathbf{1}\left[\eta_{ij} < \lambda_j + P_{ij}'\psi_\lambda + \exp\big(\kappa_j + P_{ij}'\psi_\kappa\big)\, Z_{ij}\right]. && (7) \end{aligned}$$

Outcomes

I jointly model cognitive outcomes at the end of preschool and at the end of kindergarten. For treatment status $d \in \{0,1\}$, the potential outcome for child i in site j at time t is

$$Y_{tij}(d) = \alpha_{dtj} + P_{ij}'\psi_{dt} + \varepsilon_{dtij}. \tag{8}$$

Here, $t \in \{P, K\}$, where P denotes the end of preschool and K denotes the end of kindergarten. The parameter α_{dtj} is the residual site-specific mean potential outcome under treatment status d at time t, conditional on observed inputs. The term $P_{ij}'\psi_{dt}$ captures systematic variation in potential outcomes associated with observed child- and site-level inputs, while ε_{dtij} captures unobserved individual-specific determinants of the potential outcome.

Equivalently, the four potential outcomes are

$$Y_{Pij}(1) = \alpha_{1Pj} + P_{ij}'\psi_{1P} + \varepsilon_{1Pij}, \tag{9}$$

$$Y_{Pij}(0) = \alpha_{0Pj} + P_{ij}'\psi_{0P} + \varepsilon_{0Pij}, \tag{10}$$

$$Y_{Kij}(1) = \alpha_{1Kj} + P_{ij}'\psi_{1K} + \varepsilon_{1Kij}, \tag{11}$$

$$Y_{Kij}(0) = \alpha_{0Kj} + P_{ij}'\psi_{0K} + \varepsilon_{0Kij}. \tag{12}$$

The observed outcome at time t is determined by the child's realized Head Start participation status:

$$Y_{tij} = D_{ij}Y_{tij}(1) + (1 - D_{ij})Y_{tij}(0), \qquad t \in \{P, K\}. \tag{13}$$

The participation decision D_{ij} is common across the two periods because it records whether the child attended Head Start during the preschool year. The model then follows the child's outcomes through the end of preschool and the end of kindergarten.

Joint Distribution of Unobservables

I assume that the unobserved participation and potential-outcome shocks are jointly normally distributed:

$$(\eta_{ij}, \varepsilon_{1Pij}, \varepsilon_{1Kij}, \varepsilon_{0Pij}, \varepsilon_{0Kij})' \sim \mathcal{N}(0, \Sigma). \tag{14}$$

The covariance matrix Σ allows the participation shock η_{ij} to be correlated with potential-outcome shocks in both treatment states and at both time points. The model also allows potential-outcome shocks to be correlated across time within each treatment state. In particular, ε_{1Pij} may be correlated with ε_{1Kij}, and ε_{0Pij} may be correlated with ε_{0Kij}. Because each child is observed in only one treatment state, cross-treatment-state covariances are not separately identified and are set to zero.

To identify the scale of the latent participation equation, I normalize the variance of the participation shock to one: $\operatorname{Var}(\eta_{ij}) = 1$.

Joint Distribution of Site-Level Parameters

Let $\theta_j = (\alpha_{1Pj}, \alpha_{0Pj}, \alpha_{1Kj}, \alpha_{0Kj}, \lambda_j, \kappa_j)'$ collect the six residual site-specific parameters. I assume that $\theta_j \sim \mathcal{N}(\theta_0, V_0)$, where

$$\theta_0 = (\alpha_{1P,0}, \alpha_{0P,0}, \alpha_{1K,0}, \alpha_{0K,0}, \lambda_0, \kappa_0)'$$

denotes the population mean of the site-specific parameters, and V_0 captures their covariance across sites.

C.2 Likelihood and Simulated Maximum Likelihood Estimation

Let $\Omega = (\theta_0, V_0, \Psi, \Sigma)$ collect the population-level model parameters, where

$$\Psi = (\psi_{1P}, \psi_{0P}, \psi_{1K}, \psi_{0K}, \psi_\lambda, \psi_\kappa)'$$

collects the coefficients on the observed inputs in the four potential-outcome equations and the two participation indices.

Conditional on the site-specific parameter vector θ_j, the individual likelihood contribution is

$$L_{ij} = L_{ij}\left(Y_{Pij}, Y_{Kij}, D_{ij} \mid Z_{ij}, P_{ij}, \theta_j, \Psi, \Sigma\right).$$

The individual likelihood can be factored into the joint density of the two observed outcomes and the probability of the observed participation decision conditional on those outcomes:

$$
\begin{aligned}
L_{ij} = & \Bigg[f_{Y_{Pij}(1),Y_{Kij}(1)}\left(y_{Pij}, y_{Kij} \mid P_{ij}, \theta_j, \Psi, \Sigma\right) \\
& \quad \times \Pr(D_{ij} = 1 \mid Y_{Pij}(1) = y_{Pij}, Y_{Kij}(1) = y_{Kij}, Z_{ij}, P_{ij}, \theta_j, \Psi, \Sigma) \Bigg]^{D_{ij}} \\
& \times \Bigg[f_{Y_{Pij}(0),Y_{Kij}(0)}\left(y_{Pij}, y_{Kij} \mid P_{ij}, \theta_j, \Psi, \Sigma\right) \\
& \quad \times \Pr(D_{ij} = 0 \mid Y_{Pij}(0) = y_{Pij}, Y_{Kij}(0) = y_{Kij}, Z_{ij}, P_{ij}, \theta_j, \Psi, \Sigma) \Bigg]^{1-D_{ij}}.
\end{aligned}
$$

Define the treatment-state-specific residual vector as

$$
\begin{aligned}
e_{dij} &= \begin{bmatrix} e_{dPij} \\ e_{dKij} \end{bmatrix} \\
&= \begin{bmatrix} y_{Pij} - \alpha_{dPj} - P'_{ij}\psi_{dP} \\ y_{Kij} - \alpha_{dKj} - P'_{ij}\psi_{dK} \end{bmatrix}, \qquad d \in \{0, 1\}.
\end{aligned}
$$

Conditional on θ_j, Ψ, and P_{ij}, observing the two potential outcomes under treatment state d is equivalent to observing $\varepsilon_{dPij} = e_{dPij}$ and $\varepsilon_{dKij} = e_{dKij}$.

Let

$$
\begin{aligned}
\Sigma_{dd} &= \mathrm{Var} \begin{bmatrix} \varepsilon_{dPij} \\ \varepsilon_{dKij} \end{bmatrix} \\
&= \begin{bmatrix} \sigma^2_{dP} & \rho_{d,PK}\sigma_{dP}\sigma_{dK} \\ \rho_{d,PK}\sigma_{dP}\sigma_{dK} & \sigma^2_{dK} \end{bmatrix}.
\end{aligned}
$$

The joint potential-outcome density under treatment state d is therefore

$$
\begin{aligned}
& f_{Y_{Pij}(d),Y_{Kij}(d)}\left(y_{Pij}, y_{Kij} \mid P_{ij}, \theta_j, \Psi, \Sigma\right) \\
& \quad = \frac{1}{\sigma_{dP}\sigma_{dK}} \phi_2\left(\frac{e_{dPij}}{\sigma_{dP}}, \frac{e_{dKij}}{\sigma_{dK}}; \rho_{d,PK} \right),
\end{aligned}
$$

where $\phi_2(a, b; \rho)$ denotes the standard bivariate normal density evaluated at (a, b) with correlation ρ.

From the threshold-crossing participation equation,

$$
D_{ij} = \mathbf{1}\left[\eta_{ij} < c_{ij}\right],
$$

where the child-specific participation cutoff is

$$c_{ij} = \lambda_j + P'_{ij}\psi_\lambda + \exp\big(\kappa_j + P'_{ij}\psi_\kappa\big)\, Z_{ij}.$$

Therefore,

$$\Pr(D_{ij} = 1 \mid \cdot) = \Pr(\eta_{ij} < c_{ij} \mid \cdot),$$
$$\Pr(D_{ij} = 0 \mid \cdot) = \Pr(\eta_{ij} \geq c_{ij} \mid \cdot).$$

The individual likelihood can thus be written as

$$\begin{aligned} L_{ij} = \Bigg[& \frac{1}{\sigma_{1P}\sigma_{1K}} \phi_2\bigg(\frac{e_{1Pij}}{\sigma_{1P}}, \frac{e_{1Kij}}{\sigma_{1K}}; \rho_{1,PK} \bigg) \\ & \times \Pr(\eta_{ij} < c_{ij} \mid \varepsilon_{1Pij} = e_{1Pij}, \varepsilon_{1Kij} = e_{1Kij}) \Bigg]^{D_{ij}} \\ \times \Bigg[& \frac{1}{\sigma_{0P}\sigma_{0K}} \phi_2\bigg(\frac{e_{0Pij}}{\sigma_{0P}}, \frac{e_{0Kij}}{\sigma_{0K}}; \rho_{0,PK} \bigg) \\ & \times \Pr(\eta_{ij} \geq c_{ij} \mid \varepsilon_{0Pij} = e_{0Pij}, \varepsilon_{0Kij} = e_{0Kij}) \Bigg]^{1-D_{ij}}. \end{aligned}$$

Conditional Normal Distributions

To evaluate the conditional participation probabilities, define

$$\begin{aligned} \Sigma_{\eta d} &= \operatorname{Cov}\left(\eta_{ij}, \begin{bmatrix} \varepsilon_{dPij} \\ \varepsilon_{dKij} \end{bmatrix} \right) \\ &= \begin{bmatrix} \sigma_{\eta,dP} & \sigma_{\eta,dK} \end{bmatrix}, \end{aligned}$$

and let

$$\Sigma_{d\eta} = \Sigma'_{\eta d}.$$

Under the joint normality assumption and the normalization

$$\operatorname{Var}(\eta_{ij}) = 1,$$

the conditional distribution of the participation shock given the two potential-outcome shocks under treatment state d is

$$\eta_{ij} \mid \varepsilon_{dPij}, \varepsilon_{dKij} \sim \mathcal{N}\left(\mu_{\eta|d,ij}, \sigma^2_{\eta|d}\right),$$

where

$$\mu_{\eta|d,ij} = \Sigma_{\eta d}\Sigma_{dd}^{-1}\begin{bmatrix}\varepsilon_{dPij}\\ \varepsilon_{dKij}\end{bmatrix},$$

and

$$\sigma_{\eta|d}^2 = 1 - \Sigma_{\eta d}\Sigma_{dd}^{-1}\Sigma_{d\eta}.$$

Substituting the observed residual vector gives

$$\mu_{\eta|d,ij} = \Sigma_{\eta d}\Sigma_{dd}^{-1}e_{dij}.$$

For an observed Head Start participant, the conditional participation probability is

$$\begin{aligned}\Pr(\eta_{ij} < c_{ij} \mid \varepsilon_{1Pij} = e_{1Pij}, \varepsilon_{1Kij} = e_{1Kij})\\ = \Phi\left(\frac{c_{ij} - \Sigma_{\eta 1}\Sigma_{11}^{-1}e_{1ij}}{\sqrt{1 - \Sigma_{\eta 1}\Sigma_{11}^{-1}\Sigma_{1\eta}}}\right).\end{aligned}$$

For an observed non-participant, the conditional non-participation probability is

$$\begin{aligned}\Pr(\eta_{ij} \geq c_{ij} \mid \varepsilon_{0Pij} = e_{0Pij}, \varepsilon_{0Kij} = e_{0Kij})\\ = 1 - \Phi\left(\frac{c_{ij} - \Sigma_{\eta 0}\Sigma_{00}^{-1}e_{0ij}}{\sqrt{1 - \Sigma_{\eta 0}\Sigma_{00}^{-1}\Sigma_{0\eta}}}\right).\end{aligned}$$

Using the identity

$$1 - \Phi(a) = \Phi(-a),$$

the non-participation probability can equivalently be written as

$$\begin{aligned}\Pr(\eta_{ij} \geq c_{ij} \mid \varepsilon_{0Pij} = e_{0Pij}, \varepsilon_{0Kij} = e_{0Kij})\\ = \Phi\left(\frac{\Sigma_{\eta 0}\Sigma_{00}^{-1}e_{0ij} - c_{ij}}{\sqrt{1 - \Sigma_{\eta 0}\Sigma_{00}^{-1}\Sigma_{0\eta}}}\right).\end{aligned}$$

Substituting these conditional probabilities into the individual likelihood yields

$$
\begin{aligned}
L_{ij} = \Bigg[& \frac{1}{\sigma_{1P}\sigma_{1K}} \phi_2\left(\frac{e_{1Pij}}{\sigma_{1P}}, \frac{e_{1Kij}}{\sigma_{1K}}; \rho_{1,PK}\right) \\
& \times \Phi\left(\frac{c_{ij} - \Sigma_{\eta 1}\Sigma_{11}^{-1} e_{1ij}}{\sqrt{1 - \Sigma_{\eta 1}\Sigma_{11}^{-1}\Sigma_{1\eta}}}\right) \Bigg]^{D_{ij}} \\
\times \Bigg[& \frac{1}{\sigma_{0P}\sigma_{0K}} \phi_2\left(\frac{e_{0Pij}}{\sigma_{0P}}, \frac{e_{0Kij}}{\sigma_{0K}}; \rho_{0,PK}\right) \\
& \times \Phi\left(\frac{\Sigma_{\eta 0}\Sigma_{00}^{-1} e_{0ij} - c_{ij}}{\sqrt{1 - \Sigma_{\eta 0}\Sigma_{00}^{-1}\Sigma_{0\eta}}}\right) \Bigg]^{1-D_{ij}},
\end{aligned}
$$

where

$$
e_{dij} = \begin{bmatrix} y_{Pij} - \alpha_{dPj} - P_{ij}'\psi_{dP} \\ y_{Kij} - \alpha_{dKj} - P_{ij}'\psi_{dK} \end{bmatrix}
$$

and

$$
c_{ij} = \lambda_j + P_{ij}'\psi_\lambda + \exp\left(\kappa_j + P_{ij}'\psi_\kappa\right) Z_{ij}.
$$

Weighted Site-Level Likelihood

Let w_{ij} denote the analysis weight assigned to child i in site j. Before estimation, I normalize the weights to have mean one in the estimation sample. This preserves their relative values while preventing the arbitrary scale of the weights from changing the strength of the site-specific likelihood relative to the population distribution of site-level parameters.

The weighted conditional likelihood for site j is

$$
L_j^C\left(\mathcal{D}_j \mid \theta_j, \Psi, \Sigma\right) = \prod_{i=1}^{N_j} L_{ij}\left(Y_{Pij}, Y_{Kij}, D_{ij} \mid Z_{ij}, P_{ij}, \theta_j, \Psi, \Sigma\right)^{w_{ij}},
$$

where

$$
\mathcal{D}_j = \{Y_{Pij}, Y_{Kij}, D_{ij}, Z_{ij}, P_{ij}, w_{ij}\}_{i=1}^{N_j}.
$$

Equivalently, the weighted conditional log-likelihood is

$$
\log L_j^C\left(\mathcal{D}_j \mid \theta_j, \Psi, \Sigma\right) = \sum_{i=1}^{N_j} w_{ij} \log L_{ij}.
$$

Site-Level Integration

The site-specific parameter vector follows

$$\theta_j \sim \mathcal{N}(\theta_0, V_0),$$

where

$$\theta_j = (\alpha_{1Pj}, \alpha_{0Pj}, \alpha_{1Kj}, \alpha_{0Kj}, \lambda_j, \kappa_j)'.$$

The marginal likelihood for site j integrates over the six-dimensional vector of site-specific parameters:

$$\begin{aligned} &L_j^M(\mathcal{D}_j \mid \theta_0, V_0, \Psi, \Sigma) \\ &\quad = \int_{\mathbb{R}^6} L_j^C(\mathcal{D}_j \mid \theta_j, \Psi, \Sigma)\, \varphi_6(\theta_j \mid \theta_0, V_0)\; d\theta_j, \end{aligned}$$

where $\varphi_6(\cdot \mid \theta_0, V_0)$ denotes the density of a six-dimensional normal distribution with mean θ_0 and covariance matrix V_0.

The full-sample marginal log-likelihood is

$$\begin{aligned} \mathcal{L}(\theta_0, V_0, \Psi, \Sigma) &= \sum_{j=1}^{J} \log L_j^M(\mathcal{D}_j \mid \theta_0, V_0, \Psi, \Sigma) \\ &= \sum_{j=1}^{J} \log \Bigg[\int_{\mathbb{R}^6} L_j^C(\mathcal{D}_j \mid \theta_j, \Psi, \Sigma) \\ &\qquad\qquad \times \varphi_6(\theta_j \mid \theta_0, V_0)\; d\theta_j \Bigg]. \end{aligned}$$

Simulated Maximum Likelihood

Because the six-dimensional integral defining the marginal likelihood has no closed-form solution, I approximate it using simulated maximum likelihood.

Let C be an upper-triangular Cholesky factor satisfying

$$V_0 = C'C.$$

For each site j, draw

$$\xi_{js} \sim \mathcal{N}(0, I_6), \qquad s = 1, \ldots, S,$$

and construct the simulated site-specific parameter draw

$$\theta_{js} = \theta_0 + C'\xi_{js}.$$

Each draw has the form

$$\theta_{js} = \left(\alpha_{1Pjs}, \alpha_{0Pjs}, \alpha_{1Kjs}, \alpha_{0Kjs}, \lambda_{js}, \kappa_{js}\right)'.$$

For each draw s, I evaluate the weighted conditional site likelihood

$$\begin{aligned} L^C_{js} &= L^C_j\left(\mathcal{D}_j \mid \theta_{js}, \Psi, \Sigma\right) \\ &= \prod_{i=1}^{N_j} L_{ij}\left(Y_{Pij}, Y_{Kij}, D_{ij} \mid Z_{ij}, P_{ij}, \theta_{js}, \Psi, \Sigma\right)^{w_{ij}}. \end{aligned}$$

The simulated marginal likelihood is

$$\widehat{L}^M_j\left(\mathcal{D}_j \mid \theta_0, V_0, \Psi, \Sigma\right) = \frac{1}{S}\sum_{s=1}^{S} L^C_{js}.$$

The simulated full-sample log-likelihood is

$$\widehat{\mathcal{L}}\left(\theta_0, V_0, \Psi, \Sigma\right) = \sum_{j=1}^{J} \log\left[\frac{1}{S}\sum_{s=1}^{S} L^C_{js}\right].$$

I estimate $(\theta_0, V_0, \Psi, \Sigma)$ by maximizing the simulated log-likelihood. As $S \to \infty$, the simulated marginal likelihood converges to the corresponding exact marginal likelihood for fixed parameter values.

C.3 Posterior Moments and Empirical Bayes Shrinkage

By Bayes' rule, the posterior distribution of the site-specific parameter vector θ_j is

$$\begin{aligned} &p\left(\theta_j \mid \mathcal{D}_j, \theta_0, V_0, \Psi, \Sigma\right) \\ &\qquad = \frac{L^C_j\left(\mathcal{D}_j \mid \theta_j, \Psi, \Sigma\right)\varphi_6\left(\theta_j \mid \theta_0, V_0\right)}{\displaystyle\int_{\mathbb{R}^6} L^C_j\left(\mathcal{D}_j \mid \theta, \Psi, \Sigma\right)\varphi_6\left(\theta \mid \theta_0, V_0\right)\, d\theta}, \end{aligned}$$

where $\varphi_6(\cdot \mid \theta_0, V_0)$ denotes the density of a six-dimensional normal distribution with mean θ_0 and covariance matrix V_0.

After estimating

$$\widehat{\Omega} = \left(\widehat{\theta}_0, \widehat{V}_0, \widehat{\Psi}, \widehat{\Sigma}\right)$$

by simulated maximum likelihood, I evaluate posterior moments at the estimated population-level parameters. For any policy-relevant function $h(\theta_j)$, the empirical Bayes posterior moment is

$$\begin{aligned}
\widehat{h}_j^{EB} &= \mathbb{E}\left[h\left(\theta_j\right) \mid \mathcal{D}_j, \widehat{\Omega}\right] \\
&= \frac{\displaystyle\int_{\mathbb{R}^6} h\left(\theta\right) L_j^C\left(\mathcal{D}_j \mid \theta, \widehat{\Psi}, \widehat{\Sigma}\right) \varphi_6\left(\theta \mid \widehat{\theta}_0, \widehat{V}_0\right) d\theta}{\displaystyle\int_{\mathbb{R}^6} L_j^C\left(\mathcal{D}_j \mid \theta, \widehat{\Psi}, \widehat{\Sigma}\right) \varphi_6\left(\theta \mid \widehat{\theta}_0, \widehat{V}_0\right) d\theta}.
\end{aligned}$$

This formulation applies both to the site-specific parameters themselves and to policy moments constructed from those parameters. When $h(\theta_j) = \theta_j$, the posterior moment becomes the conventional empirical Bayes shrinkage estimator:

$$\widehat{\theta}_j^{EB} = \mathbb{E}\left[\theta_j \mid \mathcal{D}_j, \widehat{\Omega}\right].$$

For policy moments, I calculate the posterior mean of the policy function directly. In general,

$$\mathbb{E}\left[h\left(\theta_j\right) \mid \mathcal{D}_j, \widehat{\Omega}\right] \neq h\left(\mathbb{E}\left[\theta_j \mid \mathcal{D}_j, \widehat{\Omega}\right]\right)$$

when $h(\cdot)$ is nonlinear. Therefore, rather than first constructing the posterior mean of θ_j and then applying the policy function, I evaluate the policy function at each posterior draw and average the resulting policy moments using the posterior weights. The two approaches coincide when the policy function is linear in θ_j.

Monte Carlo Approximation

Because these posterior moments are ratios of integrals without closed-form solutions, I approximate them using Monte Carlo integration under the estimated cross-site distribution. I draw

$$\theta_{js} \sim \mathcal{N}\left(\widehat{\theta}_0, \widehat{V}_0\right), \qquad s = 1, \ldots, S.$$

For each draw, I evaluate the weighted within-site likelihood:

$$L_{js}^C = L_j^C\left(\mathcal{D}_j \mid \theta_{js}, \widehat{\Psi}, \widehat{\Sigma}\right).$$

I then construct the normalized posterior simulation weights

$$\omega_{js} = \frac{L_{js}^{C}}{\sum_{r=1}^{S} L_{jr}^{C}}.$$

The simulated empirical Bayes posterior mean of any function $h(\theta_j)$ is

$$\widehat{h}_j^{EB,MC} = \sum_{s=1}^{S} \omega_{js} h\left(\theta_{js}\right).$$

The simulated posterior moments are therefore likelihood-weighted averages of functions evaluated at draws from the estimated cross-site distribution. As $S \to \infty$, the simulated posterior moments converge to their corresponding exact empirical Bayes posterior moments for fixed estimated population-level parameters.