NBER WORKING PAPER SERIES

AN EVALUATION OF THE ALIEF INDEPENDENT SCHOOL DISTRICT JUMPSTART PROGRAM: USING A MODEL TO RECOVER MECHANISMS FROM AN RCT

Flavio Cunha
Qinyou Hu
Andrea Salvati
Kenneth Wolpin
Rui Zeng

Working Paper 33537
http://www.nber.org/papers/w33537

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2025

We thank the Family and Community Engagement Department of the Alief Independent School Districts for its support of the study, especially Shanceler Terry and Laura Escamilla. We also thank the Alief Family Liaisons for their help in recruiting families for the study and Yanitza Velez, our project coordinator. A grant by the Laura and John Arnold Foundation supported this research. Cunha also acknowledges support from the National Institutes of Health, grant 1R01HD073221-01A1. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

An Evaluation of the Alief Independent School District Jumpstart Program: Using a Model to Recover Mechanisms from an RCT
Flavio Cunha, Qinyou Hu, Andrea Salvati, Kenneth Wolpin, and Rui Zeng
NBER Working Paper No. 33537
March 2025
JEL No. C51, I24, I32, J15

ABSTRACT

This paper evaluates the Jumpstart Program (JSP), a parenting intervention implemented by a school district in the Houston area to enhance school readiness among economically disadvantaged three-year-old children. Unlike many early childhood programs typically tested in controlled research settings, JSP leverages existing school district resources for scalability and practical application. We conducted a three-year randomized controlled trial to measure the program's impact on child cognitive outcomes, parental engagement, and mechanisms of change. The results indicate improvements in children's performance on curriculum-aligned assessments and modest gains in general cognitive readiness as measured by the Bracken School Readiness Assessment. Furthermore, treatment group parents demonstrated increased reading frequency with their children, underscoring enhanced parental involvement as a crucial mechanism behind the program's success. We employed a structural model to analyze both the direct effects of JSP and its indirect effects through changes in the marginal productivity of investments or preferences via habit formation. Our analysis concludes that 75% of the program's impact is attributed to direct effects, while 25% is mediated through changes in habit formation in parental investments. Our research underscores the potential of scalable, real-world interventions to bridge socio-economic gaps in early childhood development and inform the design of effective educational policies.

Flavio Cunha
Department of Economics
Rice University
P.O. Box 1892
Houston, TX 77251-1892
and NBER
flaviocunha@rice.edu

Qinyou Hu
Department of Economics
School of Business
Aalto University
PO Box 21210
Aalto
Finland
hdfhqy@gmail.com

Andrea Salvati
Geneva School of Economics and Management
Institute of Economics and Econometrics
University of Geneva
Uni-Mail, 40 Boulevard du Pont d'Arve
CH-1211 Geneva 4
Switzerland
andrea.salvati@unige.ch

Kenneth Wolpin
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6209
kennethwolpin@rice.edu

Rui Zeng
Department of Economics
Rice University
6100 Main St.
Houston, TX 77005
rz32@rice.edu

1 Introduction

Recent research highlights persistent disparities in school readiness at the onset of kindergarten, particularly across socio-economic groups. Reardon and Portilla (2016) estimate that among children born in 2010, income-related gaps in reading and math achievement are approximately one standard deviation. Moreover, these achievement disparities remain remarkably stable as children progress through school grades (Bond and Lang, 2013). The difficulty of addressing these early academic deficits after formal schooling begins has motivated researchers and policymakers to focus on programs designed to foster early childhood development.

A growing body of empirical evidence suggests that these income-related disparities in early academic achievement are partially rooted in differences in the home environment. Despite some progress over the past two decades, disadvantaged children continue to face significant deficits in both the quantity and quality of early learning opportunities. These gaps are evident in the time devoted to learning activities (Kalil et al., 2012), access to learning materials (Bassok et al., 2016), and engagement in routine activities that promote quality interactions with parents and other adults, such as family dinners (Putnam, 2016). Theoretical research underscores the potential of early interventions targeting malleable, foundational skills during sensitive developmental periods in early childhood to produce significant long-term impacts on human capital formation (Cunha and Heckman, 2007). Supporting this view, empirical studies have demonstrated that small-scale parenting interventions, when implemented by highly trained staff, can enhance parental investments and substantially improve child development outcomes (e.g., Gertler et al., 2014, 2021).

In contrast, evidence about the impact of large-scale parenting interventions is mixed (see Kalil, 2014, for a summary of the literature). Heckman et al. (2017) analyzed the Nurse Family Partnership Program in Memphis, TN, and found that by the program's conclusion (when children were two years old), mothers served by nurses demonstrated higher levels of home investment, better parenting attitudes, and improved mental health. These changes translated into positive impacts on many measures of child development at age 6 and as late as age 12. Additionally, recent studies focused on the translation and cultural adaptation of the Jamaican Program show promise that large-scale parenting programs can positively impact child development and the quality of the home environment (e.g., Meghir et al., 2023; Zhou et al., 2023).

However, such successes are not the rule for studies conducted in the United States. For example, Love et al. (2005) found that the home visitation component of Early Head Start achieved only modest impacts on family investments or child development.[1] Similarly, St. Pierre et al. (2003) report that the Department of Education's Even Start Program, which costs $13,674 per family, yielded no significant impacts on parental investments or child development outcomes. Evaluations of the Home Instruction for Parents of Pre-school Youngsters (HIPPY)—a widely adopted two-year program implemented in over 100 U.S. locations and 14 countries, costing between $1,500 and $2,000 per family annually—also show mixed results. While one small-scale randomized trial involving 69 HIPPY participants demonstrated substantial improvements in test scores, these results were not replicated in a randomized trial involving a similar cohort of families entering the program the following year (Baker et al., 1998, 1999).

In this paper, we evaluate a parenting program independently implemented by a school district using its own staff and resources. The program under review is the Jumpstart Program (JSP) from the Houston-area Alief Independent School District (AISD). JSP is a parent-directed group- and center-based program that provides training and resources to support home-based learning that aims to prepare three-year-old children for entry into AISD's Pre-Kindergarten program. The program lasts seven months and the families meet three times per month with the AISD's Family Liaison at their local elementary school campus. The children also attend one meeting per month, and, during that meeting, they interact with the Family Liaison. Unlike pilot programs typically implemented by highly trained research staff, the JSP represents a scalable intervention executed by a school district with standard resources and personnel. Thus, our paper fills an important gap in the literature by asking: Can scalable programs implemented in real-world settings improve the home environment and enhance the school readiness of economically disadvantaged children? And, if so, what are the mechanisms through which such a program operates?

We conducted a three-year randomized controlled trial (RCT) to answer these questions. The RCT evaluation of the JSP included families residing within the catchment areas of all 24 elementary schools in AISD. Over the evaluation period, the program served an average of approximately 200 families per year, with an additional 200 families assigned to the control group. At baseline, families were offered twenty-five dollars to participate in the pre-program assessments, with 90–95% of those registering for the Jumpstart Program agreeing to participate. Families were offered fifty dollars to participate in the post-program assessments; depending on the year, between 77-85 percent of those families agreed to participate. Not only was attrition relatively low, but the treatment and control families in the post-program sample did not differ in observable characteristics. Importantly, all families who registered for the program were included in the lottery randomization, regardless of whether they consented to participate in the study.

[1] See also Gelber and Isen (2013).

Our study used two assessments at baseline and endline. One assessment, designed by AISD staff, is based explicitly on the Jumpstart curriculum. The second, the Bracken School Readiness Assessment, is a nationally normed test of five concepts, only one of which coincides with the Jumpstart curriculum content. The additional assessment, specific to the evaluation, provides evidence on whether, due to the Jumpstart training, parents choose to teach their children more advanced concepts and skills than those in the Jumpstart curriculum.

The importance of a comparison group in evaluating the JSP is especially clear in the case of early childhood interventions. Young children are continually learning new things. In the case of the material covered in the Jumpstart curriculum, the test scores of the control group children increased from 50 percent at the baseline to 68 percent at the end of the program eight months later. Thus, even if the JSP had no impact, one would expect an 18 percentage point increase in the score of those in the program. Based on a difference-in-difference regression, the gain in the score of the children in the program was 7 percentage points higher than for the control group children. Program gains were particularly large for two of the Jumpstart modules, name recognition and book handling, with the treatment effect being over 20 percentage points for each. We show that our findings about the impact of JSP on the AISD assessment tool is robust to attrition.

The effect of the program on Bracken test scores was considerably smaller in absolute terms. Over the program period, the Bracken score increased from 23 to 34 percent for the control group. Our analysis showed that the JSP increased the Bracken score by a little over 2 percentage points. Using a different metric, the JSP reduced the percent of children defined by the Bracken score as either delayed or very delayed by 4 percentage points, about double the reduction for the control group (about one-half of the control group children were in this category at the baseline). Thus, there was some spillover in learning beyond the JSP curriculum. Unfortunately, these results are not robust when we consider flexible approaches to account for attrition.

In addition to the child assessments, the parents enrolled in the evaluation study also completed a questionnaire both at baseline and endline. Besides the usual demographic information, we collected data on the number of days the parent read to the child, and on whether family members helped the child learn the alphabet, colors, numbers, and shapes. At baseline, on average, control-group parents read to the child three days a week, and between 80 and 90 percent of the families reported helping their child with the alphabet, colors, and numbers, and 67 percent with shapes. Control-group families increased their average number of days spent reading to the child by a little over one day, and the other activities increased to 89, 99, 96, and 83 percent. Because family engagement in these latter activities became almost universal, there was essentially no difference between the treatment and control groups at the end of the program period. However, the number of days spent reading to the child increased by about one-half a day more for the treatment group. This result is suggestive that the increase in parent-child interactions (especially through reading to the child) may have been an important mechanism underlying the impact of the JSP. Indeed, many of the JSP curriculum items utilized children's books supplied to the parents.

Our study extends the scope of traditional evaluations by exploring the potential mechanisms driving the effectiveness of the Jumpstart Program (JSP). Specifically, we examine whether JSP has a direct effect on school readiness. Several considerations motivate this inquiry.

First, when children attend these periodic group- and center-based sessions, they gain access to a structured learning environment that fosters both academic and social development. During these sessions, children are exposed to cognitively stimulating activities, including guided play, storytelling, and interactive lessons led by trained educators. Such structured engagement helps them develop foundational skills—like language acquisition, early literacy, problem-solving, and attention regulation—that are critical for school readiness. Because these experiences occur at the center rather than at home, the developmental benefits are partly independent of how frequently parents engage in learning activities at home. In other words, by participating in these sessions, children gain an additional channel through which essential school-readiness skills can be cultivated—complementing (but not requiring) ongoing parental involvement.

Second, when parents receive training in early childhood development, they become more aware of the kinds of interactions that foster a child's cognitive, social-emotional, and language growth. This knowledge can reshape how parents plan and carry out daily routines, even if the total number of parent-child interactions stays roughly the same. For example, a parent might learn that reading time is most effective when they encourage the child to ask questions, point to pictures, and make predictions about the story—rather than just passively listening. Importantly, these qualitative improvements can enhance the richness and depth of the learning environment at home. Even if parents do not increase the *quantity* of time they spend with their child, these more thoughtfully structured sessions can have a disproportionately positive effect on the child's development, leading to stronger school readiness outcomes.

Third, the provision of tangible materials, such as books, fundamentally changes the nature of learning opportunities in the home. These materials make it easier and more appealing for both children and parents to engage in meaningful, skill-building activities. Importantly, these benefits go beyond simple measures of how *often* parents interact with their children. Even if the frequency of interactions remains unchanged, the *quality* of each interaction may improve due to the ready availability of developmentally appropriate materials. In other words, a single reading session with a new book can be more beneficial than multiple sessions with the same book.

Although our findings show that JSP parents interact with their children more frequently than do control parents, the relevance of this parental activity as a mechanism underlying the JSP treatment effect depends on the extent to which JSP induces parents to interact with their children and how these interactions, in turn, improve the child's skills. Evidence supporting such an effect appears in the literature on human capital production functions (see, for example, Todd and Wolpin, 2007; Cunha et al., 2010; Agostinelli and Wiswall, 2016; Attanasio et al., 2020). To determine the importance of this potential mechanism, we consider three indirect mechanisms operating through the parental investment channel.

First, we hypothesize that the provision of training and materials may increase the marginal productivity of parental investments. An increase in productivity can alter the frequency of parent-child interactions in two opposing ways. On the one hand, higher marginal productivity induces a *substitution effect*, motivating parents to spend more time interacting with their children since each unit of time is now more productive. On the other hand, higher productivity also leads to an *income effect*, which may reduce the time parents devote to such interactions. Because the net outcome is the sum of these two effects, the total impact on interaction frequency can be negative, zero, or positive.

Second, our paper introduces a novel mechanism in which parenting programs affect skill formation through increased parental investments. This mechanism relies on the concept of *habit formation*, wherein past investment patterns influence current preferences and, consequently, current investment choices. In our model, habit formation implies that a parent's utility from investing in their child depends not only on today's investment but also on prior investments. Thus, once established, investment routines are difficult to change. We argue this mechanism is relevant for two main reasons. First, JSP is a structured seven-month program that helps families establish parent-child interaction routines, such as regular reading times, which may be habit-forming. Second, these habits can be reinforced by the program's group-based component, as social expectations and peer interactions can further entrench such routines (sometimes referred to as "keeping up with the Joneses").

Third, we investigate whether JSP affects parent-child interactions by enhancing parental beliefs about the impact of investments on school readiness. Since the elasticity of parental investments with respect to parental beliefs can be substantial (Cunha et al., 2022), even small shifts in these beliefs can lead to more meaningful, developmentally appropriate activities at home. The current research on whether parent interventions work by increasing parental beliefs has produced mixed findings. For example, Cunha et al. (2024) find that the impact of the LENA Start Program on language interaction is mainly due to its impact on parental beliefs. In contrast, Attanasio et al. (2024) do not find evidence that parental beliefs mediates the impact on child development in their analysis of the adaptation of the Jamaica Home Visitation Program. List et al. (2021) provides evidence that the belief channel depends on the intensity of the intervention. Our data on parental beliefs suggest that JSP did not change parental expectations about the returns to investments in child development. For this reason, we do not include this channel in our structural analysis.

We develop and estimate a structural model to decompose the effects of the program. After outlining the conditions required for identification, we estimate the structural parameters necessary for mediation analysis. Our model decomposes the total effect into four components: a direct effect, an indirect effect through productivity, an indirect effect through preferences (i.e., habit formation), and an indirect effect arising from the interaction between productivity and preferences. Our findings indicate that JSP does not operate via the indirect productivity channel (alone or through its interaction with preferences). Instead, approximately 75% of the total effect is attributed to the direct effect, while about 25% is explained by the habit formation mechanism.

Our work is closely related to Attanasio et al. (2020), who examine the short-term impacts of the Jamaica home visitation program when scaled up in Colombia. Leveraging data from a randomized controlled trial (RCT), they estimate production functions for cognitive and socio-emotional skills that can differ between treatment and control groups. Notably, they find no differences in the marginal productivity of parental investments across these groups—an outcome that parallels our own findings.

We extend their analysis by introducing a model that incorporates habit formation. Our empirical results suggest that parenting programs may raise the frequency of parent-child interactions through this habit-formation channel, thereby offering a new perspective on how program-induced changes in parental behavior can persist over time.

The rest of the paper is organized as follows. Section 2 briefly describes the Alief Independent School District Jumpstart Program. Section 3 discusses the evaluation study, including the recruitment of the target population, randomization strategy, measurements, randomization results, attendance, and the analytical approach. Section 4 presents the results of the evaluation study. Section 5 develops the details of the analysis to uncover the JSP's mechanism, including the model, identification, estimation, and results. Section 6 concludes.

2 The Jumpstart Program

The JSP was created and designed by Alief Independent School District (AISD) staff. AISD is located in southwest Houston, Texas. In the 2017-18 academic year, AISD served over 46,000 students, 83% of which are economically disadvantaged and 43% of which are English Language Learners. Since 2015, our team has been working with the Department of Family and Community Engagement (FACE) in the evaluation of the JSP.

To be eligible for the JSP, parents must live within the AISD catchment area and must have a child who is between 36 and 47 months old on September 1 of each academic year. The JSP targets low-income families: across the 24 elementary schools in the AISD network, 71% of the children come from families that qualify for free or reduced-price lunch. Over 50% of the families that apply to the JSP receive food stamps. The JSP serves minority families: 57% of the parents are Hispanic, and 25% are black. The parents that apply to the JSP tend to have low education. For example, 35% of the parents have completed less than twelve years of schooling.

The twenty-two-week curriculum was developed by the district specifically to be aligned with the district's pre-K program to ensure that children have the necessary foundational skills upon entering pre-K. The week-by-week Jumpstart curriculum, which includes learning colors, acquiring fine motor skills, counting, name recognition, and book handling, is summarized in Appendix A. Three times each month, over an eight-month period, parents meet for one hour with a family liaison at their local elementary school in a group setting (there is one additional meeting at the end of the program year). The children are present at the third meeting of each month, a frequency which not only provides a chance for the child to interact directly with the liaison, but also enables the family liaison to assess parent fidelity to the curriculum. Thus, the JSP guides parents to teach foundational skills for the AISD Pre-K program and the Family Liaisons model behavior so parents can use the JSP materials with their children at home. The cost of the JSP, accounting for family liaison's pay, the cost of materials, and the opportunity cost of the parents, is a little under $450 per family per year.[2]

School districts and not-for-profit organizations implement similar parenting programs in the

[2]The program has three inputs: Family Liaison time, parental time, and materials. The program requires family liaisons to dedicate two hours per week: one hour for preparation and one hour for leading the session with the parents. Over the 22-week program, this effort equates to 44 hours per liaison, serving approximately 10 families. At the hourly rate of $9, the additional cost per family for liaison time is approximately $40. Regarding parental involvement, our data indicate that parents spend an additional 6.6 hours interacting with their children. Additionally, parents are required to spend one hour per week traveling to and from sessions, totaling 22 hours of travel time. Thus, the total time commitment from parents is 28.6 hours. We estimated log wage regressions controlling for selection into employment, and we found that $7.25 per hour is a good proxy for the parents' opportunity cost; the opportunity cost is approximately $207.35 per family. The program also requires $200 for materials. Consequently, the total estimated cost of the program per family amounts to just a little under $450.

United States. For example, the Houston Independent School District, which is the largest in the Houston area, implements the Home Instruction for the Parents of Preschool Youngesters (HIPPY). HIPPY is a home-visiting initiative designed to enhance school readiness for children aged 2 to 5 by empowering parents to be their children's first teachers. It focuses on families in underserved communities, particularly those with limited access to educational resources or where parents may lack prior teaching experience. Trained home visitors from the local community conduct regular visits, typically weekly or biweekly, providing parents with developmentally appropriate educational materials, including storybooks, structured learning activities, and role-playing exercises. The curriculum emphasizes active learning and covers areas such as language development, problem-solving, early literacy, and social-emotional skills.[3] HIPPY's cost (personnel plus materials) is around $1,548 to $2,322 per family per year.[4]

The Parents as Teachers (PAT) program is a home-visiting initiative that supports parents in fostering their child's early development and school readiness, serving families from pregnancy to kindergarten. Through personalized visits, the program provides guidance on child development, strategies for creating a nurturing learning environment, and activities to support language, social-emotional, and motor skills. PAT also offers developmental screenings to identify potential delays early, group connections for peer support, and resources to link families with local services. Its overarching goal is to strengthen parent-child relationships and empower parents as their child's first teacher. The PAT program costs approximately $2,663 per family per year.[5]

In developing countries, the Jamaican Early Childhood Development Intervention, initiated in the 1980s, was a pioneering home-visiting program aimed at enhancing the cognitive, language, and psychosocial skills of young children from disadvantaged backgrounds. This intervention involved community health aides conducting weekly home visits over a two-year period, teaching parenting skills and encouraging interactive play between mothers and their children. The program demonstrated significant long-term benefits, including improved cognitive development, educational attainment, and increased earnings in adulthood (Gertler et al., 2014, 2021). The success of the Jamaican model has inspired its adaptation and replication in various countries, including Colombia (Attanasio et al., 2014), India (Meghir et al., 2023), and China (Zhou et al., 2023), each tailored to local contexts. Zhou et al. (2023) estimates these programs cost approximately

[3]See Westheimer (2003) for details about the HIPPY Program.

[4]The website https://www.idra.org/families-and-communities/hippy/ reports that the cost of the HIPPY program ranged from $1,000 to $1,500 per family per year in August 2006 dollars. Using the Bureau of Labor Statistics CPI Inflation Calculator, we adjusted this range to reflect costs in October 2024 dollars.

[5]See Corso et al. (2022). We used the Bureau of Labor Statistics CPI Inflation Calculator to adjust for October 2024 dollars.

$661 per child per year.[6]

3 Data and Methods

3.1 Recruitment

The evaluation study of the JSP began with the 2016/17 school year and continued for an additional two years, 2017/18 and 2018/19. Prior to the start of the study, about eight district elementary schools provided the JSP to families within their catchment area. Total enrollment throughout the district amounted to about 60 families. To participate in the program, each school had to cover the cost of the materials for their program, $200 for each family, from their own budget. The cost of the program to the schools was a major impediment to a school adopting Jumpstart or expanding it to more families. As part of the evaluation study, the Laura and John Arnold Foundation provided the money needed to expand the program to all 24 elementary schools in all three years.

Prior to the study, AISD family liaisons and professionals that administer the program to the families recruited families on a first-come, first-serve basis. Eligible families were those that had a child who was three years old as of September 1 of the academic year. Recruitment ended when a pre-set target number of families, determined by the amount of money each individual school had allotted to the program, was reached. To implement the randomized controlled trial, the process of recruiting families for the study was changed.

In the first step, at the recruitment meeting with the family liaison at the school, the family representative was informed that there was an oversubscription to the program and that, to be fair, participants in the program would be chosen through a lottery. Any family desiring to be in the program had to agree to be in the lottery; we are not aware of any families declining the program because of the lottery. After agreeing to be in the lottery, a Jumpstart application form was completed by the person who came with the child, almost always the child's mother, and the child was administered a baseline assessment, the Jumpstart test (designed by AISD), that covers the content of the Jumpstart curriculum. The family was then informed that the Texas Policy Lab at Rice University was engaging in a study of the JSP to assess whether the program was achieving its aims. The family was told about their role in the study and offered twenty-five dollars as an incentive to participate. Families that agreed to participate were consented pursuant to IRB regulations. Participation in the study required the completion of a survey instrument and the administration of another child assessment, the Bracken School Readiness Assessment, a

[6]October 2024 dollars. We used the Bureau of Labor Statistics CPI Inflation Calculator to account for inflation.

nationally normed test of five concepts; only one concept coincides with the Jumpstart curriculum content.

The recruitment strategy in the first year of the study was based on the assumption that interest in the Jumpstart Program was sufficiently great in all of the 24 school catchment areas that a recruitment target of 20 families per school could be met. With the randomization, 10 families would be assigned to the treatment sample (the JSP participants) and 10 to the control sample (the program non-participants). Twenty of the twenty-four schools reached the 20-family enrollment target for the lottery; in all, 461 families enrolled in the lottery. Although the first-year recruitment in terms of the enrollment numbers was clearly successful, in part because family liaisons were given a monetary incentive to reach the 20-family goal, there was not uniformly high interest among the families across the schools. In half of the schools, at least one family who received a spot in the Jumpstart Program through the lottery never attended a Jumpstart session. Overall, 10 percent of the families never attended a Jumpstart session, and 25 percent attended five or fewer sessions out of the 22 total sessions.

In response to this outcome, we modified the recruitment strategy in the second and third years of the study. Liaisons were provided a monetary incentive if they recruited 14 families and a group incentive if they recruited additional families. In the second year, in total, 389 families entered the lottery, spanning 24 schools. In the third year, two family liaisons left the district just before the start of the academic year and were not replaced. Over the 22 remaining schools in the third year, 366 families entered the lottery. In contrast to the first year, in the second year, only 4 schools had a family that never attended a Jumpstart session, accounting for only 2 percent of the families, and only 13 percent attended fewer than 5 sessions. However, in year three, for unclear reasons, attendance declined; 8 percent of the families never attended a session, and 19 percent attended five or fewer sessions, not quite as problematic as the first year but significantly worse than the second year.

The results of the recruitment are shown in Table 1. Participation in the study was high in all three years. In the first year, 89.5 percent of control families and 90.9 percent of treatment families that agreed to be in the Jumpstart lottery also consented to be in the evaluation study.[7] There was even greater participation in the second and third years, 93.7 percent of the control families and 92.5 percent of the treatment families in year 2 and 96.0 percent of the control families and 94.2 percent of the treatment families in year 3. Cumulatively over the three years, 1,216 families entered the Jumpstart lottery, of which 1,126 participated in the study, a refusal rate of only 7.4 percent.

[7] Recall that all families, regardless of whether they agreed to be in the study, were entered into the lottery.

Table 1 also shows the results of the end-of-year (about 8 months after the start of the program) recruitment. As seen, attrition from the study fell each year. Noting that we attempted to contact only those families that had agreed to be in the study at the baseline, 71.3 percent of them in the first year participated in the end-of-year re-interview, 82.3 percent in the second year, and 84.8 percent in the third year. More importantly, in the last two years, the rate of attrition from the study, which is higher for the control group, fell significantly more for the control group than for the treatment group. In year one, the attrition rate (refusals and contact failure) was 18.5 percentage points higher for the control group. In years 2 and 3, the difference in the refusal rate fell to 9.4 and 6.0 percentage points.

3.2 Randomization

The randomization was conducted using a standard stratified block design, the same in all three years. Families were divided into four groups based on whether their score on the Jumpstart baseline test was above or below the mean of the lottery sample and on whether the child's age in months was above or below the mean of the lottery sample. Families were randomly selected within each school and blocking group. The number of families chosen within each school to participate in the program was constrained to be one-half of the total number of lottery families in the school, with the qualification that when the total number of families was an odd number, the additional family was placed into the treatment group.

3.3 Measurements

The data for this study include measures from two rounds of tests—before and after the program (baseline and endline survey)—of three cohorts. The two key measures for early childhood development are the Jumpstart test and the Bracken test, which we detail below. We also collected information on parental involvement as well as demographics on both children and parents.

Jumpstart Test Score The AISD Jumpstart Assessment Tool takes about 10-15 minutes and measures a child's knowledge of colors as well as gross and fine motor skills, which is aligned with the program content. In total, the test module has 26 questions, including the following sub-modules: colors (10 questions), fine motor skills (8 questions), counting (2 questions), sorting (1 question), name recognition (1 question), and book handling (4 questions).

Table 1: Sample Sizes

	In Jumpstart (Treatment Group)	Not In Jumpstart (Control Group)	Total
Year 1			
In Lottery			
Number	232	229	461
Percent	50.3	49.7	100
In Study			
Baseline			
Number	211	205	416
Percent	50.7	49.3	100
Percent of Lottery	90.9	89.5	90.2
End of Year			
Number	170	127	297
Percent	57.2	42.8	100
Percent of Lottery	73.3	55.4	64.4
Percent of Baseline	80.5	62.0	71.3
Year 2			
In Lottery			
Number	199	190	389
Percent	51.2	48.8	100
In Study			
Baseline			
Number	184	178	362
Percent	50.8	49.2	100
Percent of Lottery	92.5	93.7	93.1
End of Year			
Number	160	138	298
Percent	53.7	46.3	100
Percent of Lottery	80.4	72.6	76.6
Percent of Baseline	86.9	77.5	82.3
Year 3			
In Lottery			
Number	190	176	366
Percent	51.9	48.1	100
In Study			
Baseline			
Number	179	169	348
Percent	50.8	49.2	100
Percent of Lottery	94.2	96.0	95.1
End of Year			
Number	157	138	295
Percent	53.2	46.8	100
Percent of Lottery	82.6	78.4	80.6
Percent of Baseline	87.7	81.7	84.8

Note. This table shows the sample sizes for treatment group, control group, and total numbers for three years respectively. For each year, we report the number of families with lottery and the percentage with respect to the whole sample. We also report the number of families and percentages in our study sample for both baseline and endline (end of year). Notice that, for in study samples, we present the percentage of families with lottery by dividing the number of families in study sample by the number in lottery sample. For endline, we also report the percentage of families exist in the program by dividing baseline numbers.

Bracken Test Score The Jumpstart program also used the Bracken School Readiness Assessment – Third Edition (BSRA-3) designed by Bracken (2007) to assess the children's preschool readiness. BSRA-3 is an individual cognitive test for children Pre-K through second grade. It helps the parents determine if their child is ready for school by measuring a child's exposure to concepts necessary for learning at school. BSRA-3 includes five subtests (85 questions in total) to assess basic concepts related to school readiness, and all of the subtests reflect receptive and expressive language skills. The subjects include the following aspects: 1. Colors (10 questions): Identify common colors by name. 2. Letters (15 questions): Identify upper-case and lower-case letters. 3. Numbers/Counting (18 questions): Identify single- and double-digit numerals, and count objects, Size/Comparison. 4. Sizes (22 questions): Demonstrate knowledge and understanding of words used to depict size (such as tall or wide) Match or differentiate objects based on a specific characteristic. 5. Shapes (20 questions): Identify basic shapes by name.

The test is individually administered, so the length of test time depends upon the child. Concepts are tested using oral and visual means. If missing five questions in a row in one module, then the children will go directly to the next module. For both baseline and endline test results, we additionally construct two measures besides the number of correct answers. The first is the percent score, which is calculated by dividing the number of correct answers by the total number of questions. The second is an indicator of whether the child was not delayed in school readiness based on the total number of correct responses to all the questions. More specifically, if the child answered one question correctly, she gained one point. We count the total number of points the child gained. To check which kid delays and which does not, we follow Bracken (2007) with the criteria of the maximum SRC Raw score of delay based on the monthly age. Kids who score more than the number are not delayed. As an example, for a child aged between 36 and 38 months, based on the SRC Raw score of the age in the nationally representative sample, he/she will be considered NOT delayed if he/she answers more than 12 corrected questions.

Parental Involvement We collected additional survey measures for parental involvement. Families that agreed to be in the study were administered a baseline and an endline survey. The survey included questions concerning the extent to which family members engaged in learning activities with their children. The activities include the number of days/week read to the child, having 10 or more children's books, helping a child learn the alphabet, helping a child learn colors, helping a child learn numbers, and helping a child learn shapes. While the first variable is a continuous counting measure ranging from 0 to 7, the remaining variables are all indicators.

Other Demographic Information Our intervention also collected demographic information of participating families from both parents and children. For parents, we collected the participating guardian's age, completed yearly education, race (black, Hispanic, or white), language (English or Spanish), and whether they are employed. For children, we collected their monthly age at the beginning of the intervention, gender, and child language (English or Spanish). We also collected household-level measures, including the number of children in the household and an indicator of receiving food stamps.

3.4 Randomization Results

Table 2 provides information on the outcome of the randomization in each year for two separate samples. The full sample includes all families who consented to participate in the study at baseline. The analytical sample includes the families who completed both the baseline and endline surveys. For each variable shown in the first column of the table, the second and fifth columns show the mean in the control group (μ_C), the third and sixth columns display the difference in the value of the variable between the treatment group and the control group ($\Delta = \mu_T - \mu_C$), and the fourth and seven columns present the p-values associated with the hypothesis that Δ is equal to zero.

The first two variables in Table 2, the child's baseline Jumpstart test score and the child's age, are the blocking variables; the rest of the variables were not used in the randomization. As seen in the table, the differences between the treatment and control samples are small in magnitude (relative to the means) and have p-values well above conventional levels of statistical significance. This finding is generally true for the full and analytical samples. The only slight difference relate to the fraction of Hispanics (59.5% in the full sample and 62% in the analytical sample), and the percentage of families receiving food stamps (49% in the full sample versus 46.9% in the analytical sample).

3.5 Baseline Test Score Summary Statistics

Table 3 provides summary statistics of the baseline Jumpstart and Bracken raw test scores (percentage correct answers) for families in the post-test sample for the three years separately and pooled. The mean Jumpstart test score pooled over the three years is 51.0 percent. The range over the three years is only 5.1 percentage points, from a low of 48.8 percent in year 3 to a high of 53.9 percent in year 2. The standard deviation of the score is about 25 percentage points. Approximately 20 percent of the sample has a test score below 25 percent, and another 20 percent of the

Table 2: Comparison of Treatment and Control Groups: Full vs. Analytical Sample

	Full Sample			Analytical Sample		
Variable (Units)	μ_C	Δ	p-value	μ_C	Δ	p-value
Blocking Variables						
Jumpstart Baseline Test Score (%)						
Year 1	50.90	1.00	0.70	49.50	1.00	0.74
Year 2	53.40	0.20	0.93	53.70	0.50	0.86
Year 3	48.40	0.00	1.00	48.10	1.30	0.66
Pooled Years	50.90	0.40	0.78	50.40	0.90	0.59
Child Age at Test (Years)						
Year 1	3.58	0.15	0.61	3.57	0.02	0.52
Year 2	3.51	0.01	0.68	3.51	0.00	0.91
Year 3	3.46	0.00	0.92	3.48	-0.03	0.36
Pooled Years	3.52	0.01	0.65	3.52	0.00	0.96
Non-Blocking Variables						
Parent Years of Schooling						
Year 1	12.00	0.32	0.36	12.40	0.02	0.96
Year 2	12.00	0.13	0.73	11.80	0.46	0.27
Year 3	12.20	0.15	0.70	12.00	0.45	0.26
Pooled Years	12.10	0.21	0.33	12.10	0.32	0.18
Parent Hispanic (%)						
Year 1	60.50	-4.60	0.35	59.80	1.30	0.82
Year 2	59.60	-3.00	0.57	63.00	-6.10	0.29
Year 3	58.30	-0.20	0.97	64.00	-3.80	0.51
Pooled Years	59.50	-2.70	0.36	62.00	-2.90	0.39
Parent Age (Years)						
Year 1	32.10	0.21	0.75	32.70	-0.22	0.78
Year 2	32.10	0.73	0.38	32.30	0.73	0.41
Year 3	32.30	0.09	0.90	32.50	-0.15	0.84
Pooled Years	32.20	0.34	0.42	32.50	0.12	0.49
Number of Children in HH						
Year 1	2.84	0.07	0.58	2.82	0.15	0.30
Year 2	2.76	0.09	0.47	2.83	-0.03	0.84
Year 3	2.73	0.17	0.21	2.76	0.09	0.50
Pooled Years	2.78	0.11	0.15	2.80	0.07	0.36
Percent Receive Food Stamps (%)						
Year 1	53.00	-3.50	0.49	50.00	0.90	0.88
Year 2	48.30	-6.40	0.22	46.70	-8.00	0.17
Year 3	45.00	2.00	0.72	44.20	2.30	0.69
Pooled Years	49.00	-2.80	0.35	46.90	-1.50	0.66
Bracken Baseline Test Score (%)						
Year 1	24.60	1.20	0.54	23.30	0.50	0.83
Year 2	24.80	1.00	0.64	23.40	3.20	0.17
Year 3	22.30	1.00	0.60	22.50	0.80	0.70
Pooled Years	23.40	1.10	0.36	23.00	1.50	0.24
Employed (%)						
Year 1	39.20	-0.01	0.86	37.80	-0.02	0.74
Year 2	41.60	-0.06	0.27	39.10	-0.05	0.40
Year 3	29.00	0.03	0.57	26.80	0.02	0.64
Pooled Years	36.80	-0.01	0.65	34.50	-0.01	0.70
Numbers of books $>$ 10 (%)						
Year 1	38.00	0.00	0.96	39.40	-0.02	0.75
Year 2	41.00	-0.02	0.70	38.40	0.00	0.97
Year 3	43.80	-0.03	0.57	44.90	-0.04	0.54
Pooled Years	40.80	-0.02	0.58	40.90	-0.02	0.58

sample has 75 percent or higher.

The Bracken baseline test scores are considerably lower than the Jumpstart test scores. As seen, the mean baseline score on the Bracken is 23.9 percent averaged over the three years. As with the Jumpstart test, the range over the three years is narrow, 22.9 to 25.1 percent. The standard deviation of the pooled sample is 19 percent. We note that 50 percent of the children in our study are judged to be delayed or very delayed in their cognitive development.

Table 3: Baseline Test Scores: Study Participants in the Analytical Sample By Year

	Year 1	Year 2	Year 3	All Years
Jumpstart Test				
Mean Test Score (%)	50.1	53.9	48.8	51.0
Standard Deviation (%)	26.4	25.0	25.6	25.7
% Test Score<25.0%	20.6	16.4	21.7	19.7
% Test Score>= 75.0%	20.6	22.8	17.6	20.3
Bracken Test				
Mean Test Score (%)	23.6	25.1	22.9	23.9
Standard Deviation (%)	19.1	20.2	17.7	19.0
Percent Very Delayed or Delayed	52.4	48.7	51.2	50.7

Note. This table provides summary statistics of the baseline Jumpstart and Bracken raw test scores (percentage correct answers) for families in the analytical sample for the three years separately and pooled. For the Jumpstart test score, we report the mean scores, standard deviation, and percentage of observations that have scored lower than 25.0% and higher than or equal to 75%. For Bracken test scores, besides reporting mean scores and standard deviation, we report the percentage of delayed children by the criteria from Bracken (2007).

3.6 Jumpstart Attendance

As noted, Jumpstart lessons are held in group sessions within each school three times a month, with 22 lessons in total over the school year. The first two lessons each month are with the parents alone, and the third with the parents and their children. Parents who miss either or both of the first two lessons each month may make them up at another time. Parents who miss 3 lessons in a row without make-up are dropped from the JSP.

Table 4 shows selected statistics in each year from the distribution of completed (inclusive of made-up classes) weeks for the treatment group over the 22 weeks of the program. The mean number of weeks completed was 13.5 in year one, increased to 17.4 in year 2, and dropped slightly to 16.1 in year 3. These differences are also reflected in graduation rates, which require that a family complete 16 or more weeks out of the 22. In large part due to the recruitment strategy in year one, only 57.7 families completed the Jumpstart program in that year. Given the change in the recruitment strategy, 74.5 percent completed the program in year 2 and 71.6 percent in year 3. In years 1 and 2, about half of the families that didn't graduate completed 5 weeks or less, while that is true for about two-thirds of the families in year 3. More specifically, among the non-graduates, the mean number of weeks completed was 4.7, 6.4 and 4.3 for the three years .

Table 4: Jumpstart Program Completion By Year

	Year 1	Year 2	Year 3
Mean Weeks Completed	13.5	17.4	16.1
Pct. Completing			
0 weeks	10.3	1.6	8.4
1 – 5 Weeks	15.5	10.9	10.6
6 – 15 Weeks	18.5	13	8.9
Complete Program (16+ weeks)	57.7	74.5	71.6

Note. This table displays the Jumpstart program completion for three cohorts separately. We report the average weeks that families participate. We also report the percentage completing weeks: 0, 1-5, 6-15, and 16+ weeks respectively.

Table 5 compares participants who completed the Jumpstart program (i.e., attended at least 16 weeks) to those who did not. Two notable patterns emerge in the baseline test scores: completers have significantly higher Jumpstart test scores (0.53 vs. 0.45; difference $= -0.08$, $p < 0.01$) and Bracken test scores (0.26 vs. 0.21; difference $= -0.05$, $p = 0.03$). Child demographics, such as male gender (0.51 vs. 0.50) and age (3.51 vs. 3.54), exhibit no statistically significant differences. Among parent and household characteristics, factors like parental education, employment status, and age show no significant gaps. However, non-completers appear more likely to receive food stamps (50% vs. 38%; difference $= 0.12$, $p = 0.10$), are significantly less likely to have more than 10 books at home (24% vs. 40%; difference $= -0.15$, $p = 0.01$), and have an average of nearly half a child more in the household (3.21 vs. 2.73; difference $= 0.47$, $p < 0.01$). These findings suggest that completers enter the program with stronger baseline skills and a moderately more resource-rich home environment.

3.7 Evaluation Method

We use the following specification to estimate the effects of the Jumpstart intervention on school readiness and parental investments:

$$Y_{i1} = \alpha + \beta_1 T_i + \beta_2 Y_{i0} + X_i \gamma + \tau_s + \varepsilon_i, \tag{1}$$

Table 5: Characteristics of Jumpstart Program Compliers

	Incomplete	Complete	Difference	P-value
Baseline Test Scores				
Jumpstart Test	0.45	0.53	-0.08***	0.00
Bracken Test	0.21	0.26	-0.05**	0.03
Baseline Child Demographics				
Male	0.50	0.51	-0.01	0.92
Age	3.54	3.51	0.03	0.31
Baseline Parent Demographics				
Hispanic	0.57	0.58	-0.01	0.88
Black	0.28	0.24	0.04	0.44
High school graduate	0.65	0.71	-0.06	0.27
Age	28.42	28.93	-0.51	0.72
Employed	0.37	0.32	0.05	0.31
Baseline Household Demographics				
Food stamp	0.50	0.38	0.12*	0.10
Number of books > 10	0.24	0.40	-0.15**	0.01
Number of children	3.21	2.73	0.47***	0.00

This table compares individuals who completed the Jumpstart program with those who did not. We define completion as attending at least 16 weeks of the curriculum. $^{*}\mathrm{p} < 0.10,^{**}\mathrm{p} < 0.05,^{***}\mathrm{p} < 0.01$.

where Y_{i1} is a vector of test scores for child i at endline, T_i is the treatment group indicator for child i, which was assigned at the baseline, and ε_i is an error term that is independent from T_i. We also control for baseline test scores Y_{i0} and a set of demographic variables (X_i). [8] The strata fixed effects are denoted by τ_s. The main parameter we are interested in is β_1, which we interpret as the Intent-to-Treat (ITT) treatment effect parameter.[9]

4 Program Evaluation Results

4.1 The Impact of Jumpstart on Test Scores

Tables 6 and 7 present estimates of JSP's effects on the Jumpstart and Bracken test scores, respectively. In each table, the first column shows the results without controlling for lagged test scores or demographic characteristics. The second column includes a control for the baseline value of the relevant test score, and the final column adds demographic controls in addition to the baseline

[8]The demographics included in regressions are parental years of schooling, an indicator of the parent being Hispanic, parent age, an indicator of the parent being employed, number of children under 6 in the household, an indicator of getting food stamps, child age, and an indicator of a male child. Note that all variables are measured at the baseline. If missing, we replace it with a median value, and we also control for a missing indicator for that variable.

[9]Appendix C.1 presents the results from a 2SLS procedure in which the endogenous variable is the number of weeks the parent attended and the instrumental variable is the dummy variable representing random assignment to the JSP or control group.

test score.

In each table, the results are reported by "block," where each block corresponds to a different test module used as the dependent variable. The first block in every table shows the effects on the overall test (all modules combined). Next to each module name, the number of test questions is listed in parentheses, followed by the baseline and endline control group scores in brackets. As shown in Table 6, the Jumpstart test comprises twenty-six questions covering six modules: ten questions on color recognition, eight on fine motor skills (e.g., coloring a circle within the lines), two on counting, one on sorting similar objects, one on recognizing one's printed name, and four on book handling (e.g., identifying the title of a book).

The control group's mean baseline score for the overall test is 50.4 percent. The highest baseline score is obtained on the colors module, 58.4 percent, followed by the fine motor skills module, 56.8 percent. At the other extreme, only one-quarter of the children were able to recognize their printed names, and most children were able to answer only one of four questions about book handling. The counting and sorting modules lie in between, with a mean score of around 40 percent.

It should not be surprising that children in the control group learn without participating in the Jumpstart program. Indeed, the overall score for the control group increases by 18.3 percentage points, from 50.4 to 68.7 percent. The increases for the individual modules are: colors - 16.3 percentage points, Fine Motor Skills – 18.8 percentage points, Counting – 22.4 percentage points, Sorting – 24.3 percentage points – Name Recognition – 23.3 percentage points, Book Handling – 17.3 percentage points. Clearly, any before-after comparison for the children participating in the Jumpstart program would need to account for this growth.

We also report robust standard errors as well as standard errors clustered at school and family levels separately. In the last row of each block, we show the randomization inference p-value for each estimate. As seen in table 6, the three estimates of the impact of the Jumpstart program are similar in magnitude. Concentrating on the last column, the Jumpstart program is estimated to increase the overall test score by 7.2 percentage points (randomization inference p-value of .000) more than the increase in the test score of the control group. Among the modules, the impact of the program is estimated to be 3.1 percentage points (randomization inference p-value=.057) for Colors, 4.4 (randomization inference p-value=.003) for Fine Motor Skills, and 7.5 (randomization inference p-value=.015) for Sorting. The program increases all types of skills for the modules except for counting. Results are also robust using the alternative randomization inference p-values.

Most striking is the impact on the score for name recognition, 23.2 percentage points (randomization inference p-value=.000), and the impact on the score for book handling, 22.2 percentage points (randomization inference p-value=.000). Thus, for example, although less than one-half of the control group children can recognize their name on the post-test, about 70 percent of the treatment group can do so.

Table 7 shows the results for the Bracken test. The Bracken test consists of 85 questions contained in five modules that test knowledge of colors, letters, numbers, sizes, and shapes; testing in each module ends when the child has answered three consecutive questions incorrectly. Recall that the overall score on the baseline tests was only 23 percent, less than half of the score on the Jumpstart test. The percentage point gain in the Bracken test score for the control group was also smaller than that of the Jump-Start test, 11.1 percentage points overall. Only performance on recognizing colors was similar in level and gained to the Jumpstart test; the control-group baseline Bracken score was 56.4 percent, and the increase was 21.1 percentage points. The gains on the individual modules aside from colors were around 10 percentage points.

The estimated impact of the Jumpstart program on Bracken test scores is robust to controlling for various factors, although the effect sizes are generally larger without controls. Focusing on the fully controlled specification, the Jumpstart program increases the overall test score by 2.0 percentage points (randomization inference $p = 0.032$), relative to the control group's gain. Disaggregated by module, the estimated improvements are 4.3 percentage points (randomization inference $p = 0.022$) for *Colors*, 2.7 percentage points (randomization inference $p = 0.096$) for *Letters*, 2.6 percentage points (randomization inference $p = 0.107$) for *Numbers*, 1.3 percentage points (randomization inference $p = 0.778$) for *Sizes*, and 0.8 percentage points (randomization inference $p = 0.778$) for *Shapes*.

The last two blocks of Table 7 also present results for the national percentile score and for the percentage of children classified as delayed or very delayed. In the control group, the percentile score remains essentially unchanged, whereas Jumpstart participants' percentile scores rise by 3.1 percentiles (randomization inference $p = 0.025$). Meanwhile, although the proportion of delayed or very delayed children in the control group declines by 5.5 percentage points, the program is associated with a further 3.5-point reduction (randomization inference $p = 0.243$)—a decrease that is not statistically significant.

Table 6: Impact of Jumpstart Program on the Jumpstart Test (All Years)

Test Module (no. of questions):	ITT		
[control group baseline test score, control group end-of-year test score]	Model 1	Model 2	Model 3
All Modules (26): [50.4, 68.7]	7.2***	7.0***	7.2***
robust SE	(1.2)	(1.1)	(1.1)
cluster school level SE	(1.4)	(1.4)	(1.3)
cluster family level SE	(1.2)	(1.1)	(1.1)
Randomization inference p-value	0.000	0.000	0.000
Colors (10): [58.4, 74.7]	2.9*	3.0*	3.1**
robust SE	(1.7)	(1.6)	(1.5)
cluster school level SE	(1.4)	(1.5)	(1.5)
cluster family level SE	(1.7)	(1.6)	(1.5)
Randomization inference p-value	0.098	0.066	0.057
Fine Motor Skills (8): [56.9, 75.7]	4.1***	4.3***	4.4***
robust SE	(1.4)	(1.3)	(1.3)
cluster school level SE	(1.4)	(1.4)	(1.4)
cluster family level SE	(1.4)	(1.3)	(1.3)
Randomization inference p-value	0.006	0.003	0.003
Counting (2): [40.3, 62.7]	3.7	3.0	3.2
robust SE	(2.7)	(2.6)	(2.6)
cluster school level SE	(3.5)	(3.5)	(3.4)
cluster family level SE	(2.7)	(2.6)	(2.6)
Randomization inference p-value	0.185	0.270	0.238
Sorting (1): [42.9, 67.2]	7.4**	7.3**	7.5**
robust SE	(2.9)	(2.9)	(2.9)
cluster school level SE	(3.9)	(3.8)	(3.7)
cluster family level SE	(2.9)	(2.9)	(2.9)
Randomization inference p-value	0.012	0.015	0.015
Name Recognition (1): [24.1, 47.4]	23.3***	23.1***	23.2***
robust SE	(3.1)	(3.1)	(3.1)
cluster school level SE	(3.6)	(3.7)	(3.8)
cluster family level SE	(3.1)	(3.1)	(3.1)
Randomization inference p-value	0.000	0.000	0.000
Book Handling (4): [31.2,48.5]	22.2***	21.8***	22.2***
robust SE	(2.4)	(2.4)	(2.4)
cluster school level SE	(2.5)	(2.4)	(2.4)
cluster family level SE	(2.4)	(2.4)	(2.4)
Randomization inference p-value	0.000	0.000	0.000
Observations	890	890	890
Outcome variables at baseline	N	Y	Y
Demographics	N	N	Y

Note. This table shows the intention-to-treat (ITT) estimates of the program's impact on Jumpstart test scores at the endline. We report the ITT estimates on whether families are under treatment group with three models that are under different controls, with or without Jumpstart test scores at baseline and demographic variables. We report the ITT estimation results for end-of-year test scores for all modules and sub-modules separately. We also display the number of questions (in parentheses) and the baseline and endline test scores for the control group (in brackets) in the first column. For each regression, we report three different standard errors: robust, clustered at the school level, and clustered at the family level ($^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$). We also report the randomization inference p-value with the most conservative significance level. Note that for all regressions, we additionally control for the blocking group indicators (24 schools X 4 block groups).

Table 7: Impact of Jumpstart Program on the Bracken Test (All Years)

Test Module (no. of questions):	ITT		
[control group baseline test score, control group end-of-year test score]	Model 1	Model 2	Model 3
All Modules (85): [23.0, 34.1]	3.0**	2.0*	2.0*
robust SE	(1.1)	(0.9)	(0.9)
cluster school level SE	(1.2)	(1.0)	(1.0)
cluster family level SE	(1.1)	(0.9)	(0.9)
Randomization inference p-value	0.009	0.025	0.032
Colors (10): [56.4, 77.4]	4.6**	4.0**	4.3**
robust SE	(1.9)	(1.7)	(1.7)
cluster school level SE	(2.0)	(1.8)	(1.7)
cluster family level SE	(1.9)	(1.7)	(1.7)
Randomization inference p-value	0.026	0.034	0.022
Letters (8): [20.2, 28.2]	4.1*	2.7*	2.7
robust SE	(2.0)	(1.5)	(1.5)
cluster school level SE	(2.1)	(1.6)	(1.8)
cluster family level SE	(2.0)	(1.5)	(1.5)
Randomization inference p-value	0.046	0.096	0.096
Numbers (18): [11.9, 20.6]	4.3**	2.6*	2.6*
robust SE	(1.8)	(1.5)	(1.5)
cluster school level SE	(1.9)	(1.4)	(1.5)
cluster family level SE	(1.8)	(1.5)	(1.5)
Randomization inference p-value	0.025	0.105	0.107
Sizes (22): [20.3, 29.1]	1.3	1.3	1.3
robust SE	(1.2)	(1.1)	(1.1)
cluster school level SE	(1.3)	(1.2)	(1.1)
cluster family level SE	(1.2)	(1.1)	(1.1)
Randomization inference p-value	0.314	0.271	0.283
Shapes (20): [21.5, 34.9]	-0.5	1.2	0.8
robust SE	(2.9)	(2.6)	(2.6)
cluster school level SE	(2.3)	(2.6)	(2.6)
cluster family level SE	(2.9)	(2.6)	(2.6)
Randomization inference p-value	0.867	0.669	0.778
Percentile Score (85): [25.9, 26.0]	4.5**	3.1**	3.1**
robust SE	(1.7)	(1.3)	(1.3)
cluster school level SE	(1.7)	(1.3)	(1.4)
cluster family level SE	(1.7)	(1.3)	(1.3)
Randomization inference p-value	0.009	0.026	0.025
Pct. Very Delayed or Delayed (85): [51.0, 46.5]	-3.7	-3.7	-3.5
robust SE	(3.2)	(2.9)	(2.9)
cluster school level SE	(3.9)	(3.5)	(3.7)
cluster family level SE	(3.2)	(2.9)	(2.9)
Randomization inference p-value	0.249	0.217	0.243
Observations	888	888	888
Outcome variables at baseline	N	Y	Y
Demographics	N	N	Y

Note. This table shows the intention-to-treat (ITT) estimates of the program's impact on Bracken test scores at the endline. We report the ITT estimates on whether families are under treatment group with three models that are under different controls, with or without Jumpstart test scores at baseline and demographic variables. We report the ITT estimation results for end-of-year test scores for all modules and sub-modules separately. We also display the number of questions (in parentheses) and the baseline and endline test scores for the control group (in brackets) in the first column. We report two additional measures that rely on the data from the Bracken's normalizing national sample. They are the Percentile Score and the Percent of Children Delayed or Very Delayed. For each regression, we report three different standard errors: robust, clustered at the school level, and clustered at the family level ($^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$). We also report the randomization inference p-value with the most conservative significance level. Note that for all regressions, we additionally control for the blocking group indicators (24 schools X 4 block groups).

Table 8: Impact of Jumpstart Program on Parental Inputs — All Years

Input	ITT		
[control group baseline, control group end of year]	Model 1	Model 2	Model 3
Number of Days/Week Read to Child: [3.0, 4.1]	0.668***	0.567***	0.591***
robust SE	(0.150)	(0.143)	(0.143)
cluster school level SE	(0.148)	(0.126)	(0.126)
cluster family level SE	(0.150)	(0.143)	(0.143)
Randomization inference p-value	0.000	0.000	0.000
Has 10 or More Children's Books: [.41, .49]	0.216***	0.220***	0.224***
robust SE	(0.032)	(0.030)	(0.030)
cluster school level SE	(0.031)	(0.028)	(0.029)
cluster family level SE	(0.032)	(0.030)	(0.030)
Randomization inference p-value	0.000	0.000	0.000
Helps Child Learn Alphabet: [.81, .87]	0.021	0.018	0.021
robust SE	(0.021)	(0.021)	(0.021)
cluster school level SE	(0.020)	(0.020)	(0.021)
cluster family level SE	(0.021)	(0.021)	(0.021)
Randomization inference p-value	0.338	0.394	0.322
Helps Child Learn Colors: [.90, .98]	-0.008	-0.007	-0.006
robust SE	(0.011)	(0.011)	(0.011)
cluster school level SE	(0.011)	(0.011)	(0.010)
cluster family level SE	(0.011)	(0.011)	(0.011)
Randomization inference p-value	0.485	0.491	0.614
Helps Child Learn Numbers: [.88, .96]	0.003	0.004	0.006
robust SE	(0.014)	(0.014)	(0.013)
cluster school level SE	(0.015)	(0.015)	(0.014)
cluster family level SE	(0.014)	(0.014)	(0.013)
Randomization inference p-value	0.841	0.794	0.682
Help Child Learn Shapes: [.68, .83]	-0.041	-0.041	-0.040
robust SE	(0.026)	(0.026)	(0.026)
cluster school level SE	(0.024)	(0.025)	(0.023)
cluster family level SE	(0.026)	(0.026)	(0.026)
Randomization inference p-value	0.124	0.127	0.138
Observations	890	890	890
Outcome variables at baseline	N	Y	Y
Demographics	N	N	Y

Note. This table shows the intention-to-treat (ITT) estimates of the program's impact on parental inputs at the endline, which includes the number of days/weeks read to the child, whether the family has 10 or more books at home, and whether families help the child learn alphabet/colors/numbers/shapes. We report the ITT estimates on whether families are under treatment group with three models that are under different controls, with or without Jumpstart test scores at baseline and demographic variables. We also display the baseline and endline parental inputs for the control group (in brackets) in the first column. For each regression, we report three different standard errors: robust, clustered at the school level, and clustered at the family level ($^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$). We also report the randomization inference p-value with the most conservative significance level. Note that for all regressions, we additionally control for the blocking group indicators (24 schools X 4 block groups).

4.2 The Impact of Jumpstart on Parental Inputs

Our study, which features extensive data collection, allows us to investigate if participating in the Jumpstart program alters parental behaviors. Families who agreed to take part in the study completed baseline and endline surveys. These surveys collected information on the frequency of family engagement in various child learning activities. Table 8 summarizes the estimated impact of JSP on parental inputs, using a format similar to that of Tables 6 and 7. Specifically, Table 8 presents the intent-to-treat (ITT) estimates for several measures: the number of days per week parents read to their child, whether the household owns at least 10 children's books, and whether parents help the child learn the alphabet, identify colors, learn numbers, and learn shapes.

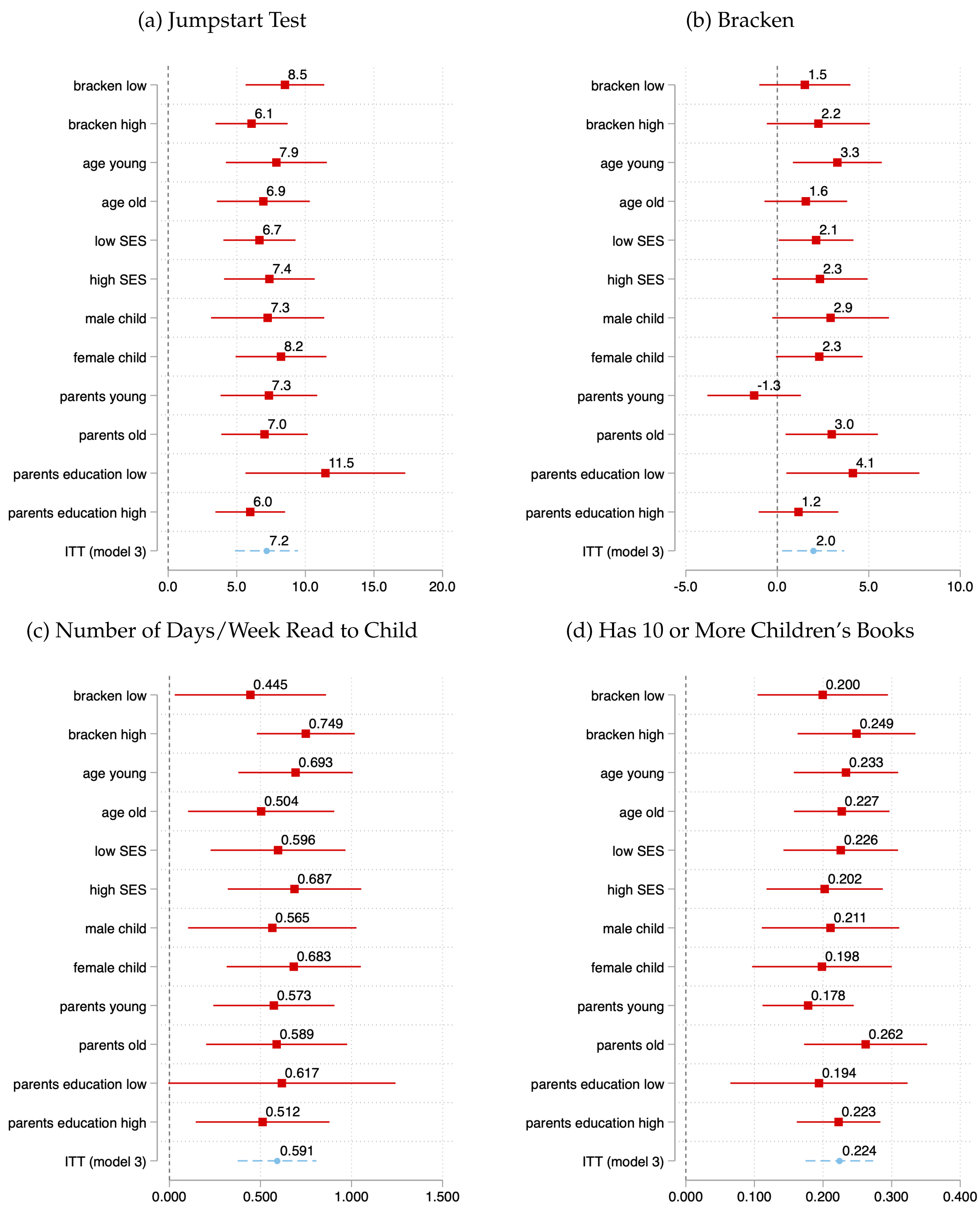
Parental inputs increased in the control group between baseline and endline. On average, the number of days someone in the family read to the child rose from 3.0 to 4.1, and the share of families owning at least 10 children's books went from 41% to 49%. In terms of parental help with specific skills, 80–90% of parents in the control group were already helping their child learn the alphabet, colors, or numbers at baseline; these figures increased by 8–10 percentage points by the end of the year. Fewer parents (about two-thirds) initially helped their children with shapes, but that proportion rose to 83.3% at endline.

As shown in Table 8, estimates across the three models remain relatively consistent. By the end of the program, JSP families spent more days reading to their children and were more likely to own at least 10 children's books compared to control families. Specifically, JSP families read to their child an average of half a day more per week, and an additional 22% of families reported owning 10 or more children's books. The latter impact is likely attributable to the JSP curriculum, which provided families with a dozen books; indeed, 72% of JSP families reported owning 10 or more books at endline. In contrast, we find essentially no impact on the other parental inputs measured. This result is not surprising, given that a high proportion of parents in both groups were already providing these activities at baseline.

4.3 Heterogeneous Treatment Effects

We examine whether JSP's effects differ according to participants' baseline characteristics. Specifically, we consider four outcomes: (i) the Jumpstart test score, (ii) the Bracken score, (iii) the number of days per week the child is read to, and (iv) whether the household has at least ten children's books. We employ the Generalized Random Forest Method (Athey et al., 2019) to identify the characteristics that most strongly predict treatment heterogeneity. We split the sample based on the covariates included in our most comprehensive specification (Model 3).

Figure 1: Heterogeneous Treatment Effects



Note. This figure shows the heterogeneous ITT estimates of JSP's impact on Jumpstart test scores, Bracken test scores, the number of days/week read to the child, and whether or not having 10 or more children's books at the endline. We report the estimates for different sub-samples of Bracken score, child age, SES, gender, parents' age, and parental education at the baseline. We also benchmark the results with the ITT estimates from Model 3 in Tables 6, 7, and 8, and the estimates are displayed at the bottom in blue with a dashed confidence interval. We plot the line of coefficient 0 as a reference. For each regression, we plot the estimated ITT (red dot) and 90% confidence intervals (red line).

Table 9: Sensitivity of the Intent-to-Treat Treatment Effect Parameter with Respect to Attrition

Outcome	Point Estimates				Lee Bounds	
	No Correction for Attrition	**Inverse Probability Weight**	**Balancing**	**Heckman Selection Model**	**Lower Bound**	**Upper Bound**
JumpStart Test Score	7.2***	7.1***	7.4***	6.2***	3.3***	11.3***
(Cluster school SE)	(1.3)	(1.4)	(1.6)	(1.8)	(1.1)	(1.5)
(Cluster Family SE)	(1.1)	(1.1)	(1.2)	(1.6)	(1.2)	(1.2)
(Robust SE)	(1.1)	(1.1)	(1.2)	(1.1)	(1.6)	(1.1)
Bracken Score	2.0*	1.9	2.3**	2.3	-1.2	4.7***
(Cluster school SE)	(1.0)	(1.2)	(1.0)	(1.9)	(1.0)	(1.1)
(Cluster Family SE)	(0.9)	(1.0)	(0.9)	(1.4)	(1.0)	(0.9)
(Robust SE)	(0.9)	(1.1)	(0.9)	(1.4)	(0.9)	(0.9)
Days/Week Parent Reads to Child	0.591***	0.587***	0.627***	0.590***	0.130	0.997***
(Cluster school SE)	(0.126)	(0.126)	(0.163)	(0.125)	(0.121)	(0.176)
(Cluster Family SE)	(0.143)	(0.143)	(0.157)	(0.144)	(0.165)	(0.163)
(Robust SE)	(0.143)	(0.143)	(0.157)	(0.144)	(0.156)	(0.156)
Family Has 10 or More Children's Books	0.224***	0.225***	0.228***	0.226***	0.137***	0.321***
(Cluster school SE)	(0.029)	(0.029)	(0.033)	(0.028)	(0.026)	(0.039)
(Cluster Family SE)	(0.030)	(0.030)	(0.033)	(0.030)	(0.035)	(0.036)
(Robust SE)	(0.030)	(0.030)	(0.033)	(0.030)	(0.032)	(0.037)

Note: This table shows the sensitivity of intention-to-treat (ITT) treatment effect parameters accounting for attrition. The first column is taken from Model 3 in Tables 6, 7, and 8. The second, third, and fourth columns report results using inverse probability weight, the new approach we proposed that deals with unbalanced attrition, and the Heckman selection model. The last two columns report results using Lee sharp bounds methods. We report the results for outcomes on JumpStart test score, Bracken score, number of days/week read to child, and whether the family has 10 or more children's books, respectively. We report three different standard errors: robust, clustered at the school level, and clustered at the family level. We report the most conservative significance level (*p < 0.10, **p < 0.05, ***p < 0.01).

Figure 1 presents the results. Panel (a) focuses on the Jumpstart test score. The most striking source of heterogeneity relates to parental education: among parents without a high school diploma, JSP increases the Jumpstart test score by 11.5 percentage points, compared to 6 percentage points among parents with at least a high school diploma. Panel (b) shows a similar pattern for the Bracken score: the impact is 4.1 percentage points for parents with low education, but only 1.2 percentage points for those with higher education. We also observe differences by parental age: the estimated effect for older parents is 3 percentage points, whereas it is –1.3 percentage points for younger parents.

Panel (c) examines heterogeneity in the weekly frequency of parent–child reading. The most notable contrast arises between children with high versus low baseline Bracken scores, at 0.749 and 0.445 additional days read to per week, respectively. Panel (d) reports heterogeneous effects on the likelihood of having at least ten children's books in the household. Here, we observe the greatest difference between older and younger parents, with effects of 0.262 and 0.178, respectively.

Overall, we find only limited evidence that the JSP generates heterogeneous impacts across observable characteristics. We return to this point below, where we explore Marginal Treatment Effects to further investigate the potential heterogeneity in JSP outcomes.

4.4 Attrition

In this section, we assess the robustness of our findings to sample attrition. As shown in Appendix Tables C6, C7, and C8, parents who attrite are, on average, younger, less likely to be Hispanic, more likely to be Black, and more frequently employed than parents who remain in the study.

We evaluate the sensitivity of the intent-to-treat (ITT) estimates using several methods, including inverse probability weighting, the Heckman selection model, and Lee bounds. We additionally adopt a procedure designed to address unbalanced attrition, which we implement in four steps:

1. **Estimate the probability of attrition.** Let $D = 0$ indicate an individual who remains in the sample, and let $D = 1$ indicate an individual who attrites. We then estimate a probit model of D on demographic characteristics and the baseline Jumpstart test score.[10]

2. **Create matched pairs.** Using the estimated probabilities of attrition, we match each parent in the treatment group with the control-group parent whose probability of attrition is closest.

[10] See Appendix Table C5 for these first-stage results. We find that higher Jumpstart test scores, being male, and identifying as Hispanic or Black are associated with a higher probability of attrition. Receiving food stamps is also correlated with attrition.

This matching is performed within each stratum, ensuring that every matched pair consists of one treatment and one control parent.

3. **Refine the sample using endline data.** If either member of a matched pair is missing from the endline dataset, we exclude the entire pair. Thus, if a control parent in a pair is missing from the endline data, the corresponding treatment parent is also removed. This results in a smaller but balanced sample, thereby eliminating any differential attrition between treatment and control groups.

4. **Estimate ITT parameters.** Finally, we re-estimate the ITT effects on this refined sample. In Table 9, we refer to this procedure as "Balancing."

Table 9 presents the results for two key measures of children's cognitive development—the Jumpstart Test Score and the Bracken Score—as well as two measures of parental investment that exhibit significant treatment effects (the number of days per week the parent reads to the child and whether the household has at least ten children's books). For clarity and conciseness, we restrict our discussion here to these four outcomes; the full set of results for all sub-items is reported in Appendix Tables C9, C10, and C11.

Overall, we find that the JSP's effects on the Jumpstart Test Score and on the home-environment measures (i.e., how frequently parents read to their child and whether the household has at least ten children's books) are generally robust to attrition. By contrast, the JSP's impact on the Bracken Score, which assesses components of school readiness not covered in the JSP curriculum, does not remain robust once we account for attrition.

5 The Mechanisms

The RCT evidence demonstrates that the JSP enhances school readiness, as measured by the Jumpstart test. However, the RCT alone does not clarify *how* these gains arise. In this section, we explore the mechanisms underlying these improvements, providing a framework that can be applied to other parenting interventions and shedding light on how public programs promote child development.

Our analysis proceeds as follows. First, we present the model. Second, we discuss measurement by linking the model's variables to our data. Third, we derive the estimating equations and relate them to two treatment-effect parameters: the Average Treatment Effect (ATE) and the Effect of the Treatment on the Treated (ETT). We then illustrate how our model decomposes the treatment effect into multiple mechanisms. Next, we address identification. Finally, we conclude the section with our empirical results.

5.1 Model

The JSP builds on the principle that parents are their children's first teachers. Accordingly, our model disentangles several potential pathways through which the JSP may operate: (a) a direct effect, (b) an indirect effect through changes in the marginal productivity of investments, (c) an indirect effect through changes in preferences (habit formation), and (d) an interaction between changes in marginal productivity and habit formation. In addition, the JSP could theoretically alter parents' beliefs about how their investments influence children's human capital formation. However, supplementary data on parental beliefs show no evidence of this mechanism (see Appendix B). Therefore, changes in parental beliefs are not included in the model.

Let G_i denote an indicator taking the value one if parent i graduates from the JumpStart Program and zero otherwise. Similarly, let Z_i be a U$[0,1]$ random variable draw that determines whether the family is randomly assigned to the control or treatment group; higher values of Z_i imply a greater likelihood of treatment assignment. Indeed, in our study, Z_i determined the random assignment to the control or treatment group. The variables $K_{i,0}$ and $K_{i,1}$ represent the child's knowledge at the beginning and end of the year, respectively. Let X_i represent the level of parental investment in the child's human capital.

The end-of-year knowledge production depends on the beginning-of-year knowledge, JSP graduation, and family investments. Specifically, it is modeled as

$$K_{i,1} = \beta_0 + \beta_1 K_{i,0} + \beta_2 G_i + \beta_3 X_i + \beta_4 G_i X_i + \epsilon_i, \tag{2}$$

Our formulation draws on the literature on value-added model in economics of education (see, e.g., Todd and Wolpin, 2007). In comparing the technology of skill formation in Cunha and Heckman (2007) with the value-added model used in economics of education, it is useful first to distinguish their respective perspectives on how skills evolve over time. The technology of skill formation treats human capital development as a multi-period, multi-dimensional process in which many dimensions of capabilities accumulate through self-productivity and dynamic complementarities. In this framework, a child's current stock of skills not only depends on parental investments and other environmental inputs, but also on the skills acquired in earlier periods. In addition, the skills that the child forms in one stage of the lifecycle (i.e., the dependent variable) may be different from the past skills (i.e., the lagged dependent variable). This explicit recognition of the interplay between early and later stages of development stands in contrast to the approach of the value-added model, which focuses on a single dimension of achievement and provides a framework to study how current outcomes depend on past outcomes and a teacher, program, or school fixed effect.

The two approaches differ in their goals and typical applications. The technology of skill formation is designed to inform long-term policy initiatives by clarifying the interplay among early interventions, family investments, and subsequent stages of development. Because it posits a dynamic production function for multiple skill types, it enables policymakers and researchers to simulate how particular interventions at various ages might affect future outcomes, such as educational attainment or labor market earnings. The value-added model, by contrast, is widely deployed in educational accountability and teacher evaluation programs. Policymakers often favor value-added measures because they provide straightforward estimates of teacher, school program, or school effectiveness based on test-score gains. This targeted focus on isolating contributions by specific educational programs makes the value-added approach appealing for our analysis.

The coefficient β_2 captures the direct effect of the JSP on the child's knowledge, net of parental investment. This effect is potentially driven by the direct impact of attending the periodic group- and center-based JSP sessions. During these sessions, children are exposed to cognitively stimulating activities, including guided play, storytelling, and interactive lessons led by trained educators. Such structured engagement fosters key skills, including language acquisition, early literacy, problem-solving, and attention regulation, which are crucial for school readiness. Since these experiences take place at the center, they may provide developmental benefits beyond those derived from parental engagement at home.

The model further allows the marginal productivity of parental investment to differ by graduation status. In the absence of the JSP, each additional unit of parental investment increases knowledge by β_3. If the parent graduates, this effect becomes $\beta_3 + \beta_4$. A positive β_4 implies that the program enhances the effectiveness of parental investments, thus representing an *indirect* channel through which the intervention improves child outcomes. This channel reflects how program resources, such as books, enhance learning opportunities at home by enriching the content and structure of parent-child interactions. These materials could make it easier and more appealing for both children and parents to engage in meaningful, skill-building activities. Indeed, even if the frequency of interactions remains unchanged, the *quality* of each interaction may improve due to the ready availability of developmentally appropriate materials.

The parent's utility depends on the end-of-year knowledge, the net benefit or cost of program graduation, and the (dis)utility from deviating from a habitual investment level. Formally, the utility is specified as

$$U(K_{i,1}, G_i, X_i) = K_{i,1} - \big(\eta_i - \alpha_0 - \alpha_1 Z_i\big)\, G_i - \frac{1}{2\alpha_2}\big(X_i - X_i^H\big)^2. \tag{3}$$

The term $K_{i,1}$ enters positively, reflecting that higher levels of child's knowledge increase the parent's utility. The expression $\big(\eta_i - \alpha_0 - \alpha_1 Z_i\big)\, G_i$ captures how both the random draw that determines assignment to the control or treatment group (i.e., Z_i), and the idiosyncratic component η_i affect the net cost or benefit of graduating from JSP. If $\eta_i - \alpha_0 - \alpha_1 Z_i$ is large and positive, graduation is costly; negative values imply that graduation is effectively subsidized.

We incorporate *habit formation* as a key determinant of parental investments in children's human capital. Specifically, parents incur a quadratic penalty for deviating from their *habitual investment* level X_i^H:

$$\frac{1}{2\,\alpha_2}\,\big(X_i - X_i^H\big)^2.$$

The parameter α_2 reflects how strongly parents dislike deviating from X_i^H. We further define the habitual investment level as:

$$X_i^H = \gamma_1\, X_{i,0} \;+\; \gamma_2\, G_i \;+\; \nu_i, \tag{4}$$

where $X_{i,0}$ is the baseline level of investment and ν_i is an idiosyncratic error term. In this framework, the Jumpstart Program (JSP) can shift parental preferences by influencing the habitual investment level. The coefficient γ_2 thus captures the extent to which JSP graduation modifies parents' "default" investment behavior.

Duesenberry (1948) seminal contribution introduced the concept of intertemporally dependent preferences, where past consumption levels significantly influence current consumption choices. This idea laid the groundwork for a deeper exploration into how past consumption affects consumer behavior over time. Ryder and Heal (1973) built directly on this foundation by analyzing how these preferences impact optimal growth models, showing that traditional assumptions about consumption smoothing and savings behavior need to be reconsidered when past consumption patterns are factored into utility functions. Similarly, Becker and Murphy (1988) applied the concept of dependent preferences to addictive behaviors, formulating a theory of rational addiction where past consumption of an addictive good increases current utility derived from that good, thereby systematically altering consumer demand patterns. Meanwhile, Constantinides (1990) explored the implications of habit formation on the stock market, positing that investors' decisions are influenced by their past consumption experiences, which in turn affects asset prices and market dynamics.

In our research, we extend the concept of habit formation to decisions about investments in the human capital of children through parent-child interactions. Habit formation acts as a fundamental mechanism through which daily behaviors become automatized, significantly influencing long-term outcomes. Studies such as Beeken et al. (2016) and Fritz et al. (2019) illustrate the efficacy of interventions based on habit formation for weight management and health promotion through simple, repeatable actions linked to consistent cues. Similarly, Cleo et al. (2019) demonstrate modest but sustained weight loss through habit-based interventions compared to controls, emphasizing the potential of such approaches in fostering enduring health-related behaviors. Extending these principles to the domain of parent-child interactions, targeted interventions can significantly alter the frequency and quality of these interactions. By implementing structured, habitual interactions driven by consistent environmental cues, such interventions can help parents more effectively engage in high-quality interactions with their children. These habit-based strategies ensure that positive interactions become part of the daily routine, reducing the cognitive load on parents and enabling a more natural and sustained engagement in nurturing behaviors that are crucial for child development.

The relevance of habit formation for programs like JSP is paramount as it spans a seven-month period, providing ample time for parents to adopt and solidify new routines. During this period, parents in the treatment group may increase the frequency of daily reading, while those in the control group continue their usual practices. Additionally, JSP's group-based sessions expose parents to peers who are learning and adopting similar strategies, reinforcing the habit-formation process through social support and shared expectations. As a result, once the formal program ends, parents might maintain these newly established routines, producing long-lasting benefits for their children's development. Modeling habit formation mathematically thus captures how a relatively short intervention can transform everyday parenting behaviors well beyond the program's conclusion.

We assume that the error terms $(\eta_i, \nu_i, \epsilon_i)$ are jointly distributed with mean zero and variance–covariance matrix Σ, conditional on the vector $(Z_i, K_{i,0}, X_{i,0})$. The conditional assumption with respect to Z_i is relatively mild because Z_i is drawn from a U$[0,1]$ distribution that determines whether each family is assigned to treatment or control. We exploit this exogenous variation in Z_i as an exclusion restriction when estimating the graduation equation.

By contrast, independence with respect to the predetermined variable $K_{i,0}$ is a standard feature of value-added models. Specifically, these models assume that once prior achievement $K_{i,0}$ and relevant covariates are controlled for, there are no remaining unobserved factors correlated with both graduation from JSP and the student outcomes captured by $K_{i,1}$.

Finally, the independence of the error vector with respect to $X_{i,0}$ is satisfied if, for example, the error terms in habit formation are not serially correlated. If lags of $X_{i,0}$ were observed, one could relax this no-serial-correlation assumption. Unfortunately, we do not have such data.

5.2 Measurement

We briefly link the model's variables with our data. We measure $K_{0,i}$ and $K_{1,i}$ with the Jumpstart test scores at baseline and endline, respectively. The graduation variable G_i is the dummy variable that is equal to one if the parent graduates and zero otherwise.

The variable Z_i is an exogenous draw from a uniform distribution on the interval $[0, 1]$ that determined whether a family was assigned to the treatment or control group in our study. We use Z_i as our exclusion restriction, rather than the treatment assignment T_i, because there are no observations for which $T_i = 0$ and $G_i = 1$. In other words, $T_i = 0$ perfectly predicts $G_i = 0$, and such complete separation would lead the probit estimation to drop all observations in the $T_i = 0$ category, leaving no valid exclusion restriction. By contrast, incorporating Z_i circumvents this issue.

Our data include three cohorts from more than twenty elementary schools, each showing substantial variation in recruitment outcomes. Moreover, the cutoff value for Z_i differs across cohorts and schools, ensuring that our analysis remains feasible. The thought experiment is an expansion of the sample size by including other campuses that did not participate in the study. This expansion is natural for a parenting program since increasing the number of participating parents in each cohort and school could introduce congestion problems related to staff and space availability.

Finally, we measure $X_{i,0}$ and $X_{i,1}$ using factor scores obtained (after oblique rotation) from all parental investment measures collected at baseline and endline, respectively. Specifically, we use Bartlett factor scores for $X_{i,1}$ and Regression factor scores for $X_{i,0}$. This asymmetry follows the recommendations in Skrondal and Laake (2001), who analyze the consistency of estimators in regressions that include factor scores.[11] Appendix Section D.1 provides additional details about our procedure.

[11] See also Heckman et al. (2022) for a discussion on the use of factor scores in regression models.

5.3 Estimating Equations

In this section, we present the model's estimating equations. In addition, we provide the mapping between the parameters in the estimating equations and structural parameters, which we will use to discuss identification.[12]

In principle, the model is semiparametrically identified, as discussed further below. However, the nature of our exclusion restriction and the constraints of our limited sample size make implementing a semiparametric estimation strategy impractical. In this section, we present the model's estimating equations when we assume that the error terms are normally distributed. We use the parametric model as our benchmark and later evaluate the robustness of our findings through the semiparametric approach. First, the probability that parent i will graduate from the JumpStart Program is:

$$\mathbf{Pr}\left(G_i = 1 | Z_i, X_{i,0}\right) = \Phi\left(a_0 + a_1 Z_i + a_2 X_{i,0}\right) = \Phi\left(\eta_i^*\right), \tag{5}$$

The optimal investment depends on the parent's decision to graduate from the JumpStart Program. Let

$$\pi_i = \begin{cases} \frac{\phi\left(\eta_i^*\right)}{1-\Phi\left(\eta_i^*\right)}, & \text{if } G_i = 0, \\ -\frac{\phi\left(\eta_i^*\right)}{\Phi\left(\eta_i^*\right)}, & \text{if } G_i = 1. \end{cases}$$

The estimating equation for investment under the normal case is:

$$X_i^* = \alpha_2\beta_3 + \gamma_1 X_{i,0} + \left(\alpha_2\beta_4 + \gamma_2\right) G_i + \gamma_3\pi_i + \xi_i, \tag{6}$$

where $\gamma_3 = \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma_*}$.

Next, define $\pi_{i,0} = (1 - G_i)\pi_i$, $\pi_{i,1} = G_i\pi_i$, and:

$$\delta_0 = \frac{\beta_3\sigma_{\eta\nu} - \beta_3\beta_4\sigma_\nu^2 + \sigma_{\eta\epsilon} - \beta_4\sigma_{\nu\epsilon}}{\sigma^*},$$
$$\delta_1 = \frac{\left(\beta_3 + \beta_4\right)\sigma_{\eta\nu} - \left(\beta_3 + \beta_4\right)\beta_4\sigma_\nu^2 + \sigma_{\eta\epsilon} - \beta_4\sigma_{\nu\epsilon}}{\sigma^*}.$$

The estimating equation for knowledge under the normal case is:

$$K_{i,1} = \tilde{\beta}_0 + \beta_1 K_{i,0} + \tilde{\beta}_2 G_i + \tilde{\beta}_3 X_{i,0} + \tilde{\beta}_4 G_i X_{i,0} + \delta_0\pi_{i,0} + \delta_1\pi_{i,1} + \zeta_i \tag{7}$$

[12] Appendix Section D.2 presents the derivations of these equations.

5.4 Estimating Equations for a Restricted Model

In our analysis, we also consider a restricted model that imposes $\beta_4 = 0$, a null hypothesis that, as demonstrated below, cannot be rejected within our framework. This constraint carries a significant drawback: it eliminates heterogeneity in treatment effects, thereby equating the Average Treatment Effect (ATE) with both the Treatment on the Treated (TT) and the Treatment on the Untreated (TU). The primary benefit of setting $\beta_4 = 0$ lies in enhancing the precision of the estimates for other crucial parameters in our model. In this section, we present the analysis using the parametric model. Additionally, we explore the semiparametric model in the robustness checks section to validate our findings.

The graduation equation in the restricted model is:

$$\mathbf{Pr}\left(G_i = 1 \middle| Z_i, X_{i,0}\right) = \Phi\left(a_0 + a_1 Z_i\right). \tag{8}$$

Because $\beta_4 = 0$, the variable $X_{i,0}$ no longer enters the decision to participate in the JSP. The corresponding investment equation becomes

$$X_i^* = \alpha_2\beta_3 + \gamma_1 X_{i,0} + \gamma_2 G_i + \gamma_3\pi_i + \xi_i. \tag{9}$$

In this restricted model, we can thus estimate γ_2 directly from equation (9). Finally, the child's knowledge equation reduces to

$$K_{i,1} = \tilde{\beta}_0 + \beta_1 K_{i,0} + \left(\beta_2 + \beta_3\gamma_2\right) G_i + \tilde{\beta}_3 X_{i,0} + \delta\pi_i + \zeta_i. \tag{10}$$

By eliminating the interaction $G_i\, X_{i,0}$, this specification simplifies the coefficient on G_i in equation (10), which in turn sharpens the estimates of β_2 and β_3.

In the following section, we describe how the structural parameters estimated from both models can be used to recover other treatment-effect parameters and to inform our decomposition analysis.

5.5 Effects and Decomposition Analysis

We use the structural model to estimate various treatment effects of the JSP. Let $K_{i,1}^0$ and $K_{i,1}^1$ denote the end-of-year knowledge for children in families not participating and participating in the JSP, respectively. By combining the knowledge production function (2), the optimal investment rule (6), and the graduation decision (5), we derive the following expressions for the relevant treatment effects:

$$\begin{aligned}
\mathbf{ATE} &= \beta_2 + \alpha_2\beta_3\beta_4 + (\beta_3 + \beta_4)(\gamma_2 + \alpha_2\beta_4) + \gamma_1\beta_4 X_0, \\
\mathbf{TT} &= \beta_2 + \alpha_2\beta_3\beta_4 + (\beta_3 + \beta_4)(\gamma_2 + \alpha_2\beta_4) + \gamma_1\beta_4 X_0 + (\delta_1 - \delta_0)\,\pi_{i,1}, \\
\mathbf{TU} &= \beta_2 + \alpha_2\beta_3\beta_4 + (\beta_3 + \beta_4)(\gamma_2 + \alpha_2\beta_4) + \gamma_1\beta_4 X_0 + (\delta_1 - \delta_0)\,\pi_{i,0},
\end{aligned}$$

The term **ATE** represents the *total* average treatment effect, encompassing both the direct effect of the JSP and any indirect effects that operate through changes in investment behavior. Similarly, **TT** and **TU** denote the total effects on the treated and untreated populations, respectively, and also reflect these direct and indirect channels.

To isolate the role of each channel, we can systematically "turn off" specific parameters in our model. We begin by noting that the **ATE** can be rewritten as the sum of four components: the intervention's direct effect, its indirect effect through preferences, its indirect effect through the marginal productivity of investment, and its indirect effect through the interaction of preferences and productivity:

$$\mathbf{ATE} = \underbrace{\beta_2}_{\text{Direct Effect}} + \underbrace{\beta_3\gamma_2}_{\text{Indirect Effect through Preferences}} + \underbrace{\gamma_1\beta_4 X_0}_{\text{Indirect Effect through Productivity}} + \underbrace{2\alpha_2\beta_3\beta_4 + \gamma_2\beta_4.}_{\text{Indirect Effect through Interaction}}$$

In this formulation, β_2 captures the *direct effect* of the intervention on the child's knowledge, holding preferences and productivity constant. We obtain this term by setting the preference parameters α_2 and γ_2 and the productivity parameter β_4 to zero.

Next, $\beta_3\gamma_2$ isolates the *indirect effect* that operates *exclusively through preferences*. Here, the intervention alters the decision maker's preferences, which in turn affects investment choices and ultimately the child's knowledge. We obtain this term by setting the direct effect parameter β_2 and the productivity parameter β_4 to zero.

The component $\gamma_1\beta_4 X_0$ captures the *indirect effect* that works *exclusively through the marginal productivity of investment*, reflecting how the intervention modifies the effectiveness of each unit of investment after it is introduced. We obtain this term by setting the direct effect parameter β_2 and the preference parameters α_2 and γ_2 to zero.

Finally, the expression $2\alpha_2\beta_3\beta_4 + \gamma_2\beta_4$ measures how *interactions* between preferences and productivity contribute to the total effect. In other words, the intervention can change both the decision maker's baseline preferences and the productivity of any subsequent investments, with these changes either reinforcing or offsetting one another.

The restricted model reduces the potential channels to the direct effect and the indirect effect via preferences:[13]

$$\mathbf{ATE}_{Restricted} = \underbrace{\beta_2}_{\text{Direct Effect}} + \underbrace{\beta_3\gamma_2}_{\text{Indirect Effect through Preferences}}$$

By selectively shutting down these parameters and comparing how **ATE** changes, we can parse out the contribution of each mechanism to the overall treatment effect. This decomposition provides a clearer picture of not only whether the intervention is effective but also the precise *pathways* through which it exerts its influence.

5.6 Identification

In this section, we discuss the identification of our model. Our goal is to recover the parameters that are crucial for the decomposition of the treatment effects. Our approach generates a semiparametric single-index binary dependent variable model, aligning with frameworks considered by Gallant and Nychka (1987) and Klein and Spady (1993). A critical aspect to discuss is a limitation associated with our methodology regarding the exclusion variable Z_i, which dictates the assignment to control and treatment groups.

Consider a hypothetical scenario involving only one school and one cohort. In this simplified setting, a deterministic relationship would be observed between the random variable Z_i and the group assignment variable T_i. Specifically, the assignment rules would be $T_i = 0$ if $Z_i < z$ and $T_i = 1$ if $Z_i \geq z$. However, our actual experimental setup incorporates multiple randomization clusters, including over 20 schools annually, across several years. This complexity leads to a stochastic relationship between T_i and Z_i, diverging from the deterministic links typical in simpler randomized controlled trials. Consequently, this design allows us to estimate the structural model without the need to rely on the assumptions of a linear probability model, by exploiting the inherent randomness in T_i given Z_i.[14]

This identification argument necessitates the exclusion of strata fixed effects from our analysis, given the deterministic relationship between Z_i and T_i within each stratum. However, to assess the robustness of our findings, we can estimate models that incorporate controls for the child's baseline knowledge and age at enrollment. These variables were instrumental in constructing

[13] In addition, the restricted model eliminates heterogeneity in treatment effects. Thus, $\mathbf{TT}_{Restricted} = \mathbf{TU}_{Restricted} = \mathbf{ATE}_{Restricted}$.

[14] If we used T_i as the exclusion restriction, we would be compelled to employ the linear probability model because if $T_i = 0$, then $G_i = 0$, and $T_i = 0$ perfectly predicts $G_i = 0$ in both probit and logit models.

the blocks within schools, and we can include them in our estimating equations to study the sensitivity of our findings.

Furthermore, this stochastic relationship can be utilized to estimate a probability choice model. If Z_i not only provides exogenous variation, but this variation is strong enough to influence choice probabilities across the entire interval $[0,1]$ (i.e., achieving *identification at infinity*), then the propensity function derived from Z_i,$X_{i,0}$ could be used to construct control functions in a semiparametric approach. In this case, the identification of critical structural parameters of our model would follow from the analysis of identification of Generalized Roy Model (see, e.g., Heckman and Honoré, 1990).

However, despite Z_i providing exogenous variation in the probability of graduation, it may not sufficiently ensure *identification at infinity*. On the one hand, we expect $\lim_{Z\to 0} P(Z, X_0) = 0$, as an assignment to the control group guarantees that families will not be permitted to enroll in the JSP program. Conversely, the argument may or may not hold at the other extreme. For example, although families with higher values of Z_i are more likely to be assigned to the treatment group, a higher Z_i alone does not alter the costs and benefits associated with graduating—aside from the opportunity of enrollment.

Thus, we estimate our model semiparametrically to assess the feasibility of this approach. We utilize the estimator developed by Gallant and Nychka (1987). After recovering the choice probability parameters, we predict the single index and approximate the probability function using local polynomials. To implement this approximation, we consider two bandwidth values: the Rule-of-Thumb (ROT) bandwidth and three times the ROT bandwidth, denoted by b. This procedure yields the estimated graduation probability $\hat{P}_b(Z_i, X_{i,0})$, where the subscript b indicates the dependence on the bandwidth. We implement this procedure for both the Full and Restricted models.

Figure 2 presents the results of our semiparametric estimation of the graduation probability. The solid line represents the semiparametric estimates, while the dashed line corresponds to the probit model predictions. Notably, the greatest divergence between these two approaches occurs in the high probability region of graduation. Here, the probit model consistently yields higher estimates than the semiparametric model. However, these deviations are only statistically significant when larger bandwidth values are employed. As is well known, higher bandwidth leads to greater precision but introduces more bias. Thus, the primary takeaway from this analysis is that with optimal bandwidth choices, the differences between the probit and semiparametric approaches are generally negligible.

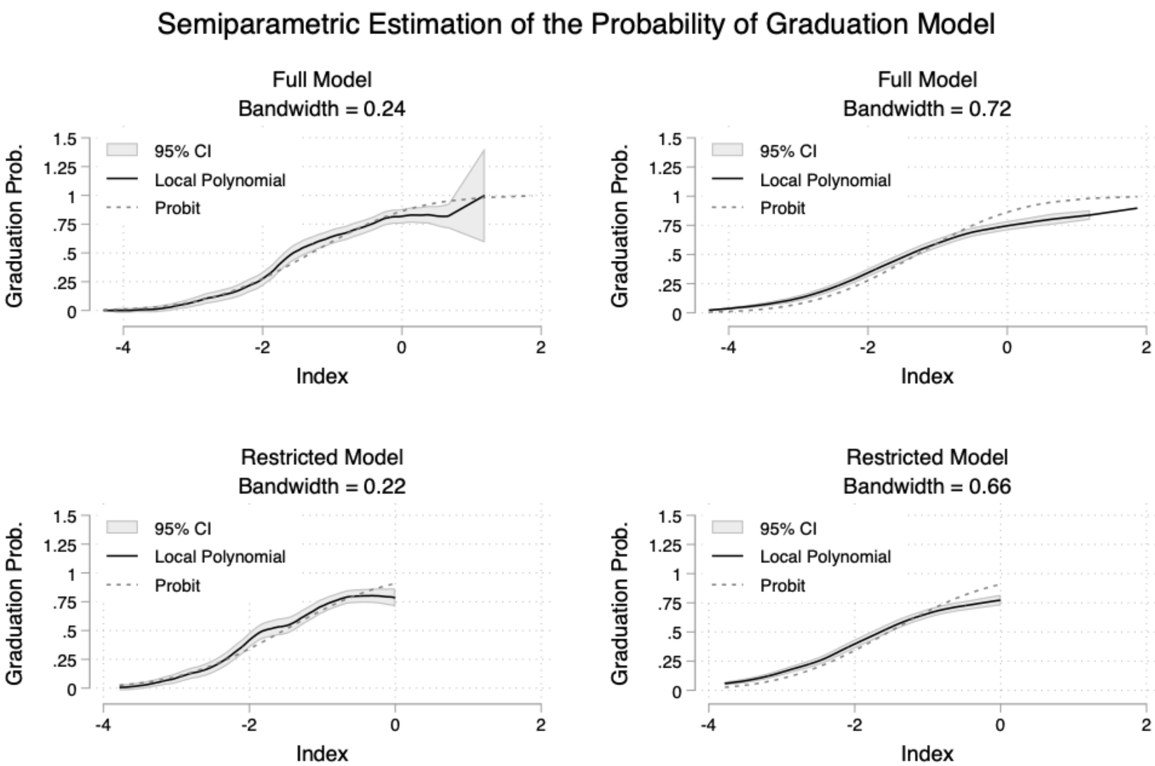


Figure 2: This figure presents four semiparametric estimates of the single-index model for the JSP graduation probability. Each panel combines the full or restricted model specification with either a low or high bandwidth. The solid line depicts the semiparametric estimates, while the dashed line shows the corresponding probit predictions. The gray region represents the 95% confidence interval for the semiparametric estimator. Overall, both approaches produce similar estimates, except for individuals whose probability of graduation is extremely high.

Secondly, whether the probability of graduation approaches one as the index values become more extreme also depends on the selected bandwidth. Using the Rule-of-Thumb (ROT) bandwidth, the probability approaches one; however, this is not the case at higher probability values. Furthermore, a comparison between the Full and Restricted models reveals that in the Restricted model, the probability does not approach one as Z_i increases. This observation implies that the identification at infinity in the Full Model is attained due to variations in $X_{i,0}$. This finding casts doubt on the reliability of the semiparametric approach for this specific application. Given that the probit model does not exhibit systematic biases in estimating choice probabilities, we favor the parametric approach. Nonetheless, we present the results from the semiparametric approach below and discuss the sensitivities of our findings.

Next, we show that we can identify all of the parameters for the estimation of the treatment effects and the decomposition analysis. Note first that estimating equation (6) allows us to identify γ_1. Once γ_1 is known, we can use the estimates of $\tilde{\beta}_3$ and $\tilde{\beta}_4$ from equation (22) to pin down β_3 and β_4. In turn, recovering these two parameters enables us to separately identify α_2 and γ_2 in equation (6).

Moreover, estimation of equation (22) recovers δ_0 and δ_1 and provides an estimate of $\tilde{\beta}_2$. Since we have already identified the remaining parameters that enter $\tilde{\beta}_2$, we can then back out the value of β_2. Identification of these parameters are sufficient for the estimation of the treatment effects and their decomposition.[15]

5.7 Estimation Results

Our empirical approach proceeds in two stages. First, we estimate the graduation probability function (5) via probit and use the resulting parameters to construct the relevant control functions. Second, we estimate equations (6) and (22) jointly by seemingly unrelated regressions, thereby recovering the structural parameters and their standard errors. Table 10 presents the estimation results.

Table 10 presents the estimated coefficients and corresponding standard errors for three key equations in our framework: the JumpStart Program Graduation Equation (Panel A), the Investment Equation (Panel B), and the Knowledge Production Function (Panel C). For each equation, we report results from both the *full model* and a *restricted model* in which we set $\beta_4 = 0$. By comparing the two model specifications, we gain insights for the decomposition analysis we report later in this section. In the paragraphs below, we discuss each panel in more detail, focusing on

[15] It is possible to identify additional parameters of this model. See Appendix Section D.4 for details.

the parameters most critical to our analysis.

In Panel A, a key finding is the statistically significant coefficient on Z_i (denoted by a_1), which serves as an exclusion restriction in both the full and restricted models. In the full model, a_1 is estimated at -3.435 (standard error 0.209), and in the restricted model, -3.286 (standard error 0.200). The fact that this parameter is consistently significant and precisely estimated allays potential concerns about *weak instruments* and supports the validity of using Z_i to identify the graduation decision in the JumpStart Program.

In Panel B, the parameter γ_1 is of particular interest, as it measures the persistence of past investment levels—one dimension of habit formation in our model. Its estimate is 0.368 (standard error 0.025) in the full model and 0.369 (standard error 0.026) in the restricted model. These results indicate that initial investment ($X_{0,i}$) strongly predicts subsequent investment choices, providing evidence that families tend to continue along the investment trajectories they have already established.

Turning to Panel C, β_1, the coefficient on the child's baseline knowledge $K_{0,i}$, remains a core driver of knowledge outcomes in both the full and restricted models. Its estimates are 0.390 (standard error 0.026) and 0.390 (standard error 0.025), respectively, underscoring the importance of initial knowledge in shaping subsequent achievement. Moreover, the parameters δ_0 and δ_1 are small in magnitude and statistically insignificant in the full model, suggesting that selection on unobservable factors is unlikely to bias our estimates.

A notable difference between the full and restricted specifications lies in the precision of the estimates for β_2 and β_3. In particular, when $\beta_4 = 0$ (the restricted model), the standard errors for these parameters decrease relative to those in the full model, reflecting the simplifying assumption that the JumpStart Program does not affect the marginal productivity of investment. While this restriction rules out certain mechanisms (such as treatment heterogeneity), it can sharpen estimates of other key parameters in the model.

Table 11 presents the estimates and standard errors of key *structural* parameters (Panel A) and *treatment-effect* parameters (Panel B) from both the full model and a restricted model in which we set $\beta_4 = 0$. This table is a pivotal element of our analysis, as it simultaneously characterizes the core structural parameters that drive household investment decisions and the overall effectiveness of the JumpStart Program. Below, we describe each panel and discuss the central findings.

In the top panel, α_2 captures the *penalty* associated with deviating from a household's habitual investment level. A smaller estimate of α_2 implies a stronger penalty for shifting away from prior investment choices, emphasizing the importance of habit formation. The full and restricted

Table 10: Parameter Estimates and Standard Errors

Panel A: JumpStart Program Graduation Equation (5)					
		Full Model		Restricted Model	
Parameter	**Description**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
a_0	Intercept	1.063	0.121	1.330	0.100
a_1	Coefficient on Z_i	-3.435	0.206	-3.286	0.200
a_2	Coefficient on $X_{i,0}$	0.101	0.022	–	–
Panel B: Investment Equation (6)					
		Full Model		**Restricted Model**	
Parameter	**Description**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
$\alpha_2\beta_3$	Intercept	2.830	0.133	2.830	0.144
$\alpha_2\beta_4 + \gamma_2$	Coefficient on G_i	1.014	0.204	1.013	0.206
γ_1	Coefficient on $X_{i,0}$	0.368	0.025	0.369	0.026
γ_3	Coefficient on π_i	0.009	0.142	0.007	0.142
Panel C: Knowledge Production Function (22)					
		Full Model		**Restricted Model**	
Parameter	**Description**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
$\tilde{\beta}_0$	Intercept	0.469	0.018	0.473	0.017
β_1	Coefficient on $K_{i,0}$	0.390	0.026	0.390	0.025
$\tilde{\beta}_2$	Coefficient on G_i	0.082	0.027	0.078	0.019
$\tilde{\beta}_3$	Coefficient on $X_{i,0}$	0.008	0.004	0.007	0.003
$\tilde{\beta}_4$	Coefficient on $G_i X_{i,0}$	-0.001	0.005	–	–
δ_0	Coefficient on $\pi_{i,0}$	-0.009	0.020	–	–
δ_1	Coefficient on $\pi_{i,1}$	-0.013	0.018	–	–
δ	Coefficient on π_i	–	–	-0.010	0.014

models yield estimates that are of similar magnitude (129.156 vs. 151.622), with standard errors (57.311 vs. 59.065). Nonetheless, these parameter values are statistically significant, suggesting that the penalty for deviating from habitual investment remains meaningful under both model specifications.

Next, γ_2 measures the JumpStart Program's impact on families' *habitual* investment. The estimates hover around 1.0 in both models (1.026 vs. 1.013), indicating that graduation from the program significantly increases household investment over and above initial levels. The parameters β_2 and β_3 jointly govern how the formation of knowledge respond to graduation from the program and investment, respectively. β_2 represents the *direct* effect of the program on a child's knowledge, whereas β_3 reflects the *marginal productivity* of investment in the control group. Notably, the restricted model delivers tighter standard errors for β_2 and β_3 (particularly β_2), consistent with our

Table 11: Estimates and Standard Errors of Structural and Treatment-Effect Parameters

Panel A: Structural Parameters					
		Full Model		**Restricted Model**	
Parameter	**Description**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
α_2	Habit Deviation Penalty	129.156	57.311	151.622	59.065
γ_2	Habitual Inv.	1.026	0.209	1.013	0.206
β_2	JSP Graduation	0.084	0.097	0.059	0.021
β_3	Investment	0.022	0.010	0.019	0.007
β_4	Interaction Grad. and Inv.	-0.004	0.014	–	–
Panel B: Estimated Treatment Effect Parameters					
		Full Model		**Restricted Model**	
Parameter	**Description**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
ATE	Average Treatment Effect	7.7	1.9	7.8	1.9
TT	Effect of Treat. on the Treated	8.2	3.3	7.8	1.9
TU	Effect of Treat. on the Untreated	7.4	2.9	7.8	1.9

earlier discussion about the trade-off between model complexity and estimation precision. Finally, β_4 captures the JumpStart Program's impact on the *marginal productivity* of investment. In the full model, it is estimated at -0.004 (with a standard error of 0.014), suggesting a small and imprecise effect, which the restricted model excludes altogether by setting $\beta_4 = 0$. The small point estimate for β_4 provides evidence in favor of the restricted model.

The bottom panel highlights the *Average Treatment Effect* (ATE), the *Treatment on the Treated* (TT), and the *Treatment on the Untreated* (TU), which represent different ways of quantifying the overall benefits of the JumpStart Program. In both the full and restricted models, the ATE is approximately 7.7–7.8, indicating a positive and substantial average impact on child knowledge. As noted above, the restricted model imposes $ATE = TT = TU$, thus ruling out any heterogeneity in treatment effects.

By contrast, the full model shows a modest difference across these measures (TT of 8.2 vs. TU of 7.4). However, formal statistical tests do not reject the null hypothesis that the three treatment effects are equal. This finding provides additional evidence that the simpler, restricted specification—in which $ATE = TT = TU$—does not miss any meaningful variation in program impact. Below, we provide further evidence that essential heterogeneity, if it exists in our study, is small.

Table 12 reports the decomposition of the Average Treatment Effect (ATE) into its direct and indirect components. By comparing results from the *full model* and a *restricted model* (which sets

Table 12: ATE Decomposition Analysis

	Full Model		Restricted Model	
Decomposition Components	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
Total Effect (ATE)	7.7	1.9	7.8	1.9
Direct Effect	8.4	9.7	5.9	2.1
Indirect Effect through Preferences	2.2	1.2	1.9	0.8
Indirect Effect through Marg. Prod.	-0.5	1.7	–	–
Indirect Effect through Interaction	-2.4	8.2	–	–

$\beta_4 = 0$), the table clarifies the relative importance of each channel through which the JumpStart Program operates. Below, we first summarize the decomposition for the full model and then turn to the restricted specification.

Under the full model, the total effect (ATE) is estimated at 7.7 (standard error 1.9), and it comprises a relatively large, though imprecisely estimated, *direct effect* of 8.4 (standard error 9.7) together with a *small and imprecisely estimated indirect effect through preferences* of 2.2 (standard error 1.2). The indirect effects that operate through the *marginal productivity of investment*, whether exclusively (-0.005, with a standard error of 0.017) or in *interaction* with preferences (-2.4, standard error 8.2), are both small and not statistically significant. Hence, these latter channels do not appear to contribute meaningfully to the program's overall impact in the full model.

Turning to the restricted specification (which shuts down any program impact on marginal productivity, i.e., $\beta_4 = 0$), the total effect is estimated at 7.8 (standard error 1.9), almost identical to that in the full model. In contrast to the full model, however, the *direct effect* here is more precisely estimated at 5.9 (standard error 2.1), representing approximately *75%* of the total impact. Meanwhile, the *indirect effect through preferences* stands at 1.9 (standard error 0.8), corresponding to about *25%* of the total. This simpler decomposition confirms that even after restricting out a possible effect on investment productivity, the bulk of the JumpStart Program's impact arises from its direct channel, with a modest but non-negligible contribution from preference changes.

5.8 Sensitivity Analysis

In this section, we conduct sensitivity analyses of our findings. We use the semiparametric estimator of the probability of graduating from the JSP to consider two alternative specifications for the control function. The first specification is a linear function of $P_b\left(Z_i, X_{i,0}\right)$ while the second specification assumes a cubic function of $P_b\left(Z_i, X_{i,0}\right)$. Once the model is estimated with the semiparametric control function, we recover the parameters of interest and decompose the Average

Treatment Effect into its direct and indirect components.

Table D13 presents the results. Overall, the semiparametric approach generates higher estimates of the JSP's effect on knowledge compared to the parametric model. Specifically, the parametric model produces an ATE of 7.7 percentage points (with a standard error of 1.9), while the semiparametric approach yields estimates ranging from 11 percentage points (with a standard error of about 2.7) to 20.4 percentage points (with a standard error of about 11.5).

In general, semiparametric models that specify the control function as a cubic polynomial of the graduation probability tend to produce higher ATE estimates but are more sensitive to the choice of bandwidth. Moreover, increasing the bandwidth results in greater imprecision. By contrast, models that assume a linear specification for the control function are more stable and less sensitive to the bandwidth choice.

As in the parametric full model, the semiparametric estimation of the full model indicates that the direct effect equals the entire ATE. In contrast, for the restricted model, the semiparametric approach with a linear control function yields a decomposition that closely aligns with the parametric restricted model: roughly 80% of the total effect stems from the direct effect, and the remaining 20% is attributable to preferences.

When we discussed the identification of the model, we established that our strategy requires excluding strata fixed effects from our structural analysis due to the deterministic relationship between the treatment assignment and the covariates within each stratum. Although we cannot incorporate strata fixed effects directly, we can control for the child's baseline knowledge and age. These variables were instrumental in forming the blocks within schools.

Table 13 presents the results of this approach. When we adjust for baseline knowledge and age at enrollment, the treatment effect parameters appear slightly higher, and the decomposition remains quantitatively similar. This analysis suggests that the exclusion of strata fixed effects does not significantly impact our conclusions, affirming the robustness of our findings.

As we discussed above, the restricted model eliminates essential heterogeneity and equates the ATE to the TT and TU. We test this prediction of the restricted model by estimating a Generalized Roy Model semiparametrically to recover the MTE and compare it to the ATE. Figure 3 shows the estimated MTE and its confidence interval. We find that the ATE (7.4 percentage points, standard error equal to 1.9) is within the 95% confidence band for the MTE, which also shows little variation. This finding suggests that the estimates of the restricted model estimated via the parametric approach produces credible inference.

Table 13: Estimates and Standard Errors of Treatment Effect Parameters and Decomposition When We Control for Baseline Knowledge and Age of the Test at Endline

	Full Model		Restricted Model	
	Estimate	Std. Error	Estimate	Std. Error
Panel A: Treatment Effect Parameters				
Average Treatment Effect (ATE)	7.9	1.9	8.1	1.9
Effect of the Treatment on the Treated	9.1	2.5	8.1	1.9
Effect of the Treatment on the Untreated	5.7	4.2	8.1	1.9
Panel B: Decomposition of the ATE				
Total Effect (ATE)	7.9	1.9	8.1	1.9
Direct Effect	9.5	10.5	6.0	2.1
Indirect Effect through Preferences	2.5	1.3	2.0	0.9
Indirect Effect through Marg. Prod.	-0.6	1.6		
Indirect Effect through Interaction	-3.6	9.3		

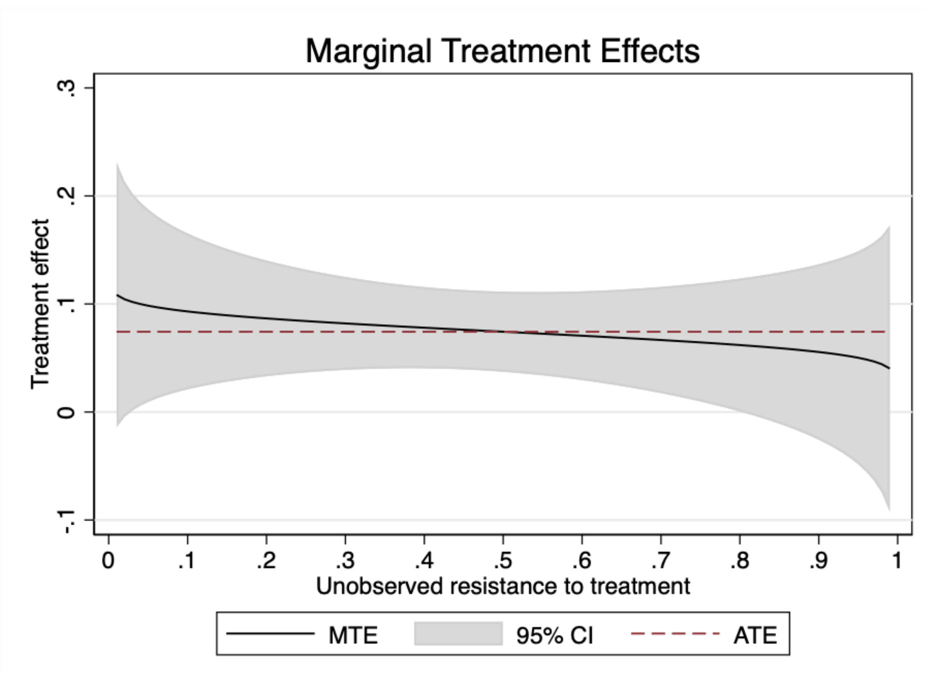


Figure 3: This figure shows the estimates of the Marginal Treatment Effect (MTE, solid line), its 95% interval (gray region), and the Average Treatment Effect (ATE, dashed line) as recovered by the semiparametric estimation of a Generalized Roy Model. The estimated ATE is equal to 7.4 percentage points with a standard error of 1.9 percentage points. We note that the ATE is within the confidence region for the MTE, which suggests that the heterogeneity in treatment effects is negligible in our study.

6 Conclusion

This paper presents the results of an experimental study evaluating JSP, a parent-directed school-readiness development program implemented by the Alief Independent School District in the Houston area. Our work contributes to the literature on parenting programs and child development by addressing both a scientific and a policy gap.

On the scientific side, we go beyond measuring treatment effects alone by developing and estimating a model of parent–child interactions shaped by JSP graduation, habit formation, and group-level heterogeneity in the marginal productivity of investments. We show that the structural parameters of this model can be used to recover several treatment-effect parameters of interest. Moreover, by mapping these structural parameters to the treatment-effect parameters, we naturally decompose JSP's total effect into a direct effect and indirect effects operating through changes in marginal productivity, habit formation, and their interaction.

Much of the existing evaluation of parenting programs in the United States has focused on efficacy studies, which estimate the impacts of programs implemented under ideal conditions. For example, the evaluation studies of the Perry Preschool Program and the Abecedarian Program constitute some of the strongest cases for early childhood interventions (e.g., see Conti et al., 2016; Heckman, 2024). However, these are small-scale pilot programs implemented under arguably ideal conditions. In contrast, we examine the impact of a parenting program implemented by a school district under realistic conditions, providing valuable evidence of its effectiveness in a real-world setting. This approach addresses a critical policy gap by demonstrating how parenting programs perform outside controlled environments.

We designed and implemented an RCT to answer these scientific and policy questions. We found that the program produced modest impacts on the acquisition of the skills targeted by the program curriculum, about 0.25 of a standard deviation on raw test scores. We also found some evidence of spillovers to content not directly covered by the curriculum as measured by the Bracken Assessment of School Readiness scale. However, the estimated coefficients are small and not robust to attrition. The program increased parental investments as measured by the frequency parents read to their children and the probability that there are at least 10 children's books in the household. We do not find evidence of meaningful heterogeneity in treatment effects, whether we do so by breaking down the study sample into different groups or by comparing different treatment effect parameters.

The estimation of the structural model produced significant insights for the literature on the economics of parenting. Several influential studies have highlighted the role of parental beliefs in shaping investment decisions. For instance, Cunha et al. (2022) demonstrate how changes in parental beliefs about the returns to investments can drive variations in parental behavior and child outcomes. The literature, however, offers mixed results. On the one hand, Cunha et al. (2024) find that the implementation of a parent-directed language program with education, coaching, and objective feedback impacts parental investments by influencing parental beliefs. On the other hand, our findings diverge from this literature and are consistent with those reported by Attanasio et al. (2024), who find that a parenting program in Colombia did not impact parental beliefs, despite increasing parental investments. Perhaps the mixed findings relate to the intensity of the various interventions, as conjectured by List et al. (2021).

Our work also distinguishes itself by directly addressing the mechanism of productivity gains in the investment process. While studies such as Almond and Currie (2011) emphasize the importance of the quality of parental investments, few have explicitly modeled or tested the adoption of more effective production functions. Furthermore, according to our model, this heterogeneity in the marginal productivity generates essential heterogeneity as defined by Heckman and Vytlacil (2005). However, we find no evidence that the marginal productivity of investments differ between the two groups. This finding is similar to the ones reported by Attanasio et al. (2020).

Our analysis suggests that JSP improves school readiness through both a direct effect and an indirect effect operating via habit formation. The direct effect likely arises from children's structured exposure to enriching activities, as well as from parents learning practical strategies to support their child's development. The habit-formation channel, on the other hand, highlights the program's potential to establish or reinforce consistent, high-quality interaction routines over the program's seven-month duration. This result is particularly noteworthy because it points to persistent changes in parental behavior that extend beyond the formal intervention period.

If habit formation is a central mechanism driving parenting program's impacts, we would anticipate that light-touch interventions—characterized by minimal training, fewer resources, or shorter durations—are unlikely to foster lasting changes in parental behavior. Habit formation requires repeated exposure, consistent reinforcement, and the supportive social environment provided by group-based programs that last multiple months, as with JSP. In the absence of these elements, light-touch interventions might yield only direct effects, which may not be sufficient to establish sustained, high-quality interactions between parents and their children. Future research should explore the economic significance of habit formation in shaping investments in children's human capital and as a key pathway through which parenting programs exert their effects, as such insights are crucial for the optimal design of these interventions.

References

Agostinelli, Francesco and Matthew Wiswall, "Estimating the Technology of Children's Skill Formation: A Flexible Latent Factor Approach," Working Paper 22442, National Bureau of Economic Research July 2016.

Almond, Douglas and Janet Currie, "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 2011, *25* (3), 153–172.

Athey, Susan, Julie Tibshirani, and Stefan Wager, "Generalized random forests," *The Annals of Statistics*, 2019, *47* (2), 1148–1178.

Attanasio, Orazio, Flávio Cunha, and Pamela Jervis, "Subjective Parental Beliefs: Their Measurement and Role," *NBER Working Paper No. 26516*, 2024.

Attanasio, Orazio P., Emla O. Fernández, Christine Fitzsimons, Sally M. Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina, "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial," *BMJ*, 2014, *349*, g5785. Erratum in: BMJ. 2014;349:g6126.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina, "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia," *American Economic Review*, 2020, *110* (1), 48–85.

Baker, Amy JL, Chaya S Piotrkowski, and Jeanne Brooks-Gunn, "The effects of the Home Instruction Program for Preschool Youngsters (HIPPY) on children's school performance at the end of the program and one year later," *Early Childhood Research Quarterly*, 1998, *13* (4), 571–588.

_ , _ , and _ , "The home instruction program for preschool youngsters (HIPPY)," *The Future of Children*, 1999, pp. 116–133.

Bassok, Daphna, Jenna E Finch, RaeHyuck Lee, Sean F Reardon, and Jane Waldfogel, "Socioeconomic gaps in early childhood experiences: 1998 to 2010," *Aera Open*, 2016, *2* (3), 2332858416653924.

Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96* (4), 675–700.

Beeken, R J, B Leurent, V Vickerstaff, R Wilson, H Croker, S Morris, R Z Omar, I Nazareth, and J Wardle, "A brief intervention for weight control based on habit-formation theory delivered through primary care: Results from a randomised controlled trial," *International Journal of Obesity*, 2016.

Bond, Timothy N and Kevin Lang, "The evolution of the Black-White test score gap in Grades K–3: The fragility of results," *Review of Economics and Statistics*, 2013, *95* (5), 1468–1479.

Bracken, Bruce A, *Bracken school readiness assessment*, Psychological Corporation, 2007.

Cleo, G, E Isenring, R Thomas, and P Glasziou, "Efficacy of Habit-Based Weight Loss Interventions: A Systematic Review and Meta-Analysis," *Journal of Behavioral Medicine*, 2019, *42* (5), 831–849.

Constantinides, George M., "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 1990, *98* (3), 519–543.

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

Corso, Phaedra S., Justin B. Ingels, and Rebecca L. Walcott, "Costs of Evidence-Based Early Childhood Home Visiting: Results from the Mother and Infant Home Visiting Evaluation," OPRE Report 2022-01 2022.

Cunha, Flavio and James Heckman, "The technology of skill formation," *American Economic Review*, 2007, *97* (2), 31–47.

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

Cunha, Flavio, Marsha Gerdes, Qinyou Hu, and Snejana Nihtianova, "Language environment and maternal expectations: An evaluation of the LENA start program," *Journal of Human Capital*, 2024, *18* (1), 000–000.

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

Duesenberry, James S., "Income, Saving, and the Theory of Consumer Behavior." PhD dissertation, Harvard University, Cambridge, MA 1948. PhD dissertation.

Fritz, Heather, Yi-Ling Hu, Kevin Gahman, Chen Almacen, and Jon Ottolini, "Intervention to Modify Habits: A Scoping Review," *OTJR: Occupation, Participation and Health*, 2019, *39* (1), 5–14.

Gallant, A. Ronald and Douglas W. Nychka, "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 1987, *55* (2), 363–390.

Gelber, Alexander and Adam Isen, "Children's schooling and parents' behavior: Evidence from the Head Start Impact Study," *Journal of Public Economics*, 2013, *101*, 25–38.

Gertler, Paul, James J. Heckman, Rodrigo Pinto, Ariana Zanolini, Christel Vermeerch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor, "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 2014, *30* (344), 998–1001.

__ , James J Heckman, Rodrigo Pinto, Susan M Chang, Sally Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika Wright, "Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31," Working Paper 29292, National Bureau of Economic Research September 2021.

Heckman, James J., "Perry Preschool at 50: What Lessons Should Be Drawn and Which Ones Should Be Forgotten?," *National Bureau of Economic Research Working Paper Series*, September 2024, *No. 32972*.

Heckman, James J and Bo E Honoré, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58* (5), 1121–1149.

Heckman, James J. and Edward J. Vytlacil, "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 2005, *73* (3), 669–738.

Heckman, James J, Margaret L Holland, Kevin K Makino, Rodrigo Pinto, and Maria Rosales-Rueda, "An analysis of the Memphis nurse-family partnership program," Technical Report, National Bureau of Economic Research 2017.

Heckman, James J., Susanne M. Schennach, and Benjamin Williams, "Matching on Proxy Variables," 2022. Working paper, University of Chicago.

Kalil, Ariel, "Inequality begins at home: The role of parenting in the diverging destinies of rich and poor children," in "Families in an era of increasing inequality: Diverging destinies," Springer, 2014, pp. 63–82.

__ , Rebecca Ryan, and Michael Corey, "Diverging destinies: Maternal education and the developmental gradient in time with children," *Demography*, 2012, *49* (4), 1361–1383.

Klein, Roger W and Richard H Spady, "An efficient semiparametric estimator for binary response models," *Econometrica*, 1993, *61* (2), 387–421.

List, John A., Julie Pernaudet, and Dana L. Suskind, "Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes," *Nature Communications*, 2021, *12*, 5765.

Love, John M, Ellen Eliason Kisker, Christine Ross, Helen Raikes, Jill Constantine, Kimberly Boller, Jeanne Brooks-Gunn, Rachel Chazan-Cohen, Louisa Banks Tarullo, Christy Brady-Smith et al., "The effectiveness of early head start for 3-year-old children and their parents: lessons for policy and programs.," *Developmental Psychology*, 2005, *41* (6), 885.

Meghir, Costas, Orazio Attanasio, Pamela Jervis, Monimalika Day, Prerna Makkar, Jere Behrman, Prachi Gupta, Rashim Pal, Angus Phimister, Nisha Vernekar, and Sally Grantham-McGregor, "Early Stimulation and Enhanced Preschool: A Randomized Trial," *Pediatrics*, 05 2023, *151* (Supplement 2), e2023060221H.

Pierre, Robert St., Anne Ricciuti, Fumiyo Tao, Cindy Creps, Janet Swartz, Wang Lee, Amanda Parsad, and Tracy Rimdzius, "Third National Even Start Evaluation: Program Impacts and Implications for Improvement," Technical Report, U.S. Department of Education, Planning and Evaluation Service, Elementary and Secondary Division, Washington, DC 2003.

Putnam, Robert D, *Our kids: The American dream in crisis*, Simon and Schuster, 2016.

Reardon, Sean F and Ximena A Portilla, "Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry," *Aera Open*, 2016, *2* (3), 2332858416657343.

Ryder, Harl E. Jr. and Geoffrey M. Heal, "Optimal Growth with Intertemporally Dependent Preferences," *Review of Economic Studies*, 1973, *40* (1), 1–31.

Skrondal, Anders and Peter Laake, "Regression among factor scores," *Psychometrika*, 2001, *66* (4), 563–575.

Todd, Petra E and Kenneth I Wolpin, "The production of cognitive achievement in children: Home, school, and racial test score gaps," *Journal of Human Capital*, 2007, *1* (1), 91–136.

Westheimer, Miriam, *Parents Making a Difference: International Research on the Home Instruction for Parents of Pre School Youngsters (Hippy) Program*, The Hebrew University Magnes Press, 2003.

Zhou, Jin, James J. Heckman, Bei Liu, Mai Lu, Susan M. Chang, and Sally Grantham-McGregor, "Comparing China REACH and the Jamaica Home Visiting Program," *Pediatrics*, 05 2023, *151* (Supplement 2), e2023060221I.

Appendix A Alief Jumpstart Curriculum Weeks 1-22

Table A1: JumpStart Program Curriculum

Week	Month	Child Present?	Book	CD	Other Materials
1	October	No	No, David!	The Color Songs	Play Dough, Laminated Sentence Strip, Laminated Construction Mat, Expo Marker
2	October	No	The Pigeon Needs a Bath	The Color Songs, Brush a Brush	Play Dough, Laminated Sentence Strip, Laminated Construction Mat, Expo Marker, Hand Pump Soap, Tooth Brush, Tooth Paste, Small Pack of Tissues
3	October	Yes	No, David!	Name Song, Brush a Brush	Hand Sanitizer, Pre-made Laminated Sentence Strip with Child's Name
4	November	No	Brown Bear, Brown Bear	The Color Songs	Red Construction Paper, Glue Bottles, Leaf Outline, Apple Outline
5	November	No	Brown Bear, Brown Bear	The Color Songs	Blue Construction Paper, Glue Bottles, Butterfly Outline, Bird Outline
6	November	Yes	Brown Bear, Brown Bear	The Color Songs	Pre-made Laminated Sentence Strip with Child's Name Written in Blue and Red, Red and Blue Beads, Red and Blue Markers
7	December	No	Chicka Chicka, Boom Boom	The Color Songs	Scissors, Green, Red, and Blue Strips of Construction Paper
8	December	No	Dr. Suess's ABC: An Amazing Alphabet Book!	The Color Songs	Play Dough, Scissors, Green Construction Paper, Green Sticker Dots, Frog Outline, Alligator Outline, Glue
9	December	Yes	Dr. Suess's ABC: An Amazing Alphabet Book!	The Color Songs	Pre-made Laminated Sentence Strip with Child's Name Written in Green, Tree Outline, Scissors, Green, Blue, and Red Sticker Dots
10	January	No	Rainbow Fish	The Color Songs, Five Little Fish, Five Little Monkeys	Large Crayons, White Copy Paper, Three White Die-Cut Fish, Yellow Construction Paper, Ocean Scene Outline, Cotton Balls, Glue
11	January	No	Ten Black Dots	The Color Songs, Five Little Fish, Five Little Monkeys	Large Crayons, Dot-to-Dot Paper, Spider-to-Dot Paper, Cutting Line Paper, Spider Web, Five Black Spiders
12	January	Yes	Ten Black Dots	The Color Songs, Five Little Monkeys	Pre-made Laminated Sentence Strip with Child's Name Written in Black, Bee Outline, Yellow Construction Paper, Scissors, Glue, Wax Paper Wings, Black Crayon, Black Sticker Dots

Table A2: JumpStart Program Curriculum, Continuation

Week	Month	Child Present?	Book	CD	Other Materials
13	February	No	Harold and the Purple Crayon	The Color Songs	Tongs, Pom Poms, Sorting Mat, Clear Plastic Cup
14	February	No	The Three Bears	The Color Songs	Goldilocks Cutting Paper, Play Dough, Die Cut Counting Bears, Sorting Mat
15	February	Yes	The Three Bears	The Color Songs	Pre-made Laminated Sentence Strip with Child's Name Written in Purple, Goldilocks Rhyme, Popsicle Sticks, Bear Outline, Goldilocks Die Cut, Purple Marker
16	March	No	From Head to Toe	The Color Songs, Head, Shoulders, Knees, and Toes	Simon Says Poster
17	March	No	If You're Happy and You Know It	The Color Songs, Head, Shoulders, Knees, and Toes	Circle Outline, Glue, Scissors, Crayons, Black and Blue Dots, Copy Paper, Play Dough, Counting Mat, Brown Construction Paper
18	March	Yes	From Head to Toe	The Color Songs, Head, Shoulders, Knees, and Toes	Pre-made Laminated Sentence Strip with Child's Name Written in Orange
19	April	No	Three Little Pigs	The Color Songs	Beads, Yarn, Cut-Up Straws, Pipe Cleaners, Buttons, Pink Construction Paper, Crayons, Pig Outline
20	April	No	Three Little Pigs	The Color Songs, Five Little Monkeys	Spinning Tops, Play Coins, Crayons, Pink Coloring Page, Cookie Cutters, Sentence Strip
21	April	Yes	Three Little Pigs	The Color Songs, Head, Shoulders, Knees, and Toes	Pre-made Laminated Sentence Strip with Child's Name Written in Pink, Name Puzzle, Pink Construction Paper, Glue
22	May	Yes	Mary Had a Little Lamb Rhyme Sheet, Humpty Dumpty Rhyme Sheet	The Color Songs	Lamb Outline, Glue, Cotton Balls, Tweezers, Counting Mat, Red Construction Paper, Crayons, Scissors, Egg Outline, Black Crayon, Humpty Dumpty's Wall Outline

Appendix B The Effect of Jumpstart Program on Parental Beliefs

In this Appendix, we analyze the impact of the Jumpstart Program (JSP) on parental beliefs. Our study directly tests this channel by eliciting parental beliefs from participants in the third cohort. Specifically, we asked parents five pairs of belief questions at both baseline and endline. Consistent with the literature, the belief questions present a hypothetical scenario. In our study, the scenario is as follows:

> *Emma and Olivia are two cousins who have just turned three years old. So far, they have been raised in similar ways by their parents. But now that they have turned three years old, their parents are planning to do different things. At Emma's home, the parents are planning to spend 30 minutes per day reading books, teaching Emma about letters, colors, and numbers, teaching them to count things, practicing writing her own name, sorting objects, making balls and snakes with play dough. At Olivia's home, the adults are planning to spend only 10 minutes per day teaching Olivia about the same things.*

Following this scenario, we ask parents to report their expectations regarding Emma's and Olivia's developmental progress by the time they reach age four. Respondents choose from three options: below average, average, or above average. A response indicating that Emma will develop to a higher level than Olivia is scored as $+1$; if both children are expected to reach the same level, the score is 0; and if Olivia is expected to achieve a higher level than Emma, the score is -1. A parental belief score is then computed by summing the scores from all five questions.

In our data, some respondents did not answer all five pairs of questions. To account for this, we calculate two belief scores. **Score 1** is computed by summing the scores for families who answered at least one pair of questions and dividing by the number of pairs answered. **Score 2** is calculated only for families that answered all five pairs, by summing their scores.

Table B3 reports the effect of JSP treatment on parental belief scores under different controls and with various standard error specifications. For each outcome, three models are estimated: one without any controls, one controlling for the baseline belief score, and one controlling for both baseline belief scores and demographic variables. For each model, results are provided using robust standard errors, standard errors clustered at the school level, and standard errors clustered at the family level.

Overall, our findings show that baseline belief scores are highly correlated with endline scores. Crucially, regardless of the specification used, we find no significant treatment effect of JSP on parental beliefs. This result suggests that the program does not operate by altering parents' beliefs about the productivity of their investments in early childhood human capital.

Table B3: Impact of Jumpstart Program on Parental Belief Scores

Panel A: Belief Score 1 Endline

	Model 1	Model 2	Model 3
Treat	0.032	0.155	0.112
Robust SE	(0.256)	(0.257)	(0.258)
Clustered at School Level	(0.282)	(0.308)	(0.300)
Clustered at Family Level	(0.256)	(0.257)	(0.258)
Belief Score 1 Baseline		0.350**	0.328*
Robust SE		(0.115)	(0.126)
Clustered at School Level		(0.151)	(0.158)
Clustered at Family Level		(0.115)	(0.126)
Constant	2.645	0.812	-1.413
Robust SE	(1.764)	(1.879)	(3.926)
Clustered at School Level	(0.188)	(0.773)	(4.330)
Clustered at Family Level	(1.764)	(1.879)	(3.926)
Observations	293	286	286
Outcome variables at baseline	N	Y	Y
Demographics	N	N	Y
R-squared	0.186	0.249	0.290

Panel B: Belief Score 2 Endline

	Model 1	Model 2	Model 2
Treat	0.018	0.050	0.027
Robust SE	(0.270)	(0.291)	(0.299)
Clustered at School Level	(0.303)	(0.366)	(0.369)
Clustered at Family Level	(0.270)	(0.291)	(0.299)
Belief Score 2 Baseline		0.337*	0.321*
Robust SE		(0.125)	(0.132)
Clustered at School Level		(0.165)	(0.161)
Clustered at Family Level		(0.125)	(0.132)
Constant	2.655	0.947	-0.836
Robust SE	(1.778)	(1.912)	(4.303)
Clustered at School Level	(0.202)	(0.839)	(5.128)
Clustered at Family Level	(1.778)	(1.912)	(4.303)
Observations	279	256	256
Outcome variables at baseline	N	Y	Y
Demographics	N	N	Y
R-squared	0.188	0.263	0.310

Note: This table reports the effect of the Jumpstart program on parental belief scores. Outcome variables are constructed in two ways: using all five pairs of belief questions and using at least one pair. For each outcome variable, we estimate three models that differ by the inclusion of baseline outcome measures and demographic controls. For each model, standard errors are reported in three ways: robust, clustered at the school level, and clustered at the family level ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Appendix C Additional Analyses on the Impact of JSP

C.1 Weekly Attendance

The in-study sample contains families who are willing to join the evaluation of the Jumpstart Program. In other words, our evaluation faces a self-selection issue of participation in the evaluation study. In this section, we check if our results are driven by this self-selection problem by using distance from home to campus as an instrumental variable for the outcome variables. More specifically, we obtained families' home addresses and the corresponding participating campuses' addresses. We then calculated the driving distance from a typical routine at 7:30 am.

The intuition to validate our instrumental variables is that participating in the Jumpstart program requires parents to travel to campus to learn the class and materials. Therefore, we believe families who live far away from school may be less likely to participate without a direct effect on the outcome variables.

To rule out self-selection, we implement a Two Stage Least Squares (2SLS) analysis with the following regressions:

$$Y_{i1} = \beta_0 + \beta_1 W_i + \beta_2 \hat{D}_i + \beta_3 B_{i0} + \beta_{\mathbf{4}} \mathbf{X_i} + \epsilon_i, \tag{11}$$

$$D_i = \alpha_0 + \alpha_1 T_i + \alpha_2 B_{i0} + \alpha_{\mathbf{3}} \mathbf{X_i} + \eta_i. \tag{12}$$

Equation 12 is the first stage for 2SLS, where T_i is the treatment indicator, B_i is the Jumpstart Baseline Score, $\mathbf{X_i}$ is the set of demographic controls which includes baseline child monthly age, whether the child is male, whether parent completes high school, whether the parent is Hispanic, parent age, number of children in household, whether family receives food stamp, whether parent is employed and, finally, an indicator on whether missing demographics. The second stage for our 2SLS is demonstrated in equation 11 where Y_{i0} is the outcome variables we are interested in, which includes the endline Jumpstart testing score. W_i is, which is instrumented by $\hat{D}_i$, the predicted value of distance by running regression from equation 12. We also include the baseline JS score and the same set of demographic controls in equation 11.

Table C4 reports 2SLS results under different controls represented by various Models. Under each model, columns named (1) for the second row report the results of the first stage and (2) for the second stage. Overall, the results are very stable under different controls. Distance is associated with a negative sign, which means parents living far away from schools are less likely to participate, even though it's not significant. Families that are in the treatment group and achieve higher Jumpstart scores are more likely to join more weeks.

From the second stage, even after controlling for the endogeneity of participation, we still see a significant positive effect of Jumpstart. We also find that residuals do not have a significant effect on predicting the endline Jumpstart score under different controls, suggesting no significant selection issue.

Table C4: Self-Selection: 2SLS results

	Model 1		Model 2		Model 3		Model 4	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Distance(km)	-0.016		-0.028		-0.020		-0.031	
	(0.056)		(0.053)		(0.061)		(0.058)	
Treat	17.265***		17.255***		17.323***		17.311***	
	(0.416)		(0.414)		(0.409)		(0.407)	
Baseline Jumpstart Score	1.964**	0.423***	2.423**	0.397***	1.774*	0.400***	2.227**	0.372***
	(0.862)	(0.045)	(0.877)	(0.047)	(0.888)	(0.041)	(0.880)	(0.043)
Completed Weeks		0.004***		0.004***		0.004***		0.004***
		(0.001)		(0.001)		(0.001)		(0.001)
Residuals		0.002		0.002		0.001		0.002
		(0.002)		(0.002)		(0.002)		(0.002)
Constant	-0.913**	0.474***	4.510*	0.207***	1.993	0.408***	7.254**	0.133*
	(0.432)	(0.033)	(2.584)	(0.069)	(1.314)	(0.043)	(3.078)	(0.074)
Children Demographic Controls	NO	NO	YES	YES	NO	NO	YES	YES
Parents Demographic Controls	NO	NO	NO	NO	YES	YES	YES	YES
Observations	890	890	890	890	890	890	890	890

Note. This table shows the Two-Stage Least Squares (2SLS) estimation for self-selection in the Jumpstart program. Models 1-4 represent the estimation results under different controls: with or without children/parents demographics. For each model, Odd columns labeled with (1) report the first-stage results from estimating equation 12, while even columns labeled with (2) report the second-stage results from estimating equation 11. Standard errors (in parentheses) are clustered at the school level (*$p < 0.10$,**$p < 0.05$,***$p < 0.01$).

C.2 Attrition

Table C5: First-stage Results of Probit Estimation of Attrition

	Attrit
VARIABLES	
Baseline Test Scores	
Jumpstart Test	1.784***
	(0.273)
Baseline Child Demographics	
Male	0.698***
	(0.118)
Age	-0.070
	(0.145)
Baseline Parent Demographics	
Hispanic	0.604***
	(0.178)
Black	0.593***
	(0.120)
High school graduate	-0.013
	(0.020)
Age	-0.025
	(0.018)
Employed	0.066
	(0.103)
Baseline Household Demographics	
Food stamp	0.233**
	(0.110)
Number of books >10	0.032
	(0.126)
Number of children	0.034
	(0.065)
Missing indicator	1.171***
	(0.167)
Constant	-2.472***
	(0.653)
Observations	1,126
R-squared	0.458

Note: This table shows the results of the first-stage Probit estimation when we deal with attrition using the inverse probability weighting, the new approach we described, and the Heckman selection model. Attrit = 1 if the observation appears at the baseline but not the endline. Standard errors are clustered at the school level ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Table C6: Characteristics of Jumpstart Program Attrition

	Non-Attrit	Attrit	Difference	P-value
Baseline Test Scores				
Jumpstart Test	0.51	0.52	-0.01	0.59
Bracken Test	0.24	0.27	-0.03**	0.04
Baseline Child Demographics				
Male	0.49	0.47	0.02	0.51
Age	3.52	3.55	-0.03	0.21
Baseline Parent Demographics				
Hispanic	0.60	0.49	0.11***	0.002
Black	0.26	0.34	-0.08**	0.01
High school graduate	0.68	0.64	0.04	0.27
Age	32.60	31.41	1.19**	0.02
Employed	0.34	0.45	-0.11***	0.001
Baseline Household Demographics				
Food stamp	0.46	0.53	-0.07*	0.05
Number of books > 10	0.40	0.40	0.001	0.97
Number of children	2.84	2.80	0.03	0.71

Note: This table compares the attrition and non-attrition samples for families who joined the Jumpstart program. $^{*}\mathrm{p} < 0.10,^{**}\mathrm{p} < 0.05,^{***}\mathrm{p} < 0.01$.

Table C7: Characteristics of Jumpstart Program Attrition for Control Group

	Non-Attrit	Attrit	Difference	P-value
Baseline Test Scores				
Jumpstart Test	0.50	0.52	-0.02	0.46
Bracken Test	0.23	0.26	-0.03*	0.06
Baseline Child Demographics				
Male	0.48	0.54	-0.06	0.21
Age	3.51	3.52	-0.004	0.89
Baseline Parent Demographics				
Hispanic	0.62	0.53	0.09**	0.04
Black	0.25	0.31	-0.06	0.18
High school graduate	0.64	0.68	-0.04	0.35
Age	32.50	31.18	1.31*	0.05
Employed	0.34	0.43	-0.09*	0.06
Baseline Household Demographics				
Food stamp	0.47	0.55	-0.08	0.10
Number of books > 10	0.41	0.40	0.007	0.89
Number of children	2.80	2.72	0.08	0.48

Note: This table compares the attrition and non-attrition samples for families who joined the Jumpstart program and we focus on those in the control group. $^{*}\mathrm{p} < 0.10,^{**}\mathrm{p} < 0.05,^{***}\mathrm{p} < 0.01$.

Table C8: Characteristics of Jumpstart Program Attrition for Treated Group

	Non-Attrit	Attrit	Difference	P-value
Baseline Test Scores				
Jumpstart Test	0.51	0.51	0.00	1.00
Bracken Test	0.25	0.28	-0.03	0.20
Baseline Child Demographics				
Male	0.51	0.36	0.15***	0.009
Age	3.52	3.58	-0.07*	0.05
Baseline Parent Demographics				
Hispanic	0.59	0.44	0.15***	0.007
Black	0.26	0.39	-0.13**	0.01
High school graduate	0.71	0.56	0.14***	0.007
Age	32.69	31.79	0.89	0.27
Employed	0.33	0.48	-0.15***	0.007
Baseline Household Demographics				
Food stamp	0.45	0.50	-0.05	0.38
Number of books > 10	0.39	0.39	0.00	0.99
Number of children	2.87	2.95	-0.08	0.60

Note: This table compares the attrition and non-attrition samples for families who joined the Jumpstart program and we focus on those in the treated group. $^{*}\text{p} < 0.10,^{**}\text{p} < 0.05,^{***}\text{p} < 0.01$.

Table C9: Sensitivity of the Intent-to-Treat Treatment Effect Parameter with Respect to Attrition

Jumpstart Submodules

Outcome	**Point Estimates**				**Lee Bounds**	
	No Correction for Attrition	**Inverse Probability Weight**	**Balancing**	**Heckman Selection Model**	**Lower Bound**	**Upper Bound**
Colors	3.1**	2.0	2.7	2.1	-2.3	9.2***
(Cluster school SE)	(1.5)	(2.5)	(1.7)	(2.6)	(1.5)	(1.8)
(Cluster Family SE)	(1.5)	(2.3)	(1.7)	(2.4)	(1.6)	(1.7)
(Robust SE)	(1.5)	(2.3)	(1.7)	(2.4)	(1.7)	(1.6)
Fine Motor Skills	4.4***	3.7*	4.6***	3.8*	0.4	9.3***
(Cluster school SE)	(1.4)	(2.0)	(1.6)	(2.0)	(1.1)	(1.7)
(Cluster Family SE)	(1.3)	(2.0)	(1.4)	(2.0)	(1.5)	(1.4)
(Robust SE)	(1.3)	(2.0)	(1.4)	(2.0)	(1.4)	(1.5)
Counting	3.2	2.3	3.3	2.8	-4.8	13.4***
(Cluster school SE)	(3.4)	(3.3)	(4.1)	(3.1)	(3.2)	(3.6)
(Cluster Family SE)	(2.6)	(3.7)	(2.9)	(3.7)	(2.9)	(3.0)
(Robust SE)	(2.6)	(3.7)	(2.9)	(3.7)	(2.7)	(2.9)
Sorting	7.5**	9.8*	8.0**	9.5*	-0.5	19.4***
(Cluster school SE)	(3.7)	(5.1)	(3.8)	(5.1)	(3.6)	(4.4)
(Cluster Family SE)	(2.9)	(4.2)	(3.2)	(4.2)	(3.3)	(3.5)
(Robust SE)	(2.9)	(4.2)	(3.2)	(4.2)	(3.2)	(3.6)
Name Recognition	23.2***	23.5***	25.5***	22.2***	13.6***	34.0***
(Cluster school SE)	(3.8)	(5.1)	(4.0)	(5.1)	(3.5)	(4.5)
(Cluster Family SE)	(3.1)	(4.5)	(3.4)	(4.5)	(3.5)	(3.7)
(Robust SE)	(3.1)	(4.5)	(3.4)	(4.5)	(3.4)	(3.8)
Book Handling	22.2***	18.8***	22.6***	18.9***	14.3***	30.2***
(Cluster school SE)	(2.4)	(3.4)	(2.7)	(3.5)	(2.2)	(3.0)
(Cluster Family SE)	(2.4)	(3.4)	(2.6)	(3.4)	(2.5)	(2.7)
(Robust SE)	(2.4)	(3.4)	(2.6)	(3.4)	(2.4)	(2.7)

Note: This table shows the sensitivity of intention-to-treat (ITT) treatment effect parameters accounting for attrition for Jumpstart Test Submodules. The first column is taken from Model 3 in Table 6. The second and third columns report results using inverse probability weight and the Heckman selection model. The last two columns report results using Lee sharp bounds methods. We report three different standard errors: robust, clustered at the school level, and clustered at the family level. We report the most conservative significance level (*p < 0.10, **p < 0.05, ***p < 0.01).

Table C10: Sensitivity of the Intent-to-Treat Treatment Effect Parameter with Respect to Attrition

Bracken Submodules

Outcome	**Point Estimates**				**Lee Bounds**	
	No Correction for Attrition	**Inverse Probability Weight**	**Balancing**	**Heckman Selection Model**	**Lower Bound**	**Upper Bound**
Colors	4.3***	5.1	4.8**	5.1	-1.5	11.0***
(Cluster school SE)	(1.7)	(4.0)	(2.1)	(3.9)	(1.7)	(1.9)
(Cluster Family SE)	(1.7)	(2.6)	(1.9)	(2.5)	(1.9)	(1.8)
(Robust SE)	(1.7)	(2.6)	(1.9)	(2.5)	(1.9)	(1.8)
Letters	2.7	3.7	1.6	3.9	-3.6*	7.2***
(Cluster school SE)	(1.8)	(2.4)	(2.1)	(2.4)	(1.8)	(1.9)
(Cluster Family SE)	(1.5)	(2.3)	(1.6)	(2.3)	(1.6)	(1.7)
(Robust SE)	(1.5)	(2.3)	(1.6)	(2.3)	(1.6)	(1.6)
Numbers	2.6*	1.6	3.6**	2.0	-3.5**	7.0***
(Cluster school SE)	(1.5)	(2.6)	(1.6)	(2.6)	(1.6)	(1.6)
(Cluster Family SE)	(1.5)	(2.3)	(1.6)	(2.3)	(1.6)	(1.6)
(Robust SE)	(1.5)	(2.3)	(1.6)	(2.3)	(1.6)	(1.5)
Sizes	1.3	1.9	1.9	2.0	-3.0**	4.8***
(Cluster school SE)	(1.1)	(1.6)	(1.1)	(1.6)	(1.2)	(1.4)
(Cluster Family SE)	(1.1)	(1.7)	(1.2)	(1.7)	(1.2)	(1.2)
(Robust SE)	(1.1)	(1.7)	(1.2)	(1.7)	(1.2)	(1.3)
Shapes	0.8	0.5	1.1	0.7	-6.1**	8.2***
(Cluster school SE)	(2.6)	(3.5)	(2.7)	(3.6)	(2.4)	(3.0)
(Cluster Family SE)	(2.6)	(2.9)	(2.9)	(3.1)	(2.9)	(2.3)
(Robust SE)	(2.6)	(2.9)	(2.9)	(3.1)	(2.6)	(2.2)

Note: This table shows the sensitivity of intention-to-treat (ITT) treatment effect parameters accounting for attrition for Bracken Submodules. The first column is taken from Model 3 in Table 7. The second and third columns report results using inverse probability weight and the Heckman selection model. The last two columns report results using Lee sharp bounds methods. We report three different standard errors: robust, clustered at the school level, and clustered at the family level. We report the most conservative significance level ($^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Table C11: Sensitivity of the Intent-to-Treat Treatment Effect Parameter with Respect to Attrition

Parental Inputs Subitems

	Point Estimates				Lee Bounds	
Outcome	**No Correction for Attrition**	**Inverse Probability Weight**	**Balancing**	**Heckman Selection Model**	**Lower Bound**	**Upper Bound**
Helps Child Learn Alphabet	0.021	0.022	0.007	0.022	-0.023	0.106***
(Cluster school SE)	(0.021)	(0.021)	(0.023)	(0.022)	(0.021)	(0.016)
(Cluster Family SE)	(0.021)	(0.021)	(0.023)	(0.021)	(0.021)	(0.018)
(Robust SE)	(0.021)	(0.021)	(0.023)	(0.021)	(0.022)	(0.017)
Helps Child Learn Colors	-0.006	-0.005	-0.017	-0.006	-0.022**	0.025***
(Cluster school SE)	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.008)
(Cluster Family SE)	(0.011)	(0.011)	(0.012)	(0.011)	(0.011)	(0.007)
(Robust SE)	(0.011)	(0.011)	(0.012)	(0.011)	(0.010)	(0.007)
Helps Child Learn Numbers	0.006	0.005	0.012	0.006	-0.022	0.049***
(Cluster school SE)	(0.014)	(0.014)	(0.015)	(0.014)	(0.012)	(0.012)
(Cluster Family SE)	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)	(0.010)
(Robust SE)	(0.013)	(0.013)	(0.014)	(0.013)	(0.014)	(0.011)
Help Child Learn Shapes	-0.040	-0.040	-0.051*	-0.039	-0.097***	0.071**
(Cluster school SE)	(0.023)	(0.022)	(0.028)	(0.023)	(0.026)	(0.031)
(Cluster Family SE)	(0.026)	(0.026)	(0.029)	(0.026)	(0.027)	(0.032)
(Robust SE)	(0.026)	(0.026)	(0.029)	(0.026)	(0.028)	(0.032)

Note: This table shows the sensitivity of intention-to-treat (ITT) treatment effect parameters accounting for attrition for Bracken Submodules. The first column is taken from Model 3 in Table 8. The second and third columns report results using inverse probability weight and the Heckman selection model. The last two columns report results using Lee sharp bounds methods. We report three different standard errors: robust, clustered at the school level, and clustered at the family level. We report the most conservative significance level ($^{*}\mathrm{p} < 0.10$, $^{**}\mathrm{p} < 0.05$, $^{***}\mathrm{p} < 0.01$).

Appendix D Details about the Decomposition Model

D.1 Construction of the Investment Measure

In the first stage, we conduct a factor analysis using the following variables: (i) the number of days per week the parent reads to the child, (ii) the number of minutes per day devoted to reading, (iii) the number of hours the child watches television, (iv) a binary indicator for whether there are at least ten children's books in the household, and (v) four items indicating whether the parent helps the child learn numbers, the alphabet, colors, or shapes. We estimate this factor model separately for both the baseline and endline data. This estimation yields three orthogonal factors, but only one of which strongly loads on all items.

In the second stage, we apply an oblique rotation. Rotating the factors helps clarify the structure by encouraging each variable to load most strongly on a single factor, thereby making the factors more interpretable. Unlike an orthogonal rotation (which constrains factors to be uncorrelated), an oblique rotation allows factors to correlate with one another, reflecting the idea that different dimensions of parental engagement and the home environment may be interrelated in practice. The rotation method uses a two-stage procedure (commonly beginning with a varimax rotation) followed by a power transformation with an exponent of 5, sharpening the factor loadings and yielding a simpler, more interpretable factor structure. The factor rotation generates three correlated factors with one of which that loads more heavily on the frequency that the parent reads to the child and the dummy variable that indicates that there are at least 10 children's books in the household. These analyses allow us to extract a single underlying latent factor that captures common dimensions of the home environment and parental support for early learning. Table D12 presents the factor loadings and the variances of the uniquenesses before and after the rotation, for the baseline and endline investment data.

Table D12: Factor Loadings and Variances of the Uniqueness Before and After Rotation

Measure Description	**Before Rotation**				**After Rotation**			
	Factor 1	**Factor 2**	**Factor 3**	**Uniqueness**	**Factor 1**	**Factor 2**	**Factor 3**	**Uniqueness**
Panel A: Baseline								
Number of days the parent reads to the child	0.451	0.398	0.026	0.637	0.005	0.556	0.130	0.637
Number of minutes per day the parent reads	0.301	0.381	-0.123	0.750	-0.028	0.523	-0.053	0.750
Number of hours the child watches TV	-0.028	-0.035	0.216	0.951	-0.098	-0.084	0.234	0.951
Number of children's books is at least 10	0.276	0.104	0.284	0.833	-0.008	0.137	0.359	0.833
Parent helps the child learn numbers	0.562	-0.126	-0.067	0.664	0.588	0.017	-0.037	0.664
Parent helps the child learn the alphabet	0.621	-0.136	-0.057	0.593	0.638	0.019	-0.023	0.593
Parent helps the child learn colors	0.611	-0.186	-0.047	0.590	0.669	-0.041	-0.019	0.590
Parent helps the child learn shapes	0.544	-0.102	0.100	0.684	0.471	0.010	0.151	0.684
Panel B: Endline								
Number of days the parent reads to the child	0.418	-0.281	-0.023	0.746	0.515	0.005	-0.029	0.746
Number of minutes per day the parent reads	0.155	-0.137	0.048	0.955	0.200	-0.047	0.050	0.955
Number of hours the child watches TV	-0.120	0.182	0.121	0.938	-0.288	0.045	0.152	0.938
Number of children's books is at least 10	0.326	-0.272	-0.005	0.820	0.446	-0.047	-0.014	0.820
Parent helps the child learn numbers	0.329	0.316	-0.059	0.788	-0.081	0.485	-0.003	0.788
Parent helps the child learn the alphabet	0.434	0.204	0.071	0.765	0.033	0.401	0.141	0.765
Parent helps the child learn colors	0.358	0.192	-0.086	0.828	0.063	0.404	-0.047	0.828
Parent helps the child learn shapes	0.402	-0.030	0.093	0.829	0.225	0.171	0.134	0.829

In the third stage, we generate predicted factor scores using the Bartlett prediction formula for the endline investment factor, and the regression-based prediction formula for the baseline investment factor. We then relocate and rescale each factor so that its mean and variance match those of the weekly reading frequency measure for the corresponding round of data collection.

D.2 Derivation of the Estimating Equations

We derive the main estimation equations by solving for parents' optimal investment choices under both program graduation and non-graduation. We begin by substituting the knowledge production function (2) and the habitual investment specification (4) into the utility function (3). This yields

$$U\left(X_i, G_i\right)=\beta_0+\beta_1 K_{i,0}+\beta_3 X_i+\beta_4 G_i X_i-\left(\eta_i-\alpha_0-\beta_2-\alpha_1 Z_i\right) G_i-\frac{1}{2\alpha_2}\left(X_i-\gamma_1 X_{i,0}-\gamma_2 G_i-\nu_i\right)^2 \tag{13}$$

Taking the first-order condition with respect to X_i and rearranging the resulting expression leads to:

$$X_i=\alpha_2\beta_3+\gamma_1 X_{i,0}+\left(\gamma_2+\alpha_2\beta_4\right) G_i+\nu_i \tag{14}$$

Substituting (14) back into (13) yields

$$U\left(G_i\right)=U_{i,0}+a_0 G_i+\alpha_1 Z_i G_i+\gamma_1\beta_4 X_{i,0} G_i+\left(\beta_4\nu_i-\eta_i\right), \tag{15}$$

where:

$$\begin{aligned}
U_{i,0} &= \left(\beta_0+\alpha_2\beta_3-\frac{\alpha_2}{2}\beta_3^2\right)+\beta_1 K_{i,0}+\gamma_1\beta_3 X_{i,0}+\beta_3\nu_i+\epsilon_i, \\
a_0 &= \alpha_0+\beta_2+\left(\beta_3+\beta_4\right)\left(\gamma_2+\alpha_2\beta_4\right)-\frac{\alpha_2\beta_3^2}{2}
\end{aligned}$$

A parent chooses to participate in the JSP if and only if, $U\left(G_i=1\right) \geq U\left(G_i=0\right)$, or:

$$G_i=1 \iff \eta_i-\beta_4\nu_i \leq \alpha_0+\alpha_1 Z_i+\gamma_1\beta_4 X_{i,0}. \tag{16}$$

Because $\eta_i-\beta_4\,\nu_i$ is normally distributed with mean zero and variance

$$\begin{gathered}
\left(\sigma^*\right)^2=\sigma_\eta^2+\beta_4\sigma_{\eta\nu}+\beta_4^2\sigma_\nu^2 \\
\mathbf{Pr}\left(G_i=1 \mid Z_i, X_{i,0}\right)=\Phi\left(\frac{a_0+\alpha_1 Z_i+\gamma_1\beta_4 X_{i,0}}{\sigma^*}\right),
\end{gathered}$$

which corresponds to the choice probability equation (5) in the main text.

Next, let X_i^0 and X_i^1 denote the optimal levels of investment under non-graduation ($G_i = 0$) and graduation ($G_i = 1$), respectively. From (14), we have

$$\mathbf{E}\left(X_i^0 \mid X_{i,0}, G_i = 0\right) = \alpha_2\beta_3 + \gamma_1 X_{i,0} + \mathbf{E}\left(\nu_i \mid X_{i,0}, G_i = 0\right) \tag{17}$$

However, $\mathbb{E}\big(\nu_i \,\big|\, X_{i,0}, G_i = 0\big) \neq 0$ because ν_i influences the graduation decision, thereby creating selection bias. To address this issue, we exploit the joint normality of ν_i and $\eta_i - \beta_4\,\nu_i$. By expressing ν_i as a linear function of $\eta_i - \beta_4\,\nu_i$ plus an independent noise term, it follows that

$$\begin{aligned}
\nu_i &= \frac{\mathbf{Cov}\left(\nu_i, \eta_i - \beta_4\nu_i\right)}{\mathbf{Var}\left(\eta_i - \beta_4\nu_i\right)}\left(\eta_i - \beta_4\nu_i\right) + \xi_{i,0} \Rightarrow \\
\nu_i &= \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\left(\sigma^*\right)^2}\left(\eta_i - \beta_4\nu_i\right) + \xi_{i,0} \Rightarrow \\
\nu_i &= \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma^*}\frac{\left(\eta_i - \beta_4\nu_i\right)}{\sigma^*} + \xi_{i,0},
\end{aligned}$$

where $\xi_{i,0}$ is normally distributed with mean zero, variance $\sigma_\nu^2 - \frac{\left(\sigma_{\eta\nu} - \beta_4\sigma_\nu^2\right)^2}{\left(\sigma^*\right)^2}$, and independently from $(\eta_i - \beta_4\nu_i)$. Next, note that:

$$\begin{aligned}
\mathbf{E}\left(\nu_i \mid X_{i,0}, G_i = 0\right) &= \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma^*}\mathbf{E}\left(\frac{\eta_i - \beta_4\nu_i}{\sigma^*}\,\middle|\, X_{i,0}, G_i = 0\right) \Rightarrow \\
&= \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma^*}\mathbf{E}\left(\frac{\eta_i - \beta_4\nu_i}{\sigma^*}\,\middle|\, X_{i,0}, \frac{\eta_i - \beta_4\nu_i}{\sigma^*} > \frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right) \Rightarrow \\
&= \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma^*}\frac{\phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right)}{1 - \Phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right)}
\end{aligned}$$

Substituting this back into (17) gives

$$\mathbf{E}\left(X_i^0 \mid X_{i,0}, G_i = 0\right) = \alpha_2\beta_3 + \gamma_1 X_{i,0} + \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma^*}\frac{\phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right)}{1 - \Phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right)}$$

A similar derivation implies that

$$\mathbf{E}\left(X_i^1 \mid X_{i,0}, G_i = 1\right) = \alpha_2\beta_3 + \left(\gamma_2 + \alpha_2\beta_4\right) + \gamma_1 X_{i,0} - \frac{\sigma_{\eta\nu} - \beta_4\sigma_\nu^2}{\sigma^*}\frac{\phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right)}{\Phi\left(\frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}\right)}$$

Next, note that $X_i = \left(1 - G_i\right) X_i^0 + G_i X_i^1$. If we use the definition of π_i in the main paper, we obtain:

$$X_i = \alpha_2\beta_3 + (\gamma_2 + \alpha_2\beta_4)\, G_i + \gamma_1 X_{i,0} - \frac{\sigma_{\eta\nu} + \beta_4\sigma_\nu^2}{\sigma^*}\pi_i + \xi_i$$

Finally, the child's knowledge equation can also be expressed in reduced form by substituting (14) into the knowledge production function (2). This yields

$$K_{i,1} = \tilde{\beta}_0 + \beta_1 K_{i,1} + \tilde{\beta}_2 G_i + \tilde{\beta}_3 X_{i,0} + \tilde{\beta}_4 G_i X_{i,0} + (\beta_3 + \beta_4 G_i)\,\nu_i + \epsilon_i$$

For non-participants and participants, respectively, end-of-year knowledge is

$$\mathbf{E}\left(K_{i,1}^0 \mid K_{i,0}, X_{i,0}, G_i = 0\right) = \tilde{\beta}_0 + \beta_1 K_{i,1} + \tilde{\beta}_3 X_{i,0} + \mathbf{E}\left(\beta_3\nu_i + \epsilon_i \middle|\, K_{i,0}, X_{i,0}, G_i = 0\right) \tag{18}$$

Because of selection, the term $\mathbf{E}\left(\nu_i \middle|\, K_{i,0}, X_{i,0}, G_i = 0\right) \neq 0$. Thus, we must also add the control function to account for the selection bias. If we use the definitions of $\pi_{i,0}$ and $\pi_{i,1}$ from the main text, and follow the steps we derived above, we show that:

$$\begin{aligned}
\mathbf{E}\left(K_{i,1}^0 \middle|\, K_{i,0}, X_{i,0}, G_i = 0\right) &= \tilde{\beta}_0 + \beta_1 K_{i,1} + \tilde{\beta}_3 X_{i,0} + \delta_0 \pi_{i,0} \\
\mathbf{E}\left(K_{i,1}^1 \middle|\, K_{i,0}, X_{i,0}, G_i = 1\right) &= \tilde{\beta}_0 + \tilde{\beta}_2 + \beta_1 K_{i,1} + \left(\tilde{\beta}_3 + \tilde{\beta}_4\right) X_{i,0} + \delta_1 \pi_{i,1}
\end{aligned}$$

where:

$$\begin{aligned}
\delta_0 &= \frac{\beta_3\sigma_{\eta\nu} - \beta_3\beta_4\sigma_\nu^2 + \sigma_{\eta\epsilon} - \beta_4\sigma_{\nu\epsilon}}{\sigma^*} \\
\delta_1 &= \frac{(\beta_3 + \beta_4)\,\sigma_{\eta\nu} - \beta_4\,(\beta_3 + \beta 4)\,\sigma_\nu^2 + \sigma_{\eta\epsilon} - \beta_4\sigma_{\nu\epsilon}}{\sigma^*}
\end{aligned}$$

Now, define the error term ζ_i such that:

$$\zeta_i = K_{i,1} - (1 - G_i)\,\mathbf{E}\left(K_{i,1}^0 \middle|\, K_{i,0}, X_{i,0}, G_i = 0\right) - G_i \mathbf{E}\left(K_{i,1}^1 \middle|\, K_{i,0}, X_{i,0}, G_i = 1\right).$$

Then,

$$K_{i,1} = \tilde{\beta}_0 + \beta_1 K_{i,1} + \tilde{\beta}_2 G_i + \tilde{\beta}_3 X_{i,0} + \tilde{\beta}_4 G_i X_{i,0} + \zeta_i \tag{19}$$

D.3 More on Identification

Although not necessary for our analysis, we can identify additional structural parameters. For example, turning to the graduation equation (5), we identify $\frac{\gamma_1\beta_4}{\sigma^*}$. Because γ_1 and β_4 are known by this stage, we can infer the value of σ^*. This, in turn, allows us to recover a_0 and α_1 when estimating equation (5). We can then use our knowledge of a_0, β_0, α_2 and β_3 to recover α_0.

In addition, the estimation of equation (6) identifies the variance of ξ. This information is useful because the variance of ν and the variance of ξ are linked by the following equation:

$$\sigma_\nu^2 = \frac{\delta_0^2}{(\sigma^*)^2} + \sigma_\xi^2$$

where we already know δ_0, σ^*, and σ_ξ^2. Thus, σ_ν^2 is identified. A similar argument identifies σ_ϵ^2, which is the variance of the error term in equation (22). Furthermore, consider the following difference:

$$\sigma^* (\delta_1 - \delta_0) = \beta_3 \sigma_{\eta\nu} - \beta_4^2 \sigma_\nu^2.$$

Thus, we can solve for $\sigma_{\eta\nu}$. Indeed:

$$\sigma_{\eta\nu} = \frac{\sigma^* (\delta_1 - \delta_0) + \beta_4^2 \sigma_\nu^2}{\beta_3}$$

Finally, note that $(\sigma^*)^2 = \sigma_\eta^2 - \beta_4 \sigma_{\eta\nu} + \beta_4^2 \sigma_\nu^2$. Thus, we can identify σ_η^2. Note, however, that we do not have information to identify the remaining terms $\sigma_{\eta\epsilon}$ and $\sigma_{\nu\epsilon}$ of the variance-covariance matrix Σ. However, identification of these parameters is not necessary for the goals of our analysis.

D.4 Semiparametric Approach

In this section, we present the estimating equations without invoking assumptions about the joint distribution of the error terms. First, the probability that parent i will participate in the JumpStart Program is:

$$\mathbf{Pr}\left(G_i = 1 | \, Z_i, X_{i,0}\right) = P\left(a_0 + a_1 Z_i + a_2 X_{i,0}\right) = P\left(Z_i, X_{i,0}\right), \tag{20}$$

where $a_0 = \frac{\alpha_0 + \beta_0 + \alpha_2\beta_3 - \frac{\alpha_2}{2}\beta_3^2}{\sigma^*}$, $a_1 = \frac{\alpha_1}{\sigma^*}$, $a_2 = \frac{\gamma_1\beta_4}{\sigma^*}$ and $\eta_i^* = \frac{\alpha_0 + \alpha_1 Z_i + \gamma_1\beta_4 X_{i,0}}{\sigma^*}$. As we discuss below, we cannot separately identify the unknown function P from the intercept a_0. This intercept, however, is identified in the parametric approach because then the function P is known. In any case, knowledge of the intercept a_0 is not necessary for the goals of our analysis.

The estimating equation for the optimal investment depends on the parent's decision to graduate from the JumpStart Program. We can write the investment equation as:

$$X_i^* = \alpha_2\beta_3 + \gamma_1 X_{i,0} + (\alpha_2\beta_4 + \gamma_2) G_i + CF_X(Z_i, X_{i,0}) + \xi_i, \tag{21}$$

Where $CF_X(Z_i, X_{i,0})$ is the following control function:

$$CF_X(Z_i, X_{i,0}) = (1 - G_i) Q_{X,0}(P(Z_i, X_{i,0})) + G_i Q_{X,1}(P(Z_i, X_{i,0}))$$

and $Q_{X,0}$ and $Q_{X,1}$ are polynomial functions of $P(Z_i, X_{i,0})$.

Next, define:

$$\begin{aligned}
\tilde{\beta}_0 &= \beta_0 + \alpha_2 \beta_3^2, \\
\tilde{\beta}_2 &= \beta_2 + \alpha_2 \beta_3 \beta_4 + (\beta_3 + \beta_4)(\gamma_2 + \alpha_2 \beta_4), \\
\tilde{\beta}_3 &= \gamma_1 \beta_3, \\
\tilde{\beta}_4 &= \gamma_1 \beta_4.
\end{aligned}$$

These expressions are useful to derive the reduced-form equation for knowledge, which, once we include a control function to account for selection, which we denote by $CF_K(Z_i, X_{i,0})$, is:

$$K_{i,1} = \tilde{\beta}_0 + \beta_1 K_{i,0} + \tilde{\beta}_2 G_i + \tilde{\beta}_3 X_{i,0} + \tilde{\beta}_4 G_i X_{i,0} + CF_K(Z_i, X_{i,0}) + \zeta_i \tag{22}$$

D.5 Decomposition Using the Semiparametric Model

Table D13: Decomposition of Average Treatment Effect According to the Semiparametric Model

Panel A: Full Model

	Control Function is Poly. of Deg. 1				**Control Function is Poly. of Deg. 3**			
	Bandwidth = 0.24		Bandwidth = 0.72		Bandwidth = 0.24		Bandwidth = 0.72	
	Est.	Std. Dev.	Est.	Std. Dev.	Est.	Std. Dev.	Est.	Std. Dev.
Total Effect (ATE)	11.085	2.612	11.173	3.032	19.232	7.428	16.226	10.493
Direct Effect	11.100	9.826	11.137	9.397	19.974	11.782	19.109	11.727
Indirect Effect through Preferences	2.757	1.457	2.889	1.563	0.324	1.870	-0.138	3.162
Indirect Effect through Marg. Prod.	-0.442	1.778	-0.450	1.717	-0.186	1.671	-0.524	1.756
Indirect Effect through Interaction	-2.329	8.735	-2.403	8.556	-0.880	7.631	-2.222	6.694

Panel B: Restricted Model

	Control Function is Poly. of Deg. 1				**Control Function is Poly. of Deg. 3**			
	Bandwidth = 0.22		Bandwidth = 0.66		Bandwidth = 0.22		Bandwidth = 0.66	
	Est.	Std. Dev.	Est.	Std. Dev.	Est.	Std. Dev.	Est.	Std. Dev.
Total Effect (ATE)	11.071	2.668	11.029	3.047	15.220	6.485	20.420	11.485
Direct Effect	8.759	2.897	8.483	3.283	12.627	6.337	16.911	11.277
Indirect Effect through Preferences	2.311	1.084	2.547	1.201	2.592	1.686	3.509	2.728