Reducing the Academic Achievement Gap through a Summer Pre-Kindergarten Program

Roland Fryer[1], Steven Levitt[2], John List[2] and Anya Samek[3]

[1]Harvard University and NBER
[2]University of Chicago and NBER
[3]University of California, San Diego and NBER

February 2021

Abstract

We present the results of an early childhood intervention in which mostly disadvantaged 3-4-year-old children were randomized to a typical preschool program (9 months), a novel shortened summer program (2 months) or a control group. Both programs incorporated a parenting component and focused on improving academic and executive functioning skills. Both programs significantly improved academic test scores by one quarter of a standard deviation relative to the control group, but neither had a significant impact on executive functioning. Despite its significantly lower cost, the shortened program was equally as effective as the typical program. This suggests a potentially cost-effective way to reduce the academic achievement gap and improve the Kindergarten-readiness skills of disadvantaged children. We are now collecting data on standardized test scores to understand the long-term impacts of these programs.

JEL codes: J24, C93
Keywords: preschool education, human capital, field experiments

Corresponding author: Anya Samek, University of California, San Diego and NBER (anyasamek@gmail.com). We thank the Kenneth and Anne Griffin Foundation for funding this research. Samek was funded under National Institutes of Health (NIH) grant 5R01DK114238 during the write-up of this paper. For excellent research assistance, we thank Edie Dobrez, Kristin Troutman, Eric Andersen, Reuben Bauer, Amanda Chuan, Claire Ha, Alannah Hoefler, Justin Holz, Shreemayi Samujjwala, Andrew Simon, Graham Tierney, Mattie Toma, Jeannine van Reeken, and Mina Zhang.

1. Introduction

Early life conditions have profound effects later in life. Intervening in early childhood may thus be an important way to reduce economic disparities between high- and low-income children and reduce the academic achievement gap. Despite a growing literature on the early determinants of human capital development, little is known about the most cost-effective ways to intervene (Currie and Almond, 2011). Since the majority of children in the U.S. and other high-income nations spend a large part of their time in care outside the home (Blau and Currie, 2006), one point of intervention is to target children directly through high-quality preschool programs. Parental inputs also play a major role in intergenerational inequality (Becker and Tomes, 1979); therefore, a second and potentially very important point of intervention is to target parents through parenting interventions (see, e.g., List, Samek, and Suskind, 2018).

Research has also called for interventions that focus on both academic and executive functioning skills. Academic skills – as measured by achievement in reading, writing, and math – are important for Kindergarten readiness and future success in school. Executive functioning skills – which involve motivation, perseverance, and self-control – are also thought to affect all aspects of academic achievement and future labor market outcomes (Heckman, Stixrud, and Urzua, 2006; Cunha and Heckman, 2008). Yet, the majority of studies investigating the impact of preschool have focused only on academic skills. And the existing literature on executive functioning skills mostly highlights the correlation between these skills and life outcomes, but there is little evidence that executive functioning can be affected through interventions (Fryer, Levitt, and List, 2015 is one counterexample).

Our contribution in our study is the development and evaluation of a preschool program aimed at targeting both academic and executive functioning skills. The program that we developed

– called *Cog-X* – focuses on math, literacy, self-regulation and socio-emotional learning, integrating reading, math, social studies, science and socio-emotional activities throughout the day. Importantly, the program puts a strong emphasis on parental involvement through bi-weekly "Parent Empowerment" workshops. During these workshops, parents learn from their child's own teacher about how to scaffold the child's learning at home. A novel aspect of our study is that it evaluates *Cog-X* delivered either as a standard, academic-year 9-month preschool or as a shortened, summer 2-month version (which we call Kinderprep). The shortened preschool has not been studied previously but represents a potentially cost-effective way to improve the Kindergarten readiness of disadvantaged children.

We evaluated *Cog-X* as part of a larger set of interventions through the Chicago Heights Early Childhood Center (CHECC) field experiment. We randomized nearly 600 3-4-year-old children in Chicago Heights, Illinois and the surrounding area to an academic-year (9 month) *Cog-X* program (which we call Preschool), a shortened summer (2 months) version of the *Cog-X* program (which we call Kinderprep), or to a control group that did not receive any intervention from us. The program was implemented in a school district with a high poverty rate. This was also a high-minority school district.

The programs significantly improved academic test scores by about one quarter of a standard deviation relative to the control group at the end of the year. Most of the gains in the academic-year program were concentrated among children who started out with below median test scores. In addition, the gains largely occurred within the first 4 months, and the shortened version of the program was equally as effective as the academic-year program. We also observed small and insignificant gains in executive function skills. The difficulty in moving executive function skills highlights an important methodological point: if executive function skills are difficult to

move, then discussions of the importance of executive function skills on life outcomes may not be as actionable for improving outcomes.

The remainder of our paper is structured as follows. Section 2 provides a background. Section 3 describes the program and data collected. Section 4 summarizes the results. Section 5 discusses several robustness tests of the main results. Section 6 provides a discussion and concludes.

2. Background

There is a well-documented gap in academic achievement between low-income, minority students and their higher-income counterparts. List, Pernadaut, and Suskind (2020) provide several cuts of data that lend insights into the underpinnings for such outcome disparities. Their analysis shows that the cycle starts with parental beliefs in that there are key differences in parental beliefs across SES families. For example, low-SES parents largely believe their children;s trajectory is immutable, whereas high-SES parents believe their investments matter for their children's outcomes. List, Pernadaut, and Suskind (2020) show that the belief disparities map onto parental inputs and ultimately child outcomes. This reasoning suggests that interventions in the early years involving both parents and children may be the key to long-term success. For instance, Kindergarten readiness is an important predictor of long-term academic success (Duncan et al., 2007). Our work joins an important academic and policy debate about the best ways to intervene at this critical age using both parental and child interventions.

Our basic framework assumes the following education production function: $A=f(S,P,C)$, where A represents academic achievement, S is schooling inputs, P is parental inputs, and C is a variable representing the student's ability and initial level of learning prior to entry into the

preschool. The *Cog-X* program directly affects *A* through increases in schooling inputs *S* and parental inputs *P* through its focus on preschool and parent education. Education production is a dynamic process such that early investments lead to increased productivity of later investments through increases in child ability *C.* Therefore, early interventions are key for generating improved educational outcomes. As noted by Currie and Almond (2011) as well as Heckman (2000), characteristics measured before age 5 have a pronounced impact on long-term outcomes. However, the shape of the education production function (i.e., when and for how long within the preschool years to invest) is less clear.

Our paper joins several landmark RCTs in early childhood education. The Perry/HighScope Preschool Project[1] and the Carolina Abecedarian Project[2] in the 1960s-70s each randomized ~100 children to either preschool with home visitation or a control group. These studies found that children randomized to preschool significantly outperformed children in the control group on school achievement tests and had significantly higher levels of educational attainment (Schweinhart et al., 2005; Heckman et al., 2010; Campbell et al., 2002). These studies were small and samples were relatively homogeneous (e.g., Perry and Abecedarian included 99 students, 98% of whom were black students from very disadvantaged backgrounds). By contrast, our preschool program recruited an ethnically diverse population from Chicago Heights, Illinois and surrounding areas.

Most early education programs have been evaluated based on their short-term impact on academic test scores.[3] Yet researchers argue that much of the gains of early childhood education

[1] https://highscope.org/perry-preschool-project/

2 https://abc.fpg.unc.edu

[3] One exception is evaluations of *Tools of the Mind*, a curriculum aimed at improving executive functioning skills (Bodrova and Leong, 2006). *Tools of the Mind* was found to reduce problem behaviors relative to an existing curriculum in one small study (Barnett et al., 2008), yet another study across several school districts found no impact of *Tools of the Mind* relative to an existing curriculum on executive functions – gains were similar in both groups

appear in the form of improved executive function skills later in life, which lead to improved related outcomes such as lower incarceration rates and better health (Heckman, 2000). One key contribution of our work is to investigate whether we can evaluate short-term improvements in executive function skills using tests that we administer. Such an analysis would allow us to directly link improvements in these skills to the potential of future, long-term improvements in related outcomes.

This project is also related to literature on improving parental investment in children. Fryer et al. (2015) used a different sample of CHECC children (from earlier waves of the CHECC program) to investigate parenting programs with incentives.[4] The authors found large and statistically significant positive impacts on academic and executive function test scores of Hispanics and whites, but no impact on Blacks. Additional interventions that have included incentivizing parents include Opportunity NYC (Riccio et al., 2013; Fryer, 2011) and Fryer, Devi, and Holden's field experiment in Houston public schools (Fryer, Devi, and Holden, 2012).

Intervening on one dimension of the education production function (e.g., schooling, *S*) may not have a straightforward effect on academic achievement *A* due to potential complementarities between *S* and *P.* If *S* and *P* are strategic substitutes, for example, it is possible that increasing *S* causes parents to decrease *P.* Therefore, programs that focus only on schooling inputs may cause a reduction in parental investment, limiting the overall impact on child academic achievement. Of course, the opposite may also be true – parents could increase their effort in response to an increase in *S.* To reduce the probability that parents respond to our preschool intervention by reducing their

(Wilson and Farran, 2012). We are also in the process of evaluating the impact of *Tools of the Mind* with a separate sample of children from our programs.

[4] The programs reported on in Fryer, Levitt, and List (2015) come from the same CHECC program but use a different sample of children who attended different CHECC interventions in prior years.

own effort, the *Cog-X* program intervenes on both *S* and *P* simultaneously. However, we cannot credibly identify the separate impact of increasing *S* versus *P* due to this design choice.

Finally, our research is related to a literature on student learning throughout the year. This research finds that students, especially students from disadvantaged backgrounds, tend to lose knowledge and skills over the summer break (Downey, Von Hippel, and Broh, 2004; Cooper et al., 1996).[5] We were motivated by this research to design the Kinderprep program to occur during the critical summer months where "summer loss" might commonly take place. We are also the first to consider how learning proceeds throughout an academic year by considering the impact of our program during both the mid-year (January) evaluation period and the end-year (May) evaluation period.

3. Experimental Design and Implementation

3.1 Procedures

Our experiments were conducted at the Chicago Heights Early Childhood Center (CHECC) in 2012-2014. We founded CHECC in 2010 in Chicago Heights, Illinois in partnership with Chicago Heights School District 170 (thereafter, SD170). Chicago Heights, IL is a low-income suburb of Chicago, IL with a per capita income of just over $19,000 and a 29% poverty rate.[6] The Chicago Heights, IL population is 42% Black, 34% Hispanic, and 38% Caucasian. The high school graduation rate in the area is 47%. This economic and demographic make-up is similar to urban

[5] There are few rigorous studies on variation in school calendars, and many programs have been evaluated using non-experimental methods (e.g., Saam and Nowak, 2005; Larson, 2003; Gullo, 2000; Koopmans, 1991). Of the few studies of school year duration that use experimental methods, many have not been able to utilize a comparable control group (for a review of this literature, see Cooper et al., 2003; Aronson, Zimmerman, and Carlos, 1998; Patall et al., 2010).

[6] Data from 2010 Census, http://www.census.gov/quickfacts/table/PST045215/1714026

inner-city schools.[7]

The implementation of CHECC included securing space for the programs and hiring staff – we hired a preschool teacher and an assistant for each classroom for a total of 20 instructional staff members.[8] To support recruiting efforts, we ran a local marketing campaign each year, which included direct mailings, automated phone calls to families with children enrolled in the district, and information booths at community events in and around the district. We also distributed program information through district leadership staff in SD170 and surrounding districts and encouraged administrative assistants at schools to collect and submit registration forms to us.

All children ages 3-4.99 (as measured on September 1 of any year[9]) were eligible to sign up for our program, regardless of socio-economic status. Children were eligible for the program if they lived within SD170 or outside of it. The only children who were excluded were those who had been identified by the district or by our staff as requiring special education, since we did not have the capability to serve these students in our program.

Parents registered their child(ren) for the program by completing a registration form, signing a consent form, and participating in a pre-assessment in the summer prior to the start of each academic year. The pre-assessment consisted of an hour-long battery of academic and executive function tests that each child completed one-on-one with an experimenter. While they waited, parents completed a detailed survey about their child and household.

---

[7] For a discussion of urban city school demographic make-up, see https://nces.ed.gov/pubs2007/minoritytrends/ind_2_7.asp

[8] All preschool teachers had at least a Bachelor's degree and 04 EC teaching certification. Teaching assistants had a paraprofessional certification. In Year 4, instructional staff decreased to 10 since the Preschool program was only administered in one building.

[9] All children were required to start Kindergarten if they were at least 5 years old on September 1 of any year.

3.2 Experimental Design

Households were randomized to one of two treatments or to a control group, as summarized in Table 1. In the Preschool program, children were offered a half-day (morning) of instruction with an optional free afternoon daycare. This amounts to 486 total hours of instruction during the academic year. The Preschool used the *Cog-X* curriculum, which we developed with a focus on math, literacy, self-regulation, and socio-emotional learning. *Cog-X* integrates reading, math, social studies, science, and socio-emotional activities throughout the day. Appendix A provides more detail on the curriculum.

[Table 1 Here]

Cog-X also puts a strong emphasis on parental involvement, primarily through bi-weekly Parent Empowerment workshops. During Parent Empowerment, parents attend a group parent class to learn from their own child's teacher how to scaffold the child's learning through activities in the home. Parents whose children were randomized to the Preschool program received 21 parent contact hours during the group classes.

In the Kinderprep program, we used the same *Cog-X* curriculum, but children attended a half-day (morning) instruction with no option for free daycare, which constitutes 115 hours of instruction. Parents of children randomized to Kinderprep received 6-7 parent contact hours during the group classes.

Parents in Preschool and parents in Kinderprep received $50 for attendance at each session ($25 if they arrived late). Payment was made via debit cards that could be utilized at any merchant where credit cards are accepted. Households randomized to the control group did not receive any educational interventions from us, but did receive incentives for participating in assessments, as described below.

The randomization was conducted at two points in time each year. During the summer prior to each year of our study, we randomized entrants to either Preschool, Kinderprep, or the control group (we call this the "main" randomization). Since Kinderprep occurs nearly one year after the initial randomization, we also conducted a second randomization in the spring (we call this the "late" randomization). We allowed eligible households who missed the initial deadline to continue registering for the program until the following spring, and then randomized entrants to either Kinderprep or the control group.

Our main randomization followed a blocked approach. In each randomization, we created matched groupings of children based on gender, race (white, Hispanic, or Black), and age (within ½ year intervals). Then, we randomly assigned each child in the grouping to a treatment or control group. This means that some children who registered were not placed into the randomization (because a matched grouping could not be created for them). We constrain our analysis to the matched groupings, but provide robustness checks using the full sample in Appendix B.

Our main outcome measures are a battery of assessments administered at the beginning (June/July), middle (January/February), and end of each academic year (May/June), as well as at the end of the summer (August). Students who attended a school within SD170 in later years also completed the assessments at the end of Kindergarten, 1st grade, and 2nd grade (April-May each year). Each assessment was administered by a team of assessors who all held at least a Bachelor's degree, received training specific to the assessment, and were blind to the child's treatment. Children were assessed by either being brought to the school by their parents (who received an incentive between $25 and $100 for attendance, depending on the assessment), or by being pulled out of class if they were attending the school during that time. The assessment took about 1 hour.

The assessments include academic and executive function components and are summarized

in Table 2. For the academic component, we used a series of nationally normed tests focused on reading, writing, and mathematics and that can be administered to children as young as age 2, including the Peabody Picture Vocabulary Test (PPVT) (Dunn et al., 1965) and the Woodcock-Johnson III Test of Achievement (Woodcock et al., 2003). For the executive function component, we used a combination of tests of executive functions, including working memory, inhibitory control, and attention shifting, developed by Blair and Willoughby (Blair and Willoughby, 2006a, 2006b, 2006c) and a questionnaire completed by the assessor focused on attention and emotion (Smith-Donald et al., 2007). For our main analysis, we create a academic index made up of averaged percentile scores on each academic construct and a executive function index made up of averaged percent correct scores (or points earned in the case of PPVT) on each executive function construct (equal weights) which are then converted to standard deviation units for our sample (see Appendix A).

[Table 2 Here]

Our assessment data is complemented with Kindergarten readiness tests administered by SD170 to all students in the district at the beginning of the Kindergarten year. These tests are a useful robustness check since they involve all students in the district; hence, unlike our assessments which sometimes rely on parental involvement, we do not have a selection problem. They are also useful because they are administered independently by the district and test skills that the district considers important for Kindergarten readiness, including math, literacy, and social-emotional learning. However, these tests only include the sub-sample of children attending SD170 since we did not have similar data-sharing agreements with other districts.

A difficulty with experiments such as ours is keeping the control group engaged with assessments and continuing in the study. We used a number of innovative strategies to encourage

control group participation. First, we referred to the control group as the "Family Group" rather than the "control group" to increase the appeal of being in this group. Second, this group received 3 family parties for the family each year, including free food and small gifts for the children.[10] Differential attrition between treatment and control groups is discussed further in Section 5.

4. Results

4.1 Summary of Sample

621 children participated in the randomization and were matched across both years of the experiment. Table 3 provides information about the number of children randomized to the program, by treatment group.[11]

[Table 3 Here]

The average age of children entering the study was 3.3 years old (s.d.=1.4). Children were 50% female, 44% Hispanic, 48% Black, and 9% white. The average combined academic ability score was the 34.9 percentile (s.d.=23.6 percentiles), showing that many of these children were performing below average for their age.

Tables 4 and 5 provide balance checks for each analysis sample, including for each sample that data is available, where the pre-assessment sample is equivalent to the randomization sample (Appendix Tables B1 and B2 do the same for follow-up tests). There are two tables since the correct comparison group for Kinderprep differs from that of Preschool. Kinderprep children are restricted to the children who are at least 4 years old and will start Kindergarten in the following

[10] At the fall party, we hosted outdoor games and provided families with free ice cream. At the winter party, children could get a photo with Santa Claus and received a small gift. At the spring party, we organized an Easter egg hunt and the chance to get photographed with the Easter Bunny. None of the parties included an educational component.

[11] The full sample reported in Appendix B is 791 children, see Table B3.

year, while Preschool children could be either 3 or 4 years old to join the program. We report the averages of individual deviations from the mean for each matched pair, and conduct t-tests to compare these residuals. We find that we are balanced on the observables that we collected.

[TABLES 4, 5 Here]

As displayed in Table 6, 75% of the children assigned to Preschool attended at least one day, and 66% of children assigned to Kinderprep attended at least one day. Conditional on attending, children attended over 84% of instructional days. We were careful to exclude all children assigned to the control group from participation in the treatments, and none attended Preschool or Kinderprep. Our main analysis uses intent-to-treat (ITT) estimates, including all children who were initially randomized to the programs, regardless of whether they eventually attended them. Interested readers can convert our estimates to treatment-on-treated (ToT) by dividing the relevant coefficient estimate by the attendance probability.

[TABLE 6 Here]

4.2 Main Treatment Effects

Table 7 provides ordinary least squares (OLS) estimates for intent-to-treat effects, regressing treatment assignment on academic and executive function scores, with controls for pre-assessment academic and executive function scores and demographic characteristics (race, gender, age at test date, and home language), year of baseline tests, matched pair grouping, test form, number of previous assessments, age dummy at randomization (3 or 4 years old, which determines year of school entry), and year of randomization).[12] We do this analysis separately for the mid-

[12] Notice that one can also use Table 7 to estimate the local-average-treatment effect (LATE); that is, the average effect of attending one of our programs (Angrist and Imbens, 1995). To estimate LATE, each treatment effect coefficient should be divided by the attendance probability for that sub-group available in Table 6. A few assumptions are needed to estimate LATE. These are, 1) treatment assignment is random, 2) being selected for treatment has a monotonic impact on attending, and 3) being selected for treatment affects outcomes through its effect on attendance

year assessment (January/February), the end of year assessment (May/June), and the end of summer assessment (August). The comparison group for Preschool is the sub-set of control children in the main randomization, while the comparison group for Kinderprep is the sub-set of control children who would have been eligible for the summer program (i.e., were at least 4 years old on September 1 of the academic year following the program), including both main randomized and late randomized children with a dummy to control for randomization period.

[TABLE 7 – Mean Effect Sizes]

We find a large and significant impact of our programs on academic achievement. The impact of Preschool on academic achievement is large and statistically significant – students assigned to this program have a academic score that is 0.23σ higher than the control group at the end of the year (p-value<0.01). The impact of the Kinderprep on academic achievement is also large and statistically significant -- students assigned to this program have a academic score that is 0.18σ higher than the control group at the end of the summer (p-value<0.01).[13]

Next, we investigate the shape of the education production function for the academic-year program by comparing academic test scores obtained at mid-year to those obtained at the end of the academic year. We see that most gains from Preschool occur within the first 4 months of the program (p-value comparing mid-year and end-year scores = 0.78). This finding provides additional support for the potential effectiveness of shorter programs like Kinderprep. At the end of summer, scores in the Preschool group are somewhat below end of academic year scores, but the difference is not statistically significant (p-value=0.65).

The impact of our programs on executive function skills is more limited and not statistically

(Angrist and Imbens, 1995).

[13] The robustness checks in Appendix B Tables B9-B11 confirm the large and significant impact on academic skills. Evidence of the impact on executive function skills is weak just as in the main data.

significant. The estimate for the impact on executive function skills of being assigned to Preschool at the end of the program is 0.08σ (p-value=0.39), and of being assigned to Kinderprep at the end of the program is 0.09σ (p-value=0.32).

We also investigate which skills were most improved by the programs. In Table 8, we split the academic and executive function scores into their sub-tests. We find that the program significantly improved most academic skills (relative to the control group), including letter and word identification (WJ Letter-Word), applied problems (WJ Applied Problems), and math (WJ Quantitative Concepts). Receptive vocabulary and spelling were less affected.[14] As for executive function skills, we find evidence that the Preschool program affected working memory (Operation Span) but neither treatment affected inhibitory control (Spatial Conflict) or attention shifting (Same Game). There was also no impact on our assessor-administered self-regulation score – impulse control and emotion – as measured by the Preschool Self-Regulation Assessment (PSRA).

[TABLE 8 ABOUT HERE: MEAN EFFECT SIZES SUBSCORES]

Our results suggest that executive function skills are much harder to move through preschool education than academic skills. An alternative explanation is that our tests do not measure this construct very well. However, Fryer et al. (2015) used similar tests in their analysis and found a strong positive impact of incentivized parenting programs on executive function skills, suggesting that the tests do measure a skill that is measurable and can be moved. If parenting is an important input to executive function skills, we may have expected to find effects as well. However, our parenting programs are much less intensive than the parenting programs discussed in Fryer, Levitt, and List (2015), which may have contributed to our limited effects.

[14] The same result is observed in the robustness checks, see Appendix B, Table B5.

4.3 Heterogeneous Treatment Effects

An open question in the academic production function is how child skills *C* affect the production process. Particularly, we were interested to know whether there were heterogeneous treatment effects when comparing higher or lower academic and executive function ability children. Tables 9 and 10 report on the impact of the Preschool and Kinderprep, grouping students by pre-treatment academic and executive function skills.

In Table 9, we observe that the greatest improvement in academic skills is concentrated among students who start out with below median academic test scores. Moreover, it does not matter whether the students also show low or high executive function test scores. These children improve by 0.50σ (p-value<0.01) relative to children who were above median who do not improve (-0.01σ, p-value=0.96). Among Kinderprep children in Table 10, we also find large improvements on academic scores from children who start below the median on academic test scores (0.27σ, p-value<0.02) and insignificant improvement among kids who start above the median on academic skills (0.12σ, p-value=0.19).

The same result is observed in Appendix B Tables B6 and B7, which shows that in the overall sample, children with the lowest academic test scores improve the most. Preschool children improve by 0.44σ (p-value<0.01) while children above the median in academic test scores improve by 0.08σ (p-value=0.40). The corresponding values for Kinderprep are 0.27σ (p-value=0.01) and 0.14σ (p-value=0.12), which are lower in magnitude but similar in relative size.

[TABLES 9 & 10 - Mean Effects Score Subgroups about here]

5. Robustness Checks

In this section, we explore the robustness of these results under potential threats to the interpretation of the data.

5.1 Attrition

A first concern is that our estimates might be biased based on the sample of students who complete the battery of tests described above at each point in the school year, either during treatment or while enrolled in SD170 during follow-up years. If treatment affects selection into this sample, our results may be biased. Table 9, 10, and 11 account for this selective attrition by estimating inverse probability weighted (IPW) treatment effects. IPWs are estimated from a linear probability model of a binary variable indicating whether the academic/non- academic index score is available for the post-treatment wave regressed on the baseline measure, a dummy variable indicating whether the baseline measure is present or not, gender, race, and age at assessment. These are calculated separately for Preschool and Kinderprep treatment arms. The results remain largely robust upon weighing the coefficients by IPWs across the overall mean effect (Table 9) and within Preschool and Kinderprep pre-score subgroups (Table 10 and Table 11).

5.2 Small Sample Estimation

Given our relatively small sample size, it is possible that the data violates the assumption that the estimated treatment effects are normally distributed, which could affect inference. To account for this, we conduct a non-parametric permutation test and calculate exact p-values (Fisher 1960).

The sample is re-randomized 10,000 times within matched blocks (i.e. identical to the

original randomization). We re-calculate the main results using these new, synthetic treatment assignments and record the new treatment effects. The exact p-value is the proportion of simulated treatment effects that are larger than the actual observed treatment effect (in absolute value).

Figure 1 plots the actual observed treatment effect against the distribution of simulated treatment effects for various outcomes. All results that are statistically significant in the main analysis remain significant.

5.3 Multiple Hypothesis Testing

We test two different treatments along two (sometimes more) outcome measures; therefore, it is necessary to adjust for multiple hypothesis testing (MHT) (List, Shaikh, and Xu, 2019). In Table 7, we have two treatments and two outcome variables, which gives us a Bonferroni adjusted p-value threshold of 0.05/4=0.0125. Our results on the impact of programs on academic scores continue to hold for Preschool at the end of the year and for Kinderprep at the end of the summer. We conclude that our main results are robust to MHT.

In Tables 9 and 10, we consider whether the programs are more effective for children who start with higher/lower abilities (4 subgroups and 2 outcome variables in each table, for a Bonferroni adjusted threshold of 0.05/8=0.00625). Our finding that Preschool children starting at below median scores significantly improve (Table 9) continues to hold up to MHT. The result does not hold up for Kinderprep (Table 10), potentially due to the smaller sample size. Results in Tables 8 and 10 are less robust to MHT.[15]

[15] In Table 8, we consider the impact of the programs on different sub-scores. With 8 sub-tests, our Bonferroni adjusted p-value threshold is 0.05/16=0.0038. Results that still hold for Kinderprep are the significant impact on Letter-Word and Quantitative Concepts. Results that still hold for Preschool are Spelling at the end of the summer and Quantitative Concepts at the end of the year. In Table 10, if we control for testing in each year as a separate hypothesis, results do not hold; but if we control for 4 hypotheses as in Table 7, results continue to hold.

6. Conclusion

Providing children with the best opportunities to reach their potentials remains one of the most important educational problems today. In this study, we find that preschool has a large and significant impact on the academic skills of children enrolled. Importantly, the shorter and cheaper Kinderprep program also has similar effects. This provides a clear policy implication, since the children we studied are low-income and started out with lower academic achievement scores relative to the nation. We also explored the impact of our program on executive function skills, finding that executive function skills are much more challenging to change. Finally, and perhaps most importantly, we traced out the shape of the education production function, finding that most gains occur in the first 4 months (with large gains in even 2 months, as evidenced by the Kinderprep program). Our finding about *when* gains occur is highly relevant for policy-makers, since we show that a cheaper, shorter summer program can rival an academic-year program in the short-term.

The finding that executive function skills are difficult to move through schooling alone is important given that a large literature is devoted to studying the correlates between executive function skills and life outcomes. The related work of Fryer, Levitt, and List (2015), which did show an impact of parenting programs on executive function skills, points to the possibility that inputs outside of schooling (i.e., parents) are highly important for this skill.

Our programs were most effective for those children who needed it most: those below the median in academic skills improved substantially, while those above the median did not significantly improve. In future work, we will continue follow-up with these children to understand the progression of treatment effects as the children age into middle childhood, adolescence, and beyond.

A limitation of this work is that we have thus far evaluated the impact of the programs in the short-term. The next step is to collect additional data on the CHECC children going forward so that the effects in the medium term can be evaluated. We are now working on this and have obtained agreement from the State of Illinois to share data, and have also received data from schools. Many related programs document substantial fade-out in the years following the program (Almond et al., 2018). Despite this fade-out, programs such as Perry Preschool and Abecedarian have shown substantial lasting impact into adulthood (Schweinhart et al., 2005). Hence, understanding the trajectories of CHECC children will be important going forward.

References

Almond, Douglas, Janet Currie, and Valentina Duque. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56, no. 4 (2018): 1360-1446. doi: 10.1257/jel.20171164.

Aronson, J., J. Zimmerman, and L. Carlos. "Improving student achievement by extending school time: Is it just a matter of time." West Education, San Francisco, CA Available from (http://www. wested. org/online_pubs/po-98-02. pdf) (1998)

Angrist, Joshua D., and Guido W. Imbens. Identification and estimation of local average treatment effects. No. t0118. National Bureau of Economic Research, 1995.

Barnett, W. Steven, Kwanghee Jung, Donald J. Yarosz, Jessica Thomas, Amy Hornbeck, Robert Stechuk, and Susan Burns. "Educational effects of the Tools of the Mind curriculum: A randomized trial." Early childhood research quarterly 23, no. 3 (2008): 299-313.

Bodrova, Elena, and Deborah J. Leong. *Tools of the mind*. Pearson Australia Pty Limited, 2006.

Currie, Janet, and Douglas Almond. "Human capital development before age five." In *Handbook of labor economics*, vol. 4, pp. 1315-1486. Elsevier, 2011.

Becker, Gary S., and Nigel Tomes. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of political Economy* 87, no. 6 (1979): 1153-1189.

Blair, C., and M. T. Willoughby. "Measuring executive function in young children: Operation Span." *Chapel Hill, NC: Pennsylvania State University and University of North Carolina* (2006a).

Blair, C. B., and M. T. Willoughby. "Measuring executive function in young children: Spatial Conflict II: Arrows." *Chapel Hill, NC: The Pennsylvania State University and the University of North Carolina at Chapel Hill* (2006b).

Blair, C.B. and Willoughby, M.T. (2006c). *Measuring Executive Function in Young Children: Item Selection*. Chapel Hill, NC: The Pennsylvania State University and The University of North Carolina at Chapel Hill (2006c)

Blau, David, and Janet Currie. "Pre-school, day care, and after-school care: who's minding the kids?." *Handbook of the Economics of Education* 2 (2006): 1163-1278.

Campbell, Frances A., Craig T. Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson. "Early childhood education: Young adult outcomes from the Abecedarian Project." *Applied Developmental Science* 6, no. 1 (2002): 42-57.

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse. "The effects of summer vacation on achievement test scores: A narrative and meta-analytic review." *Review of educational research* 66, no. 3 (1996): 227-268.

Cooper, Harris, Jeffrey C. Valentine, Kelly Charlton, and April Melson. "The effects of modified school calendars on student achievement and on school and community attitudes." *Review of Educational Research* 73, no. 1 (2003): 1-52.

Cunha, Flavio, and James J. Heckman. "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation." *Journal of human resources* 43, no. 4 (2008): 738-782.

Downey, Douglas B., Paul T. Von Hippel, and Beckett A. Broh. "Are schools the great equalizer? Cognitive inequality during the summer months and the school year." *American Sociological Review* 69, no. 5 (2004): 613-635.

Duncan, Greg J., Chantelle J. Dowsett, Amy Claessens, Katherine Magnuson, Aletha C. Huston, Pamela Klebanov, Linda S. Pagani et al. "School readiness and later achievement." *Developmental psychology* 43, no. 6 (2007): 1428.

Dunn, Lloyd M., Leota M. Dunn, Stephan Bulheller, and Hartmut Häcker. *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service, 1965.

Fisher, Ronald Aylmer. "The design of experiments." *The design of experiments.* 7th Ed (1960).

Fryer Jr, Roland G., Steven D. Levitt, and John A. List. *Parental incentives and early childhood achievement: a field experiment in Chicago heights*. No. w21477. National Bureau of Economic Research, 2015.

Fryer Jr, Roland G., Tanaya Devi, and Richard T. Holden. *Vertical versus horizontal incentives in education: Evidence from randomized trials*. No. w17752. National Bureau of Economic Research, 2012.

Fryer Jr, Roland G. "Financial incentives and student achievement: Evidence from randomized trials." *The Quarterly Journal of Economics* 126, no. 4 (2011): 1755-1798.

Gullo, Dominic F. "The long term educational effects of half-day vs full-day kindergarten." Early Child Development and Care 160, no. 1 (2000): 17-24.

Heckman, James J. "Policies to foster human capital." *Research in economics* 54, no. 1 (2000): 3-56.

Heckman, James J., Jora Stixrud, and Sergio Urzua. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor economics* 24, no. 3 (2006): 411-482.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. "The rate of return to the HighScope Perry Preschool Program." *Journal of public Economics* 94, no. 1 (2010): 114-128.

Koopmans, Matthijs. "A Study of the Longitudinal Effects of All-Day Kindergarten Attendance on Achievement." (1991)

Larson, Bruce E. "Comparing face-to-face discussion and electronic discussion: A case study from high school social studies." Theory & Research in Social Education 31, no. 3 (2003): 347-365.

Lee, David S. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76, no. 3 (2009): 1071-1102.

List, John A., Julie Pernadaut, and Dana Suskind. "It All Starts with Beliefs: Addressing the Roots of Educational Inequities by Changing Parental Beliefs," working paper, University of Chicago, 2020

List, John A., Anya Samek, and Dana L. Suskind. "Combining behavioral economics and field experiments to reimagine early childhood education." *Behavioural Public Policy* 2, no. 1 (2018): 1-21.

List, John A., Azeem M. Shaikh, and Yang Xu. "Multiple hypothesis testing in experimental economics." *Experimental Economics* 22, no. 4 (2019): 773-793.

Patall, Erika A., Harris Cooper, and Ashley Batts Allen. "Extending the school day or school year a systematic review of research (1985–2009)." *Review of educational research* 80, no. 3 (2010): 401-436.

Riccio, J., Dechausay, N., Miller, C., Nunez, S., Verma, N., & Yang, E. 2013. Conditional Cash Transfers in New York City: The Continuing Story of the Opportunity NYC-Family Rewards Demonstration. *MDRC*.

Riccio, James, Nadine Dechausay, Cynthia Miller, Stephen Nuñez, Nandita Verma, and Edith Yang. "Conditional Cash Transfers in New York City: The Continuing Story of the

Opportunity NYC-Family Rewards Demonstration." *MDRC* (2013).

Saam, Julie, and Jeffrey A. Nowak. "The effects of full-day versus half-day kindergarten on the achievement of students with low/moderate income status." Journal of Research in Childhood Education 20, no. 1 (2005): 27-35.

Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, William S. Barnett, Clive R. Belfield, and Milagros Nores. "Lifetime effects: the High/Scope Perry Preschool study through age 40." (2005).

Smith-Donald, Radiah, et al. "Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research."*Early Childhood Research Quarterly* 22.2 (2007): 173-187.

Wilson, Sandra Jo, and Dale C. Farran. "Experimental Evaluation of the Tools of the Mind Preschool Curriculum." *Society for Research on Educational Effectiveness* (2012).

Woodcock, Richard W., Kevin S. McGrew, Nancy Mather, and Fredrick A. Schrank. "Woodcock-Johnson III diagnostic supplement to the tests of cognitive abilities." *Itasca, IL: Riverside* 10 (2003): 003435520104400407.

Table 1: Experimental Design

	Child Component	**Parent Empowerment Component**
Preschool (9 months)	Free, morning preschool, and afternoon daycare 486 hours of instruction	2x month Parent Empowerment classes 21 parent contact hours $50 for attendance
Kinderprep (2 months)	Free, morning preschool 115 hours of instruction	2x month Parent Empowerment classes 6-7 parent contact hours $50 for attendance
Control	None	None

Table 2: Assessments

	Construct Measured	**Name of Test**
Academic	Receptive vocabulary	Peabody Picture Vocabulary Test
	Letter/word recognition	Letter-Word subtest of Woodcock-Johnson III Test of Achievement
	Writing/spelling	Spelling subtest of Woodcock-Johnson III Test of Achievement
	Math/problem solving	Applied Problems subtest of Woodcock-Johnson III Test of Achievement
	Math/problem solving	Quantitative Concepts subtest of Woodcock-Johnson III Test of Achievement
Executive Function	Working memory	Operation Span test of Blair/Willoughby Measures of Executive Function
	Inhibitory control	Spatial Conflict test of Blair/Willoughby Measures of Executive Function
	Attention shifting	Same Game test of Blair/Willoughby Measures of Executive Function
	Attention	Attention factor constructed from Preschool Self-Regulation Assessment (questionnaire completed by assessor)
	Emotion	Emotion factor constructed from Preschool Self-Regulation Assessment (questionnaire completed by assessor)

Table 3: Randomization

	Main Randomization	Late Randomization	*Total*
Preschool	237	N/A	237
Kinderprep	130	39	169
Control	175	40	215
Total	542	79	621

Most children were randomized in the main randomization each summer. However, a small group of children were randomized in the following spring either to control or Kinderprep. The sample includes every child who appears in Table 7: Mean Effect Sizes.

Table 4: Summary Statistics: Preschool-Plus

	Pre Sample		Mid Sample		Post Sample		Summerloss Sample	
	Preschool	p-value	Preschool	p-value	Preschool	p-value	Preschool	p-value
Academic Baseline	-0.517	0.797	-4.313	0.090	-2.288	0.301	-0.494	0.848
	(2.006)		(2.537)		(2.207)		(2.580)	
Missing Academic Baseline	-0.006	0.611	-0.020	0.171	-0.004	0.766	-0.007	0.703
	(0.013)		(0.015)		(0.014)		(0.018)	
Executive Function Baseline	-0.916	0.571	-4.816	0.043	-2.209	0.222	-0.890	0.652
	(1.616)		(2.368)		(1.805)		(1.969)	
Missing Executive Function Baseline	0.033	0.242	0.039	0.240	0.044	0.135	0.051	0.151
	(0.028)		(0.033)		(0.029)		(0.036)	
Age Baseline	-0.683	0.117	0.103	0.906	-0.244	0.754	-0.103	0.918
	(0.435)		(0.876)		(0.778)		(0.998)	
Male	-0.006	0.898	-0.034	0.601	-0.013	0.812	-0.033	0.606
	(0.050)		(0.065)		(0.056)		(0.065)	
White	-0.001	0.981	0.028	0.424	0.007	0.834	-0.006	0.869
	(0.029)		(0.035)		(0.031)		(0.038)	
Black	-0.010	0.846	0.079	0.221	0.036	0.517	0.025	0.692
	(0.050)		(0.065)		(0.055)		(0.064)	
Hispanic	-0.000	0.999	-0.107	0.099	-0.051	0.359	-0.020	0.753
	(0.049)		(0.064)		(0.055)		(0.064)	
Home Language - English	-0.031	0.501	0.064	0.303	0.031	0.553	-0.001	0.991
	(0.046)		(0.062)		(0.053)		(0.062)	
Home Language - Spanish	0.058	0.156	-0.050	0.382	0.015	0.750	-0.003	0.965
	(0.041)		(0.057)		(0.047)		(0.057)	
Home Language - Spanish and English	-0.020	0.489	-0.014	0.709	-0.038	0.258	0.016	0.646
	(0.029)		(0.037)		(0.033)		(0.035)	
Treatment Observations	249		208		222		167	
Control Observations	225		106		157		127	

This table reports summary statistics and balance tests. Each panel is restricted to children who took the corresponding assessment. Each row reports the coefficient from a regression of the listed variable on treatment and matched pair controls. Late randomized children are treated as a single matched pair. The p-value columns report the p-value from the treatment coefficient. The observation rows provides the number of children in each panel who were used in any of the above regressions. The p-value for a ttest on the residuals from a regression of the executive function index on a second degree age polynomial and matched pair controls is 0.693.

Table 5: Summary Statistics: Kinderprep

	Pre Sample		Mid Sample		Post Sample		Summerloss Sample	
	Kinderprep	p-value	Kinderprep	p-value	Kinderprep	p-value	Kinderprep	p-value
Academic Baseline	-1.244	0.631	-2.375	0.480	-4.293	0.168	-3.265	0.313
	(2.589)		(3.358)		(3.102)		(3.227)	
Missing Academic Baseline	-0.016	0.172	-0.017	0.406	-0.007	0.565	-0.008	0.557
	(0.011)		(0.020)		(0.012)		(0.014)	
Executive Function Baseline	-0.882	0.643	-3.547	0.147	-2.493	0.251	-3.244	0.169
	(1.898)		(2.434)		(2.165)		(2.349)	
Missing Executive Function Baseline	-0.016	0.545	0.058	0.151	0.009	0.733	0.033	0.272
	(0.027)		(0.040)		(0.027)		(0.030)	
Age Baseline	0.104	0.841	0.615	0.562	0.561	0.394	0.103	0.892
	(0.518)		(1.059)		(0.657)		(0.757)	
Male	0.001	0.989	-0.098	0.209	-0.022	0.707	-0.017	0.784
	(0.057)		(0.078)		(0.059)		(0.064)	
White	-0.000	0.989	0.041	0.330	0.029	0.369	0.029	0.408
	(0.034)		(0.042)		(0.032)		(0.035)	
Black	-0.025	0.656	0.032	0.671	-0.017	0.776	0.008	0.899
	(0.057)		(0.076)		(0.059)		(0.063)	
Hispanic	0.012	0.831	-0.087	0.259	-0.024	0.688	-0.052	0.421
	(0.057)		(0.077)		(0.059)		(0.064)	
Home Language - English	-0.049	0.348	0.095	0.194	0.039	0.485	0.038	0.535
	(0.052)		(0.073)		(0.056)		(0.061)	
Home Language - Spanish	0.050	0.247	-0.093	0.153	-0.043	0.375	-0.039	0.473
	(0.043)		(0.064)		(0.048)		(0.054)	
Home Language - Spanish and English	-0.001	0.967	-0.012	0.805	0.003	0.937	0.003	0.932
	(0.036)		(0.050)		(0.036)		(0.040)	
Treatment Observations	193		98		145		135	
Control Observations	197		75		145		116	

This table reports summary statistics and balance tests. Each panel is restricted to children who took the corresponding assessment. Each row reports the coefficient from a regression of the listed variable on treatment and matched pair controls. Late randomized children are treated as a single matched pair. The p-value columns report the p-value from the treatment coefficient. The observation rows provides the number of children in each panel who were used in any of the above regressions. The p-value for a ttest on the residuals from a regression of the executive function index on a second degree age polynomial and matched pair controls is 0.548.

Table 6: Attendance

	Proportion of kids who ever attended	**Proportion of class attended, conditional on ever attending**	**Proportion of class attended, unconditional**
Preschool	75%	86%	65%
Kinderprep	66%	80%	53%
Total	71%	84%	60%

The sample includes every child who appears in Table 7: Mean Effect Sizes.

Table 7: Mean Effect Sizes

	ITT		
	Mid Year	End Year	End Summer
A. Standardized Academic Score			
Preschool:	0.214**	0.234***	0.204**
	(0.088)	(0.078)	(0.089)
Kinderprep:			0.184***
			(0.070)
Preschool v. Kinderprep:			0.831
B. Standardized Executive Function Score			
Preschool:	0.018	0.078	0.087
	(0.101)	(0.091)	(0.097)
Kinderprep:			0.087
			(0.087)
Preschool v. Kinderprep:			0.997

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores. The academic and executive function indexes are calculated as the mean of the standardized values of each of the subtests. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. The row Preschool v. Kinderprep reports the p-value of a chi quared test of the equality of coefficients reported at summerloss. Asterisks indicate significance at the 10/5/1 percent levels.

Table 8: Mean Effect Sizes Subscores

	ITT		
	Mid Year	End Year	End Summer
A. Picture Vocabulary			
Preschool-Plus:	-1.543	2.075	2.685
	(2.763)	(2.500)	(2.729)
Kinderprep:			-1.266
			(2.662)
B. WJ Letter-Word			
Preschool-Plus:	4.324	7.275***	7.278**
	(2.983)	(2.579)	(2.921)
Kinderprep:			6.881***
			(2.231)
C. WJ Spelling			
Preschool-Plus:	6.050*	4.810*	8.709***
	(3.312)	(2.823)	(3.110)
Kinderprep:			4.857**
			(2.289)
D. WJ Applied Problems			
Preschool-Plus:	5.097**	4.843**	0.317
	(2.536)	(2.248)	(2.498)
Kinderprep:			0.782
			(2.194)
E. WJ Quantitative Concepts			
Preschool-Plus:	7.211***	6.470***	1.703
	(2.693)	(2.301)	(2.759)
Kinderprep:			7.897***
			(2.535)
F. Operation Span			
Preschool-Plus:	0.145	0.209**	0.319***
	(0.123)	(0.094)	(0.116)
Kinderprep:			0.129
			(0.107)
G. Spatial Conflict			
Preschool-Plus:	0.023	-0.004	0.041
	(0.114)	(0.092)	(0.099)
Kinderprep:			0.055
			(0.096)
H. Same Game			
Preschool-Plus:	-0.002	-0.026	-0.104
	(0.121)	(0.101)	(0.114)
Kinderprep:			0.094
			(0.102)
H. Preschool Self Regulatory Assessment			
Preschool-Plus:	0.006	-0.073	-0.105
	(0.102)	(0.111)	(0.115)
Kinderprep:			-0.083
			(0.096)

This table reports ITT coefficient estimates of treatment effects on test scores. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Asterisks indicate significance at the 10/5/1 percent levels.

Table 9: Mean Effect Sizes on Academic and Executive Function Indices within Prek Pre-Score Subgroups

	Academic		Executive Function	
	Coefficient	N	Coefficient	N
Above Median Acad Baseline	-0.006	194	-0.087	194
	(0.103)		(0.122)	
Below Median Acad Baseline	0.496***	185	0.261**	183
	(0.108)		(0.128)	
Above Median Exec Baseline	0.113	174	0.053	174
	(0.111)		(0.130)	
Below Median Exec Baseline	0.332***	205	0.095	203
	(0.106)		(0.125)	
Above Median Acad Above Median Exec Baseline	-0.023	115	0.042	115
	(0.134)		(0.157)	
Above Median Acad Below Median Exec Baseline	0.017	79	-0.290	79
	(0.164)		(0.193)	
Below Median Acad Above Median Exec Baseline	0.392**	59	0.081	59
	(0.184)		(0.216)	
Below Median Acad Below Median Exec Baseline	0.533***	126	0.349**	124
	(0.133)		(0.157)	

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores. The academic and executive function indexes are calculated as the mean of the age-standardized values of each of the subtests. The covariates' coefficients are fixed in rows 1 and 2, 3 and 4, and 5 through 8 by regressing the outcome on treatment interacted with an indicator for if the child is above or below the median score along with all other controls. The observation column reports the number of kids in the regression in the given baseline group. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Kinderprep does not include controls for matched pairs because of small sample sizes. Asterisks indicate significance at the 10/5/1 percent levels. The median standardized academic score is -0.159 and the median standardized executive function score is 0.047.

Table 10: Mean Effect Sizes on Academic and Executive Function Indices within Kinderprep Pre-Score Subgroups

	Academic		Executive Function	
	Coefficient	N	Coefficient	N
Above Median Acad Baseline	0.121	143	0.181	141
	(0.093)		(0.116)	
Below Median Acad Baseline	0.272**	107	-0.040	107
	(0.106)		(0.132)	
Above Median Exec Baseline	0.183*	129	0.064	130
	(0.096)		(0.119)	
Below Median Exec Baseline	0.187*	121	0.107	118
	(0.104)		(0.130)	
Above Median Acad Above Median Exec Baseline	0.067	75	0.068	75
	(0.128)		(0.157)	
Above Median Acad Below Median Exec Baseline	0.177	68	0.294*	66
	(0.142)		(0.177)	
Below Median Acad Above Median Exec Baseline	0.327**	54	0.019	55
	(0.151)		(0.186)	
Below Median Acad Below Median Exec Baseline	0.215	53	-0.165	52
	(0.155)		(0.192)	

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores. The academic and executive function indexes are calculated as the mean of the age-standardized values of each of the subtests. The covariates' coefficients are fixed in rows 1 and 2, 3 and 4, and 5 through 8 by regressing the outcome on treatment interacted with an indicator for if the child is above or below the median score along with all other controls. The observation column reports the number of kids in the regression in the given baseline group. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Kinderprep does not include controls for matched pairs because of small sample sizes. Asterisks indicate significance at the 10/5/1 percent levels. The median standardized academic score is -0.179 and the median standardized executive function score is 0.317.

Appendix A: Description of Programs & Assessments

A.1: Cog-X

Goals of Cog-X
Cog-X not only prepares preschool students for kindergarten—it aims to provide students the skills necessary to succeed in all grades. By the time students enter kindergarten, they should:

- Display academic curiosity through oral and written language.
- Exhibit self-regulation, as well as empathy toward others.
- Have a foundation of academic skills, including letter and number recognition, the ability to respond to texts, and basic mathematical concepts.

Early childhood professionals developed the Cog-X curriculum in alignment with the Illinois State Board of Education Kindergarten Standards and the Common Core Kindergarten State Standards.

Inclusion of content to be taught with intentionality and integration
Cog-*X* integrates reading, math, social studies, science, and socio-emotional activities throughout the day, and is divided into four major thematic units that motivate classroom work: My Family, My Community, and My World; The Scientific Method; The Past and Present; and The Natural World. Cog-X puts a strong emphasis on parental involvement. In addition to the half-day curriculum, teachers lead biweekly Parent Empowerment workshops.

Provision for child initiation and engagement
The five domains of early childhood development: language, academic skills, social, emotional, and physical are all fully developed and thoroughly integrated into the daily schedule of the Cog-X curriculum. The Cog-X curriculum focuses on both the child's academic learning and social-emotional growth. The following curricular components support the domains of learning:

- Critical Conversations: provides students the opportunity to explore and discuss social-emotional topics relating to the thematic units. Also supports the students development of self-regulation and critical thinking skills
- Morning Infrastructure Meetings: incorporates and develops a variety of learning opportunities including a daily greeting, daily calendar activities, shared writing, singing, dancing and thematic read-alouds.
- Blueprints/Fundamental Play: Student directed with opportunities for community building, written expression, peer conversations, sensory exploration, self-expression, role-playing and theme development.
- Daily Debriefs: provides teachers with the opportunity to conference individually with students during Fundamental Play. Allows both teacher and student to be reflective on their learning.
- Math Small Groups: provides direct instruction, collaborative and independent activities to encourage and engage students to think as a mathematician.
- Literacy and Language Small Groups: provides focused reading experiences with thematic books and also direct instruction, collaborative and independent activities to encourage vocabulary, pre-reading and pre-writing skills.

Provision for parent involvement, through meaningful communication with families
The Cog-X curriculum puts a strong emphasis on parental involvement. In additional to the preschool curriculum, teaching teams and administrators lead bimonthly Parent Empowerment workshops that

instruct and reinforce the academic and social-emotional learning goals. Each session, parents learn how to extend the current lessons from school to the home. The structure of the workshops invites parents into the classroom and gives them the tools necessary to be their child's first teacher. It also gives them a welcoming school environment to ask questions and collaborate with other parents in their child's class. In addition to the Parent Empowerment program, the preschool teaching teams plan Family Theme Activities days that invite parents into the classroom to work with their students on literacy and math activity that relate to the current theme.

Alignment with an authentic assessment tool that is ongoing and comprehensive

The Cog-X curriculum provides teachers with many long term and daily assessment tools to help focus the instruction in the classroom to support the five domains of learning. In implementing the Cog-X curriculum all attending students are assessed with standard preschool screening tools, including the DIAL 3, Pre-IPT and TOPEL (Test of Preschool Early Literacy). Assessments are designed to provide timely feedback to teachers, administrators, students and parents. The curriculum has weekly skills based literacy and math assessments that are completed through teacher observation and small group opportunities. There is also a unit assessment that is completed by the teacher individually with each student at the end of a theme to determine the student's competency. The blueprint created by the student is included each day as part of the Fundamental Play block and serves as an on-going assessment tool for the teacher in regards to the student's development in the following areas: pre-writing, fine motor, language, social-emotional and peer relationships. Blueprints are added to the yearlong portfolio of the student's work. Additionally, the curriculum supports the creation of Professional Learning Communities (PLC) with teaching teams to present, discuss and plan for on-going student assessment and instruction. The PLC framework allows for teaching teams to plan re-teaching and enrichment instruction based on student assessments.

Consideration of the child's linguistic and cultural background

According to the 2013 State Report Card, 23 % of students in the Chicago Heights School District 170 have limited English proficiencies. The intent of the Cog-X curriculum is to support both the bilingual student and their family in the school environment. From the classroom level, many of the teachers have their ESL certification as required by the state of Illinois for the 2014- 2015 school year; remaining teachers are currently attending programs to receive certification. Bilingual assistants support students throughout the instructional day in all areas of the curriculum. To support best practices in bilingual early childhood education, vocabulary and songs are taught in multiple languages. To support the bilingual families, all school-wide and classroom communication is translated into multiple languages. To support attendance and participation in the Parent Empowerment Workshop, bilingual classes are offered. This gives all families the opportunity to participate in a school community at their comfort level.

Consideration of the range of experience and qualifications of early childhood teachers

All of the Cog-X preschool teachers have the required Type 04 Early Childhood Education Certification. The Cog-X curriculum was written with the classroom preschool teacher in mind and recognizes that they are creating and promoting the first school experience for most of their young students. Additionally, the curriculum developers and program administrators have provided staff development through a peer-coaching model, theme collaboration, peer-to-peer focus walks, professional learning communities and mentoring.

Consideration of a wide range of children's abilities, including those of children with IEPs.

The Cog-X curriculum was created with the intent to provide developmentally appropriate instruction to all preschool students. To meet the academic and social-emotional needs of all students, classrooms are blended by age group and student ability. A DIAL 3 screening is held at the start of the school year to provide a baseline as to where a child will be starting school at from an academic and social level. In partnership with the school district, we work with the Early Childhood Special Education Support Services

staff to assist with identifying, supporting and monitoring students that may need additional classroom supports in place. We are able to provide those services at the classroom level with no disruption to the student's school routine.

Table A1: CogX Sample Weekly Schedule

Suggested times	Monday	Tuesday	Wednesday	Thursday
8:00-8:30 **Arrival and Morning Infrastructure**	Greeting Calendar & Weather Thematic Read-Aloud	Greeting Calendar & Weather Thematic Read-Aloud	Greeting Calendar & Weather Thematic Read-Aloud	Greeting Calendar & Weather Thematic Read-Aloud
8:30-9:10 **Small Literacy Groups**	Small Group Language/ Print Knowledge (2 group rotation)	Small Group Literacy/ Print Knowledge (2 group rotation)	Small Group Language/ Print Knowledge (2 group rotation)	Small Group Literacy/ Print Knowledge (2 group rotation)
9:10-9:35 **Breakfast**	Question of the Day Critical Conversation	Question of the Day Critical Conversation	Question of the Day Critical Conversation	Question of the Day Critical Conversation
9:35-10:35 **Play**	My Blueprint & Fundamental Play	My Blueprint & Fundamental Play	My Blueprint & Fundamental Play	My Blueprint & Fundamental Play
10:35-11:15 **Math**	Small Group Math (3 group rotation: students should go to **1** direct or collaborative activity and **1** independent activity)	Small Group Math (3 group rotation: students should go to **1** direct or collaborative activity and **1** independent activity)	Small Group Math (3 group rotation: students should go to **1** direct or collaborative activity and **1** independent activity)	Small Group Math (3 group rotation: students should go to **1** direct or collaborative activity and **1** independent activity)
11:15-11:35 **Outdoor Play**				
11:35-11:50 **Large Group Literacy**	Whole group literacy activity	Whole group literacy activity	Whole group literacy activity	Whole group literacy activity

Small Group Language	Small Group Literacy	Small Group Math
1. Teacher: Language (Phonological Awareness Skills) 2. Assistant: Handwriting & Phonics or Independent Group	1. Teacher: Literacy (Reading Comprehension Skills) 2. Assistant: Handwriting & Phonics or Independent Group	1. Teacher: Direct Instruction 2. Assistant: Collaborative Activity 3. Independent Activity

Table A2: Preschool Thematic Units and Concepts

Theme (Each theme lasts 9 weeks)		**Literacy**		**Math**	**Socio-emotional**
		Literacy & Language	***Print Knowledge & Vocabulary***		
My Family, My Community, My World	All About Me (school, family, and friends)	Print concepts Book handling Picture walk Responding to questions	P, M, K	Colors & shapes	Following instructions
	My Community & My World (where we live, the world around)	Rhyming Compound Words Character empathy Visualizations	H, W, B, T	Counting & numbers	Positive interactions with peers and adults
The Scientific Method	Weather & Seasons	Compound words Syllables Connections Predictions	N, D, F, A	Counting & numbers Spatial Sense Review	Curiosity
	Health & Nutrition	Syllables Prediction Review Rhyming, Compound Words, & Syllables Retelling & Sequencing	G, O, R, L	Arithmetic operations	Responsibility
Past and Present	Stories from Our Past (Fairy Tales, Fables, Nursery Rhymes)	Onset & rhyme Isolating Sounds Story Elements	S, E, V, C	Geometry Review	Empathy & sympathy
	Transportation	Phonemic awareness (PA): Isolating sounds PA: Blending & segmenting Genre Review	Z, I, Q, X	Logical reasoning /patterns Arithmetic operations	Problem solving/conflict resolution
The Natural World	From Seed to Plant	PA: Blending & Segmenting Review Onset & Rime, Isolating Sounds Evaluating	Y, U, J	Measurement & data	Independence
	Wildlife	All concepts review	All letters review	All concepts review	All concepts review

Table A3: Kinderprep Thematic Units

Theme (Each theme lasts 4 weeks)		Literacy		Math	Socio-emotional
		Literacy & Language	*Print Knowledge & Vocabulary*		
Exploring the Arts	What is Art?	Enjoyment of reading Print concepts Book handling Rhyming Syllables Onset and rime	P, M, K H, W, B	Colors & shapes Counting & numbers	Definition and development of personal goals
	Places & Spaces	Picture walk Predictions Responding to questions Isolating sounds Blending & segmenting	T, N, D F, G, A	Arithmetic operations	
Creating Art	Visual Arts	Retelling Sequencing Story elements Substituting and deleting sounds	O, R, L S, E, V, J	Spatial sense & geometry Logical reasoning & patterns	Use concrete steps to work towards and achieve personal goals
	Performing Arts	Character empathy Connections Evaluating texts Genre Review: phonemic awareness and phonological skills	C, Z, I Q, X, Y, U	Measurement & data	

A.2 Assessments

Assessments were always carried out in the order listed below, by assessors that we hired and trained. All assessors held at least a Bachelor's degree and had experience with children. Assessors were always supervised by a head assessor or project manager for CHECC and were blind to the child's treatment assignment.

Main Battery of Assessments

The Peabody Picture Vocabulary Test III (Receptive Vocabulary, 5-10 minutes)
An untimed test of verbal ability and receptive vocabulary, the Peabody Picture Vocabulary Test III (PPVT-III) is a norm-referenced standardized assessment that can be used with subjects ages 2-90+. While the test is said to take 10-15 minutes for adult subjects, early childhood experts estimate that the test takes 5-10 minutes to complete for that population. The PPVT-III is available in both English and Spanish. The PPVT includes an A and B form, and A and B were alternated in each testing period to reduce potential for learning.

Woodcock Johnson III Tests of Achievement (Academic achievement, 5-10 minutes per test)
This norm-referenced assessment allows researchers to track subjects from ages 2 to 85+ and contain most of the key outcomes highlighted by the experts with whom we have spoken. Of the 22 tests that compose the WJ-III ACH, there are four specific tests that are applicable for a preschool population that we used. The WJ includes an A and B form, and A and B were alternated in each testing period to reduce potential for learning.

* *Letter Word Identification*: Measures the subject's ability to identify letters and words.
* *Spelling*: Measures the subject's ability to write orally presented words correctly.
* *Applied Problems*: Measures the subject's ability to analyze and solve math problems.
* *Quantitative Concepts*: Measures the subject's knowledge of mathematical concepts, symbols, and vocabulary.

All WJ-III ACH tests are available in both English and Spanish.

Executive function assessments (attention shifting, working memory, inhibitory control, 15 minutes)
These tests are used to understand levels of executive function among young children. In the first task, which tests attention, children are asked to identify the direction arrows are pointing over a series of different examples. In the second task, which tests working memory, children are asked to remember the colors and animals pictured on the previous page over a series of different examples. In the third task, which tests attention shifting, children are asked to identify which aspects of 2-3 objects are similar based on color, shape or size. These tests were developed by Professor Clancy Blair at NYU (see Blair and Willoughby, 2006a, b, c).

Preschool Self Regulation Assessment - Assessor Report (Self Regulation, 10 minutes)
Designed to assess self-regulation in emotional, attentional, and behavioral domains. The Assessor report is a global report of children's behavior during a series of tasks. In our assessment, the assessor reports on subject behavior during the assessment battery.

Kindergarten Readiness Assessments

Kindergarten Achievement Test (KGAT) (Kindergarten readiness -- academic, 20 minutes)
Administered by the school district. This test requires preschoolers to compete a series of tasks related to math (simple counting and addition), literacy (rhyming, recognition of sounds and letters), and social-emotional learning (following directions, getting along with others).

Head, Knees, Toes & Shoulders (HKTS) (Kindergarten readiness – executive function)
This test asks children to do the opposite of what the instructor asks of them. For example, if the instructor says to "Touch your toes," children should touch their head. It evaluates students on if they touch the right body part and how quickly they arrive at the correct answer.

Follow-up assessments

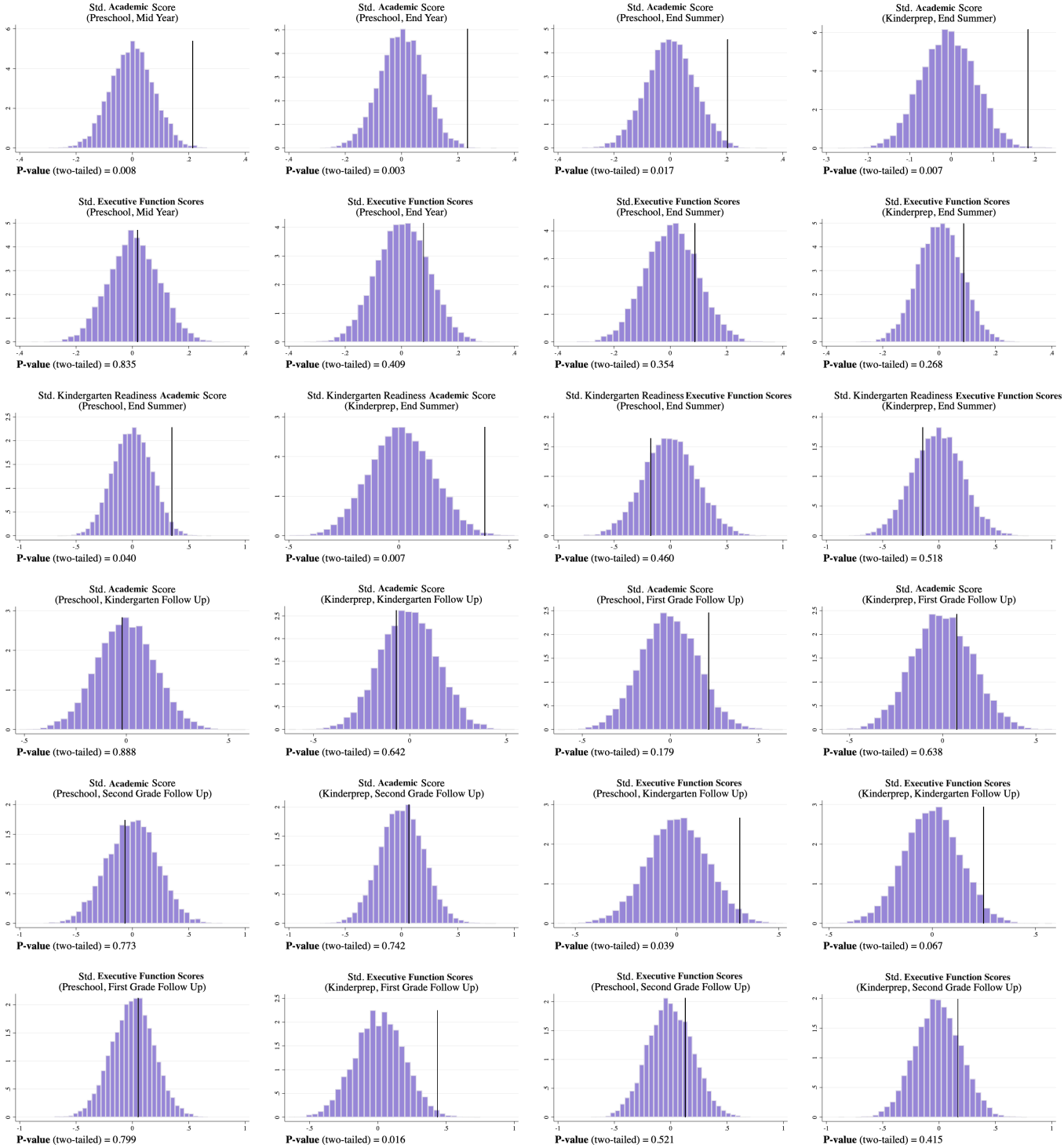
Academic: The academic assessments continued to be administered throughout the follow-up assessment since they are suitable through age 85.

Executive function: Since Blair and Willoughby tests are designed for preschool, we created new tests. Children in kindergarten through second grade were given three tests, measuring the same executive function skills as previously, including inhibitory control, working memory and attention shifting. We

increased the difficulty of the Blair and Willoughby inhibitory control and working memory tests by adding more items. For the attention shifting test, we created a variant of the Wisconsin Card Sort for our age group. The children are given a set of cards with multiple colored shapes and tasked to sort them based on color, shape, or number of objects on the card.

Appendix B: Additional Tables and Robustness Tests

Figure B1: Permutation Tests



Notes: These figures plot the distribution of simulated treatment coefficients obtained by conducting OLS regressions on 10000 re-randomized samples. The black vertical line indicates the true observed beta coefficient Re-randomization is done analogously to the original assignment, within matched blocks, 10000 times. The main specifications are re-run using this simulated treatment assignment and the simulated betas are stored. The exact two-sided p-value is the number of simulated betas that are greater than the observed beta in absolute value.

Table B1: Summary Statistics: Preschool-Plus

	Pre Sample		Mid Sample		Post Sample		Summerloss Sample	
	Preschool	p-value	Preschool	p-value	Preschool	p-value	Preschool	p-value
Academic Baseline	-1.184	0.537	-4.464	0.059	-2.526	0.231	-0.558	0.823
	(1.920)		(2.352)		(2.107)		(2.491)	
Missing Academic Baseline	0.008	0.559	-0.010	0.536	0.003	0.825	0.016	0.414
	(0.014)		(0.016)		(0.014)		(0.020)	
Executive Function Baseline	-0.720	0.668	-4.865	0.032	-1.884	0.297	-0.816	0.682
	(1.679)		(2.256)		(1.802)		(1.988)	
Missing Executive Function Baseline	0.049	0.059	0.031	0.334	0.050	0.073	0.061	0.083
	(0.026)		(0.032)		(0.028)		(0.035)	
Age Baseline	-0.603	0.198	-0.491	0.600	-0.448	0.570	-0.938	0.370
	(0.467)		(0.935)		(0.788)		(1.044)	
Male	0.011	0.814	-0.002	0.979	0.025	0.639	-0.020	0.738
	(0.047)		(0.061)		(0.053)		(0.061)	
White	0.010	0.728	0.024	0.469	0.015	0.622	0.010	0.780
	(0.029)		(0.033)		(0.030)		(0.037)	
Black	-0.001	0.975	0.097	0.105	0.036	0.493	0.044	0.465
	(0.047)		(0.060)		(0.052)		(0.060)	
Hispanic	-0.017	0.714	-0.120	0.045	-0.059	0.254	-0.058	0.340
	(0.047)		(0.060)		(0.052)		(0.061)	
Home Language - English	-0.032	0.465	0.054	0.353	0.023	0.650	0.022	0.709
	(0.043)		(0.058)		(0.050)		(0.058)	
Home Language - Spanish	0.055	0.142	-0.054	0.312	0.016	0.726	-0.012	0.828
	(0.038)		(0.053)		(0.045)		(0.055)	
Home Language - Spanish and English	-0.017	0.540	0.003	0.942	-0.029	0.364	0.001	0.969
	(0.028)		(0.035)		(0.032)		(0.032)	
Treatment Observations	263		218		233		174	
Control Observations	605		160		247		169	

This table reports summary statistics and balance tests. Each panel is restricted to children who took the corresponding assessment. The control and preschool columns report means as well as standard errors in parentheses. The p-value columns report the p-value from a two sided t-test between control and preschool. The F-test p-value row reports the p-value from a joint F-test that all differences between means are equal to zero. The observations column provides the number of children in each panel who have scores for either the academic or the executive function assessment. The p-value for a ttest on the residuals from a regression of the executive function index on a second degree age polynomial is 0.729.

Table B2: Summary Statistics: Kinderprep

	Pre Sample		Mid Sample		Post Sample		Summerloss Sample	
	Kinderprep	p-value	Kinderprep	p-value	Kinderprep	p-value	Kinderprep	p-value
Academic Baseline	-2.098	0.409	-2.943	0.377	-4.703	0.122	-3.608	0.263
	(2.537)		(3.322)		(3.027)		(3.217)	
Missing Academic Baseline	-0.007	0.634	-0.018	0.442	-0.008	0.579	-0.008	0.586
	(0.015)		(0.024)		(0.014)		(0.015)	
Executive Function Baseline	-1.438	0.462	-3.810	0.128	-2.735	0.219	-3.701	0.116
	(1.951)		(2.492)		(2.220)		(2.342)	
Missing Executive Function Baseline	-0.001	0.974	0.074	0.095	0.018	0.505	0.032	0.303
	(0.029)		(0.044)		(0.027)		(0.031)	
Age Baseline	-0.136	0.796	0.511	0.680	0.371	0.617	-0.091	0.909
	(0.523)		(1.236)		(0.739)		(0.794)	
Male	-0.005	0.932	-0.103	0.183	-0.019	0.732	-0.001	0.990
	(0.056)		(0.077)		(0.057)		(0.062)	
White	-0.002	0.950	0.040	0.300	0.027	0.377	0.029	0.424
	(0.033)		(0.038)		(0.031)		(0.036)	
Black	-0.015	0.795	0.030	0.685	-0.034	0.550	-0.020	0.749
	(0.056)		(0.075)		(0.057)		(0.061)	
Hispanic	0.004	0.944	-0.083	0.270	0.004	0.942	-0.032	0.598
	(0.056)		(0.075)		(0.056)		(0.061)	
Home Language - English	-0.032	0.530	0.105	0.147	0.032	0.547	0.013	0.820
	(0.051)		(0.072)		(0.053)		(0.059)	
Home Language - Spanish	0.042	0.337	-0.100	0.124	-0.020	0.664	-0.016	0.760
	(0.043)		(0.065)		(0.046)		(0.051)	
Home Language - Spanish and English	-0.010	0.780	-0.015	0.758	0.003	0.919	0.004	0.902
	(0.034)		(0.047)		(0.034)		(0.036)	
Treatment Observations	204		100		154		143	
Control Observations	299		104		182		144	

This table reports summary statistics and balance tests. Each panel is restricted to children who took the corresponding assessment. The control and kinderprep columns report means as well as standard errors in parentheses. The p-value columns report the p-value from a two sided t-test between control and kinderprep. The F-test p-value row reports the p-value from a joint F-test that all differences between means are equal to zero. The observations column provides the number of children in each panel who have scores for either the academic or the executive function assessment. The p-value for a ttest on the residuals from a regression of the executive function index on a second degree age polynomial is 0.516.

Table B3: Randomization

	Main Randomization	Late Randomization	*Total*
Preschool	249	N/A	249
Kinderprep	131	45	176
Control	311	55	366
Total	691	100	791

Most children were randomized in the main randomization each summer. However, a small group of children were randomized in the following spring either to control or Kinderprep. The sample includes every child who appears in Table B4: Mean Effect Sizes.

Table B4: Mean Effect Sizes

	ITT		
	Mid Year	End Year	End Summer
A. Standardized Academic Score			
Preschool:	0.244***	0.256***	0.174**
	(0.078)	(0.074)	(0.085)
Kinderprep:			0.192***
			(0.069)
Preschool v. Kinderprep:			0.846
B. Standardized Executive Function Score			
Preschool:	-0.001	0.078	0.105
	(0.093)	(0.088)	(0.090)
Kinderprep:			0.112
			(0.086)
Preschool v. Kinderprep:			0.941

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores. The academic and executive function indexes are calculated as the mean of the standardized values of each of the subtests. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. The row Preschool v. Kinderprep reports the p-value of a chi quared test of the equality of coefficients reported at summerloss.Asterisks indicate significance at the 10/5/1 percent levels.

Table B5: Mean Effect Sizes Subscores

	ITT		
	Mid Year	End Year	End Summer
A. Picture Vocabulary			
Preschool-Plus:	-0.249	2.956	2.111
	(2.545)	(2.391)	(2.577)
Kinderprep:			-1.489
			(2.547)
B. WJ Letter-Word			
Preschool-Plus:	4.630*	7.328***	5.458*
	(2.690)	(2.383)	(2.794)
Kinderprep:			7.226***
			(2.172)
C. WJ Spelling			
Preschool-Plus:	7.327**	6.067**	7.875***
	(3.033)	(2.583)	(2.945)
Kinderprep:			5.551**
			(2.222)
D. WJ Applied Problems			
Preschool-Plus:	6.085***	5.012**	0.577
	(2.309)	(2.161)	(2.401)
Kinderprep:			1.231
			(2.082)
E. WJ Quantitative Concepts			
Preschool-Plus:	6.847***	6.457***	1.616
	(2.407)	(2.153)	(2.695)
	0.005	0.003	0.549
	375	480	342
Kinderprep:			7.584***
			(2.461)
F. Operation Span			
Preschool-Plus:	0.080	0.218**	0.292***
	(0.114)	(0.091)	(0.110)
Kinderprep:			0.153
			(0.107)
G. Spatial Conflict			
Preschool-Plus:	0.031	0.044	0.088
	(0.103)	(0.091)	(0.092)
Kinderprep:			0.117
			(0.093)
			0.213
			285
H. Same Game			
Preschool-Plus:	-0.017	-0.072	-0.091
	(0.113)	(0.094)	(0.111)
Kinderprep:			0.099
			(0.100)
H. Preschool Self Regulatory Assessment			
Preschool-Plus:	0.031	-0.021	-0.050
	(0.096)	(0.106)	(0.107)
Kinderprep:			-0.080
			(0.092)

This table reports ITT coefficient estimates of treatment effects on test scores. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Asterisks indicate significance at the 10/5/1 percent levels.

Table B6: Mean Effect Sizes on Academic and Executive Function Indices within Prek Pre-Score Subgroups

	Academic		Executive Function	
	Coefficient	N	Coefficient	N
Above Median Acad Baseline	0.080	239	-0.075	239
	(0.096)		(0.115)	
Below Median Acad Baseline	0.437***	241	0.237**	239
	(0.097)		(0.117)	
Above Median Exec Baseline	0.252***	268	0.056	268
	(0.093)		(0.111)	
Below Median Exec Baseline	0.255**	212	0.112	210
	(0.108)		(0.130)	
Above Median Acad Above Median Exec Baseline	0.054	171	-0.006	171
	(0.113)		(0.137)	
Above Median Acad Below Median Exec Baseline	0.113	68	-0.231	68
	(0.173)		(0.208)	
Below Median Acad Above Median Exec Baseline	0.593***	97	0.160	97
	(0.144)		(0.174)	
Below Median Acad Below Median Exec Baseline	0.316**	144	0.293*	142
	(0.127)		(0.154)	

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores. The academic and executive function indexes are calculated as the mean of the age-standardized values of each of the subtests. The covariates' coefficients are fixed in rows 1 and 2, 3 and 4, and 5 through 8 by regressing the outcome on treatment interacted with an indicator for if the child is above or below the median score along with all other controls. The observation column reports the number of kids in the regression in the given baseline group. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Kinderprep does not include controls for matched pairs because of small sample sizes. Asterisks indicate significance at the 10/5/1 percent levels. The median standardized academic score is -0.164 and the median standardized executive function score is -0.088.

Table B7: Mean Effect Sizes on Academic and Executive Function Indices within Kinderprep Pre-Score Subgroups

	Academic		Executive Function	
	Coefficient	N	Coefficient	N
Above Median Acad Baseline	0.141	164	0.203*	162
	(0.090)		(0.113)	
Below Median Acad Baseline	0.265**	120	-0.010	120
	(0.103)		(0.129)	
Above Median Exec Baseline	0.157**	216	0.128	217
	(0.078)		(0.097)	
Below Median Exec Baseline	0.309**	68	0.052	65
	(0.140)		(0.178)	
Above Median Acad Above Median Exec Baseline	0.079	135	0.137	135
	(0.100)		(0.124)	
Above Median Acad Below Median Exec Baseline	0.386*	29	0.483*	27
	(0.215)		(0.275)	
Below Median Acad Above Median Exec Baseline	0.270**	81	0.101	82
	(0.124)		(0.154)	
Below Median Acad Below Median Exec Baseline	0.275	39	-0.267	38
	(0.182)		(0.228)	

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores. The academic and executive function indexes are calculated as the mean of the age-standardized values of each of the subtests. The covariates' coefficients are fixed in rows 1 and 2, 3 and 4, and 5 through 8 by regressing the outcome on treatment interacted with an indicator for if the child is above or below the median score along with all other controls. The observation column reports the number of kids in the regression in the given baseline group. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Kinderprep does not include controls for matched pairs because of small sample sizes. Asterisks indicate significance at the 10/5/1 percent levels. The median standardized academic score is -0.206 and the median standardized executive function score is 0.294.

Table B8: Sample Sizes by Treatment and Assessment

	ITT		
	Mid Year	End Year	End Summer
A. Picture Vocabulary			
Preschool-Plus:	206	221	167
Kinderprep:			134
Control:	105	195	154
B. WJ Letter-Word			
Preschool-Plus:	208	222	167
Kinderprep:			134
Control:	105	197	153
C. WJ Spelling			
Preschool-Plus:	208	222	167
Kinderprep:			134
Control:	105	197	154
D. WJ Applied Problems			
Preschool-Plus:	208	222	167
Kinderprep:			134
Control:	105	197	154
E. WJ Quantitative Concepts			
Preschool-Plus:	207	222	167
Kinderprep:			134
Control:	105	197	153
F. Operation Span			
Preschool-Plus:	186	222	167
Kinderprep:			134
Control:	94	197	154
G. Spatial Conflict			
Preschool-Plus:	186	222	167
Kinderprep:			135
Control:	99	197	154
H. Same Game			
Preschool-Plus:	195	221	166
Kinderprep:			135
Control:	99	179	154
H. Preschool Self Regulatory Assessment			
Preschool-Plus:	208	221	167
Kinderprep:			133
Control:	105	196	153

This table reports sample sizes for various assessments.

Table B9: Mean Effect Sizes with IPW

	ITT		
	Mid Year	End Year	End Summer
A. Standardized Academic Score			
Preschool:	0.199**	0.215**	0.167*
	(0.084)	(0.073)	(0.087)
Kinderprep:			0.166**
			(0.077)
B. Standardized Executive Function Score			
Preschool:	0.001	0.083	0.072
	(0.105)	(0.093)	(0.097)
Kinderprep:			0.077
			(0.089)

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores weighted by inverse probability weights. IPW weights are estimated from a probit model of a binary variable indicating whether the measure is available for the post-treatment wave regressed on the baseline measure, a dummy variable indicating whether the baseline measure is present, gender, race and age at assessment. The academic and executive function indexes are calculated as the mean of the standardized values of each of the subtests. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. The row Preschool v. Kinderprep reports the p-value of a chi quared test of the equality of coefficients reported at summerloss.Asterisks indicate significance at the 10/5/1 percent levels.

Table B10: Mean Effect Sizes on Academic and Executive Function Indices within Prek Pre-Score Subgroups with IPW

	Academic		Executive Function	
	Coefficient	N	Coefficient	N
Above Median Acad Baseline	-0.005	194	-0.086	194
	(0.094)		(0.113)	
Below Median Acad Baseline	0.493***	185	0.269*	183
	(0.102)		(0.139)	
Above Median Exec Baseline	0.114	174	0.064	174
	(0.094)		(0.103)	
Below Median Exec Baseline	0.335***	205	0.096	203
	(0.106)		(0.143)	
Above Median Acad Above Median Exec Baseline	-0.018	115	0.052	115
	(0.112)		(0.129)	
Above Median Acad Below Median Exec Baseline	0.015	79	-0.292	79
	(0.163)		(0.198)	
Below Median Acad Above Median Exec Baseline	0.384**	59	0.090	59
	(0.157)		(0.168)	
Below Median Acad Below Median Exec Baseline	0.533***	126	0.355*	124
	(0.132)		(0.192)	

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores weighted by inverse probability weights. IPW weights are estimated from a probit model of a binary variable indicating whether the measure is available for the post-treatment wave regressed on the baseline measure, dummy variable indicating whether the baseline measure is present, gender, race and age at assessment. The academic and executive function indexes are calculated as the mean of the age-standardized values of each of the subtests. The covariates' coefficients are fixed in rows 1 and 2, 3 and 4, and 5 through 8 by regressing the outcome on treatment interacted with an indicator for if the child is above or below the median score along with all other controls. The observation column reports the number of kids in the regression in the given baseline group. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Kinderprep does not include controls for matched pairs because of small sample sizes. Asterisks indicate significance at the 10/5/1 percent levels. The median standardized academic score is -0.159 and the median standardized executive function score is 0.047.

Table B11: Mean Effect Sizes on Academic and Executive Function Indices within Kinderprep Pre-Score Subgroups with IPW

	Academic		Executive Function	
	Coefficient	N	Coefficient	N
Above Median Acad Baseline	0.102	143	0.183*	141
	(0.121)		(0.102)	
Below Median Acad Baseline	0.258***	107	-0.063	107
	(0.078)		(0.161)	
Above Median Exec Baseline	0.179**	129	0.046	130
	(0.076)		(0.097)	
Below Median Exec Baseline	0.154	121	0.113	118
	(0.142)		(0.162)	
Above Median Acad Above Median Exec Baseline	0.066	75	0.068	75
	(0.107)		(0.112)	
Above Median Acad Below Median Exec Baseline	0.135	68	0.302	66
	(0.242)		(0.189)	
Below Median Acad Above Median Exec Baseline	0.321***	54	-0.022	55
	(0.106)		(0.182)	
Below Median Acad Below Median Exec Baseline	0.196	53	-0.151	52
	(0.125)		(0.265)	

This table reports ITT coefficient estimates of treatment effects on academic and executive function index scores weighted by inverse probability weights. IPW weights are estimated from a probit model of a binary variable indicating whether the measure is available for the post-treatment wave regressed on the baseline measure, dummy variable indicating whether the baseline measure is present, gender, race and age at assessment. The academic and executive function indexes are calculated as the mean of the age-standardized values of each of the subtests. The covariates' coefficients are fixed in rows 1 and 2, 3 and 4, and 5 through 8 by regressing the outcome on treatment interacted with an indicator for if the child is above or below the median score along with all other controls. The observation column reports the number of kids in the regression in the given baseline group. All regressions control for academic baseline scores, executive function baseline scores, gender, race, home language, age at test date, matched pair grouping, test form, number of previous assessments, year of randomization, an indicator for aged above or below 4 at randomization, mother age at child birth, birthweight, and an indicator if the pretest is from a previous year. Kinderprep regressions also include a dummy for late randomization. Kinderprep does not include controls for matched pairs because of small sample sizes. Asterisks indicate significance at the 10/5/1 percent levels. The median standardized academic score is -0.179 and the median standardized executive function score is 0.317.