Utah High-Quality School Readiness Expansion (HQSR-E) Program Evaluation

2018-2019 Findings

Submitted to the Utah State Board of Education
October 2019

Evaluation and Training Institute
100 Corporate Pointe, Suite 387
Culver City, CA 90230
www.eticonsulting.org

Table of Contents

Executive Summary

EVALUATION PURPOSE AND RESEARCH QUESTIONS

The Utah State Board of Education (USBE) hired the Evalution and Training Institute (ETI) to conduct a three-year evaluation of the High-Quality School Readiness Expansion (HQSR-E) program to determine how participation in high-quality preschool programs impacted children's school readiness. In 2018-2019, the third year of the evaluation, we focused on answering the following four overarching research questions:

1. Did high-quality preschool children have better early literacy skills at entry to kindergarten compared to children who were not enrolled in a high-quality preschool program?
2. Did high-quality preschool children have better early math skills at entry to kindergarten compared to children who were not enrolled in a high-quality preschool program?
3. What were the social-emotional development skills of program children at the conclusion of the preschool year?
4. Did children who received a high-quality preschool program result in stronger school-based literacy outcomes during beginning, middle, and end of first grade compared to a group of peers matched in terms of demographics characteristics who were not enrolled in a high-quality preschool program?

EVALUATION APPROACH

Cohort 3 Preschool Study. The purpose of the HQSR-E preschool program evaluation was to determine the impacts of the three high-quality preschool implementation models on children's school readiness. The implementation models under investigation were: public, private and home/computer-based preschools ("UPSTART"). We measured three types of school readiness outcomes for each program model: literacy, social-emotional development (SED) and math achievement. Literacy outcomes were our main focus, but we also investigated SED and math measures with our samples. We used a pre-test/post-test quasi-experimental research design to measure the school readiness skills of program students ("treatment") and non-program students ("comparison") at two points in time (beginning and end of the school year). The majority of children in the comparison samples attended preschools that were not part of the HQSR-E program.

In an effort to control for pre-existing differences between our treatment and comparison samples, we matched comparison students to treatment students on literacy and math achievement before the program started. We studied the program effects of Cohort 3 (2018-2019). The matched groups created equal baseline levels of literacy and math

achievement, and allowed us to have groups balanced across these skill levels prior to conducting any statistical analysis. We used different types of analyses, including ordinary least squares (OLS) regression analysis and descriptive analysis of outcome data.

First Grade Study. We also investigated the extended impacts of the HQSR-E program on literacy achievement for UPSTART and public preschool providers by studying the differences in literacy outcomes of 2016-2017 HQSR-E and non-HQSR-E participants as they completed first grade in 2018-2019. The treatment and comparison groups were matched on demographics, but not on baseline scores.

COHORT 3 FINDINGS

The majority of high-quality preschool students across all three program models entered kindergarten with literacy quotient levels rated as average or above average when compared to a nationally normed sample of their peers. Beyond this descriptive analysis, students in the at-home, computer-based treatment condition (UPSTART) had higher scores on tests of literacy, on average, than their matched comparison students. There were no significant differences between the public and private high-quality group and their respective comparison groups in most instances, however, attendance in a high-quality private preschool program resulted in lower scores on some literacy subscales than their matched peers. Our key findings are summarized below by area of study:

Literacy Skill Development:

- ✓ The majority of high-quality preschool students across all three program models entered kindergarten with literacy quotient levels rated as average or above average when compared to a nationally normed sample of their peers: high-quality public preschoolers had the largest percentage of participants, with 73% average or above average literacy quotients at the end of preschool, followed by 68% of high-quality private preschoolers and 66% of UPSTART participants.
- ✓ Children using UPSTART had higher overall literacy test scores at program exit than the comparison children. Specifically, UPSTART students had significantly higher scores on subtests that measured letter knowledge and phonological awareness.
- ✓ Participation in high-quality private preschools resulted in significantly lower scores on subtests that measured listening comprehension. Enrollment in high-quality public schools did not result in significantly better literacy outcomes when compared to a group of similar children not attending high-quality preschools.

Social Emotional Development (SED):

- ✓ All three treatment groups had similar SED development by the end of preschool, which was determined through a comparison of mean SED scale scores collected at the end of the program year.

Math Skill Development:

- ✓ Math findings varied by the skills being measured, but in general there were no positive program impacts on math test scores. There were no significant differences found between the UPSTART, public, and private preschool groups and control groups' average test scores.

FIRST GRADE FINDINGS

- ✓ The first grade analyses suggested that UPSTART may have had an extended impact on students' learning through first grade. Students who used UPSTART in 2016-2017 had higher Acadience Reading test scores throughout kindergarden as well as a higher percentage of students with scores that fell within the At or Above benchmark range compared to a group of students with similar demographics.
- ✓ High-quality public programs did not have an impact on student learning through first grade compared to a group of similar students matched on demograhic characteristics.

RECOMMENDATIONS

Our recommendations were developed in response to the study findings, theoretical considerations and limitations discussed in the body of this report. Our central recommendation is to continue the program, along with the evaluation to identify its impacts over time, particularly its long-term impacts, which is crucial information for policy decisions.

Previous research has shown the value of preschool for preparing students for kindergarten but understanding the unique impacts of different types of preschool models is difficult in a program evaluation context due to the myriad of implementation details not accounted for in a study of this type. We recommend adding a preschool provider survey as a requirement to the program to help collect information about their unique preschool curriculum. These program implementation details can be used to more accurately examine the potential benefits of high-quality preschool, particularly in the public and private preschool models. Potential program implementation details to capture could include specifics about program curricula (i.e. literacy skills taught, etc.), and areas of developmental focus (i.e. emphasizing free play and discovery, or structured learning, etc.).

We have not seen strong impacts for the public and private high-quality preschool programs in past studies. It is possible that the comparison students used to study the program impacts also had access to quality preschool program providers that had not been identified as such by the state. Future research should utilize preschools that scored below the high-quality threshold on the ECERS-3 and ECERS-E, the tool used by the state to identify high-quality sites, to ensure comparison students did not receive the same quality of education as program students. With enough available data, future research could also explore the differences in program impacts for different levels of

quality. For example, the study could compare literacy or numeracy outcomes for students who attended preschool sites and that scored differently on the ECERS.

We also recommend implementing the UPSTART program in public and private HQSR-E program sites where possible. Traditional high-quality preschool offers a three-dimensional social landscape, but our results show that it does not have the same positive impacts on specific literacy skills that the UPSTART program does. A hybrid high-quality preschool model that combines an adaptive, computer-based learning program with a traditional classroom-based preschool, including all the peer and adult social experiences, could give Utah families the full spectrum of benefits. Future evaluations will be needed to better understand the impact of the different preschool program models on school readiness.

Our final recommendation is to consider broadening the evaluation measures to use data from the newly implemented Utah Preschool Entry and Exit Profile (PEEP) and Utah Kindergarten Entry and Exit Profile (KEEP). The PEEP and KEEP were not available when this evaluation was designed, and we believe it may offer additional information about HQSR-E program impacts beyond the measures of literacy, math and social-emotional development that we currently use. As the foundational measure of school readiness in the state, we believe that using PEEP and KEEP scores alongside our established measures of early literacy and math would add another view and help to better understand the relationship between HQSR-E models and their impacts on students.

HQSR-E 2018-2019 Report

Evaluation Purpose & Research Questions

The Evaluation and Training Institute (ETI) conducted a three-year evaluation of the High-Quality School Readiness Expansion (HQSR-E) program. In its third and final year, the goal of the evaluation was to understand the program's impact on students' school readiness skills across three high-quality preschool program implementation models ("preschool study") and determine if the program had lasting impacts through the first grade ("first grade study"). We studied the following three hiqh-quality prechool models: in public preschool settings, private preschool settings and through an at-home, computer administered software program (known as UPSTART: "Utah Preparing Students Today for a Rewarding Tomorrow"). Where possible, children's outcome scores from each program model were compared to scores from children who were not in a high-quality preschool setting ("control group" or "comparison group"). Children in the control group were not enrolled in a preschool designated by the state as meeting high-quality ("HQ") program criteria, but approximately 80% of the control students were enrolled in a preschool program of some type. Specific research questions used to guide the direction of our evaluation included:

Preschool Study

- Did high-quality preschool children have better **early literacy** skills at entry to kindergarten compared to control group children?
- Did high-quality preschool children have better **early math** skills at entry to kindergarten compared to control group children?
- What was the **social-emotional development** profile of program children at the conclusion of the preschool year?

First Grade Study

- Did the use of a high-quality preschool program result in stronger **school-based literacy outcomes during the beginning, middle, and end of first grade** compared to a group of peers matched in terms of demographic characteristics who were not enrolled in a high-quality preschool program?

Program Background

The Utah State Legislature provided grant funding for a multi-year project to expand access to high-quality preschool programs to economically disadvantaged children in March of 2016. To receive grant funding and qualify for high-quality status, Local Education Agencies (LEAs) and private preschool providers were scored on a rubric to ensure their programs met certain standards for high-quality programs. The process involved the submission of a grant application, an interview in which the program provider reviewed supplementary materials submitted as evidence of high-quality elements, and classroom observations. In the grant applications, the program provider described their need for funding, plans for recruiting students, and how their program met each high-quality element listed in the legislation, including, staff with at least a child development associate certification or an associate or bachelor's degree in an early childhood education related field, evidence-based curriculum that aligns to the Utah Early Childhood Standards adopted by the State Board of Education, and class sizes that did not exceed 20 students, among other criteria. Classroom observations were conducted using the Early Childhood Environment Rating Scale (ECERS-3 and ECERS-E), an observational tool that rates the quality of early education programs on the following six areas: space and furnishings, personal care routines, language and literacy, learning activities, interaction, and program structure. Local Education Agencies (LEAs) must achieve a minimum overall ECERS score of three or four, depending on the program length, as well as meet the additional rubric criteria. Private Providers must achieve an overall score of 3.67 and a score of at least a 4 in the subscales of Literacy and Numeracy and Teacher/Student interactions.

In 2018-2019, there were eleven high-quality public-school districts[1], eight private high-quality preschool providers, and one computer/home-based high-quality provider, Waterford's UPSTART program. In addition to expanding access to high-quality preschool programs, the legislation provided funding to increase the quality of early childcare professionals through a professional development program that helps educators earn their Child Development Certificate.

Evaluation Methods

In this section, we provided an overview of our research design, study measures, the evaluation sample, and analyses methods (please see **Appendices A-C** for more details).

COHORT 3

[1] HQSR-E districts included: Cache, Davis Community, Duchesne, Iron, Jordan, Logan, Murray, Salt Lake City, Sevier, Wasatch, Washington, and Weber.

Research Design. We used a repeated measures quasi-experimental research design to measure the school readiness skills of high-quality preschool program students[2] (the "treatment group") and a comparison group of children enrolled in preschool programs that were not designated as high-quality (the "Non-HQ comparison group") before the respective programs started (the "pretest") and after the programs ended (and before students enrolled in kindergarten; the "posttest"). **Table 1** depicts the evaluation design.

Table 1. Preschool Study Yearly Testing and Data Collection

	Summer/Fall 2018	Program	Summer 2019
Treatment group	Pre-K Obs 1	HQSR-E Program	Pre-K Obs 2
Non-HQ comparison group	Pre-K Obs 1	non-HQ Program	Pre-K Obs 2

Measures. Our school readiness measures included several aspects of early literacy, along with the specific areas of early math achievement and social and emotional development (SED). We measured early reading with the Brigance Inventory for Early Child Development (IED II and III; Brigance 2010, 2013, respectively), the Bader Reading and Language Inventory (Bader & Pearce, 2008), and the Preschool Early Literacy Indicator: Comprehension section (PELI; Kaminski, Abbott, Bravo-Aguayo, 2018). Consequently, our early literacy measures consisted of the following instruments:

- Global Literacy (Brigance), an overall norm-referenced composite of early literacy achievement (vocabulary, letter knowledge, print concepts, decoding);
- Letter Recognition (Brigance), a domain-specific composite that measures letter recognition, the ability to recite the alphabet, and knowledge of letter-letter sound correspondence;
- Phonological Awareness (Bader), a measure that assesses children's rhyme recognition, and phonemic blending/segmenting skills;
- Listening Comprehension (PELI), a norm-referenced assessment of listening comprehension skills, including recollection, inference, and prediction.

We measured math and social and emotional development with the Brigance IED III. Selected scales from the Brigance Math measured children's ability to count by route, read numerals, and identify missing numerals in a sequence. The social and emotional development scale was a parent survey and assessed parental perceptions of their children's interpersonal and self-regulatory skills. Based on recommendations outlined in the Cohort 1 report (Evaluation and Training Institute, 2017), we modified the socio-emotional development measure from a dichotomous response scale to a five-point Likert response scale to increase variability and reduce measurement ceiling effects.

[2] "Program students" were operationally defined as economically disadvantaged students who attended a high-quality preschool.

Lastly, we collected background information on participants' home environment through a parent intake survey developed by the ETI project team. See **Appendix A** for more details about the measures used in the evaluation.

Sample. We recruited and tested students from four treatment groups: high-quality public preschool (Public), high-quality private preschool (Private), high-quality computer/home-based preschool (UPSTART), and students who did not attend preschool identified as "high quality" by the state (non-HQ comparison students; control). Our sample consisted of children from high-quality public districts[3] and high-quality private sites[4], as well as children enrolled in UPSTART from various geographic locations. Children in the UPSTART and control samples attended preschools that were not part of the HQSR-E program (**see Appendix B** for specific details on data collection procedures)[5]. **Table 2** displays the number of children tested in Cohort 3.

Table 2. HQSR-E Cohort 3 Number Tested

Experimental group	Number Tested Cohort 3
Public	77
Private	77
UPSTART	134
Non-HQ comparison	107

**Note.* Children moved from one experimental group to another from pre-to-posttest.

Recognizing that the lack of random assignment means that pre-existing differences in literacy skills or demographics may be present in our sample, we used a statistical matching process called Coarsened Exact Matching (CEM) to match treatment students who participated in a High-Quality preschool model with comparison students who were enrolled in traditional preschools. We matched each treatment group with a group of similar comparison students on the basis of their pre-test scores on our outcomes of interest (literacy composite, letter knowledge, phonological awareness, comprehension, and math). This process resulted in the creation of 15 different analytic samples that ensured that we were comparing, for example, high-quality public preschool children with similar listening comprehension pre-test scores as their comparison counterparts.

After removing cases with incomplete or missing data, cases in which the program or control condition changed (e.g. control children who enrolled in a high-quality preschool after the pre-test), and creating statistically matched groups with CEM, the final analytic samples for each measure and treatment group were listed in **Table 3** (see **Appendix C** for sample demographic characteristics).

---

[3] High-quality public districts tested in Cohort 3 were Davis Community, Jordan, Logan, and Weber.
[4] Cohort 3 high-quality private sites were Discovery Clubhouse, Head Start (UCA), Hiltop Christian, Mountainland Head Start, Oasis Montessori, Smart Kids Salt Lake, and YMCA of Northern Utah.
[5] Approximately 80% of the control group and 32% of the UPSTART children participants attended a non-high-quality preschool.

Table 3. HQSR-E Analytic Sample

Group	Literacy (Brigance)	Letter Knowledge (Brigance)	Phonological Awareness (Bader)	Listening Comprehension (PELI)	Math (Brigance)
HQ Public	63	57	69	63	63
Non-HQ	63	57	69	63	63
HQ Private	63	59	70	60	60
Non-HQ	63	59	70	60	60
UPSTART	89	89	107	80	84
Non-HQ	89	89	107	80	84

Analysis. We conducted three types of analyses to answer our research questions: Ordinary Least Squares (OLS) regression analyses, effect sizes, descriptive statistics, such as frequencies and percentages, and an analysis using norm-referenced data.

OLS. Inferential statistics, such as OLS regression, allowed us to control for differences that might affect the outcome scores of our treatment and control groups. By accounting for these differences, we are better able to determine if the outcomes were a product of the treatment (e.g. program use) or due to other factors unrelated to the program. Statistical significance testing also allowed us to determine the likelihood that a finding was a result of chance, or due to the treatment effect.

Effect Sizes. We calculated effect sizes using Hedges G, which allowed us to compare effects between different samples. **We interpreted effect sizes using a threshold of .26 to identify meaningful impacts on a student: coefficients larger than this are above average for similar program research.** This threshold was created as a way to benchmark the strength of our findings against those found in similar studies. Please see **Appendix D** for details on how we arrived at this benchmark.

Descriptive Statistics & Benchmarks. To present our findings in an intuitive and applicable context, we used norm-referenced data from the PELI and Brigance to determine quotients and create age equivalency benchmarks. Analysis of normative outcome data was conducted using descriptive statistics, which did not allow for the control of pre-existing differences between groups, and need to be interpreted with caution.

Interpreting Study Findings: Of the two types of analysis methods (inferential and descriptive statistics), OLS regression was the more rigorous of the two methods. OLS regression analyses enabled us to control for pre-existing group differences, such as student demographics, and apply significance testing to determine the likelihood of an effect resulting from the program, or due to chance occurrence. In contrast, descriptive analyses do not allow us to control for pre-existing differences between groups, and the results for these analyses should be interpreted with this in mind.

First Grade

Research Design. We studied the UPSTART and High-Quality preschool programs using a nonequivalent group, post-program only research design. We utilized statistical match techniques (CEM) to equate the two groups and minimize the presence of preexisting differences. We matched treatment and control groups on the demographic variables of ethnicity, language, low income status, Title 1 enrollment, and English Learner and special education status. We did not have pre-test scores at preschool entrance for the sample of hiqh-quality students who participated in 2016-2017, and therefore we were not able to create equal groups based on baseline scores.

Measures. We used Acadience Reading[6] scores as our outcome measure, a standardized measure of literacy achievement for elementary school students. The Acadience Reading was administered to students in Grades K-3 in schools throughout the state. At the beginning of the year of kindergarten (BOY), Acadience Reading measured children's competency with the alphabetic principles and basic phonics with the Letter Naming Fluency and First Sound Fluency subtests. The subtests were administered in the second half of kindergarten (middle of year - MOY and end of year - EOY) and beginning of first grade (BOY) , middle (MOY), and end of year (EOY) and assessed children's letter knowledge, phonics and word attack skills with the following measures: Letter Naming Fluency, Phoneme Segmenting Fluency, and Nonsense Word Fluency (see **Table 4**).

Table 4
Acadience Reading **Next Subscales by Administration Period**

	Kindergarten BOY	Kindergarten MOY	Kindergarten EOY	First Grade BOY	First Grade MOY	First Grade EOY
First Sound Fluency	X	X				
Letter Naming Fluency	X	X	X	X	X	X
Phoneme Segmentation Fluency		X	X	X	X	X
Nonsense Word Fluency		X	X	X	X	X

The Acadience Reading Composite score is an overall measure of children's early literacy ability and was calculated by summing the subtest scores associated with each test administration period. The Acadience Reading Composite scores in Grades K-1 served as our outcome measures.

Sample. We relied on data from four different sources to create our final dataset and complete our analyses, including demographic data, Acadience Reading literacy achievement scores, UPSTART usage information, and a list of ID's for students who participated in high-quality preschool programs in 2016-2017. The USBE provided demographic data for students enrolled in first grade during the 2018-19 academic year. The demographic data consisted of student-level information such as gender, race, socioeconomic status, English language learner status, primary language, and Title 1 school status.

[6] Acadience Reading was formerly known as DIBELS Next instrument.

Analysis. We conducted two analyses to answer our research questions: effect sizes and descriptive statistics, such as frequencies and percentages.

Effect Sizes. We calculated effect sizes using Hedges G, which allowed us to compare effects between different samples. **We interpreted effect sizes using a threshold of .26 to identify meaningful impacts on a student: coefficients larger than this are above average for similar program research.** This threshold was created as a way to benchmark the strength of our findings against those found in similar studies. Please see **Appendix D** for details on how we arrived at this benchmark.

Benchmarks and Demographics. We used Acadience Reading composite benchmark classifications measured at multiple time points in kindergarten and first grade, as well as demographic characteristics to present a more complete composite of the participants.

Cohort 3 Preschool Study Findings

In this section, we present findings for each category of school readiness, including early literacy, socio-emotional development (SED), math, and first grade findings. We provide context for each area of school readiness and organize the findings in question and answer format to make them user-friendly to a wide-range of audiences with different backgrounds. Findings that were statistically significant, meaning it is unlikely the effect occurred due to chance, were identified using asterisks. More detailed findings for technical readers were provided for each high-quality preschool program in **Appendix E.**

Literacy

The development of early literacy skills in preschool is a crucial component of school readiness. Research on early literacy stresses the importance of understanding alphabetic principles and having oral language skills in order for literacy to emerge (Woolfolk, 2016). Children who are behind their peers in such reading skills at entry into kindergarten might become struggling readers, something that could have a negative impact on their long term academic success. Literacy development begins before formal instruction in school and future success in reading is predicated on mastering early literacy and communication skills (Shonkoff & Phillips, 2000). High-quality early education settings that use age-appropriate curricula with clearly articulated goals can contribute to improvements in literacy so that children have the skills for academic success at school entry (Phillips et al., 2017).

Question 1: Was enrollment in the HQSR-E program associated with higher scores on measures of early literacy skills when compared to children not enrolled in the program?

We used an OLS regression model to determine the impact of the different implementation models on children's overall literacy development while controlling for prior achievement, preschool duration, and family socioeconomic factors. **Table 5** presents the effect size estimates for the Brigance Global Literacy composite, an overall measure of alphabet knowledge, vocabulary, phonics, and language concepts. An asterisk denotes a statistically significant finding and negative effect sizes indicate the control group performed better than the treatment group.

Table 5. Post-test Analysis of Literacy Composite Effect Sizes, OLS Regression Model

Construct	UPSTART	Public	Private
Global Literacy	**0.383***	NS	NS

Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; any ES above .26 is higher than the average ES seen in similar education evaluations; NS indicates result did not reach statistical significance.

As shown in **Table 5**, the UPSTART group was the only program group to produce a statistically significant positive effect on the Brigance Global Literacy composite. The UPSTART group produced a meaningful impact on the development of early literacy skills when compared to children who were enrolled in traditional non-HQ preschools, with an effect size of .383 that exceeds our .26 benchmark for practical significance. There were no significant differences between children enrolled in high-quality public or private preschools and their comparison counterparts on the Brigance Literacy composite.

Question 2: Did high-quality preschool children have better early literacy skills at entry to kindergarten on specific literacy domains compared to control group children?

We examined children's literacy skills through the lens of individual literacy domains that have a strong relationship with future reading success. Perhaps the most fundamental early literacy skill that is predictive of later reading achievement is *letter knowledge*, or the knowledge of the names and sounds associated with printed letters (Wood & McLemore, 2001). Another key precursor of literacy acquisition is *phonological awareness*, or the ability to recognize, identify, manipulate, the smaller sound units within words, independent of their meaning (Cassady & Smith, 2004). Finally, *listening comprehension* involves the active process of unearthing and constructing meaning from oral text and develops the strategies competent readers need to comprehend written texts (Moore & Hall, 2012).

Table 6 presents the effect size estimates for skills measuring letter knowledge, oral comprehension, and phonological awareness for each high-quality preschool model. There were few statistically significant positive differences between program and control group children in literacy domain areas, with the exception of UPSTART, which had small to medium effects in areas measuring phonological awareness (ES = .462) and small effects in areas measuring letter knowledge (ES = 0.270). The comparison group scored higher than the high-quality private preschool group on measures of students' listening comprehension skills (ES = -.644). There were no significant or meaningful differences between children enrolled in high-quality public or private preschools and comparison students on scales measuring letter knowledge or phonological awareness or for UPSTART and high-quality public preschool on scales measuring listening comprehension.

Table 6. Effect Size Estimates by Literacy Domain

Literacy Domain	Skills Tested	UPSTART	Public	Private	Instrument
Letter Knowledge	Letter Sounds	**0.270***	NS	NS	Brigance
	Recites Alphabet				
	Letter Knowledge				
Listening Comprehension	Inference/Prediction	NS	NS	**-0.644****	PELI
	Recollection				
	Cloze				
Phonological Awareness	Rhyme Recognition	**0.462***	NS	NS	Bader
	Phonemic Blending				
	Phonemic Segmenting				

Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; any ES above .26 is higher than the average ES seen in similar education evaluations; NS indicates result did not reach statistical significance.

Question 3: To what extent were children ready to learn, based on their literacy skill development, at entry into kindergarten?

Literacy Quotient Levels. While raw scores, or the number of test items that a child has answered correctly, are ideal for measuring change over a specific period of time and comparing the performance of one group with another, there are some limitations with their use. For example, it is difficult to compare raw scores across measures with different numbers of items and raw scores do not factor in the impact of chronological age (i.e., older children are expected to know more than younger children). We can overcome some shortcomings of raw scores by using the Brigance literacy standardized scores that are based on a nationally representative sample of children. The standardized scores produce quotients and age equivalents for the Brigance literacy measure and provide a direct comparison with a representative sample of similarly aged children entering into kindergarten.

Analogous to intelligence quotients (IQ), Brigance literacy quotients divide the distribution of performance on the Brigance literacy scale into even, easily interpretable units with a mean of 100 and a standard deviation of 15. A score of 100 indicates that a child's performance on the Brigance Literacy falls at the mean of the standardized sample of children at a similar age. Quotients can be interpreted qualitatively with the following category levels:

<70-89	Below average
90-110	Average
111-130+	Above average

A child that has a quotient level below average does not possess the same level of literacy skills as a similarly aged child and may require additional instructional support. Conversely, a child who is average or above enters kindergarten ready to learn, with a basic understanding of fundamental early literacy concepts.

Table 7 presents the percentage of children in each program group who had literacy quotient levels that fell into the categories of below average, average, and above average at the conclusion of the HQSR-E Program.

Table 7. Post-Test Literacy Quotient Levels by HQSR-E Program Group

Quotient Level	UPSTART (N = 133)	Public (N = 75)	Private (N = 75)
Below average	34%	27%	32%
Average	19%	30%	25%
Above average	47%	43%	43%

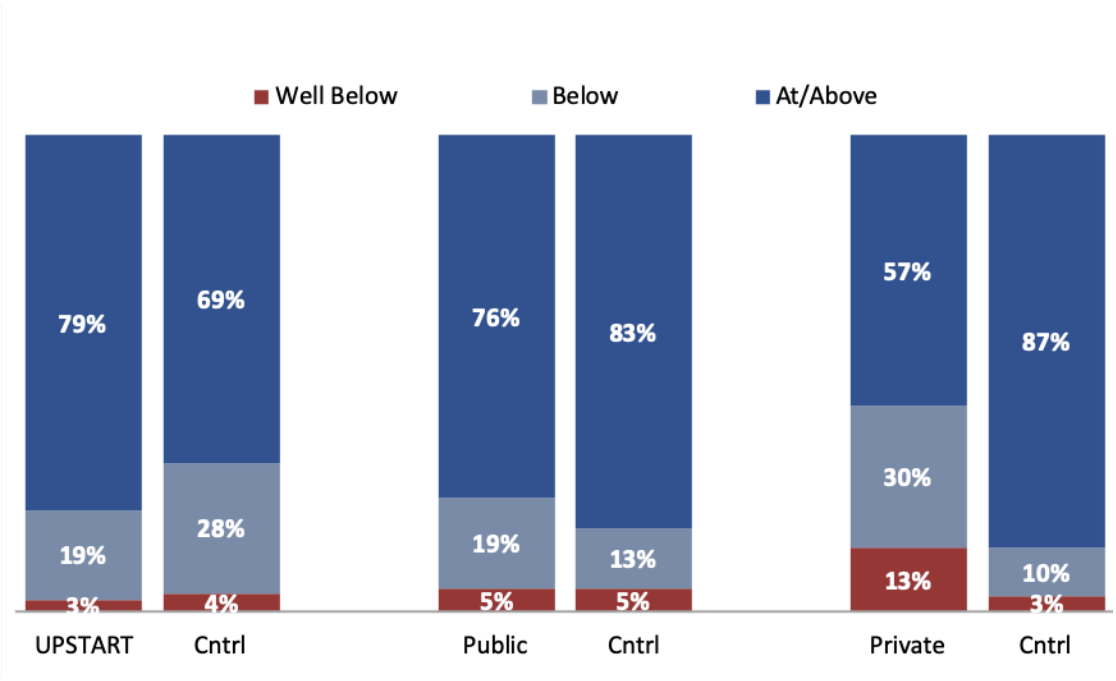
Overall, regardless of HQSR-E program model, the majority of students were entering kindergarten with literacy quotient levels of average or above, indicating that they were on target and ready to learn. Sixty-six percent of the children who participated in UPSTART had quotient scores rated as average or above at the end of the preschool year, with 47% of children scoring literacy quotients of above average.

Moreover, approximately 73% of students enrolled in high-quality public and 68% enrolled in high-quality private preschool scored average or above on the Brigance literacy measure before entering kindergarten. Students attending high-quality private preschools and high-quality public preschools equally scored above average on the Brigance (43%).

PELI Comprehension Benchmark Goals. Benchmark goals from the PELI listening comprehension scale were criterion-referenced target scores that represented adequate yearly progress for preschoolers and provided additional insight into the performance of children in the three models of the HQSR-E program. Students at or above benchmark were likely to need core support to reach future literacy goals, whereas children below or well below benchmark goals were likely to need strategic or intensive support.

Figure 1 presents post-test PELI listening comprehension benchmark levels for children enrolled in the three high-quality preschool models and corresponding matched control samples. The majority of children had listening comprehension scores that were at or above benchmark, with the UPSTART program having the highest percentage of on-target children (79%), followed by high-quality public preschools (76%) and high-quality private preschools (57%). UPSTART was the only high-quality group to outperform the comparison group at post-test.

Figure 1. Listening Comprehension Benchmark Status at Post-test



Question 4: How did children's level of learning change from pre-to-posttest compared to age specific norm group achievement levels?

The Brigance literacy scale is associated with age equivalents based on a child's literacy knowledge compared to other children at a particular age. If a child scored below the average knowledge age for his/her biological age, he/she might be at risk of struggling in kindergarten. The risk went up for students who performed far below their age equivalent norm group comparisons. For example, if a child performed at 90% of his normed achievement level age, he or she was closer to the content knowledge target than a child who performed only at 70% or 80% of his/her biological age.

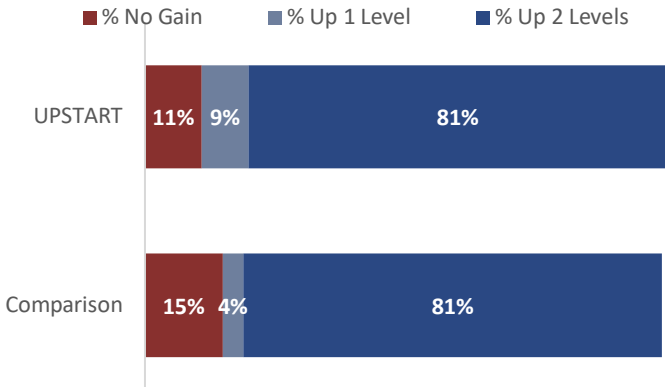
We created a ratio of children's knowledge age (KA) to their biological age (BA) for high-quality treatment preschoolers and their non-high-quality preschool comparison counterparts at pre-test and at post-test. A ratio of 100% indicated that a child's knowledge age matched his or her biological age, and a ratio greater than 100% designates a knowledge age that exceeded the child's biological age. Conversely, children with a knowledge age of 80% or below their biological age were classified as at risk for needing additional assistance to succeed academically.

High-quality preschool interventions that use age-appropriate curricula can contribute to improvements in academic areas, particularly for children who lack the skills that predict readiness for kindergarten (Phillips et al., 2017). We measured the extent to which children whose literacy content knowledge age was at or below 80% of their biological age at pre-test, and had mastered enough literacy concepts during the course of the preschool year to improve their knowledge age at post-test. If children moved from a

knowledge age at or below 80% of their biological age to a knowledge age close to their actual biological age (90% or above), they were on target with similarly aged peers and were more prepared to enter formal school settings.

Figure 2, follows the 57 UPSTART children and their 48 non-high-quality comparison counterparts who were classified as 80% or below their biological age at pre-test and examines their performance at post-test. Findings were based on descriptive analysis and comparisons to norm groups.

Figure 2. Change in Knowledge Age, UPSTART and Comparison Preschoolers

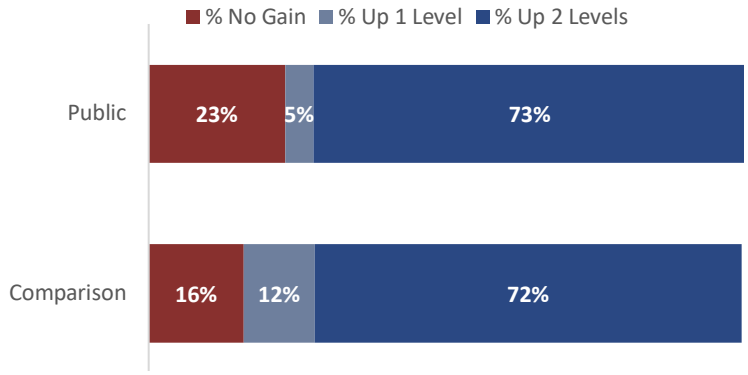


N= UPSTART (57); Comparison (48)
Students had a content knowledge age at or below 80% of their biological age (BA) at pre-test.
% No gain: Stayed at or below 80% BA; % Up 1 Level: 81-90% below BA; Up 2 levels: 91% to or at/above BA

As shown in **Figure 2**, eighty-one percent of at-risk UPSTART and comparison children increased their knowledge age two levels, raising their knowledge age from 80% or below their biological age (i.e., at-risk) to 91% or above their biological age (i.e., on target) during the course of the program. Eleven percent of UPSTART children and 15% of non-high-quality comparison children remained at 80% or below their biological age at the end of the preschool year.

Figure 3, displays the change in knowledge age for the 22 high-quality public preschoolers and 25 comparison children that were classified as at-risk at pre-test (i.e., had a literacy knowledge age of 80% or below their biological age).

Figure 3. Change in Knowledge Age, high-quality Public and Comparison Preschoolers

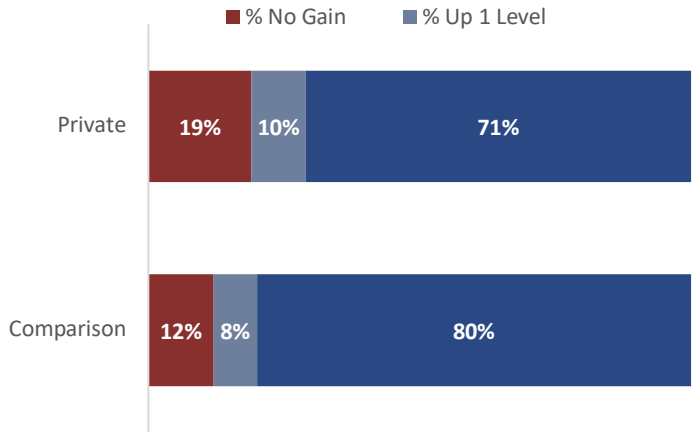


N= Public (22); Comparison (23)
Students had a content knowledge age at or below 80% of their biological age (BA) at pre-test.
% No gain: Stayed at or below 80% BA; % Up 1 Level: 81-90% below BA; Up 2 levels: 91% to or at/above BA

Seventy-three percent of the high-quality public preschoolers classified as at risk moved up two levels to their target knowledge age, compared to 72% of at-risk children who were enrolled in non-high-quality preschools (see **Figure 3**). Twenty-three of the high-quality public at-risk children had a literacy knowledge age that remained at 80% or below their biological age after a year or preschool, while 16% of at-risk comparison children saw no change in their knowledge age.

Figure 4, displays the change in knowledge age for the 31 high-quality public preschoolers and 25 comparison children who were classified as at-risk at pre-test (i.e., had a literacy knowledge age of 80% or below their biological age).

Figure 4. Change in Knowledge Age, High-quality Private & Comparison Preschoolers



N= Public (31); Comparison (25)
Students had a content knowledge age at or below 80% of their biological age (BA) at pre-test.
% No gain: Stayed at or below 80% BA; % Up 1 Level: 81-90% below BA; Up 2 levels: 91% to or at/above BA

Seventy-one percent of high-quality private preschoolers with a literacy knowledge age of 80% or below their chronological age increased their knowledge age two levels to 90% and above their chronological age, compared to 80% of comparison non-high-quality children. Nineteen percent of high-quality private preschoolers, compared to 12% of non-HQ comparison children, remained at 80% or below their biological age at the end of the preschool year.

Social-emotional Development

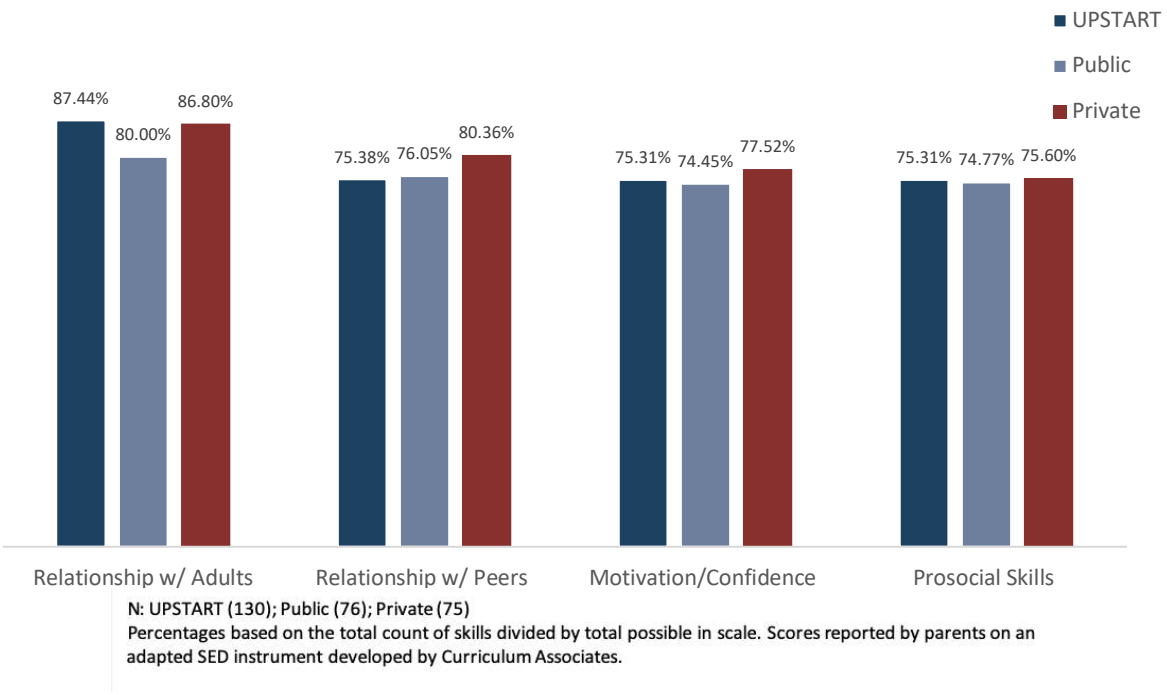
Social-emotional development includes the child's experience, expression, and management of emotions and the ability to establish positive and rewarding relationships with others (Early Education and Support Division, 2017). Children's social and emotional development were important skillsets for determining if children are emotionally ready to learn and interact with their peers in a school setting. For example, Denham (2006) explained that "compared with less socially competent peers, children who enter school with greater competence show a number of benefits related to their relationships with teachers and classmates, readiness to learn, school engagement, and overall academic adjustment." We administered a survey to parents that measured their children's social-emotional development in four areas, including areas measuring their interpersonal skills (relationships with adults and peers) and self-regulatory skills (prosocial skills and confidence and motivation). Our findings are presented below.

Question 5: What were the social-emotional development (SED) skills of children at the conclusion of the HQSR-E program?

Parents reported children's skill development on four aspects of social-emotional development at the conclusion of the preschool program: relationships with adults, play and peer relationships, prosocial skills and motivation and self-confidence.

Figure 5 presents the percent of possible skills developed in these areas, on average, for each of the four social-emotional development constructs by HQSR-E program group. In general, parents rated the SED skills of their children similarly, with little variation on SED subscales between children enrolled in the different high-quality program models.

Figure 5. Post-program Social-emotional Development Skills by Program Group



N: UPSTART (130); Public (76); Private (75)
Percentages based on the total count of skills divided by total possible in scale. Scores reported by parents on an adapted SED instrument developed by Curriculum Associates.

Parents indicated that children had the strongest SED skills in behaviors related to relationships with adults (80-87%), followed by relationships with peers (75-80%), motivation and self-confidence (74-78%), and prosocial skills (75-76%). Most children across HQSR-E program groups developed the social-emotional development skills appropriate to their age, indicating they were ready for kindergarten in this area.

Math

A meta-analysis conducted by Duncan and colleagues (2007) revealed the importance of establishing a solid, early foundation in mathematics to give children the best chance for later academic success. Educators begin teaching foundational math skills in kindergarten and children must reach certain milestones prior to learning more advanced skills in later grades. Preschool can provide an opportunity for children to get a head start in building these foundational skills through helping children understand numeracy concepts (Bisanz, 2011).

Question 6: Did high-quality preschool children have better early math skills at entry to kindergarten compared to control group children?

Table 8 depicts the differences between treatment and control groups on a composite measure of math that assesses specific early numeracy skills. An asterisk denotes a statistically significant finding and negative effect sizes indicate the control group performed better than the treatment group.

Table 8. Posttest Analysis of Math Effect Sizes, OLS Regression

Scale	Skills Tested	UPSTART	Public	Private	Instrument
Math Composite	Counts by rote	NS	NS	NS	Brigance
	Reads numerals				
	Identifies missing numbers				

Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; any ES above .26 is higher than the average ES seen in similar education evaluations; NS indicates result did not reach statistical significance.

As depicted in **Table 8**, UPSTART,[7] the high-quality public program, and high-quality private program did not have a statistically significant positive impact on children's math skill development when compared to similar children who did not participate in the HQSR-E program.

[7] As a computer-based program, not all children received the same curriculum, and children may not have participated in the math component of UPSTART.

First Grade Impacts

Evaluations of the UPSTART program have consistently shown a medium to strong impact on improving children's early literacy skills prior to entering kindergarten. Recent evaluations of preschool programs conducted after program completion show similar evidence of increased skills in both early literacy and mathematics (Weiland & Yoshikawa, 2013), suggesting that high-quality preschool programs can foster school readiness and prepare children to meet the demands of kindergarten.

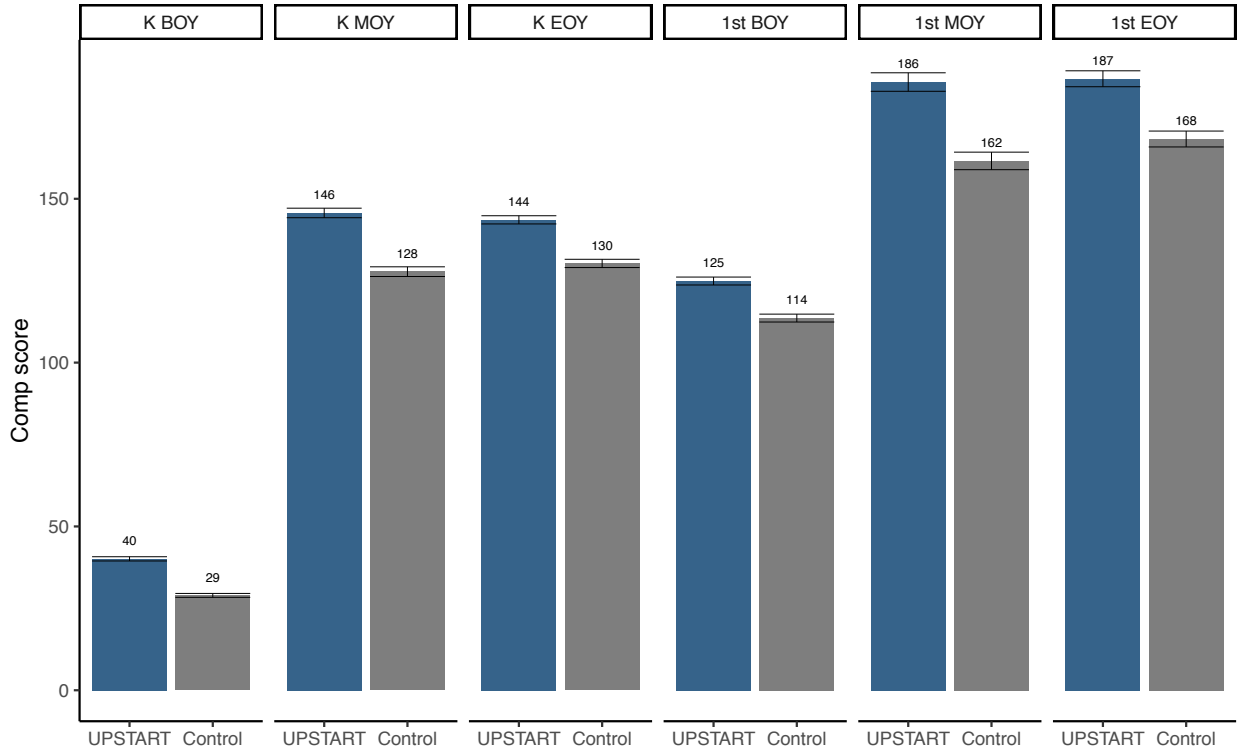
Looking at the long-term impact of preschool participation, some research points to continuing benefits of high-quality preschool experiences on cognitive outcomes into adolescence (Vandell, Belsky, Burchinal, Vandergrift, & Steinberg, 2010), while other researchers have found evidence of a "preschool fadeout" (Smith et al., 2016), with the benefits of preschool diminishing in elementary school, and in some cases as soon as by kindergarten or first grade (Puma, Bell, Cook, & Heid, 2010). A variety of factors may be involved in the convergence of preschool attendees' and non-attendees' test scores, including as schooling fails to build on the gains created by early childhood education or teachers who focus their attention on catching non-attendees up to the level of their preschool attendee counterparts (Yoshikawa et al., 2013)

Program stakeholders were interested in the long-term impact of high-quality preschool programs on students and whether program benefits present upon entry to kindergarten can be sustained once children begin elementary school. We examined whether students who participated in the UPSTART and public High-Quality preschool programs performed better in kindergarten and first grade compared to children who were not enrolled in a high-quality preschool program.

Question 7: Did the use of a home-based, computer supported literacy skill training program in preschool result in stronger school-based literacy outcomes during beginning, middle, and end of kindergarten through first grade compared to a group of peers matched terms of demographics characteristics who were not enrolled in a high-quality preschool program?

The first set of analyses looked at the impact of enrollment in the UPSTART preschool program on literacy outcomes. When compared to a group of comparison students matched on demographic characteristics, we found evidence that first grade beginning of year (BOY) Acadience Reading scores were significantly higher for children who were enrolled in the UPSTART preschool program. Specifically, UPSTART students had higher average Acadience Reading composite scores throughout kindergraden and first grade compared to the average score of control students, although only kindergarten scores were statistically signficant (**Figure 6**).

Figure 6. Acadience Reading Composite Scores for UPSTART and Control Groups



The effect sizes of the composite scores can be found in **Table 9.** An analysis of Acadience Reading composite scores at testing periods at the beginning, middle, and end of kindergarten and at the beginning, middle, and end of first grade using independent t-tests indicated that UPSTART children performed significantly higher on the Acadience Reading composite throughout kindergarten, but this trend did not continue throughout first grade when compared to a group of control children who did not participate in UPSTART.

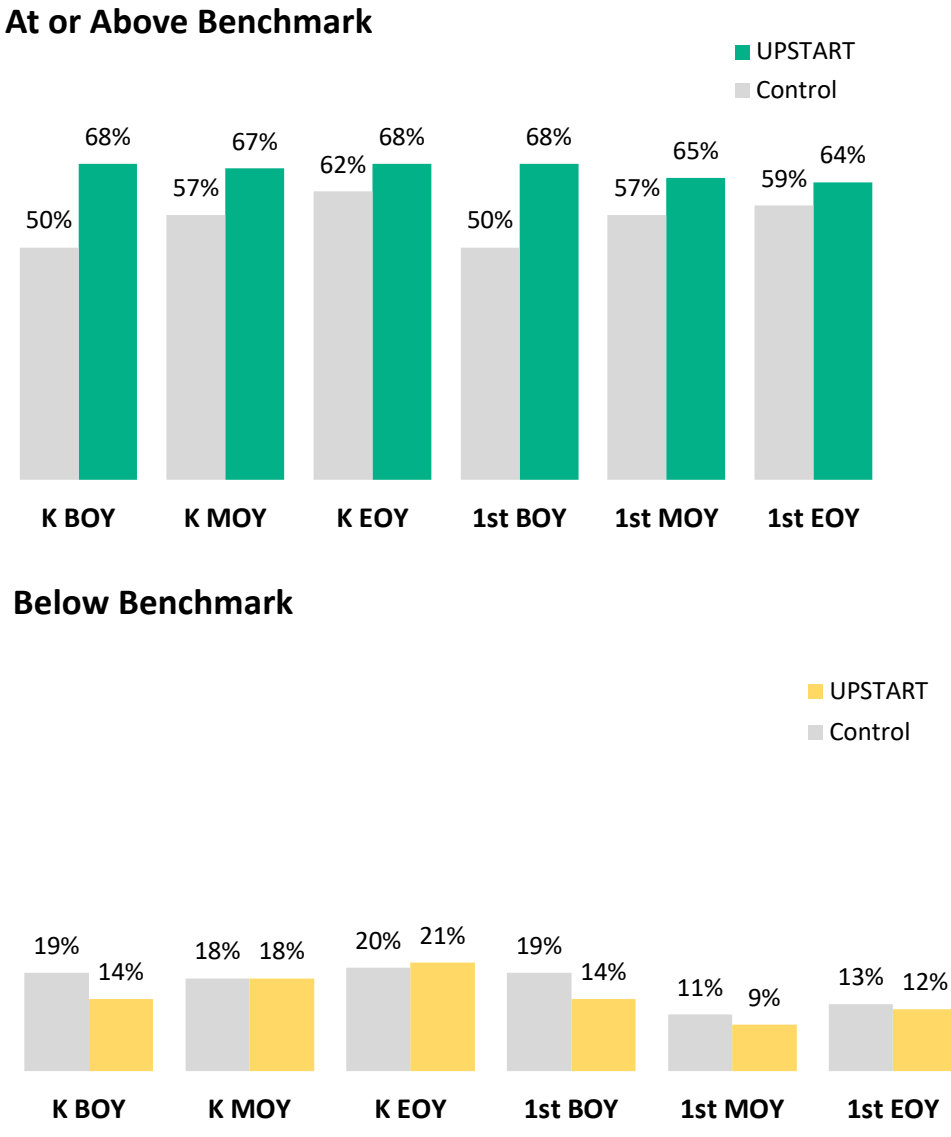
Table 9. Effect Sizes of UPSTART and Comparsion Students on Composite Scores

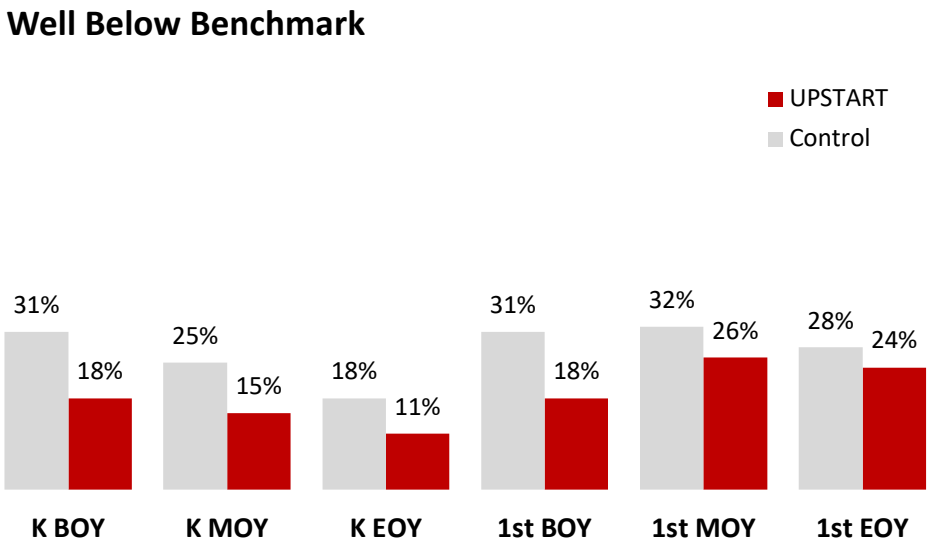
Year of Test	ES
Kindergraden BOY	0.450**
Kindergraden MOY	0.308**
Kindergraden EOY	0.269*
1st Grade BOY	NS
1st Grade MOY	NS
1st Grade EOY	NS

*$p < .05$, **$p \leq .01$, ***$p \leq .001$

Figure 7 shows the performance of children who participated in the UPSTART program with children who were not UPSTART participants on the Acadience Reading composite benchmark classifications that were measured at multiple time points in kindergarten and the beginning of first grade. Acadience Reading benchmarks were empirically derived cut points which indicate adequate reading skill for a particular grade and time of year. Benchmark categories included: at or above benchmark, below benchmark, and well below benchmark. Children who received instruction from UPSTART outperformed similar comparison students throughout kindergarten and into first grade. As seen in the **Figure 7** bar graph, UPSTART children were more likely to be classified as at or above benchmark at each assessment period than comparison students who did not participate in UPSTART and were less likely to be classified as below or well below literacy benchmarks.

Figure 7. Literacy Benchmarks Over Time: UPSTART only and Comparison Students





Question 8: Did the use of a high-quality public preschool program result in stronger school-based literacy outcomes during beginning, middle, and end of kindergarten and first grade compared to a group of peers matched in terms of demographic characteristics who were not enrolled in a high-quality public preschool program?

The second set of analyses investigated the effects of high-quality public preschools on the Acadience Reading composite benchmark classifications from kindergarten through first grade. As shown in **Table 11**, mean scores on the kindergarten and first grade Acadience Reading composite scores did not differ between the High-Quality preschool children and the control participants with nonsignificant or negligible effect sizes.

The bar graphs in **Figure 8** show the overall performance of children who participated in the High Quality preschool programs with children who did not participate in high-qulaity preschool programs on the Acadience Reading composite benchmark classifications measured throughout kindergarten and first grade.

Figure 8. Acadience Reading Composite Scores for High-Quality and Control Groups

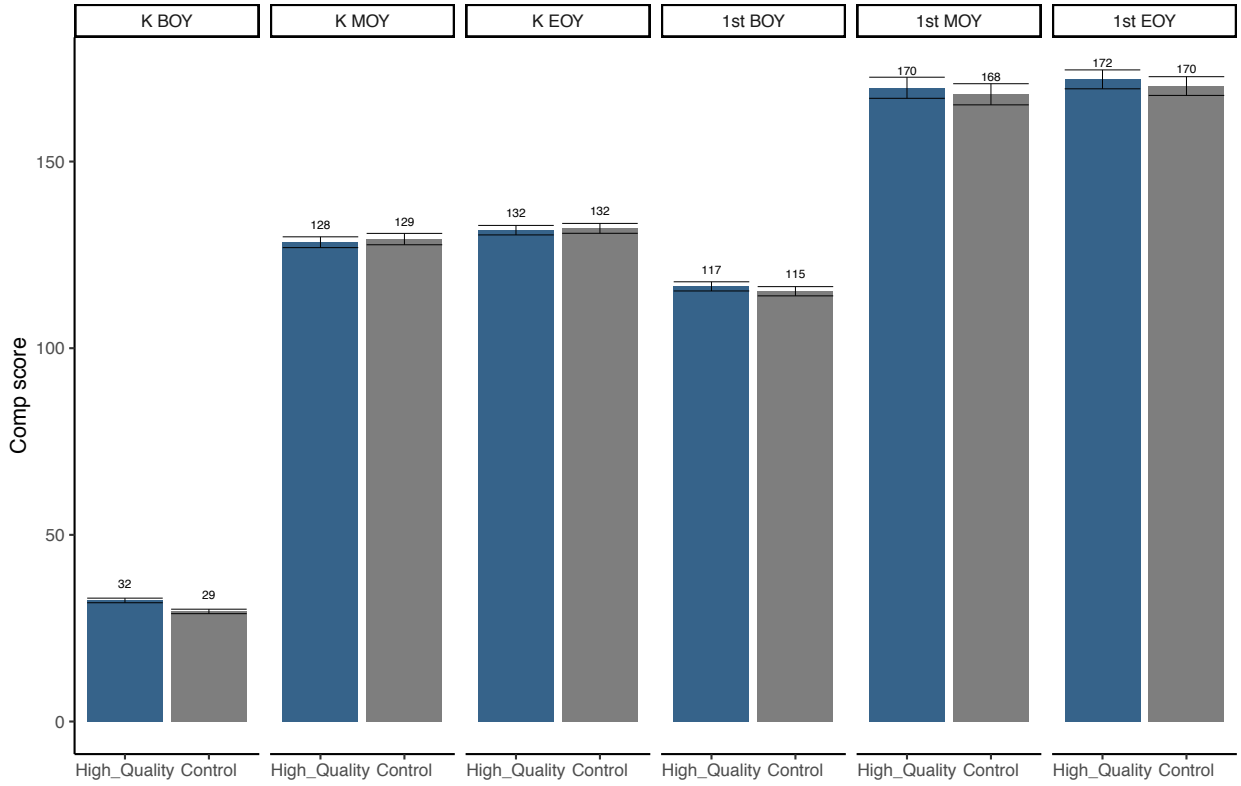
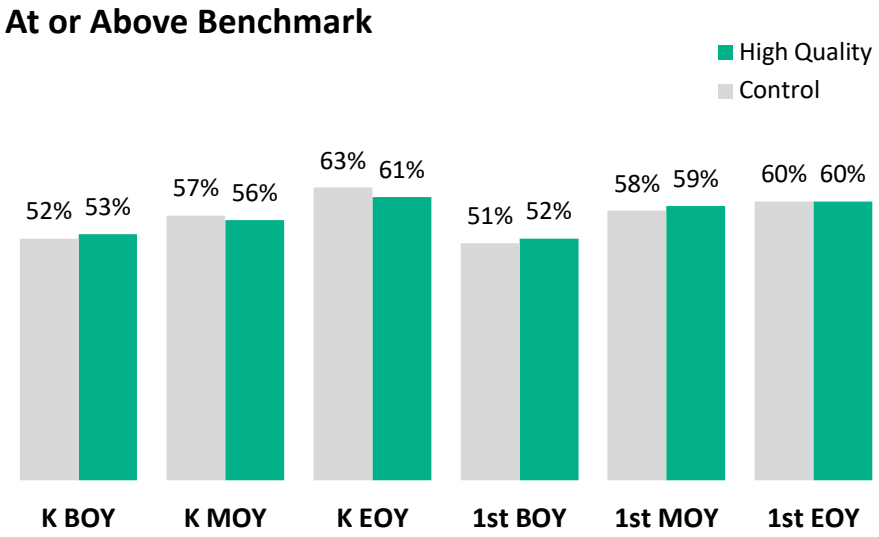
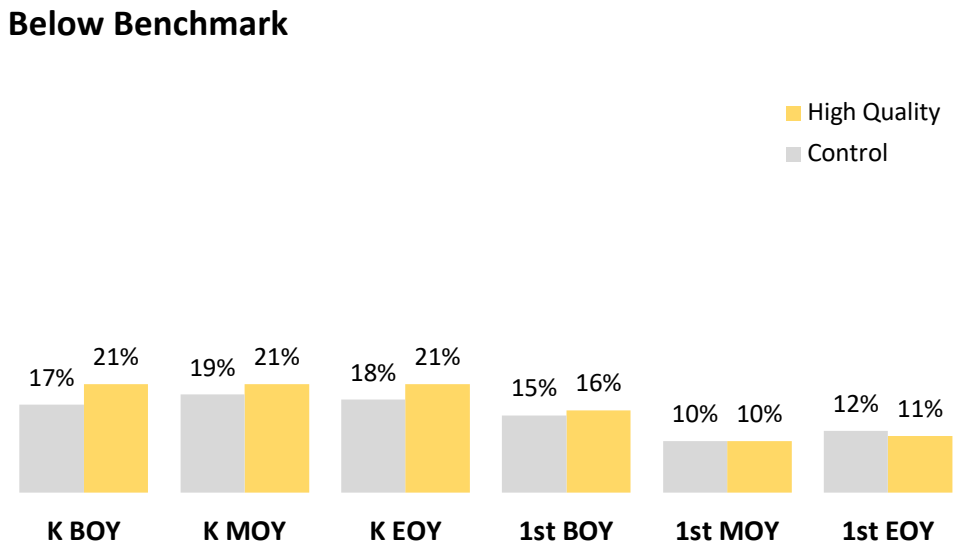
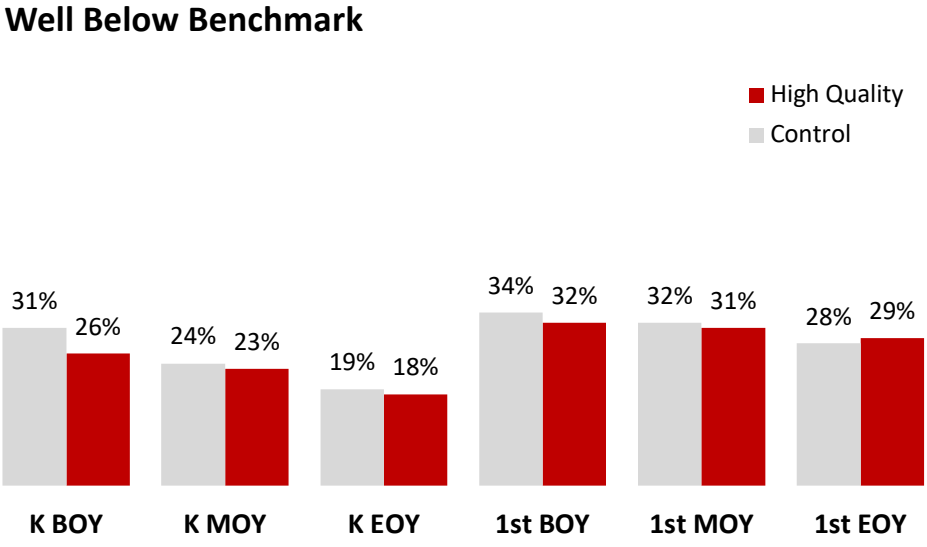


Figure 9 compares the percentage of students in high-quality public programs to students who were not enrolled in a high-quality preschool program on the Acadience Reading composite benchmark classifications. Unlike the results of the UPSTART analysis, children who were enrolled in a high-quality public preschool had similar classifications to children who were not enrolled in a high-quality preschool. As depicted in the **Figure 9** bar graph, high-quality and comparison students were more likely to be classified as at or above benchmark at each assessment period and were less likely to be classified as below or well below literacy benchmarks.

Figure 9. Literacy Benchmarks Over Time: High-Quality Public and Comparison Students







Discussion, Limitations & Recommendations

Students who participated in the UPSTART high-quality preschool program had higher scores than their matched comparison peers on measures of literacy, including global literacy, letter knowledge, and phonological awareness, but not listening comprehension or math, and there were no positive treatment impacts found for the public and private high-quality preschool programs in math. In addition, the longitudinal data from the first grade analysis indicated that benefits continued to last throughout Kindergarden and first grade for students who used UPSTART. We caution readers from interpreting the absence of significant differences between the public and private treatment and comparison student samples to mean that these programs were ineffective for encouraging students' academic growth and development. A majority of students in the comparison group were also attending preschool programs and findings should be interpreted with this in mind. In addition, although these findings were not optimal for a large part of the state's HQSR-E program, based on a comparison to nationally normed student data, **a majority of high-quality preschool program participants across all three program models entered kindergarten with literacy and social-emotional development skills that were on target with similarly aged children**.

Taken as a whole, this means that the UPSTART program has clear, strong, and lasting positive impact for specific measures of literacy, but on average all students are entering into kindergarten at age-appropriate literacy skill levels and with normal social developmental profiles. It is not surprising that a home-based computer program that focuses on early reading skills has a strong effect on specific literacy skills. The UPSTART program uses adaptive learning algorithms that are programmed to provide individualized learning and feedback to reinforce concepts. It would be difficult to have this level of focus and precision teaching reading concepts in a traditional classroom setting with multiple learners. On the other hand, traditional preschool classrooms offer rich social environments with peer and adult interactions that are important foundations for the development of the "whole child." While our study did not focus on deep measures of social and cognitive development, it is hard to argue against their importance.

When considering that our evaluation was centrally focused on early academic achievement as a predictor of school success, the discussion of unmeasured aspects of children's development is important. These unmeasured factors may be impacting outcomes. Executive function, for example, is a wide-ranging umbrella construct that pertains to the cognitive and behavioral processes that "organize and direct all cognitive activity, emotional response, and overt behavior" (Isquith, Crawford, Esby, & Gioia, 2005). Executive functioning qualities include the ability to selectively focus on a particular task (attentional control), switch focus (cognitive flexibility), hold and use front of mind information (working memory), and control impulses (inhibitory control) (Ackerman & Friedman-Krauss, 2017). It is possible that the presence or lack of executive functioning skills may be influencing outcomes as opposed to the participation in high-quality preschools. However, if we measure executive functioning in future

cohorts, we will be able to control for these skills in our statistical models and increase our confidence that they are not responsible for program effects (or lack thereof).

Evaluation Limitations

The findings should be tempered with an understanding of the study limitations. All research projects have limitations, and our study is no exception. During the third year of our study, we found limitations that should be considered when interpreting the findings, including timing of data collection, sampling sizes, outcome measures alignment, and classification of “high quality” programs.

Timing of Data Collection and sampling sizes. In many pre-posttest research designs, the pretest occurs prior to any exposure to the intervention, or program, being studied. However, this timeline is not always practical in its application, especially when it comes to measuring school-related outcomes. For example, the timing between receiving a list of program participants and when testing may begin can be affected by late school registration, staffing constraints and other challenges. It was our priority to test children as quickly as possible before they received too much exposure to the program, and we completed most of our testing by the first week of October 2018. This reduces the window of time between pretest and posttest, and could possibly diminish growth if pretest (baseline) scores are already being influenced by the program.

Outcome Measure Alignment. Our outcome measures were chosen to work across program models and represent key early childhood educational and developmental indicators of school readiness. Our measures, however, did not encompass the entire spectrum of early childhood development, and growth in executive functioning and other cognitive abilities could not be tested given the current research design and scope of work. If specific program models emphasized skill development in areas other than literacy and math (and social-emotional development), then we could not test it given the current research design. In these cases, it is possible that our measures may not have been strongly aligned with preschool program objectives.

High Quality Program Classification. High-quality preschools were identified and included in the study, and they form the basis of our treatment condition. Equally important, preschools without the classification form the basis of our control condition. In other words, an accurate classification is imperative for valid study samples. The state rated classrooms using the ECERS-3 and ECERS-E assessment tools, and, if a classroom met or exceeded pre-established criteria (scores on the measures), they were considered to meet high-quality preschool standards. These ratings, however, were passed up to the district level (or, in the case of a private preschool, to the organization level), which meant that quite a few classrooms designated as high-quality were not directly observed by the state. As high-quality status is awarded to the LEA or private provider, children in classrooms who were not directly observed by the state were included in our treatment samples (because they were considered high-quality

sites). Although it is likely that program curriculum and other factors were similar across classrooms and preschools within an LEA, it is also possible that certain classrooms or preschools may not have scored as highly on the ECERS had they been directly observed. On the otherhand, study comparison students may have been enrolled in great preschool programs that did not apply for (and were not labelled as) “high quality” status by the state. For example, approximately 80% of our comparison students were enrolled in some type of preschool, and it is possible that these students had equally rich preschool environments that stimulated their academic and social development. It would be too difficult to control for all preschool childhood experiences in our control group, so we are left with little knowledge of what unobserved – but potentially important – activities these control group children engaged in during their preschool year.

Recommendations

We recommend continuing the high-quality preschool program and advise tracking additional program implementation details that can be used to more accurately examine the potential benefits of high-quality preschool. This is especially true for the public and private preschool models, but it would also be good to develop an understanding of how the UPSTART program activities might impact development beyond academic achievement. Additional program implementation details would include specifics about program curricula (i.e. literacy skills taught, etc.), and areas of developmental focus (i.e. emphasizing free play and discovery or structured learning, for example.). In addition, we recommend adopting a whole-child measurement approach that would include instruments that focus on executive functioning and measure cognitive abilities that support goal directed behaviors (e.g., inhibitory control, attentional control, working memory, cognitive flexibility, self-regulation, and others).

Our second recommendation is to consider implementing the UPSTART program in public and private HQSR-E program sites where possible. For a third year, the evaluation findings have consistently shown that the UPSTART program positively impacts early literacy skills. While previous research has shown the value of preschool for preparing students for kindergarten (Yoshikawa et al., 2013), studying the differences between types of preschool environments are difficult due to the myriad of implementation details that are not accounted for in a study of this type. Given the complexity of comparing preschool models, in our third year (of the three-year evaluation) we are supportive of a new approach to improving high quality preschool experiences for Utah families. Traditional high-quality preschool offers a three-dimensional social landscape, but our results showed that it does not have the same positive impacts on specific literacy skills that the UPSTART program does. A hybrid high-quality preschool model that combines an adaptive, computer-based learning program with a traditional classroom-based preschool, including all the peer and adult social experiences, could give Utah families the full spectrum of benefits. Future evaluations will be needed to better understand the impact of the different preschool program models on school readiness.

Our final recommendation is to consider broadening the evaluation measures to use data from the newly implemented Utah Preschool Entry and Exit Profile (PEEP) and Kindergarten Entry and Exit Profile (KEEP). The PEEP and KEEP were not available when this evaluation was designed, and we believe it may offer additional information about HQSR-E program impacts beyond the measures of literacy, math and social-emotional development that we currently use. In addition, statewide administration of both instruments would all us to create a matched control group with a greater number of participants. As the foundational measure of school readiness in the state, we believe that using PEEP and KEEP scores alongside our established measures of early literacy and math would add another view and help to better understand the relationship between HQSR-E models and their impacts on students.

References

Bader, L. A., & Pearce, D. L. (2008). *Bader Reading and Language Inventory* (6th ed.). New York, NY: Pearson.

Bravo Aguayo, K., Abbott, M., and Kaminski, R. (2016). *Preschool Early Literacy Indicators: Time for Bed*. Eugene, OR: Dynamic Measurement group.

Bisanz, J. (2011, May). Numeracy. In *Encyclopedia of Early Childhood Development*. Retrieved September 14, 2017, from http://www.childencyclopedia.com/numeracy/introduction

Brigance, A. (2010). *Inventory of Early Development II - Standardized (IED-II)*. North Billerica, MA: Curriculum Associates.

Brigance, A. (2013). *Inventory of Early Development III - Standardized (IED-III)*. North Billerica, MA: Curriculum Associates.

Cassady, J. C., & Smith, L. L. (2004). The impact of a reading-focused integrated learning system on phonological awareness in kindergarten. *Journal of Literacy Research*, *35*, 947–964.

Concepts, L. (2015, July 22). *Norm-Referenced Test Definition*. Retrieved October 2017, from http://edglossary.org/norm-referenced-test/

Denham, S. A. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, 17, 57–89.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & ...Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446. doi:10.1037/0012-1649.43.6.1428

Early Education and Support Division. (2017). Social-Emotional Development Domain: California Infant/Toddler Learning & Development Foundations. In California Department of Education. Retrieved from http://www.cde.ca.gov/sp/cd/re/itf09socemodev.asp.

Evaluation and Training Institute (2017, October). *Utah High Quality School Readiness Expansion (HQSR-E) Program Evaluation: Cohort 1 2016-2017 Findings*. Culver City, CA: Author.

IBM Corp. Released 2013. IBM SPSS Statistics for Mac, Version 22.0. Armonk, NY: IBM Corp.

Harms, T., Clifford, R. M., & Cryer, D. (2014). Early Childhood Environment Rating Scale, third edition (ECERS-3). New York, NY: Teachers College Press.

Isquith, P. K., Crawford, J. S., Espy, K. A., & Gioia, G. A. (2005). Assessment of Executive Function in Preschool-Aged Children. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(3), 209–215. http://doi.org/10.1002/mrdd.20075

Kaminski, R.A., Abbott, M., Bravo-Aguayo, K., & Good, R.H. (2018). *Preschool Early Literacy Indicators*. Eugene, OR: Dynamic Measurement Group.

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington DC: Institute of Education Sciences.

Moore, M. R., & Hall, S. (). Listening and reading comprehension at story time: How to build habits of the mind. *Dimensions of Early Childhood*, *40*(2), 24-31.

Nugent, P. (2013). Age equivalent scale. In *PsychologyDictionary.org*, April 7, 2013, https://psychologydictionary.org/age-equivalent-scale/ (accessed October 18, 2017).

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., & Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. A consensus statement*. Washington, DC: The Brookings Institution. Retrieved from http://fpg.unc.edu/node/9046.

Shonkoff, J., & Phillips, D. (Eds.). (2000). *From neurons to neighborhoods*. Washington, DC: National Academy Press.

StataCorp (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Wood, J., & McLemore, B. (2001). Critical components in early literacy: Knowledge of the letters of the alphabet and phonics instruction. *Florida Reading Quarterly*, 38(2), 1-8. Retrieved at https://www.unf.edu/uploadedFiles/aa/fie/Woodarticle.pdf

Woolfolk, A. (2016). *Educational Psychology*. Boston, MA: Allyn & Bacon.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinoza, L. M., Gormley, W. T., Ludwig, J., Magnuson, K. A., Phillips, D., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education*. New York, NY: Foundation for Child Development.

Appendices

Appendix A. Study Measures

Participating children were administered a battery of assessments in the beginning of preschool and were re-tested at the end of preschool. We measured early reading, numeracy and social and emotional development[8] using subscales from the Brigance Inventory for Early Child Development (IED II/III) (Brigance 2010, 2013), and the Preschool Early Literacy Indicator (PELI). **Table A1** details the instrument subscales used and maps them against their respective outcome measures.

Table A1
Preschool Outcome Measures

Measure	Subtest	Phonemic Awareness	Comprehension/ Vocabulary	Alphabet & Language Concepts	Social & Emotional Development	Numeracy
PELI	Comprehension		X			
Brigance IEDII Literacy	Expressive Objects		X			
	Visual Discrimination			X		
	Auditory Discrimination			X		
	Alphabet Knowledge			X		
	Lowercase Letter Sounds			X		
	Survival Sight Words	X				
	Pre-primer Vocabulary	X				
	Experience with Books and Text (IED-III)			X		
Brigance IEDIII SE Development*	Interpersonal				X	
	Self-Regulatory				X	
Brigance IEDIII Academic Skills: Math	Count by rote					X
	Reads numerals					X
	Missing numerals & sequences					X

[8] Children's social and emotional development will be measured using a parent survey.

Appendix B. Data Collection and Analyses Methods

Scheduling and Recruitment. ETI worked with the site contacts to transfer a contact list of treatment (e.g. study participants) families who may be eligible to participate in the study. ETI's team of schedulers contacted parents about study participation, which consisted of introducing the study, its purpose and benefits, determining eligibility to participate, and scheduling families once the potential participant was deemed eligible and willing. To be eligible, treatment families were screened based on the following criteria:

- Parent must have a preschool aged child who is not entering kindergarten until 2019;
- Child must be English proficient and at least one parent must be English proficient;
- Child must not be diagnosed with a disability that could affect the testing;
- Child must not be participating in UPSTART (unless they are in the UPSTART treatment sample).

Control families were screened based on the above criteria in addition to being screened for education level (no higher than a Bachelor's degree) and income level to target control participants with similar backgrounds to our treatment sample.

We used a variety of strategies to recruit control families, such as working with school district staff; placing informational flyers in libraries, community centers, and social programs; passing out flyers and describing the study at community events; and posting announcements on Facebook parent pages and other social media sites.

Preschool Testing. Participating families were tested on-site during regular program hours and on Saturdays at nearby locations within the community by a trained assessor. At the time of their appointment, parents signed a consent form agreeing to participate in the study and then went on to complete a survey about their family and their child's social and emotional development. Preschool children were tested on measures of early literacy and math. Assessments were completed in 30 minutes, and parents received a gift card to thank them for their time. To ensure the assessment staff were adhering to evaluation protocols, senior ETI staff members conducted comprehensive trainings and observed each assessor in the field. Completed test packets were reviewed periodically for accuracy throughout the data collection period.

Appendix B. Data Collection and Analyses Methods

Cohort 3 Testing Timeline. After receiving the contact lists from participating sites in the 2018-2019 program year, ETI promptly began contacting families to schedule them to complete the testing. The computer/homebased preschool program, UPSTART, provided a contact list early in the summer and pre-test data collection was completed by July of 2018. ETI received the final list of public and private high-quality preschool

sites in May of 2018 and began data collection at the first of those sites on August 11th, 2018[9]. Sites provided ETI with contact lists from August 3rd through September 18th, 2018, and sites with earlier preschool start dates were given priority. A majority of our pre-test data had been collected by October 6th, 2018 (90% of our data).

Analysis. We conducted three types of analyses to answer our research questions: Ordinary Least Squares (OLS) regression analyses, descriptive statistics, such as frequencies and percentages, and an analysis using norm-referenced data. We describe each method below:

1. **Method:** Ordinary Least Squares (OLS) regression analyses

   Domains Studied: Literacy; Math

   Research Questions:

   - Was enrollment in the HQSR-E program associated with higher scores on measures of early literacy skills when compared to control group children?
   - Did high-quality preschool children have better early literacy skills at entry to kindergarten on specific literacy domains compared to control group children?
   - Did high-quality preschool children have better early math skills at entry to kindergarten compared to control group children?

Description: We used ordinary least squares (OLS) regression models to compare differences between each of the three treatment conditions and a control group on outcomes measuring early literacy and math, while controlling for differences in baseline scores, age in months at post-test and family characteristics (e.g. parent education level, poverty status, months in preschool). The control group of students were used in each regression model. We conducted significance[10] testing to identify findings in which we could not reject the null hypotheses: that outcomes were not a result of the treatment condition. In other words, we used this method to determine if findings were more likely to be due to the treatment effects, or due to chance occurrence (e.g. Type I error). We reported the predicted means for each group based on OLS regression results as well as effect sizes in the body of the report and/or Appendices. **Effect sizes above the .26 threshold that are statistically significant are presented in bold throughout our findings.** This threshold was created as a way to benchmark the strength of our findings against those found in similar intervention programs. The purpose and use of effect size is described in more detail in **Appendix E** to assist with interpreting the results.

2. **Method:** Norm-referenced descriptive analyses

[9] A slight revision to the final list of preschool sites was made on August 24th, 2018 after Salt Lake City Public Schools notified the State of Utah that they would no longer be participating in the HQSR-E grant.
[10] Asterisks are used to indicate statistically significant findings based on P-values within our results: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$.

Domains Studied: Literacy

Research Questions:

- How did children's level of learning change from pre-to-post compared to age specific norm group achievement levels?
- To what extent were children ready to learn, based on their literacy skill development, at entry into kindergarten?

Description: Analogous to intelligence quotients (IQ), Brigance literacy quotients divide the distribution of performance on the Brigance literacy scale into even, easily interpretable units with a mean of 100 and a standard deviation of 15. A score of 100 indicates that a child's performance on the Brigance Literacy falls at the mean of the standardized sample at his or her age level. Quotients can be interpreted qualitatively with the following category levels:

<70-89	Below average
90-110	Average
111-130+	Above average

A child that has a quotient level below average does not possess the same level of literacy skills as a similarly aged child and may require additional instructional support. Conversely, a child that is average or above enters kindergarten ready to learn, with a basic understanding of fundamental early literacy concepts.

The Brigance literacy composite scale also offers norm group referenced data, from which we can calculate if a child was performing at, above or below their normative counterparts at the age of entry into kindergarten. We used the norm-referenced data to calculate children's content knowledge age (or age equivalency score) based on their raw scores at pretest and posttest. An age equivalent score is *"used most often to average the test scores of all students at a certain age or grade level in order to determine a norm expectation of academic achievement for such a group of individuals"* (Nugent, 2013).

Children's content knowledge scores can also be used to determine to what extent a child is ready to learn as they matriculate into kindergarten. From norm group referenced data, we can calculate if a child is performing at, above or below their normative counterparts at the age of entry into kindergarten. If a child scores below the average content knowledge for their biological age, they might be at risk of struggling in kindergarten. The risk goes up for students who performed far below their age equivalent norm group comparisons. For example, if a child is performing at 90% of his/her normed achievement level, he/she is closer to the content knowledge target than a child performing only at 60 or 70%. We examined the differences in growth from pre-to-posttest among the four experimental groups and for children who we identified as at-risk (e.g. content knowledge age was 80% or below their biological age ranges).

3. **Method:** Descriptive analyses

Domains Studied: Interpersonal and Self-Regulatory Skills

Research Questions:

- What effects did the program have on the social-emotional development (SED) of program children, when compared to a group of children who did not use the program?
- To what extent did children gain SED skills?

Analyses Description: There were four categories measuring children's SED. For each SED category and experimental group, we calculated the percent of skills out of the total possible skills reported to have been developed in the child by the parent at pre-and-posttest (e.g. raw score/total possible score). This allowed us to visually identify the extent to which the children in each group had developed the appropriate SED skills to be considered ready for kindergarten and explore differences in SED between the different groups.

Appendix C. Cohort Sampling

Cohort 3 High-Quality Preschool Site Sampling
ETI received a list of High-Quality Preschool Program sites (public and private) and a separate list of UPSTART participants. ETI reviewed the public and private high-quality list and selected four public preschool providers and six private preschool providers to include in our Year Three sample. ETI selected the four public preschool providers based on the large number of participating preschoolers available to test, their proximity to each other and those that were not using UPSTART.

Table C3 A. Cohort 3 2018-2019 HQSR-E LEA's

LEA	Program Type
Cache	Public
Davis Community*	Public
Davis HOPE	Public
Duchesne	Public
Iron	Public
Jordan*	Public
Logan*	Public
Murray	Public
Sevier	Public
South Sanpete	Public
Washington	Public
Weber*	Public
CAP Head Start*	Private
Centro de la Familia	Private
Discovery Clubhouse*	Private
Hilltop Christian	Private
Mountainland HS*	Private
Oasis Montessori*	Private
Smart Kids Salt Lake*	Private
YMCA Northern Utah*	Private

*Indicate sites included in our analytic sample.

Student Sample

At the student level, ETI focused on treatment preschool aged students who were enrolled in high quality preschool sites, who were English proficient[11], and who were not identified with special needs that ETI could not accommodate during testing. Demographics for each sample are presented in **Tables C3 through C20**.

Table C3. Cohort 3 Treatment Demographics

Category	Indicator	Public (N=77)		Private (N=76)		UPSTART (N=135)	
		n	%	n	%	n	%
Race/ Ethnicity	% White	65	84%	54	71%	108	80%
	% Hispanic	14	18%	23	30%	26	19%
Gender	% Female	38	46%	35	46%	69	51%
Preschool	% Attended preschool	77	100%	73	96%	43	32%
Household	% Married	65	84%	48	63%	106	79%
	% Under 200% poverty	53	69%	46	61%	114	84%
	Average household size	5.5		4.5		5.9	
	Average household income	$57,447		$50,163		$45,727	

Bader Matched Samples

Table C4. Bader Treatment-Control Sample Demographics, high-quality UPSTART

Category	Indicator	UPSTART (N=107)		Control (N=107)	
		n	%	n	%
Race/ Ethnicity	% White	68	64%	77	72%
	% Hispanic	21	20%	21	20%
Gender	% Female	52	49%	58	54%
Preschool	% Attended preschool	33	31%	89	83%
Household	% Married	80	75%	78	73%
	% Under 200% poverty	90	84%	83	78%
	Average household size	5.66		5.55	
	Average household income	$44,418		$45,226	

[11] Although the assessments were completed in English, bilingual families with English proficiency were not excluded from the sample.

Table C5. Bader Treatment-Control Sample Demographics, high-quality Private Pre-K

Category	Indicator	Private (N=70)		Control (N=70)	
		n	%	n	%
Race/ Ethnicity	% White	38	54%	54	77%
	% Hispanic	21	30%	10	14%
Gender	% Female	31	44%	29	41%
Preschool	% Attended preschool	68	97%	58	83%
Household	% Married	45	64%	56	80%
	% Under 200% poverty	42	60%	49	70%
	Average household size	4.52		5.41	
	Average household income*	$51,572		$49,133	

Table C6. Bader Treatment-Control Sample Demographics, Public Pre-K

Category	Indicator	Public (N=69)		Control (N=69)	
		n	%	n	%
Race/ Ethnicity	% White	51	74%	52	75%
	% Hispanic	13	19%	11	16%
Gender	% Female	32	46%	43	62%
Preschool	% Attended preschool	69	100%	58	84%
Household	% Married	57	83%	63	91%
	% Under 200% poverty	48	70%	44	64%
	Average household size	5.59		5.31	
	Average household income	$56,485		$52,951	

*Note: OLS regressions controlled for difference in income between the two groups, along with other demographic variables.

Brigance Literacy Matched Samples

Table C7. Brigance Treatment-Control Sample Demographics, high-quality UPSTART

Category	Indicator	UPSTART (N=89)		Control (N=89)	
		n	%	n	%
Race/ Ethnicity	% White	62	70%	68	76%
	% Hispanic	16	18%	14	16%
Gender	% Female	43	48%	42	47%
Preschool	% Attended preschool	31	35%	76	85%
Household	% Married	67	75%	70	79%
	% Under 200% poverty	77	87%	72	81%
	Average household size	5.62		5.59	
	Average household income	$42,937		$45,293	

Table C8. Brigance Treatment-Control Sample Demographics, high-quality Private Pre-K

Category	Indicator	Private (N=63)		Control (N=63)	
		n	%	n	%
Race/ Ethnicity	% White	34	54%	51	81%
	% Hispanic	19	30%	7	11%
Gender	% Female	29	46%	29	46%
Preschool	% Attended preschool	63	100%	49	78%
Household	% Married	41	65%	52	83%
	% Under 200% poverty	41	65%	43	68%
	Average household size	4.59		5.43	
	Average household income*	$51,099		$48,949	

Table C9. Brigance Treatment-Control Sample Demographics, Public Pre-K

Category	Indicator	Public (N=63)		Control (N=63)	
		n	%	n	%
Race/ Ethnicity	% White	47	75%	47	75%
	% Hispanic	13	19%	11	16%
Gender	% Female	33	52%	39	62%
Preschool	% Attended preschool	63	100%	52	83%
Household	% Married	55	87%	55	87%
	% Under 200% poverty	43	68%	43	68%
	Average household size	5.60		5.33	
	Average household income	$57,778		$52,805	

Brigance Math Matched Samples

Table C10. Brigance Math Treatment-Control Sample Demographics, high-quality UPSTART

Category	Indicator	UPSTART (N=84)		Control (N=84)	
		n	%	n	%
Race/ Ethnicity	% White	52	62%	62	74%
	% Hispanic	17	20%	16	19%
Gender	% Female	41	49%	48	57%
Preschool	% Attended preschool	28	33%	71	85%
Household	% Married	63	75%	67	80%
	% Under 200% poverty	75	89%	71	85%
	Average household size	5.48		5.55	
	Average household income	$43,104		$44,381	

Table C11. Brigance Math Treatment-Control Sample Demographics, high-quality Private Pre-K

Category	Indicator	Private (N=60)		Control (N=60)	
		n	%	n	%
Race/ Ethnicity	% White	30	54%	43	77%
	% Hispanic	17	30%	11	20%
Gender	% Female	26	46%	27	48%
Preschool	% Attended preschool	56	100%	49	88%
Household	% Married	36	64%	46	82%
	% Under 200% poverty	33	59%	41	73%
	Average household size	4.55		5.77	
	Average household income*	$53,403		$50,333	

Table C12. Brigance Math Treatment-Control Sample Demographics, Public Pre-K

Category	Indicator	Public (N=63)		Control (N=63)	
		n	%	n	%
Race/ Ethnicity	% White	43	78%	47	85%
	% Hispanic	8	15%	5	9%
Gender	% Female	26	47%	32	58%
Preschool	% Attended preschool	55	100%	46	84%
Household	% Married	47	85%	48	87%
	% Under 200% poverty	39	71%	42	76%
	Average household size	5.57		5.43	
	Average household income	$57,611		$49,872	

PELI Matched Samples

Table C13. PELI Treatment-Control Sample Demographics, high-quality UPSTART

Category	Indicator	UPSTART (N=80)		Control (N=80)	
		n	%	n	%
Race/ Ethnicity	% White	56	70%	63	79%
	% Hispanic	11	16%	13	14%
Gender	% Female	42	53%	46	58%
Preschool	% Attended preschool	28	35%	68	85%
Household	% Married	63	79%	67	84%
	% Under 200% poverty	72	90%	69	86%
	Average household size	5.79		5.53	
	Average household income	$45,460		$44,699	

Table C14. PELI Treatment-Control Sample Demographics, high-quality Private Pre-K

Category	Indicator	Private (N=60)		Control (N=60)	
		n	%	n	%
Race/ Ethnicity	% White	33	55%	51	85%
	% Hispanic	17	28%	8	13%
Gender	% Female	27	45%	32	53%
Preschool	% Attended preschool	60	100%	47	78%
Household	% Married	38	63%	50	83%
	% Under 200% poverty	35	58%	44	73%
	Average household size	4.50		5.50	
	Average household income*	$53,003		$48,551	

Table C15. PELI Treatment-Control Sample Demographics, Public Pre-K

Category	Indicator	Public (N=63)		Control (N=63)	
		n	%	n	%
Race/ Ethnicity	% White	49	78%	51	81%
	% Hispanic	10	16%	6	10%
Gender	% Female	30	48%	39	62%
Preschool	% Attended preschool	63	100%	51	81%
Household	% Married	55	87%	51	81%
	% Under 200% poverty	45	71%	44	70%
	Average household size	5.44		5.30	
	Average household income	$57,714		$51,205	

Brigance Letter Knowledge Matched Samples

Table C16. Letter Knowledge Treatment-Control Sample Demographics, high-quality UPSTART

Category	Indicator	UPSTART (N=89)		Control (N=89)	
		n	%	n	%
Race/ Ethnicity	% White	59	64%	62	67%
	% Hispanic	19	21%	22	24%
Gender	% Female	44	48%	49	53%
Preschool	% Attended preschool	29	32%	78	85%
Household	% Married	67	73%	71	77%
	% Under 200% poverty	78	85%	71	77%
	Average household size	5.67		5.58	
	Average household income	$44,919		$44,701	

Table C17. Letter Knowledge Treatment-Control Sample Demographics, high-quality Private Pre-K

Category	Indicator	Private (N=59)		Control (N=59)	
		n	%	n	%
Race/ Ethnicity	% White	31	53%	48	81%
	% Hispanic	18	31%	5	8%
Gender	% Female	29	49%	29	49%
Preschool	% Attended preschool	59	100%	48	81%
Household	% Married	36	61%	50	85%
	% Under 200% poverty	37	63%	41	69%
	Average household size	4.54		5.32	
	Average household income*	$51,079		$49,964	

First Grade Matched Samples

Table C18. Letter Knowledge Treatment-Control Sample Demographics, Public Pre-K

Category	Indicator	Public (N=57)		Control (N=57)	
		n	%	n	%
Race/ Ethnicity	% White	43	75%	48	84%
	% Hispanic	9	16%	5	9%
Gender	% Female	28	51%	33	58%
Preschool	% Attended preschool	57	100%	50	88%
Household	% Married	50	88%	48	84%
	% Under 200% poverty	40	70%	38	67%
	Average household size	5.64		5.18	
	Average household income	$58,179		$53,377	

Table 19. First Grade Treatment-Control Sample Demographics, UPSTART

Category	Indicator	Treatment (N=1598)		Control (N=1598)	
		N	%	N	%
Race/ Ethnicity	% White	1068	67%	1068	67%
	% Hispanic	377	24%	377	24%
Gender	% Female	739	47%	739	47%
Language	% English Language Learner	241	15%	241	15%
Income	% Low income	717	45%	717	45%
Title 1 School	% Yes	566	36%	566	36%
	% Targeted for Individual Students	131	8%	131	8%

Table 20. First Grade Treatment-Control Sample Demographics, High Quality

Category	Indicator	Treatment (N=1588)		Control (N=1588)	
		n	%	n	%
Race/ Ethnicity	% White	1068	67%	1068	67%
	% Hispanic	377	24%	377	24%
Gender	% Female	740	47%	740	47%
Language	% English Language Learner	240	15%	240	15%
Income	% Low income	716	45%	716	45%
Title 1 School	% Yes	566	36%	566	36%
	% Targeted for Individual Students	132	8%	132	8%

Appendix D. Using and Interpreting Effect Sizes

It is not enough to determine which group (treatment or control) performs better on our outcome measures. We also want to show the practical significance of our findings so they are relevant to policy makers and other stakeholders. For example, if an intervention program produced a mean score that was five points higher than the mean scores generated from the control group, natural follow-up questions might include, *"Is this finding meaningful?*", or, "*What is the strength of this effect?*". Effect sizes are helpful for providing a meaningful and a more easily understood way of interpreting findings (Lipsey et. al, 2012). An effect size (ES) represents the difference between two group means on an outcome variable as standard deviation units and describes the magnitude of the difference between the two groups.

There are several ways in which effect sizes can be categorized as small, medium or large effects according to the literature. For example, Cohen's (1988) general categorization of effect sizes are small (0.2), medium (0.5), and large (0.8). Others recommend using a more targeted approach based on the average ES of similar metrics, intervention programs or interventions targeting similar groups of students (Lipsey et. al, 2012). For the purposes of this report, we used the second approach and consider any effect size above .26 higher than the average effect size seen in similar education evaluations. To arrive at this benchmark, we took an average of the average effect sizes reported for similar interventions using a report from the Institute of Education Sciences (IES), in which researchers reviewed 829 effect sizes from 124 education research studies conducted on K-12 students (Lipsey et. al, 2012).

Appendix E. HQSR-E Detailed Findings

High-quality Public Preschool Results

Table E1. Bader Analysis of Treatment-Control Differences

Experimental Group	N	Treatment mean	Control mean
UPSTART	107	4.383	2.206
Private	70	2.114	3.157
Public	69	2.130	2.493

Table E1a. Bader OLS Model of Treatment-Control Differences, UPSTART

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Bader_Pre	0.58	0.21	2.729	159.3	1	7.445	0.007	0.372
Treatment	2.82	0.73	3.840	315.6	1	14.747	0.001	0.523
Months in Preschool	0.14	0.08	1.786	68.3	1	3.190	0.075	0.243
Income	0.00	0.00	1.336	38.2	1	1.785	0.183	0.182
Education	0.67	0.62	1.076	24.8	1	1.158	0.283	0.147

Residual standard error: 4.626 on 196 degrees of freedom
(12 observations deleted due to missingness)
Multiple R-squared: 0.1264, Adjusted R-squared: 0.1041
F-statistic: 5.671 on 5 and 196 DF, p-value: 6.573e-05, Hedges's g: 0.32
Bader_post~WF: 0.462

Table E1b. Bader OLS Model of Treatment-Control Differences, <u>High-quality Private Preschool</u>

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Bader_Pre	0.98	1.11	0.883	15.09	1	0.78	0.379	0.149
Treatment	-0.95	0.78	-1.21	28.36	1	1.467	0.228	0.204
Months in Preschool	-0.03	0.12	-0.215	0.90	1	0.046	0.830	0.169
Income	-0.00	0.00	-0.567	6.90	1	0.357	0.551	0.169
Education	0.93	0.48	1.948	73.39	1	3.794	0.054	0.169

Residual standard error: 4.398 on 130 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared: 0.04877, Adjusted R-squared: 0.01218
F-statistic: 1.333 on 5 and 130 DF, p-value: 0.2543, Hedges's g: 0.194
Bader_post~Private: -0.238

Table E1c. Bader OLS Model of Treatment-Control Differences, High-quality Public Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Bader_Pre	0.739	0.455	1.625	48.17	1	2.640	0.107	0.275
Treatment	-0.259	0.769	-0.337	2.07	1	0.114	0.737	0.170
Months in Preschool	-0.06	0.129	-0.495	4.46	1	0.245	0.622	0.170
Income	0.00	0.00	0.228	0.95	1	0.052	0.820	0.170
Education	0.149	0.488	0.307	1.72	1	0.094	0.760	0.170

Residual standard error: 4.272 on 130 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.02658, Adjusted R-squared: -0.01086
F-statistic: 0.7098 on 5 and 130 DF, p-value: 0.6171, Hedges's g: 0.143
Bader_post~Public: -0.085

Table E2. Brigance Literacy Analysis of Treatment-Control Differences

Experimental Group	N	Treatment mean	Control mean
UPSTART	89	106.12	90.49
Private	63	57.48	59.00
Public	63	99.87	96.30

Table E2a. Brigance Literacy OLS Model of Treatment-Control Differences, UPSTART

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance Literacy_Pre	0.792	0.09	7.950	72802	1	63.21	0.001	1.185
Treatment	20.05	5.86	3.423	13497	1	11.72	0.001	0.511
Months in Preschool	0.939	0.63	1.489	2553	1	2.22	0.138	0.222
Income	0.00	0.00	1.003	1158	1	1.01	0.318	0.150
Education	7.129	4.77	1.496	2578	1	2.24	0.137	0.223

Residual standard error: 33.94 on 166 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared: 0.3501, Adjusted R-squared: 0.3305
F-statistic: 17.88 on 5 and 166 DF, p-value: 3.65e-14, Hedges's g: 0.63
Brig_post~WF: 0.383

Table E2b. Brigance Literacy OLS Model of Treatment-Control Differences, High-quality Private Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance Literacy_Pre	0.058	0.128	4.547	24.18	1	20.677	0.001	0.805
Treatment	0.06	6.43	0.009	0	1	0.000	0.993	0.002
Months in Preschool	1.747	1.086	1.608	3022	1	2.585	0.111	0.179
Income	0.00	0.00	-0.559	365	1	0.312	0.577	0.178
Education	6.905	4.203	1.643	3155	1	2.699	0.103	0.179

Residual standard error: 34.19 on 120 degrees of freedom
Multiple R-squared: 0.2182, Adjusted R-squared: 0.1856
F-statistic: 6.698 on 5 and 120 DF, p-value: 1.544e-05, Hedges's g: 0.458
Brig_post~Public: 0.093

Table E2c. Brigance Literacy OLS Model of Treatment-Control Differences, High-quality Public Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance Literacy_Pre	0.739	0.455	1.625	48.17	1	2.640	0.107	0.275
Treatment	-0.259	0.769	-0.337	2.07	1	0.114	0.737	0.170
Months in Preschool	-0.06	0.129	-0.495	4.46	1	0.245	0.622	0.170
Income	0.00	0.00	0.228	0.95	1	0.052	0.820	0.170
Education	0.149	0.488	0.307	1.72	1	0.094	0.760	0.170

Residual standard error: 4.272 on 130 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.02658, Adjusted R-squared: -0.01086
F-statistic: 0.7098 on 5 and 130 DF, p-value: 0.6171, Hedges's g: 0.143
Bader_post~Public: -0.085

Table E3. Brigance Math Analysis of Treatment-Control Differences

Experimental Group	N	Treatment mean	Control mean
UPSTART	84	23.69	23.57
Private	56	22.20	21.98
Public	55	22.51	22.25

Table E3a. Brigance Math OLS Model of Treatment-Control Differences, UPSTART

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance Math_Pre	0.792	0.09	7.950	72802	1	63.21	0.001	1.185
Treatment	20.05	5.86	3.423	13497	1	11.72	0.001	0.511
Months in Preschool	0.939	0.63	1.489	2553	1	2.22	0.138	0.222
Income	0.00	0.00	1.003	1158	1	1.01	0.318	0.150
Education	7.129	4.77	1.496	2578	1	2.24	0.137	0.223

Residual standard error: 33.94 on 166 degrees of freedom
(6 observations deleted due to missingness)
Multiple R-squared: 0.3501, Adjusted R-squared: 0.3305
F-statistic: 17.88 on 5 and 166 DF, p-value: 3.65e-14, Hedges's g: 0.63
Brig_post~WF: 0.383

Table E3b. Brigance Math OLS Model of Treatment-Control Differences, High-quality Private Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance Math_Pre	0.608	0.087	7.017	2657.6	1	49.233	0.001	1.317
Treatment	1.427	1.461	0.977	51.5	1	0.954	0.331	0.183
Months in Preschool	-0.107	0.231	-0.464	11.6	1	0.216	0.643	0.087
Income	0.00	0.00	0.799	34.5	1	0.639	0.429	0.150
Education	3.020	1.194	2.530	345.4	1	6.398	0.013	0.475

Residual standard error: 7.347 on 104 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.4393, Adjusted R-squared: 0.4123
F-statistic: 16.29 on 5 and 104 DF, p-value: 7.594e-12, Hedges's g: 0.758
Brig _Math _post~Private: 0.022

Table E3c. Brigance Math OLS Model of Treatment-Control Differences, High-quality Public Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance Math_Pre	0.492	0.109	4.528	1540.5	1	20.50	0.001	0.857
Treatment	0.397	1.813	0.219	3.6	1	0.048	0.827	0.042
Months in Preschool	-0.169	0.304	-0.557	23.3	1	0.310	0.579	0.105
Income	-0.00	0.00	-1.064	85	1	1.131	0.290	0.201
Education	1.653	1.484	1.114	93.3	1	1.241	0.268	0.211

Residual standard error: 8.669 on 102 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.2155, Adjusted R-squared: 0.1771

F-statistic: 5.605 on 5 and 102 DF, p-value: 0.0001324, Hedges's g: 0.448
Brig _Math _post~Public: 0. 026

Table E4. Brigance Letter Comprehension Analysis of Treatment-Control Differences

Experimental Group	N	Treatment mean	Control mean
UPSTART	89	49.44	40.47
Private	59	44.00	47.51
Public	57	48.33	44.42

Table E4a. Brigance Letter Comprehension OLS Model of Treatment-Control Differences, UPSTART

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance LetterComp_Pre	0.754	0.112	6.750	38461	1	45.566	0.001	1.01
Treatment	15.75	5.056	3.114	8186	1	9.698	0.002	0.465
Months in Preschool	1.234	0.546	2.263	4321	1	5.120	0.025	0.338
Income	0.00	0.00	0.452	173	1	0.205	0.652	0.068
Education	6.440	3.931	1.638	2265	1	2.684	0.103	0.245

Residual standard error: 29.05 on 167 degrees of freedom
(11 observations deleted due to missingness)
Multiple R-squared: 0.2895, Adjusted R-squared: 0.2682
F-statistic: 13.61 on 5 and 167 DF, p-value: 3.843e-11, Hedges's g: 0.551
LetterComp _post~WF: 0.270

Table E4b. Brigance Letter Comprehension OLS Model of Treatment-Control Differences, High-quality Private Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance LetterComp_Pre	0.832	0.147	5.669	28336	1	32.136	0.001	1.037
Treatment	4.727	5.738	-0.824	598	1	0.679	0.412	0.151
Months in Preschool	1.133	0.821	1.380	1678	1	1.903	0.171	0.252
Income	0.00	0.00	0.599	316	1	0.358	0.551	0.110
Education	3.865	3.916	0.987	859	1	0.974	0.326	0.181

Residual standard error: 29.69 on 110 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.2925, Adjusted R-squared: 0.2604
F-statistic: 9.097 on 5 and 110 DF, p-value: 2.898e-07, Hedges's g: 0.552
LetterComp_post~Private: -0.100

Table E4c. Brigance Letter Comprehension OLS Model of Treatment-Control Differences, High-quality Public Preschool

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Brigance LetterComp_Pre	0.502	0.152	3.307	10678	1	10.939	0.0012	0.615
Treatment	0.517	6.180	0.084	7	1	0.007	0.934	0.016
Months in Preschool	2.507	1.123	2.233	4867	1	4.986	0.028	0.415
Income	-0.00	0.00	-0.875	748	1	0.766	0.383	0.163
Education	9.012	4.495	2.005	3923	1	4.019	0.048	0.373

Residual standard error: 31.24 on 106 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.19, Adjusted R-squared: 0.1518
F-statistic: 4.974 on 5 and 106 DF, p-value: 0.000394, Hedges's g: 0.415
LetterComp_post~Public: 0.115

Table E5. PELI Analysis of Treatment-Control Differences

Experimental Group	N	Treatment mean	Control mean
UPSTART	80	18.18	17.69
Private	60	16.73	18.52
Public	63	17.75	17.73

Table E5a. PELI OLS Model of Treatment-Control Differences, UPSTART

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Peli_Pre	0.244	0.055	4.442	113.77	1	19.729	0.001	0.699
Treatment	0.325	0.429	0.758	3.31	1	0.574	0.450	0.119
Months in Preschool	-0.036	0.046	-0.763	3.36	1	0.582	0.447	0.120
Income	0.00	0.00	0.496	1.42	1	0.246	0.620	0.078
Education	0.158	0.399	0.397	0.91	1	0.157	0.692	0.062

Residual standard error: 2.401 on 154 degrees of freedom
Multiple R-squared: 0.148, Adjusted R-squared: 0.1203
F-statistic: 5.348 on 5 and 154 DF, p-value: 0.0001461, Hedges's g: 364
Peli_post~WF: 0.189

Table E5b. **PELI OLS Model of Treatment-Control Differences, High-quality Private Preschool**

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Peli_Pre	0.382	0.085	4.493	115.96	1	20.19	0.001	0.815
Treatment	-1.900	0.470	-4.041	93.80	1	16.33	0.001	0.733
Months in Preschool	5.626	0.065	0.870	4.35	1	0.757	0.386	0.158
Income	5.314	0.00	0.671	2.59	1	0.451	0.504	0.122
Education	0.729	0.321	2.276	29.75	1	5.179	0.025	0.413

Residual standard error: 2.397 on 112 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared: 0.3451, Adjusted R-squared: 0.3158
F-statistic: 11.8 on 5 and 112 DF, p-value: 3.529e-09, Hedges's g: 0.623
Peli_post~Private: 0.644

Table E5c. **PELI OLS Model of Treatment-Control Differences, High-quality Public Preschool**

Variable	Estimate	SE	t	ss	Df	F	p	H's g
Peli_Pre	0.124	0.102	1.219	12.23	1	1.485	0.225	0.216
Treatment	0.125	0.546	0.212	0.43	1	0.052	0.819	0.040
Months in Preschool	0.072	0.087	0.826	5.62	1	0.683	0.410	0.146
Income	-0.00	0.00	-0.212	0.37	1	0.045	0.832	0.038
Education	0.258	0.373	0.692	3.94	1	0.479	0.490	0.123

Residual standard error: 2.869 on 120 degrees of freedom
Multiple R-squared: 0.02411, Adjusted R-squared: -0.01655
F-statistic: 0.5929 on 5 and 120 DF, p-value: 0.7054, Hedges's g: 0.136
Peli_post~Public:0.006

Evaluation and Training Institute
100 Corporate Pointe, Suite 387
Culver City, CA 90230
www.eticonsulting.org

For more information about the
Evaluation and Training Institute, contact ETI:

Jon Hobbs, Ph.D., President
Phone: 310-473 8367
jhobbs@eticonsulting.org