An Evaluation of the Alief Independent School District Jump Start Program: Using a Model to Recover Mechanisms from an RCT *

Flávio Cunha†, Andrea Salvati†† and Kenneth I. Wolpin†††

January 31st 2021

Abstract

Recent research shows that the substantial differences in school readiness observed at the beginning of Kindergarten across socio-economic groups are partly due to disparities in children's early environment quality. Theory, consistent with a wide range of data, suggests that early interventions that target malleable, fundamental skills during sensitive periods of development in early childhood could help close these gaps. Empirical evidence shows that small-scale parenting interventions implemented by high-quality staff can improve parental investments and boost child development. Evidence about the impact of large-scale parenting interventions is more mixed. This paper reports the evaluation of a parenting intervention developed and implemented by the Alief Independent School District in Texas. The intervention's goal is to encourage and train parents to teach their children foundational skills for Pre-K. The results of a randomized controlled trial based on three yearly cohorts show that the program impacted parental investments and child development as measured by two different tests of school readiness. We go beyond reporting program impacts by building and estimating a model of parental choice of input levels. Our model allows for a production function of knowledge that features individual-specific coefficients that capture parental inputs' marginal productivity. We find that the model estimates validate the mechanism we posit for the program's impact.

JEL Codes H31,I24,I25,I28,I3,J38,O1,O15,O54

---

* 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.

† Department of Economics and Texas Policy Lab, Rice University; NBER.

†† Department of Economics, Rice University

††† Professor Emeritus of Economics, University of Pennsylvania; Professor of Economics, Rice University until July, 2020.

1. Introduction

Recent research shows that there are still large differences in school readiness at the beginning of Kindergarten across socio-economic groups. According to Reardon and Portilla (2016), the gaps in reading and math achievement by income are approximately one standard deviation among a cohort of children born in 2010. And, as children progress through school grades, income gaps in achievement remain remarkably stable (Bond and Lang, 2013). The difficulty in overcoming these deficits in academic preparation after children start formal schooling is an important reason why researchers and policymakers alike have turned their attention to programs that aim to foster early childhood development.

Empirical research suggests that income gaps in early academic achievement are partly generated by disparities in the home environment. Despite progress in the last two decades, disadvantaged children still experience substantial deficits in the quantity and quality of early learning opportunities, whether they are measured in the time devoted to learning (Kalil et al., 2012), learning materials (Bossok et al., 2016), or routine activities that provide quality interaction with parents and other adults (e.g, family dinners, Putnam, 2016). Theoretical research suggests that early intervention that targets malleable and fundamental skills that have sensitive periods of development in early childhood may generate large impacts on human capital formation (Cunha and Heckman, 2007). Indeed, empirical research has shown that small-scale parenting interventions implemented by high-quality research staff can lead to an improvement in parental investments and a boost in child development (Suskind et al., 2015; Cunha, Gerdes, and Nihtianova, 2019).

In contrast, evidence about the impact of large-scale parenting interventions is mixed. (Kalil, 2014). Heckman et al. (2017) conducted an analysis of the Nurse Family Partnership Program in Memphis, TN. They found that, by the end of the program (when children were two years old) mothers served by the nurses had higher levels of home investments, better parenting attitudes, and superior mental health. As a result, they report positive impacts for many measures of child development at age 6 and as late as age 12. However, such successes seem to be an exception in this literature. For example, Love et al. (2005) found that the home visitation component of Early Head Start achieved modest impacts on family investments or child development. Similarly, St. Pierre et al. (2003) report that the Department of Education’s Even Start Program (at a cost of $ $13,674 per family) yielded no 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 (in over 100 locations in the U.S. and in 14 countries) costing between $1,500 to $2,000 per family per year, also have found mixed results. Although one small-scale randomized trial of 69 HIPPY participants produced substantial improvements in test scores, the results were not replicated in a randomized trial of similar families who entered the program the following year (Baker, Piotrkowski, & Brooks-Gunn, 1998, 1999).

In this paper, we provide an evaluation of one such program, the Houston-area Alief Independent School District's (AISD) Jump Start Program (JSP). The goal of the JSP is to prepare three-year old children for entry into the AISD Pre-Kindergarten program by providing parents with training and materials to work with their children in their home. The evaluation is based on a three-year randomized controlled trial conducted at the request of the district. The purpose of the RCT was to provide the district with the best evidence about the impact of the JSP.

We go beyond the original purpose of the evaluation to try to understand the mechanism generating the program's impact. To that end, we develop and estimate a model of parental behavior in providing inputs into the production of children's age-relevant cognitive skills. A novel feature of the model is that, unlike in the extensive literature on the estimation of child cognitive achievement production functions, we allow for the marginal productivity of inputs to be heterogeneous across parents. The JSP is modeled as potentially shifting the distribution of parental productivity.[1] Removing the assumption that all parents have the same level of productivity (conditional perhaps on observables) raises challenging issues in the estimation of the production function parameters. The reason is that, as demonstrated in the optimizing model, inputs chosen by parents will depend on their parental productivity.

The RCT evaluation of the JSP included families living in the catchment areas of all of AISD's 24 elementary schools.[2] Overall, during the three-year evaluation period the number of families included in the program averaged about 200 per year, with an additional 200 in the control group. Families were offered 25 dollars to participate in the study. Between 90-95 percent of those registering for the Jump Start Program agreed to be in the study. All families that registered for the program were included in the lottery randomization regardless of whether they agreed to be in the study.

All children enrolled in the study were given assessments at the time the families were recruited into the program and at the end of the program. One assessment, designed by AISD staff, is based explicitly on the Jump Start curriculum. The second, the Bracken School Readiness Assessment, is a nationally normed test of five concepts, only one of which coincides with the Jump Start curriculum content. The additional assessment, specific to the evaluation, provides evidence on whether, due to the Jump Start training, parents choose to teach their children more advanced concepts and skills than those in the Jump Start curriculum. Families were offered 50 dollars to participate in the post-program assessments; depending on the year,

[1] This setting falls into the general framework of endogenous regressors with heterogeneous effects considered in Florens et al. (2008).

[2] Prior to the evaluation, the JSP was limited to about 100 families in a subset of the schools. The expansion of the program, as well as the implementation of the evaluation study, was funded by a grant from the Laura and John Arnold Foundation.

between 77-85 percent of those families agreed to participate. Not only was attrition low, but the treatment and control families in the post-program sample did not differ in observable characteristics.

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 Jump Start curriculum, the test score 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 (modified) 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 particulary large for two of the Jump Start modules, name recognition and book handling, with the treatment effect being over 20 percentage points for each.

The effect of the program on Bracken test scores were considerably smaller in absolute terms. Over the program period, the Bracken score increased from 23 to 34 percent for the control group. Based on a (modified) diff-in-diff regression, 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.

In addition to the child assessments, parents enrolled in the evaluation study also completed a questionnaire both at the time of the recruitment and at the end of the program period. Besides the usual demographic information, measures of parental interactions were also obtained in an attempt to understand the mechanisms behind the treatment impact. Data were collected 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 the 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 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.

Although JSP parents read more frequently to their child than did control parents, the relevance of this parental activity as a mechanism underlying the JSP treatment effect depends on the extent to which the JSP induced the activity and its effect on test scores. There is evidence for such an effect in the literature on educational production functions (see, for example. Todd and Wolpin (2007)). To ascertain the importance of this potential mechanism, we posit a model in which a parent decides on the frequency of reading to their child over some period of time, in this case from the start to the end of the JSP. The parent cares about the child's ending knowledge, which is determined by the child's initial knowledge and the parental input of reading time.

Adopting an explicit choice model allows for a potentially important extension of the educational production function literature. That literature has universally assumed that all parents are equally productive in producing child cognitive outcomes. Instead, we assume that parents are heterogeneous in their marginal product of reading time. We show how the RCT supplies exogenous variation that enables us to identify the parameters of the model without making a distributional assumption about the heterogeneous marginal product parameter (or any other error terms). The effect of the JSP is modeled as changing the parent marginal productivity distribution, which, in turn, optimally affects the amount of time parents spend reading to the child. We estimate the model using empirical likelihood, based on matching first- and second-moments of the data to their theoretical counterparts in the model.

We find that the mechanism we posit for the JSP is validated by the model estimates. The mean, as well as the variance, of the marginal product parameters for both the treatment and control groups are statistically significant. And, both are (statistically significantly) higher for the treatment families than for the control families. We also show how the standard approach of estimating the achievement production function leads to misleading interpretations. For example, as demonstrated by the model, the Wald estimator using the treatment assignment as an instrument for the input, not only doesn't identify the mean of the marginal product distribution, but actually is not even a function of the marginal product distribution parameters.

Our paper connects with the literature on the evaluation of parenting programs as well as with the estimation of mechanisms. Two papers are particularly relevant. First, we build on the econometric analysis of Florens et al. (2008).We estimate a production function of human capital formation that allows for individual heterogeneity in coefficients that capture the marginal productivity of inputs. Florens et al. (2008) present a rich analysis of identification, but do not implement any estimation. Because our instrument (the treatment dummy) is binary, our identification results are weaker, but we estimate a choice model and discuss the empirical importance of how heterogeneity in coefficients of marginal productivity of investments help us uncover the mechanism by which the JSP impacts children's knowledge and parental investments.

Second, our paper is related to recent work by Attanasio et al. (2015), who aim to uncover the short-term impacts of the Jamaica home visitation program as implemented in a larger scale in Colombia. The paper also takes advantage of data collected as part of a randomized controlled trial to estimate production functions of cognitive and socio-emotional skills that can differ between treatment groups. Their finding differs from ours in that they cannot reject that the treatment and control families are adopting the same production function. An important difference from our analysis is that we allow for individual-specific production function coefficients.

The rest of the paper is organized as follows. The next section briefly describes the Alief Independent School District Jump Start Program. Section 3 discusses the evaluation study, including the experimental design, the recruitment and the randomization. Section 4 presents the results from the evaluation study. Section 5 develops the details of the analysis to uncover the JSP's mechanism, including the model, identification, estimation, and results. The final section concludes.

2. The Jump Start 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 cost of the JSP, primarily for the materials, is less than $200 per family. 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 Jump Start 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 eighth-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 to enable 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.

3. The Evaluation Study

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 three-quarters of the 24 district elementary schools provided the JSP to families within their catchment area. Total enrollment throughout the district amounted to about 100 families. To participate in the program, each school had to cover the cost of the program, $150 for materials for each family, from their own budget. The cost of the program to the schools was a major impediment to a school adopting Jump Start 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.

3.1. Recruitment

Prior to the study, AISD family liaisons, paraprofessionals 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 Jump Start 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 Jump Start test (designed by AISD), that covers the content of the Jump Start curriculum. The family was then informed that 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 25 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 nationally normed test of five concepts; only one concept coincides with the Jump Start curriculum content.

The recruitment strategy in the first year of the study was based on the assumption that interest in the Jump Start 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 Jump Start Program through the lottery never attended a Jump Start session. Overall, 10 percent of the families never attended a Jump Start 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 the 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 Jump Start session, accounting for only 2 percent of the families, and only 13 percent attended fewer than 5 sessions. However, in year three, for reasons that are unclear, 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 Jump Start lottery also consented to be in the evaluation study.[3] There was even greater participation 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 Jump Start lottery of which 1,126 participated in the study, a refusal rate of only 7.4 percent.

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 in 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

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

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 block design, the same in all three years. Families were divided into four groups based on whether their score on the Jump Start 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.

Table 2 provides information on the outcome of the randomization in each year. The table excludes families that did not consent to be in the study (consistent with IRB regulations). Because all of the results of the study are based on the sample of study participants, it is most relevant to document differences in characteristics between the treatment and control groups for that sample. For each variable shown in the first column of the table, the second column shows it's mean in the study sample, the third column the difference in the value of the variable between the treatment group and the control group and the last column the p-value associated with the hypothesis that the difference shown in column three is equal to zero. The first two variables, the child's baseline Jump Start 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.

Table 3 shows the differences in the same characteristics of treatment and control groups as in table 2 for the post-program study sample. As in the baseline study sample, these differences are small and have p-values well above conventional levels of statistical significance. Comparing table 2 and 3, the three-year pooled sample of treatment and control families is almost identical in mean characteristics; the only slight difference is that Hispanics comprise about 3 percentage points more of the post-program sample than of the baseline sample.

3.3. Jump Start Program Attendance

As noted, Jump Start 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 in 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 57.2 families completed the Jump Start program in that year. Given the change in the recruitment strategy, 74.4 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 4.5, 6.8 and 4.4 for the three years.

4. Data

4.1. Baseline Test Score Summary Statistics

Table 5 provides summary statistics of the baseline Jump Start and Bracken raw test scores (percentage correct answers) for families in the post-test sample for the three years separately and pooled. The mean Jump Start 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 sample 75 percent or higher.

Bracken baseline test scores are considerably lower than Jump Start test scores. As seen, the mean baseline score on the Bracken is 23.9 percent averaged over the three years. As with the Jump Start 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. The Bracken is a nationally normed test. The children of families that applied to participate in the Jump Start program are well below the national mean. The mean percentile score, averaged over the three years, is only 27 and about 50 percent of the children are judged to be delayed or very delayed in their cognitive development.

4.2. The Impact of Jump Start on Test Scores

Tables 6 and 7 present estimates of the impact of the Jump Start program on the Jump Start and Bracken test scores using the three-year pooled sample. The first column in each table reports the end-of-year difference in the tests scores of the treatment and control families, the second the difference-in-difference estimate and the third the difference-in-difference estimate conditioning on the initial baseline test score. The first two estimates assume that the difference in test scores between the treatment and control children that did not take the end-of-year assessment would have been the same as the difference for those children that did take the end-of-year assessment. The estimate in the last column of each table assumes that treatment-control difference of those children not assessed would have been the same as for those that were assessed, conditional on having the same baseline test score.

The first row of each table reports the effects for the overall test (all modules), while the rest of the rows report effects for the separate test modules. Beside each nodule in parentheses is the number of test questions and in the brackets, the baseline and end-of-year test scores for the control group. As seen, there are 26 questions in the Jump Start test. The test includes 6 modules, 10 questions on recognizing colors, 8 testing fine motor skills (for example, coloring a circle inside the lines), 2 on counting, 1 on sorting objects that are similar, 1 on recognizing one's printed name and 4 on book handling (for example, recognizing the title of a book).

The control group 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 name and most children were able to answer only one of four questions about book handling. The counting and sorting modules lie in between, a mean score of around 40 percent.

As should not be surprising, children in the control group learn without participating in the Jump Start 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 Jump Start program would need to account for this growth.

As seen in table 6, the three estimates of the impact of the Jump Start program are similar in magnitude. Concentrating on the last column, the Jump Start program is estimated to increase the overall test score by 7.0 percentage points (p-value of .001) more than the increase in the test score of the control group. Among the modules, the impact of the program is estimated to be 3.2 percentage points (p-value= .04) for Colors, 4.0 (p-value=.004) for Fine Motor Skills, 2.8 (p-value=.30) for Counting and 6.4 (p-value=.04) for Sorting.

Most striking is the impact on the score for name recognition, 22.2 percentage points (p-value=.001), and the impact on the score for book handling, 21.5 percentage points (p-value=.001). 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 Jump Start 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 gain to the Jump Start test; the control-group baseline Bracken score was 56.4 percent and the increase 21.1 percentage points. The gains on the individual modules aside from colors was around 10 percentage points.

As was the case for the Jump Start test, the three estimates of the impact of the Jump Start program on the Bracken are similar in magnitude. Concentrating again on the last column, the Jump Start program is estimated to increase the overall test score by 2.2 percentage points (p-value=.02) more than the increase in the test score of the control group. Among the modules, the impact of the program is estimated to be 4.2 percentage points (p-value= .02) for Colors, 3.2 percentage points (p-value=.04) for Letters, 2.5 percentage points (p-value=.12) for Numbers, 1.2 percentage points (p-value=.32) for Sizes and 1.2 percentage points for Shapes (p-value=.36). The last two rows of table 7 also show the impact of the JSP on the national percentile score and percent of children delayed or very delayed. As seen, the control group's percentile score was essentially unchanged, while the percentile score of Jump Start participants increased by 3.5 points (p-value=.01). On the other hand, the impact of the program on percentage delayed or very delayed dropped by 4.1 percentage points (p-value=.16) on top of a 5.5 percentage point drop in the control group percentage.

5. The Mechanism

The RCT demonstrates that the children of parents trained to teach them the skills targeted by the JSP learned more of those skills than did children of parents not trained in the program. Moreover, relative to children of parents not participating in the JSP, the children of Jump Start parents acquired a greater level of skills not specifically targeted by the program (as measured by the Bracken). However, the RCT does not tell us how these results were achieved. To do that requires a model and additional data.

5.1. Model

To begin, we denote $T = 1$ if a family is in the treatment group (participates in the JSP) and $T = 0$ if the family is in the control group. Further, let K denote end-of-year knowledge and K_0 beginning-of-year knowledge. Assume the family maximizes a utility function subject to a knowledge production function and a budget constraint. Family i's utility is given by

$$U(C_i, K_i) = C_i + \gamma K_i \tag{1}$$

where C_i is household consumption and γ is the marginal utility of the child's end-of-year knowledge.

The end-of-year knowledge production function is assumed to satisfy the assumptions required for the value-added form (see Todd and Wolpin, 2006):

$$K_i = \kappa_0 + (1+\delta)K_{0i} + \sum_{m=1}^{M} \beta_{im} X_{im}^{\theta_m}, 0 < \theta_m < 1, \beta_{im} \geq 0 \tag{2}$$

Where the vector $X_{i,m}$ comprise the full set of inputs in the production function of knowledge and the vector β_{im} are family-specific multipliers of the marginal product of the inputs. The assumption of additive separability of the inputs is clearly strong, but it is necessary due both to data limitations and to estimation tractability. The multipliers are allowed to depend on the treatment, namely:

$$\beta_{im} = \beta_{0im}(1 - T_i) + \beta_{1im} T_i, \qquad \text{for } m = 1, \ldots, M. \tag{3}$$

The multiplier of the m^{th} input marginal product is β_{0im} for a family that did not participate in JSP and β_{1im} for a family that participated in the JSP. As captured in the model, the role of the JSP is to teach families how to help their children learn, that is, to increase their marginal product multipliers, not to provide inputs directly. Given that the JSP curriculum targets a specific subset of skills (see the difference between the Bracken and the JSP test), β_{1im} could be equal to β_{0im} for some inputs. On the other hand, there could be spillover effects of the program beyond the targeted skills if parents are induced by the program to expand the set of skills they work on with their child.

The family faces a one-period budget constraint: $C_i = Y_i - \sum_{m=1}^{M} p_m X_{im}$, where p_m is the price of input m. Maximizing utility with respect to each input yields the input demand functions

$$X_{im} = \left(\frac{\gamma \theta_m}{p_m}\right)^{\frac{1}{1-\theta_m}} (\beta_{im})^{\frac{1}{1-\theta_m}}, \qquad \text{for } m = 1, \ldots, M. \tag{4}$$

Thus, if $\beta_{1im} > \beta_{0im}$, then families in the JSP will have a higher demand for input m than control group families. Note that the strong separability assumption in (2), the lack of complementarity or substitutability of inputs, implies that there are no cross-price effects in the demand for the inputs and that only the own-input marginal product multiplier affects input demand.

Substituting the input demand function (4) into the production function (2) provides the link between Jump Start participation and ending knowledge. Specifically,

$$K_i = \kappa_0 + (1+\delta)K_{0i} + \sum_{m=1}^{M} \eta_m \, (\beta_{im})^{\frac{1}{1-\theta_m}}. \tag{5}$$

where $\eta_m = \left(\frac{\gamma\theta_m}{p_m}\right)^{\frac{1}{1-\theta_m}}$.

We can use (5) to derive the average treatment effect (ATE) obtained from the RCT. Let $K_i(0)$ and $K_i(1)$ denote, respectively, the end-of-year knowledge without and with participation in the JSP. Note that due to randomization, $\mathrm{E}[K_i(0)] = \mathrm{E}(K_i|T_i = 0)$, $\mathrm{E}[K_i(1)] = \mathrm{E}(K_i|T_i = 1)$, and $\mathrm{E}(K_{i0}|T_i = 0) = \mathrm{E}(K_{i0}|T_i = 1)$. Thus, $\mathrm{ATE} \equiv \mathrm{E}[K_i(1)] - \mathrm{E}[K_i(0)]$ is:

$$\mathrm{ATE} = \sum_{m=1}^{M} \eta_m \left\{ \mathrm{E}\left[(\beta_{1im})^{\frac{1}{1-\theta_m}}\right] - \mathrm{E}\left[(\beta_{0im})^{\frac{1}{1-\theta_m}}\right] \right\}. \tag{6}$$

Equation (6) corresponds to the parameter that would be estimated from an OLS regression of ending knowledge on the treatment dummy. The parameter itself is a composite function of the utility and production function parameters (and input prices). Understanding the mechanism that drives the impact of the JSP requires identifying the underlying parameters governing input choices.

Below we provide identification and estimation results for this model. The analysis addresses the issue of whether the mechanism proposed in the model is empirically consistent with the evidence from the RCT, that is, whether the marginal product multiplier for specific input(s) is greater for the JSP families than for the control group families and whether Jump Start families choose more of the inputs.

The most direct method to answer the first question would be to estimate the production function (2). To place (2) into an estimation framework, rewrite it as

$$K_i = \kappa_0 + (1+\delta)K_{0i} + \sum_{m=1}^{M} \bar{\beta}_m X_{im}^{\theta_m} + \sum_{m=1}^{M} \omega_{im} \tag{7}$$

where, on substituting the input demand equations (4) into ω_{im}, we obtain:

$$\omega_{im} = (\beta_{im} - \bar{\beta}_m)\eta_{im}^{\theta_m} \beta_{im}^{\frac{\theta_m}{1-\theta_m}}$$

The problem in estimating (7) is that inputs are not orthogonal to ω_{im}. Families with higher value of a given β_{im}, and thus a higher value of ω_{im}, will choose more of the input, X_{im}. Note that the strong separability assumption of the inputs in the production function together with an assumption that the $\beta_{im}'s$

are independent implies that $\mathrm{E}(X_{im}|\omega_{im'}) = 0$ for $m' \neq m$. Thus, in this case, direct estimation of (7) will give an upward biased estimate of (the means of) the productivity multiplier distribution and/or the curvature parameters.

Although without the model we cannot address whether the JSP improved the marginal productivities of families, we can determine whether, as the model implies, greater levels of inputs are chosen by JSP families. At the time of the recruitment and again at the end of the program year, families who agreed to be in the study were administered a survey. The survey included questions concerning how much family members engaged in learning activities with their children. Table 8 shows the impact of the JSP on parental inputs: days/week read to the child, having at least 10 children's books, helping the child learn the alphabet, identify colors, learn numbers and learn shapes.

As seen in the table, as with test scores, parental inputs increased in the control group between the baseline and the end of JSP year. On average, parents in the control group increased the number of days someone in the family read to the child from 3.0 to 4.1 and the percentage of families having 10 or more children's books increased from 41 percent to 49 percent. In terms of parental help with particular skills, between 80 and 90 percent of the parents were already helping their children learn the alphabet, colors or numbers at baseline and by the end of the year the parental help increased by between 8 and 10 percentage points. Somewhat fewer parents were helping their child learn shapes at baseline, two-thirds, but that figure increased to 83.3 percent at the end of the year.

As also shown in table 8, at the end of the program, JSP families spent more days reading to the child than did the control group families and were more likely to have 10 or more children's books. Specifically, relative to control group families, JSP families read to their child on average one-half-day more and 22 percent more families had 10 or more books. The latter impact is almost surely because the JSP provided families with a dozen books as part of the curriculum; indeed 72 percent of the JSP families reported having 10 or more books. The JSP had essentially no impact on the other parental inputs that we measured. This result is not surprising given the high baseline percentages of families already providing these inputs to their children.

That the JSP increases the number of days spent reading to the child is a first step in the validation of the mechanism. The natural next step would be to estimate the model parameters, particularly to determine the size of the effect of the JSP on the productivity multiplier. Given the results in table 8, we consider only the single input, the number of days spent reading to the child. If there are no omitted inputs, or no variation in them, as was essentially the case for the inputs in table 8 (we discuss the role of children's books below), the production function would look like (7) with $m = 1$. If there are relevant omitted inputs, then, denoting the included input with $m = 1$, the production function (7) becomes

$$K_i = \kappa_0 + (1+\delta)K_{0i} + \bar{\beta}_1 X_{i1}^{\theta_1} + \zeta_i$$

where $\zeta_i = \sum_{m=2}^{M} \bar{\beta}_m X_{im}^{\theta_m} + \sum_{m=1}^{M} \omega_{im}$.

Noting that X_{i1} depends only on ω_{i1}, X_{i1} will be correlated with the omitted inputs, X_{im}, only if β_{i1} is correlated with β_{im}, $m = 2, \dots, M$ (which affect the choice of the other inputs). As seen below, identification requires that this correlation is zero.

As noted, JSP families were provided children's books as part of the curriculum. Children's books are not by themselves direct inputs because they are of value only when combined, at a minimum, with the child's time. The significance of having books in the home is that they provide an opportunity for activities in which parents and children are sharing attention (that is, joint attention). Having more books presumably affects the marginal product associated with the parental reading input. Thus, the impact of the JSP on β_{i1} incorporates the provision of books.

5.2. Identification

The identification analysis, based on model moments, makes use of the input demand function (including additive measurement error) and the reduced form of the production function, specifically:

$$X_{i1} = \left(\frac{\gamma\theta_1}{p_1}\right)^{\frac{1}{1-\theta_1}} (\beta_{i1})^{\frac{1}{1-\theta_1}} + \varepsilon_{i1} \tag{8}$$

$$K_i = \kappa_0 + (1+\delta)K_{0i} + \left(\frac{\gamma\theta_1}{p_1}\right)^{\frac{\theta_1}{1-\theta_1}} (\beta_{i1})^{\frac{1}{1-\theta_1}} + u_i \tag{9}$$

where ε_{i1} is measurement error, and $u_i = \sum_{m=2}^{M} \eta_m (\beta_{im})^{\frac{1}{1-\theta_m}}$. We assume β_{i1} is orthogonal to all β_{im}, $m = 1, \dots, M$. We also assume $\mathrm{E}(u_i|T_i = 1) = \mathrm{E}(u_i|T_i = 0)$.[4]

To proceed, taking a Taylor expansion of $(\beta_{i1})^{\frac{1}{1-\theta_1}}$ around the mean of the β_i and dropping the input subscript $(m = 1)$ for convenience, we get:

$$(\beta_i)^{\frac{1}{1-\theta}} = \bar{\beta}^{\frac{1}{1-\theta}} + \frac{1}{1-\theta}\bar{\beta}^{\frac{\theta}{1-\theta}}(\beta_i - \bar{\beta}) + \frac{1}{2}\frac{\theta}{(1-\theta)^2}\bar{\beta}^{\frac{2\theta-1}{1-\theta}}(\beta_i - \bar{\beta})^2 + O^3 \tag{10}$$

[4] This assumption, as seen below, is necessary for identification when there are omitted inputs. If $E(u_i|T_i) \neq 0$, its value is incorporated into κ_0.

From (10), the (unconditional) expectation up to the second-order term yields[5]

$$\mathrm{E}\left[(\beta_i)^{\frac{1}{1-\theta}}\right] = \bar{\beta}^{\frac{1}{1-\theta}} + \frac{1}{2}\frac{\theta}{(1-\theta)^2}\bar{\beta}^{\frac{2\theta-1}{1-\theta}}\sigma_\beta^2 \equiv F\left(\bar{\beta}, \sigma_\beta^2, \theta, T_i\right) \tag{11}$$

We consider what can be identified from the first moments, the mean of the input, X, and the output, K. Recalling that the distribution of β_i depends on the treatment assignment (see equation (3)), we denote $F_0 = F\left(\bar{\beta}, \sigma_\beta^2, \theta, T_i = 0\right)$ as the expression of (11) for the control group and $F_1 = F\left(\bar{\beta}, \sigma_\beta^2, \theta, T_i = 1\right)$ as the expression for the treatment group. Then,

$$\mathrm{E}(X_i|T_i) = \left(\frac{\gamma\theta}{p}\right)^{\frac{1}{1-\theta}}[F_0(1-T_i) + F_1 T_i] \tag{12}$$

$$\mathrm{E}(K_i|T_i) = \kappa_0 + (1+\delta)K_{0i} + \left(\frac{\gamma\theta}{p}\right)^{\frac{\theta}{1-\theta}}[F_0(1-T_i) + F_1 T_i] \tag{13}$$

Given that assignment to treatment is random, it can be used as an instrumental variable for the input. The elements of the Wald estimator are:

$$\frac{\Delta\mathrm{E}(X_i|T_i)}{\Delta T_i} = \left(\frac{\gamma\theta}{p}\right)^{\frac{1}{1-\theta}}[F_1 - F_0] \text{ and } \frac{\Delta\mathrm{E}(K_i|T_i)}{\Delta T_i} = \left(\frac{\gamma\theta}{p}\right)^{\frac{\theta}{1-\theta}}[F_1 - F_0] \tag{14}$$

The Wald (IV) estimator is thus

$$\frac{\frac{\Delta\mathrm{E}(K_i|T_i)}{\Delta T_i}}{\frac{\Delta\mathrm{E}(X_i|T_i)}{\Delta T_i}} = \left(\frac{\gamma\theta}{p}\right)^{-1} \tag{15}$$

Interestingly, the Wald estimator is not a function of the parameters of the marginal productivity multiplier distribution.[6]

It turns out that θ cannot be identified without additional assumptions. The intuition for this result is that the instrument is binary, $T_i \in \{0,1\}$, and θ governs the curvature of the production function with respect to the input. Therefore, the exogenous variation in the binary instrument is not sufficient to recover the

[5] The expectation is unconditional because the distribution of β_i is allowed to depend on K_{0i}. The unconditional expectation is an iterated expectation over the conditional distribution of β_i and the distribution of K_{0i}.

[6] The usual LATE-estimator, which would be a function of the β_i distribution, is given by $\frac{\Delta E(K_i)}{\Delta T_i} \div \frac{\Delta E(X_i^{\theta})}{\Delta T_i}$. If, for example, $\theta = .5$ and we assume no measurement error, this estimator becomes $\left(\frac{\gamma}{2p}\right)^{-1}\frac{(F_1 - F_0)}{\bar{\beta}_1 - \bar{\beta}_0}$. This composite parameter has no particular interpretation.

curvature of the production function of knowledge. In the estimation we will have to choose a value for the curvature parameter and check the sensitivity of the estimates to alternative values. We thus can identify γ/p (given θ). In addition, from (12) we identify F_0 and F_1.

The identification of the rest of the model is based on second-moment conditions. To calculate the second moments, we need to take a second order expansion of $(\beta_i)^{\frac{2}{1-\theta}}$. Taking the (unconditional) expectation of that expansion gives

$$\mathrm{E}\left[(\beta_i)^{\frac{2}{1-\theta}}\right] = \bar{\beta}^{\frac{2}{1-\theta}} + \frac{1+\theta}{(1-\theta)^2}\bar{\beta}^{\frac{2\theta}{1-\theta}}\sigma_\beta^2 \equiv G(\bar{\beta},\sigma_\beta^2,\theta,T_i) \tag{16}$$

Given this approximation, the second order moments (recognizing that each moment differs for the treatment and control group) are:

$$\mathrm{Var}(X_i|T_i) = \left(\frac{\gamma\theta}{p}\right)^{\frac{\theta}{1-\theta}}\left\{G(\bar{\beta},\sigma_\beta^2,\theta,T_i) - \left[F(\bar{\beta},\sigma_\beta^2,\theta,T_i)\right]^2\right\} + \sigma_\varepsilon^2 \tag{17}$$

$$\begin{aligned}\mathrm{Var}(K_i|T_i) = (1+\delta)^2\sigma_{K_0}^2 + \left(\frac{\gamma\theta}{p}\right)^{\frac{2\theta}{1-\theta}}\left\{G(\bar{\beta},\sigma_\beta^2,\theta,T_i) - \left[F(\bar{\beta},\sigma_\beta^2,\theta,T_i)\right]^2\right\} \\ + 2(1+\delta)\left(\frac{\gamma\theta}{p}\right)^{\frac{\theta}{1-\theta}}\mathrm{Cov}\left(K_{0i},\beta_i^{\frac{1}{1-\theta}}\right) + \sigma_u^2\end{aligned} \tag{18}$$

$$\mathrm{Cov}(X_i,K_{0i}|T_i) = \left(\frac{\gamma\theta}{p}\right)^{\frac{1}{1-\theta}}\mathrm{Cov}\left(K_{0i},\beta_i^{\frac{1}{1-\theta}}\right) \tag{19}$$

$$\begin{aligned}\mathrm{Cov}(X_i,K_i|T_i) = (1+\delta)\left(\frac{\gamma\theta}{p}\right)^{\frac{1}{1-\theta}}\mathrm{Cov}\left(K_{0i},\beta_i^{\frac{1}{1-\theta}}\right) \\ + \left(\frac{\gamma\theta}{p}\right)^{\frac{1+\theta}{1-\theta}}\left\{G(\bar{\beta},\sigma_\beta^2,\theta,T_i) - \left[F(\bar{\beta},\sigma_\beta^2,\theta,T_i)\right]^2\right\}\end{aligned} \tag{20}$$

$$\mathrm{Cov}(K_i,K_i|T_i) = (1+\delta)\sigma_{K_0}^2 + \left(\frac{\gamma\theta}{p}\right)^{\frac{\theta}{1-\theta}}\mathrm{Cov}\left(K_{0i},\beta_i^{\frac{1}{1-\theta}}\right) \tag{21}$$

Given that γ/p (given θ) can be identified from the Wald estimator, $\mathrm{Cov}\left(K_{0i},\beta_i^{\frac{1}{1-\theta}}\right)$ is identified from (19), one for each treatment group. A non-zero value of this covariance could arise because "initial" knowledge directly affects the marginal productivity multiplier and/or because the marginal productivity multiplier that produced K_{0i} is correlated with the multiplier that produced K_i. Given these results, δ is identified from (21) and $\left\{G(\bar{\beta},\sigma_\beta^2,\theta,T_i) - \left[F(\bar{\beta},\sigma_\beta^2,\theta,T_i)\right]^2\right\}$ from (20), which implies that σ_ε^2 is identified from (17) and σ_u^2 from (18). Having previously identified $F(\bar{\beta}, \sigma_\beta^2, \theta, T_i)$ from the first moment conditions, then $G(\bar{\beta}, \sigma_\beta^2, \theta, T_i)$ is also identified. Then, for each F_T and G_T pair, there are two equations and two unknowns, $\bar{\beta}_T$ and $\sigma_{\beta_T}^2$. The solution to these equations is given by:

$$\bar{\beta}_T = \left[\frac{\theta G_T + (2+\theta)}{2F_T(1+\theta)}\right]^{1-\theta} \tag{22}$$

$$\sigma_{\beta_T}^2 = \frac{[\theta G_T + (2+\theta) - 2F_T G_T(1+\theta)][2F_T(1+\theta)]^{2\theta-1}}{[\theta G_T + (2+\theta)]^{2\theta}} \tag{23}$$

Recall that we have assumed that $\beta_i \geq 0$; given that the bracketed expression in (22) is always positive. We restrict the solution for $\bar{\beta}_T$ to be positive when there is both a positive and negative root (for example, when $\theta = 0.5$). The bracketed expression in (23) may not be positive. Parameter values that lead to a negative value are clearly not valid.

5.3. Estimation

Measuring Child Knowledge K and K_0

As discussed, children were administered two separate assessments, the curriculum-based JSP test and the Bracken. As illustrated in tables 6 and 7, the two assessments together cover 10 distinct skill domains. Presumably, parents, the actors in the model, care about each of those domains, though not necessarily equally. Identification of the model requires a single scalar measure of knowledge. Two issues arise: (i) how to measure performance within any domain and (ii) how to combine domains.

In the previous discussion, performance was measured by the child's raw score on the JSP and Bracken assessments, that is, the percent of questions answered correctly. That choice was made because the question that the Alief ISD wanted to answer was whether the children in the program had mastered the JSP curriculum (as taught to them by their parents). The content of the JSP assessment followed the program curriculum exactly. For example, the children were supposed to have acquired the ability to recognize ten colors. The assessment asked the children to identify exactly those ten colors. The same was true in each of the six domains (see table 6) covered in the curriculum, one question for each skill within the domain. The Bracken raw score was used as the performance measure to be comparable to the JSP measure.

The raw score is only one of several possible performance measures. Others include the percentile score for the tested population or, in the case of the Bracken, also the percentile score based on a nationally representative sample, a standardized score (e.g., mean zero, unit standard deviation), or an age-equivalence score, that is, the mean age of those with the score obtained by a given child. Although there is no strong reason to prefer one of these measures over the others, we chose the age-equivalence measure in the estimation of the model because it would be easily interpretable by parents and a measure they could form preferences over (your child is age a, but knows as much as children who are age $a + b$) and because it is a measure relative to a reasonable notion of their reference group, that is, children from other families that also applied to be in the JSP. It is also a measure that has cardinal scale that is invariant to any monotonic transformation of the scores.[7,8]

In combining performance over domains, we use a weighted average of the age-equivalence scores of the separate domains. We assume that parents weight more heavily (as a matter of preference) performance in the more difficult domains. The weights we choose are the control group baseline average proportion of wrong answers in each domain (see tables 6 and 7).[9] Those weights are used both for the baseline and end-of-year measure of knowledge.

The impact of the JSP is to increase this weighted performance score by .41 age-equivalent years, or by 4.9 months.[10,11] The equivalent treatment effect based on a weighted average of raw scores is 7.6 percentage points; the baseline weighted test score for the control group is 30.8 and the weighted end-of-year test score is 47.1, an increase of 16.3 percentage points.

5.4. Results

Estimation is by Generalized Empirical Likelihood (GEL).[12] There are eleven model parameters, five of which are assumed to be the same for each treatment group, κ_0, δ, γ/p, σ_ε^2, σ_u^2, and three of which differ by treatment group, $\bar{\beta}_T$, $\sigma_{\beta_T}^2$, and $\mathrm{Cov}\left(K_{0i}, \beta_i^{\frac{1}{1-\theta}}\right)$. There are 14 data moments matched to the theoretical moments given by equations (12), (13), and (17)-(21).

[7] The age-equivalence score is calculated by inverting a regression of the raw score for each domain on chronological age using the control group sample. Separate regressions were estimated for the baseline and end-of-year JSP and Bracken scores. The mean age equivalence for the control group at the baseline and at the end-of-year is thus the average chronological age of the control group sample at each time.
[8] Ideally, one would have access to adult outcomes data to anchor both test scores as in Cunha, Heckman, and Schennach (2010). This is not feasible as the participants for whom we observe test scores are still very young children.
[9] The weights are normalized to sum to one.
[10] The regression also includes the baseline weighted age equivalent score. The p-value for the treatment effect is less than .001.
[11] The treatment effect based on the unweighted (simple average) of the age-equivalent scores is .36 years, or 4.3 months.
[12] The estimation was carried out using the GEL routine in the R-package, 'gmm'.

The results of the estimation are shown in table 9 for values of θ between .3 and .7. Because θ is not identified, the overall fit of the model is identical for all values of θ. Neither the Lagrange multiplier test nor J-test reject the over-identifying restrictions. As implied by these tests, the difference between the actual and predicted fitted moments is small (and essentially the same for all θ); the percentage difference relative to the data moments is smallest for the means, less than .5 percent, largest for the covariances, ranging from less than one percent to 7 percent with a single outlier being 18 percent, and in-between for the variances, less than 2.5 percent. In addition, because θ is not identified, only parameters that govern the choice of the input change with θ, and the treatment effect, which is given by $\left(\frac{\gamma\theta}{p}\right)^{-1}$, is invariant to θ. The table also shows test statistics for the equality of $\bar{\beta}_T$, $\sigma^2_{\beta_T}$, and Cov $\left(K_{0i}, \beta_i^{\frac{1}{1-\theta}}\right)$ across the treatment groups; equality is rejected in every case for the mean and variance, but not rejected for the covariance in any case.

As implied by the model moments, the parameters governing the input choice vary systematically with θ. As θ increases, the preference parameter decreases (proportionately), the mean of the productivity multiplier, $\bar{\beta}_T$, increases, its variance, $\sigma^2_{\beta_T}$, decreases and the covariance of beginning knowledge and the productivity multiplier decreases. These differences are not important for validating the mechanism, but would be important in performing counterfactuals. For example, suppose there was a program that provided parents with a monetary incentive based on their child's performance. Such a program would be isomorphic to increasing the parent's preference for K, that is, increasing γ. The impact of such a program would depend on θ and could be calculated, for this population, from (13) if θ were known.

As seen in table 9, for all values of θ, both the mean and standard deviation of the marginal productivity multiplier are larger for the treatment group. In addition, the covariance between the baseline score and the multiplier is positive for both groups and larger for the treatment group, though the covariances are not statistically significant for all values of θ.

How important, in this population, is reading to one's child in affecting the acquisition of the skills in the domains covered by the JSP and Bracken assessments? It is possible to get a picture of the overall importance of the reading input in determining ending knowledge. Based on the estimated model, the proportion of the variance in ending knowledge accounted for by the reading input, that is by $\beta_i X_i^\theta$, is 42 percent for the control group and 45 percent for the treatment group.[13] In addition, given that about 10 percent is accounted for by variation in beginning knowledge, under the model assumptions less than half of the variation in ending knowledge is due to omitted inputs (and measurement error due to random test

[13] In this calculation, we ascribe the contribution of the covariance between beginning knowledge and the productivity multiplier (see equation (18)), to the reading input.

score variation). This result is more striking given that the model estimates imply that about 88 percent of the variance in the measured reading input is due to measurement error.[14] Thus, the actual standard deviation in the input is only .67 days per week in the control group and .75 in the treatment group (as opposed to 1.89 and 1.91 days for the input measure).

The JSP increased the average value of $\beta_i X_i^{\theta}$ by changing the distribution of β_i, both its mean and variance, and as a response by parents to the increase in their productivity, by changing the level of the input chosen by the JSP families. As already noted, this response was substantial; parents increased the number of days per week they read to their child on average by a little over one-half a day. We also found, based on the estimated treatment effect, that the change in the productivities of families substantially increased their children's knowledge as measured by the increase in their age equivalency (4.9 months). The model, which was not rejected by the data, provides a consistent interpretation of the mechanism behind the success of the JSP.

What conclusions might a researcher have drawn about the productivity of the parental input without the model? Standard practice would have been to estimate some version of the value-added production function under the assumption that the productivity multiplier is the same across families. It is important to recognize that the additional assumptions of the model about the orthogonality between included and omitted inputs are also necessary to obtain consistent estimates of the production function estimated outside of the model.

Table 10 provides six alternative estimates of the production function. The first three assume that the input, days per week reading to the child, has a constant marginal product. The second three allow for curvature as in the model, where we have set $\theta = 0.5$ for the purpose of this analysis. Within each set, the first specification assumes that the marginal productivity multiplier did not change with the treatment, the second allows for an effect of the JSP on the multiplier, and the third makes the same assumption as the first, but uses the treatment variable as an instrument for the input as a means of correcting for measurement error in the input.[15]

The input coefficient in each of these specifications would normally be interpreted as an estimate of the productivity multiplier, β. The first result to note from table 10 is that in both sets, there is strong evidence that the input coefficient is larger for the families in the JSP than for families in the control group. The second result to note is that the IV estimates are about an order of magnitude larger than the OLS estimates

---

[14] The input is measured as the average over the baseline and end-of-year number of days/week the parent read to the child.

[15] Given the assumption of homogeneity in the multiplier there is no other reason to use the treatment as an instrument.

(first specification in each set), which is consistent with the estimate of the measurement error variance under the model. [16]

One might be tempted to conclude from table 10 that the most credible estimate of the productivity multiplier (for the control group) comes from the IV-regression. However, in the context of the choice model, this conclusion would be incorrect. As shown, the IV-regression for the linear-X specification does not estimate the productivity multiplier, but rather the inverse of the product of the parental knowledge preference parameter (γ) and the production function curvature parameter (θ). Also, the IV-regression in the linear-sqrt X specification is a biased estimate of the LATE parameter shown in footnote 3.[17] The LATE parameter, as noted, is uninformative about any specific parameter of the model.

The specific point from this exercise is that allowing for heterogeneity in the productivity multiplier alters the interpretation of the IV-estimator in a non-trivial way. The more general point is that the interpretation of the mechanism governing the results of the RCT requires a choice model.

6. Conclusion:

This paper reports the results of a study that evaluated the Jump Start Program designed by the Alief ISD in Houston. We found that the program produced modest impacts on the acquisition of the skills targeted by the program curriculum, about .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. The program increased parental investments as measured by the frequency parents read to their children.

To go beyond simply measuring treatment effects, we built and estimated a model of parent-child interactions that featured rich individual-specific heterogeneity in the marginal productivity of investments. We developed a constructive identification analysis and implemented an estimation algorithm based on it. The model estimates implied that the program impact on children’s knowledge is due to an increase in the marginal productivity of parental time spent reading to their child, which increases parental efficiency and, in response, induces an increase in the input itself, that is, in the frequency with which parents choose to read to the child.

The JSP trained parents to engage more effectively with their children on a very limited set of skills. Because these skills are perceived to be foundational for Pre-K success by AISD teachers, it is possible that the children whose parents participated in the program will have medium- and long-term benefits. It is also

[16] The measurement error attenuation factor estimated from the model is .12.
[17] The reason for the bias is that the measurement error in the model is linear in X, not in the square root of X.

possible that these foundational skills are easily learned in the first weeks of the Pre-K program and any short-term advantage conferred by participation in the program will be short-lived. To answer these questions, future work will analyze medium- and long-term impacts of the program on performance as the children in the Jump Start study progress through school (in AISD or other school districts in Texas).

References

Attanasio, Orazio , Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2015. Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia. In *Working Papers*, edited by NBER.

Baker, Amy J.L., Chaya S. Piotrkowski, and Jeanne Brooks-Gunn. 1998. "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* 13 (4):571-588.

Baker, Amy J.L., Chaya S. Piotrkowski, and Jeanne Brooks-Gunn. 1999. "The Home Instruction Program for Preschool Youngsters (HIPPY)." *Home Visiting: Recent Program Evaluations* 9 (1):116-133.

Bassok, Daphna, Jenna E. Finch, RayHyuck Lee, Sean F. Reardon, and Jane Waldfogel. 2016. "Socioeconomic Gaps in Early Childhood Experiences: 1998 to 2010." *AERA Open* 2 (3):1-22. doi: 10.1177/2332858416653924.

Bond, Timothy N., and Kevin Lang. 2013. "The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results." *Review of Economics and Statistics* 95 (5):1468-1479.

Cunha, Flavio, Marsha Gerdes, and Snejana Nihtianova. 2019. Language Environment and Maternal Expectations: An Evaluation of the LENA Start Program. edited by Rice University.

Cunha, Flavio, and James J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2):31-47.

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

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward Vytlacil. 2008. "Identification of Treatment Effects Using Control Functions in Models with Continuous Endogenous Treatment and Heterogeneous Effects." *Econometrica* 72 (1):1191-1206.

Heckman, James J. , Margaret L. Holland, Kevin K. Makino, Rodrigo Pinto, and Maria Rosales-Rueda. 2017. An Analysis of the Memphis Nurse-Family Partnership Program. In *Working Papers*, edited by NBER.

Kalil, Ariel 2014. "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*, edited by Paul R. Amato, Alan Booth, Susan M. McHale and Jennifer Van Hook, 63-82. New York: Springer.

Kalil, Ariel, Rebecca Ryan, and Michael Corey. 2012. "Diverging destinies: maternal education and the developmental gradient in time with children." *Demography* 49 (4):1361-1383. doi: 10.1007/s13524-012-0129-5.

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, Allison Sidle Fuligni, Peter Z. Schochet, Diane Paulsell, and Cheri Vogel. 2005. "The Effectiveness of Early Head Start for 3-Year-Old Children and Their Parents: Lessons for Policy and Programs." *Developmental Psychology* 41 (6):885-901.

Putnam, Robert D. 2016. *Our Kids: The American Dream in Crisis*. New York: Simon and Schuster.

Reardon, Sean F., and Ximena A. Portilla. 2016. "Recent Trends in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry." *AERA Open* 2 (3):1-18. doi: 10.1177/2332858416657343.

St. Pierre, Robert, Anne Ricciuti, Fumiyo Tao, Cindy Creps, Janet Swartz, Wang Lee, Amanda Parsad, and Tracy Rimdzius. 2003. Third National Even Start Evaluation: Program Impacts and Implications for Improvement. Department of Education, Washington, DC.

Suskind, Dana L. , Kristin R. Leffel, Eileen Graf, Marc W. Hernandez, Elizabeth A. Gunderson, Shannon G. Sapolich, Elizabeth Suskind, Lindsey Leininger, Susan Goldin-Meadow, and Susan C. Levine. 2015. "A parent-directed language intervention for children of low socioeconomic status: a randomized controlled pilot study." *Journal of Child Language* 43 (2):366-406.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113 (485):F3-F33.

Todd, Petra E., and Kenneth I. Wolpin. 2207. "Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps." *Journal of Human Capital* 1 (1):91-113.