EDUCATION

THE UTAH STATE BOARD OF EDUCATION

Report to the Education Interim Committee

Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) Report

October 2018

Sarah Young
Coordinator for Digital Teaching and Learning
sarah.young@schools.utah.gov

Diana Suddreth
Director of Teaching and Learning
diana.suddreth@schools.utah.gov

Darin Nielsen
Assistant Superintendent of Student Learning
darin.nielsen@schools.utah.gov

Patty Norman
Deputy Superintendent of Student Achievement
patty.norman@schools.utah.gov

STATUTORY REQUIREMENT

U.C.A. Section 53F-4-407 requires the State Board of Education to make a report on UPSTART to the Education Interim Committee by November 30 each year. The State Board is required to contract with an independent evaluator to evaluate the program. Reporting on the program shall include the (i) the number of families participating in the program; (ii) the number of families requesting and furnished computers; (iii) frequency of software usage; (iv) obstacles encountered with software usage, hardware, or providing technical assistance to families; (v) student performance on assessments as detailed in statute; and (vi) any other information that is part of the independent evaluation.

Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) Report

EXECUTIVE SUMMARY

During the 2016-2017 school year, Cohort 8 participated in the Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) program. The UPSTART program uses a home-based educational technology approach to develop the school readiness of preschool children. The program is designed to give Utah four-year-olds an individualized reading, mathematics, and science curriculum with a focus on reading. Children participate in the program the year before they attend kindergarten. The UPSTART program is administered by the Waterford Institute. A total of 10,745 preschool students participated in Year 8 of the program. Students in Cohort 8 used the UPSTART program for an average of 56 hours during the program year. Students who were UPSTART graduates used the program for an average of 61 hours. The independent evaluation for Cohort 8 of the program is attached.

UPSTART Program Evaluation

Year 8 Program Results

Submitted to the Utah State Board of Education
February 2018

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

All correspondence should be directed to:
Jon Hobbs, Ph.D.
jhobbs@eticonsulting.org

Table of Contents

Executive Summary

Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) is a home-based computer preschool program developed and provided by the Waterford Institute to prepare young children for school entry and future academic success. The Evaluation and Training Institute (ETI), has prepared this report for the Utah State Board of Education (USBE) to document UPSTART's impact in its eighth year of implementation (Cohort 8, with students enrolled during the 2016-2017 program year). ETI responded to feedback and guidance from the UPSTART Advisory Committee (UAC), and this year we augmented our research design and scope to meet a higher level of accountability for the program and explore longer-term aspects of UPSTART by reporting on three different areas:

- The **Cohort 8 Evaluation** presents outcome and implementation results for Cohort 8 using a pre-test/post-test design with a statistically matched control group in order to assess the program's impact on developing children's early literacy skills. Our research findings cover two areas: (1) how the program was implemented and (2) what types of impacts the program has on children's literacy.
- The **Cross-Cohort Analysis** presents results for two areas of interests to the UAC: students who enroll in UPSTART but fail to graduate, and students who participate in UPSTART while also attending traditional preschool.
- The **First Grade Analysis** presents findings on UPSTART's continued impact on students' literacy achievement once children enter the elementary school setting. Using statewide data, we analyzed whether achievement gains from UPSTART that occur prior to school entry are sustained through kindergarten and into first grade.

This Executive Summary presents a summary of findings for each reporting area, along with selected recommendations for improving the program and future evaluation efforts.

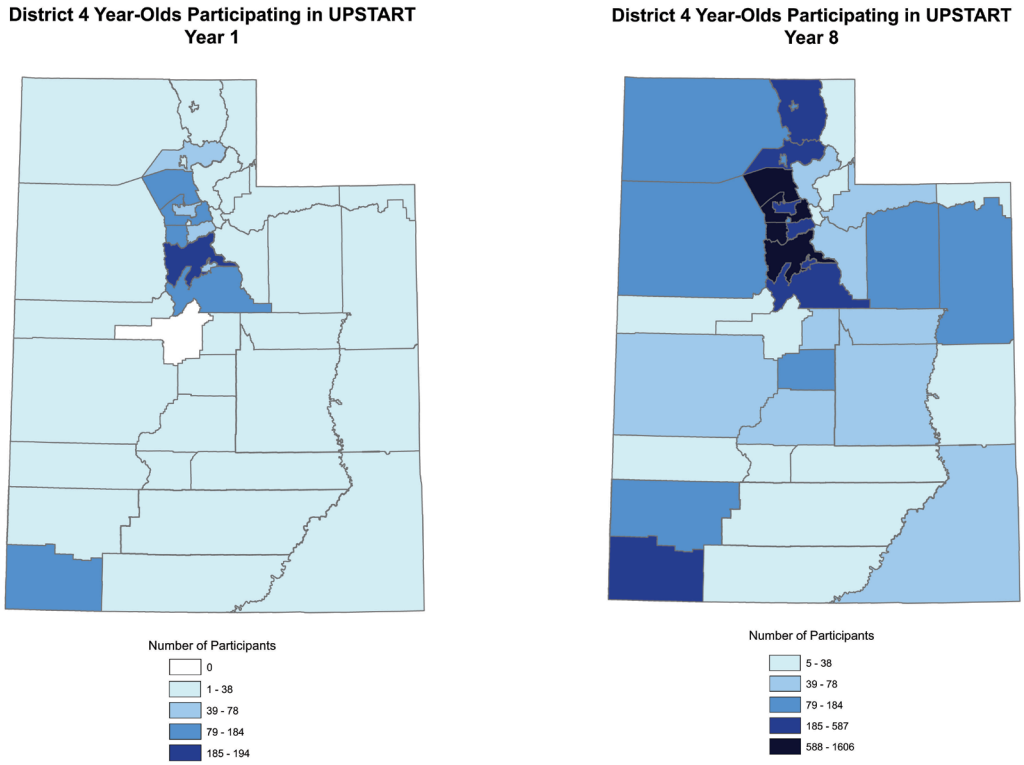
Cohort 8 Evaluation

Program Implementation

The 2016-17 program year was a continued expansion of UPSTART enrollment, as the number of preschool students participating in the program in Year 8 (N = 10,745) grew by 4,106 students from the previous year (Year 7, N = 6,639), a 62 percent increase[1]. Over the past eight years, UPSTART program participation has increased, and the program has enrolled families in urban and rural areas throughout the state of Utah. The maps depicted below showcase UPSTART program participation by school district form the inception of the program (Year 1) to the most recent program year (Year 8).

[1] UPSTART participant enrollment and program usage data used to generate program implementation findings was provided to ETI by the Waterford Institute.

Maps of UPSTART program participation in Year 1 and Year 8 by School District



Fifty percent of children enrolled in UPSTART Cohort 8 lived in families with incomes less than 200% of the federal poverty level and the majority of enrolled children were White (79%) and English speaking (90%). UPSTART enrollment increased from 6,639 children in Year 7 to 10,745 children in Year 8, an increase of 62 percent, while graduation rates reversed their downward trend (seen in Year 7) and increased from 87% to 89%.

Findings about UPSTART usage in Cohort 8 are summarized below:

- Students who used the program for the recommended amount of time (or longer) had better reading outcomes than their matched counterparts who did not use the program.
- Students in Cohort 8 used the UPSTART program for an average of 56 hours during the program year. Students who were UPSTART graduates used the program for an average of 61 hours.
- Students in Cohort 8 had an 89% graduation rate, which reverses a trend of lower graduation rates year-to-year starting in Cohort 5 (which had a graduation rate of 94%), and continued in Cohort 6 (graduation rate of 92%) and Cohort 7 (graduation rate of 87%).

- Children who did not graduate were more likely to have parents with lower levels of education, speak a language other than English, be members of an underrepresented racial or minority group, have parents who were not married, and have higher levels of household poverty than children who graduated and completed the UPSTART program.
- A positive relationship was found between UPSTART curriculum use and evaluation outcomes: as program use increased, students' scores on literacy achievement measures increased.

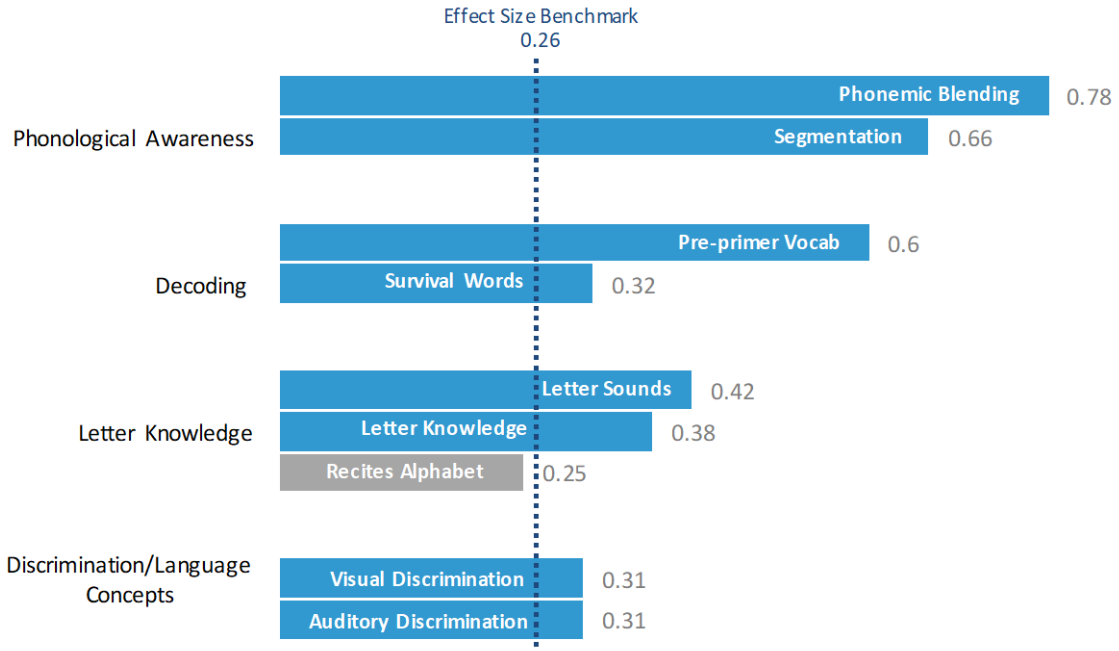
Impacts on Literacy

We present effect sizes throughout our reporting to provide additional context for our findings. An effect size (ES) takes the difference between two group means on an outcome variable and represents it in standard deviation units. Effect sizes describe the magnitude of the difference between two groups, and essentially create a standardized scale to facilitate results interpretation. Following recommendations from the What Works Clearinghouse (WWC) (What Works Clearinghouse, 2017) and a meta-analysis of similar educational interventions and studies (Lipsey et al., 2012), we set an effect size threshold of .26 to denote effects that have practical significance and are substantively important.

UPSTART had a strong impact on children's emerging literacy skills based on results from effect size and growth score analyses. Children enrolled in UPSTART produced significant positive effects (ES = .50) compared to control children on the Brigance composite, an instrument that measures decoding skills, letter knowledge, vocabulary and syntax, and pre-literacy discrimination. Similarly, UPSTART participants experienced medium effects (ES = .81) on the Bader composite, an instrument that assesses children's phonological awareness. The graph below presents effect sizes by literacy construct and provides a line marker to highlight effect sizes that fall above the predetermined threshold (.26 or higher) to showcase their practical significance.

UPSTART had significant impact on a majority of early literacy domains

Effect Size Estimates by Literacy Construct



Phonological awareness has been identified as one of the most important predictors of reading success and involves a child's facility with the sound structure of words (Phelps, 2003). Phonological skills include the ability to identify rhyming words, isolate a sound in a word, blend individual sounds, and detect word alliteration. Children's **phonological awareness** abilities were significantly improved because of their UPSTART participation.

- UPSTART students had significantly higher phonemic blending skills (ES = .78) and phoneme segmenting skills (ES = .64).
- Compared to control children, students participating in UPSTART had significantly higher increases from the pre-test to the post-test on both phonological awareness subscales (blending and segmenting).

UPSTART had a significant impact on children's **word decoding** skills. Decoding, a core reading skill that is a precursor to reading fluency, is the ability to accurately identify individual printed words. Accurate decoding results from the successful acquisition of several key pre-literacy skills, including a child's ability to recognize written letters, discern letters that correspond to phonological sounds, and blend word sounds into the generation of a single word.

- Children participating in UPSTART had significantly higher post-test scores on decoding pre-primer vocabulary words (ES = .60) and reading survival sight words (ES = .32).
- UPSTART children had stronger growth scores on reading pre-primer vocabulary words (e.g., "can", "and", "do") and survival sight words (e.g., "go", "stop", "out") compared to children who were not enrolled in the program.

Students who participated in UPSTART experienced a moderate improvement in their **letter knowledge** skills. The letter is the most basic unit of reading and familiarity with the letters of the alphabet has been shown to be a strong predictor of reading achievement. Additionally, understanding the connection between written letters and the sounds of speech is a precursor to decoding.

- UPSTART children had small to medium effects in their learning how to recite (ES = .25), identify (ES = .38), and sound out (ES = .42) letters of the alphabet.
- Compared to control students, UPSTART participants showed significantly stronger growth rates in learning how to pronounce letter sounds.

Before children can read, they need to be able to visually distinguish between shapes, letters, and words, even if they do not fully comprehend what letters represent. Similarly, children should be able to differentiate between spoken words (e.g., “fit” versus “fat”) before comprehending written words. UPSTART participants showed a moderate impact on **pre-literacy discrimination and language concepts**.

- UPSTART had a medium effect on children’s ability to discriminate between different shapes, letters, and words (ES = .31) as well as a small to medium effect on their ability to distinguish if two words sound the same (ES = .31).
- Children in UPSTART had stronger growth scores on their auditory discrimination of words when contrasted to children not enrolled in UPSTART.

The UPSTART program did not have a significant impact on children’s **vocabulary**:

- UPSTART did not have significant effects on receptive vocabulary.
- Children enrolled in UPSTART did not have significantly different growth rates on vocabulary subscales when compared to control children.

Cross-Cohort Analysis

An analysis of UPSTART students across past cohorts provided us with a sufficient sample size to look at the impact of UPSTART on subgroups of interest, namely UPSTART program graduates versus non-graduates (program graduates fulfill predetermined UPSTART usage requirements while non-graduates fail to meet program usage milestones), and students who enroll in a traditional preschool program while participating in UPSTART versus students who participate only in UPSTART. A review of findings are included below.

Impact of Graduation Status

Students who graduated from UPSTART and met usage requirements for consistency and duration had significantly higher post-test scores on the overall Brigance Composite and on four of the eight Brigance subtests compared to children who did not graduate from the program. Significant mean post-test differences between graduates and non-graduates had effect sizes that ranged from .29 to .58, indicating substantively important effects with practical significance.

Additionally, students who met all criteria for graduation showed stronger literacy growth rates from pre-test to post-test compared to non-graduates on two Brigance subtests (Survival Sight Words and Basic Vocabulary), and on the overall Brigance composite score.

Impact of Traditional Preschool

There were few observed differences between students who attended a traditional home- or center-based preschool while participating in UPSTART and students who were only enrolled in UPSTART. There was no difference between UPSTART only and UPSTART and preschool students on the Brigance Composite post-test. There were significant post-test differences on two of the Brigance subtests, but only one subtest (Letter Knowledge) had a difference that exceeded the .26 effect size threshold for practical significance.

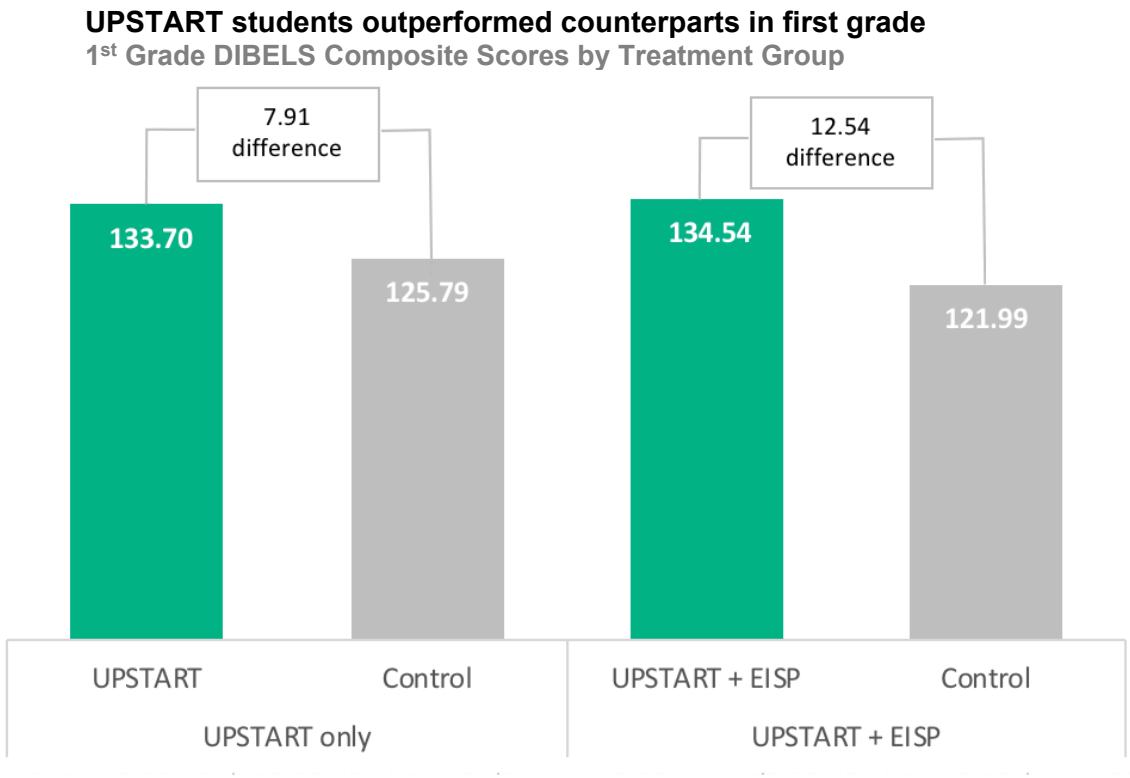
There were no significant differences in the literacy growth rates from pre-test to post-test between UPSTART only students and students who participated in UPSTART while attending a traditional preschool on the Brigance Composite or on any subtest.

First Grade Analysis

In order to determine whether or not UPSTART has a sustained benefit on children's literacy once they enter elementary school, UPSTART students and their counterparts who did not have any UPSTART experience were followed through kindergarten and first grade. To conduct this analysis, we had to address potentially confounding effects from the Early Software Intervention Program (EISP), a statewide computer-based literacy instruction software program available in grades K-3. To control for the impacts of EISP, we excluded any student who participated in EISP as a kindergartener from our control group. We utilized a post-test only design to determine if UPSTART participants had higher scores on the first grade DIBELS literacy assessment compared to similar comparison students. In an effort to isolate the effects of participating in EISP, we excluded any student who participated in EISP as a kindergartener from our control group. In addition, to give the state information about potential multiplier effects, we created two treatment groups: students who only participated in UPSTART during their preschool year (UPSTART only) and students who participated in UPSTART as preschoolers and who participated in the EISP program as kindergarteners (UPSTART + EISP).

Findings show that, on average, students who used the UPSTART program in preschool scored 7.91 points higher than comparison students on the DIBELS composite at the beginning of first grade. This difference was statistically significant and produced an effect size of .18.

Students who received continuous treatment from preschool through kindergarten did even better: Our findings show that the use of UPSTART + EISP had more of an impact on first grade reading than the UPSTART preschool program alone. Students with combined treatment in preschool and first grade scored 12.54 points higher than a group of matched comparison students, a statistically significant mean score advantage over their non-program peers that produced a .27 effect size (which is above our .26 effect size criteria to show practical significance for literacy achievement).



Summary and Recommendations

The UPSTART program continues to show success in helping preschool aged children develop literacy skills in preparation for their entry into kindergarten, and new analyses suggest that UPSTART has benefits that last into elementary school.

Program graduates had significantly higher scores on literacy outcome measures compared to children who did not meet program usage requirements. Children who did not graduate from UPSTART were more likely than program graduates to reside in households below the poverty level, have parents with lower levels of education, and be English learners – the ideal target population for UPSTART and the children that stand to benefit most from the program. Cohort 8 showed an increase in graduation rates from 87% to 89%, which is even more noteworthy considering that the overall numbers of UPSTART participants increased by 62%. Monitoring of program use and graduation requirements needs to be continual to be sure that UPSTART is being administered with fidelity so that all children can receive the full program dosage as recommended by the vendor and obtain the full cognitive benefits of the program.

Due to the strong impact on early literacy development, we recommend that the state continue to provide the UPSTART program to children. Given the importance of graduation on literacy achievement outcomes, we recommend that the program vendor continue to work with the evaluator and USBE staff to monitor program implementation carefully and be sure the trend towards higher graduation rates continues. Specifically, we recommend that the program vendor consider the following recommendations:

- The program vendor could develop new strategies for addressing falling usage and graduation rates among the most at-risk students (i.e. those with high levels of poverty and with English as a second language). Some potential strategies might include:
    - Establishing peer support systems among similar groups to discuss strategies for supporting children's program use.
    - Highlighting evaluation information that links graduation with higher literacy outcomes.
    - Developing targeted incentives for families with the highest risk factors for not meeting program usage requirements, such as monthly awards (extrinsic), being highlighted in UPSTART communications to social networks as "Gold Star Families" (intrinsic).

The fact that UPSTART children maintained their advantage over their comparison counterparts through first grade can be construed as another important benefit of the UPSTART program. Findings from the first grade analysis indicate a continued effect of the UPSTART preschool program, and this effect is increased with continued individualized computer-based literacy instruction throughout kindergarten (i.e. in conjunction with EISP). As the UPSTART program expands to reach more Utah preschoolers across the state, we recommend that USOE continues the EISP program to provide individualized instruction that builds on the gains created by UPSTART.

Introduction

The Utah State Board of Education (USBE) hired the Evaluation and Training Institute (ETI), a non-profit research and consulting firm, to conduct a multi-year evaluation of the UPSTART program to determine the effectiveness of the home-based preschool program in academically preparing children for school success.

The 2016-2017 UPSTART program year saw the program's scale increased to reach more families than in any previous cohort to date. As the program scaled-up, the evaluation had to be adapted to accommodate larger numbers of program students and the higher stakes as a result of the greater resource allocation for the program. While the scale and stakes increased, our research objectives remained constant: we continue to evaluate the program's *impact on developing children's early literacy skills* to help the state and stakeholders determine the benefits from participating in the program.

We enhanced the established evaluation design and reporting in three key ways to meet a higher level of accountability for the program, and to ensure that the program resources were having a positive impact on school readiness and beyond.

In the **Cohort 8 Evaluation**, we present outcome results for UPSTART's eighth and largest year of implementation, hereafter referred to as Cohort 8 (C8). Additionally, we document the extent to which participants used the computerized curriculum as it was intended, establish the relationship between curriculum usage and literacy outcomes, and report the program's completion or "graduation" rate. As in our evaluations with recent cohorts, the Cohort 8 evaluation included a statistically balanced match of treatment and control students. While requiring a larger sample size, the matching process enhances our ability to detect treatment effects and, in general, improve the accuracy of the evaluation results.

Second, we added a **Cross-Cohort Analysis** to the evaluation that analyzed historical data collected from past UPSTART students in Cohorts 2-7. This archival data provides a sample size that is sufficient enough to examine the impact of UPSTART on various subgroups of interest, such as children who did not meet UPSTART usage requirements or children who were enrolled in a traditional preschool setting and UPSTART.

Third, in addition to determining the impact of the UPSTART program on students' school readiness prior to the beginning of kindergarten, this report will also present findings on UPSTART's continued impact on student literacy with kindergarten and first grade literacy scores in the **First Grade Analysis**. This longitudinal study meets the provision in state law to evaluate the long-term impacts of UPSTART on students and uses DIBELS literacy data collected in schools from over 37,000 students to determine whether or not UPSTART has a lasting impact on student literacy achievement.

Each of these analyses is presented in separate sections of the report, along with an overall summary and suggestions for program recommendations. We begin with a brief overview of the UPSTART preschool program.

UPSTART Program Description

Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) is a project established by the Utah state legislature that uses a home-based education technology approach to develop the school readiness skills of preschool children. In its eighth year of operation during the 2016-17 school year, the project's implementation contractor – the Waterford Institute – enrolled 10,745 preschool children and provided them with an adaptive program of computer-based early literacy instruction to prepare them academically for kindergarten. The 10,745 children enrolled in the eighth-year cohort, hereafter referred to as Cohort 8 (C8), participated in UPSTART from September 2016 through June 2017. Cohort 8 is the largest group since the program's rollout.

The UPSTART software uses adaptive lessons, digital books, animated songs, and activities to deliver individualized early literacy content. The reading skills taught by the Waterford Early Learning Program at Level 1 of the curriculum[2] include:

- Phonological Awareness
- Phonics
- Comprehension and Vocabulary
- Language Concepts

Children are encouraged to use the UPSTART program for 15 minutes a day, 5 days a week and families are provided with parental resources and technical support from Waterford customer service representatives.

[2] Level One is the beginning point of the curriculum where the preschool child begins as a nonreader and is introduced to skills designed to teach the child to read.

Cohort 8 Evaluation

Research Questions

Our evaluation of the Cohort 8 of UPSTART users is framed by research questions. We hypothesized that if UPSTART has no effect on improving early literacy skills, then the preschool children who participated in UPSTART – the treatment group – would be expected to perform at the same level as a comparison control group (children who were not exposed to UPSTART) on post-test measures of early literacy development at the beginning of Kindergarten. If UPSTART does have an effect on improving early literacy, then the treatment group should perform significantly better than the control group on the post-test at the beginning of Kindergarten.

For purposes of triangulation, we also wanted to take a slightly different look at the data by examining growth rates from pre-test to post-test. If UPSTART shows stronger literacy growth rates, then the treatment group would be expected to show greater gain scores (post-test score minus pre-test score) relative to the comparison group on the various literacy subtests and total test scores.

With respect to concerns for school readiness, our research questions for the C8 evaluation were as follows:

> **Research Question 1.1:** Do UPSTART students have better early literacy skills at kindergarten compared to control group students?
>
> **Research Question 1.2:** Do UPSTART students show stronger literacy growth rates from preschool to kindergarten compared to control group students?

In the impact analysis, the outcomes of interest were measures of early literacy skills relevant to emerging readers such as phonological awareness, letter recognition, letter sound knowledge, and vocabulary development. Results for research questions 1.1 and 1.2 are presented in the **UPSTART Program Impacts on Literacy** section of the report.

The Utah State Board of Education (USBE) and the Utah State Legislature were also interested in outcomes related to the implementation of UPSTART. Research questions along this line included:

> **Research Question 1.3:** What was the extent of UPSTART curriculum usage in terms of the amount of exposure per participant, as measured in minutes or hours of instruction per week?
>
> **Research Question 1.4:** What percent of the participants completed the full implementation program (i.e., “graduated” as defined by the Waterford Institute)?
>
> **Research Question 1.5:** How does the level of UPSTART curriculum usage relate to reading readiness outcomes?

Data for research questions 1.3 and 1.4 were obtained from records maintained by the Waterford Institute and are answered in this report by descriptive statistics. The answer to research question 1.5 was derived from the relationship between exposure to the computer-assisted program of instruction (measured by program records documenting

minutes of computer usage for each enrolled student) and the measured literacy outcomes of interest. Results for research questions 1.3 through 1.5 are presented in the **UPSTART Program Implementation** section of the report.

Methods

The following section presents information about the research methods used to conduct the evaluation, including: the research design, creation of treatment (UPSTART students) and control (non-UPSTART students) samples, outcome measures, and ETI's data collection and analyses procedures.

Research Design

To evaluate the impact of the UPSTART program, we collected literacy data for a "treatment group" of UPSTART participants and a comparison "control group" of students who did not participate in the program. We collected pre-test and post-test data on children in each group over a 12-month interval during the year prior to enrollment in Kindergarten. Due to the legislative mandate that all children interested in enrolling in the program be allowed to participate, children could not be randomly assigned to groups, which resulted in a "quasi-experimental research design" as diagrammed below:

		Year 1		**Year 2**	
Non-Random Assignment	Treatment	Pre-Test	**UPSTART**	Post-Test	Kindergarten
	Control	Pre-Test		Post-Test	

The use of both a pre-test and a comparison group facilitated our ability to examine potential threats to validity, which could jeopardize a clear interpretation of the results (Shadish, Cook, & Campbell, 2002). Because students could not be randomly assigned to treatment or control groups, the groups began as nonequivalent by definition, and consequently selection bias could be assumed to operate to some degree in some manner. The pre-test allowed us to examine the potential for selection bias by determining the nature of the bias as well as its size and direction (i.e., which group is favored over the other by a particular inequality).

C8 Evaluation Samples

The C8 evaluation moved from a using an unmatched group seen in previous years to a new approach first adopted in the C6 evaluation that uses a statistically matched control group balanced across meaningful variables that contribute to achievement outcomes. Simply put, using a matching process to develop our treatment and control groups is a stronger method for ruling out the influence of preexisting differences between groups on program outcomes.

A matched treatment-control group is made by statistically matching control students to certain characteristics of treatment students to make two equal or "balanced" groups across a set of important predictor variables. With the appropriate resources, the matching process creates groups that are equivalent before any treatment effects are taken into account. To do this, however, students who are not matched one-to-one must be removed from the final research sample. The process depends on having a sufficiently large enough subject pool to draw from for both treatment and, especially, control students.

ETI's methods for generating the matched sample is described in more detail below.

Data Collection

The C8 study consists of data from 581 preschool children: 230 treatment group children who had enrolled in UPSTART for Year 8 of the program (the 2016-17 school year) and 351 nonparticipating control group children. The children were not randomly assigned to the treatment or control groups.

Treatment children. The 230 UPSTART children came from an initial random sample of C8 UPSTART enrollees whose families were contacted about participating in the C8 evaluation[3]. Because the legislation extending the UPSTART program gave participation priority to low-income families and non-native English speakers (Utah Code: 53A-1a-1001), we similarly prioritized recruiting low-income families in our treatment sample. The recruited 230 UPSTART children participated in pre-testing prior to entering the program over the summer of 2015 and post-tests were conducted the following year upon the conclusion of the program and before children entered kindergarten.

Control children. Data from control children consisted of panel data collected from non-UPSTART participants. The control children were recruited using a variety of strategies, including targeting preschools, daycare centers, childcare organizations, Head Start centers, parent groups, low-income housing units, and snowball sampling[4] from families who were UPSTART users.

Because the treatment and control groups were not created through random assignment, it was assumed that the two groups would be nonequivalent on factors that may influence literacy skills. Therefore, it is important to review the treatment and control demographics and pre-test scores carefully to statistically adjust for any imbalances so that accurate and fair comparisons can be made.

Table 1 presents key demographic characteristics for the unmatched treatment and control sample. As shown in **Table 1**, control families were somewhat more advantaged compared to treatment families from the standpoint of parental education and household income level.

[3] C8 treatment families were screened based on location, parental education, income level, child language, and known disabilities.

[4] Snowball sampling is when existing participants recruit future participants among their personal network of acquaintances.

Table 1
Unmatched Treatment-Control Comparisons on Key Demographics

Demographic Categories		Treatment (N=275)	Control (N=556)
Gender	Female	49%	53%
	Male	51%	47%
Ethnicity	Caucasian	89%***	78%
	Hispanic	13%	11%
Child Language	English	97%*	94%
Parent Education Level	High School Diploma	19%	16%
	Some College	65%***	46%
	Bachelor's degree	12%	27%***
Parent Marital Status	Married	88%**	80%
Household Income	Under $10,000	10%	10%
	$10k-$24,999	1%	9%
	$25k-$49,999	44%***	29%
	$50k-$74,999	29%	23%
	$75k-$99,999	7%	14%**
	$100k or more	4%	10%***
	No response	6%	5%

*p<.05, **p≤.01, ***p≤.001

There were significant differences between the two unmatched groups on family ethnicity, education level, and language spoken at home. Studies of child development have found that parents with higher levels of education spend more time with their children in ways likely to enhance their development, hold higher expectations for their children, and use varied and complex language and speech patterns (Davis-Kean, 2005; Guryan et al., 2008; Neitzel & Stright, 2004). In light of these findings, it is important to ensure that the treatment and control groups are as comparable as possible with regard to parental education before analysis or that statistical adjustments are performed to determine any impact of family characteristics on post-test literacy outcomes.

Significant differences between the treatment and control groups that favored the control group were found on both the Brigance and Bader pre-test literacy instruments. While the use of a pre-test and covariates with the full unmatched sample allows us to examine and statistically control for pre-existing literacy skills and demographic differences between the treatment and control groups, using these control methods can reduce our ability to detect treatment effects and to estimate their size. We determined that using a matched treatment and control group strategy that took into account Brigance pre-test performance and key demographic characteristics would further reduce the chance that pre-existing differences influenced our ability to statistically test for treatment effects.

Matched Treatment-Control Group Sample

To combat the limitations (cited above) of using the full unmatched C8 sample, we used a statistical process called “Coarsened Exact Matching” (CEM) to match control students to treatment students. During the CEM procedure, each treatment child is statistically matched with a control child who is most similar to them and if no matches can be made, children are removed from the sample. Additional tests are preformed to assess the balance between the treatment and control group to ensure that the groups are as similar as possible. The resulting matched treatment-control sample consists of treatment children who have a statistical control “twin”. Using CEM, we were able to construct a comparison group of control children that resembled the treatment sample as closely as possible on specific observable characteristics, such as gender, race/ethnicity, language, parental education, and performance on pre-test measures.

The CEM procedure consisted of a three-step process:

1. The C8 unmatched evaluation sample contained data from 275 treatment students from C8 and 556 comparison students who did not participate in the UPSTART program.
2. Students from the pool of potential controls were then matched to treatment students using CEM, which found an exact match—or twin—for treatment students from the group of control students in terms of:
    - Sex (Female/Male)
    - Ethnicity (White or Hispanic),
    - Language
    - Parent Education
    - Brigance Composite pre-test scores
3. Statistical tests assessed the balance between treatment and control group to ensure groups were as similar as possible at baseline (pre-test).

The matching process resulted in a data file with comparable students in each group so that we could improve our precision in estimating treatment effects. **Table 2** displays the demographic breakdown of the matched treatment and control groups. Note how the two groups in the matched sample are much more similar in terms of parental education and race than in the unmatched sample.

There is, however, some remaining imbalances between treatment control groups. Significant differences existed in marital status, with more treatment children having married parents and more control children having divorced parents, which could be a factor depressing control outcomes. On the other hand, a factor that can be presumed to improve outcomes—family income—favored the control group, with significantly more control families with higher household incomes in the upper two brackets ranging from $75,000 to over $100,000 per year. Using household income in the matching process resulted in a drastically smaller sample size which would limit our ability to test for significant effects, so we elected to retain the slight imbalance and evaluate the impact of household income during the statistical analysis process.

Table 2
Matched Treatment-Control Comparisons on Key Demographics

Demographic Categories		Treatment (N=245)	Control (N=245)
Child Gender	Female	50%	52%
	Male	50%	48%
Child Ethnicity	Caucasian	91%	91%
	Hispanic	10%	10%
Child Language	English	98%	97%
Parent Education Level	High School Diploma	18%	14%
	Some College	66%	66%
	Bachelor's degree	12%	12%
Parent Marital Status	Married	88%	87%
Household Income	Under $10,000	10%	8%
	$10k-$24,999	1%	10%***
	$25k-$49,999	46%**	29%
	$50k-$74,999	31%	26%
	$75k-$99,999	8%	13%*
	$100k or more	4%	8%*

*p<.05, **p≤.01, ***p≤.001

Outcome Measures

The reading skills taught by the Waterford Early Learning Program at Level 1 of the curriculum[5] include:

- Phonological Awareness: phonemic segmenting and blending
- Phonics: letter name knowledge, letter sound knowledge, and word reading
- Comprehension and Vocabulary: vocabulary knowledge and oral comprehension
- Language Concepts: concepts of written language from letters and pictures to basic grammar

The outcomes of interest for the UPSTART evaluation are measures of early literacy skills that are **aligned to the UPSTART curriculum and considered to be important predictors of later reading ability**, such as phonological awareness, letter knowledge, and vocabulary. In order to measure these outcomes in our treatment and control groups, we used appropriate subscales from two standardized measures of early literacy, the Brigance Inventory of Educational Development and the Bader Reading and Language Inventory.

The Brigance. The Brigance Inventory of Educational Development (Brigance, 2014) was selected as an early literacy measure of phonics and vocabulary knowledge and as a measure of pre-Kindergarten academic and cognitive skills. Ten scales were administered from the language development and academic/cognitive domains of the Brigance. Brigance subscales measured the literacy constructs of *vocabulary*, *pre-literacy discrimination*, *letter knowledge*, and *decoding* and are described in detail in **Table 3**. A composite Brigance score to create a comprehensive score of early literacy

[5] Level 1 of the UPSTART curriculum is the beginning point of the curriculum where the preschool child begins as a nonreader and is introduced to skills designed to teach the child to read.

achievement was created by adding the scores from the ten subtests. Possible scores on the Brigance composite range from a low of 0 points to a high of 240 points.

The Bader. The Bader Reading and Language Inventory (Bader, 2008) was selected as a measure of *phonological awareness*. Phonological awareness involves the child's ability to detect the sound structure of spoken words at three levels: rhyming, syllables, and phonemes. The Bader is comprised of phonological awareness subtests (phonemic blending and phoneme segmentation), along with a composite summary phonological awareness score that was calculated by adding the scores from the subtests.

Relevance of Outcome Measures. As stated previously, we selected our outcome measures based on their alignment to the UPSTART curriculum and on their ability to assess early literacy skills that are demonstrated predictors of reading success. Each outcome measure evaluates a key domain or construct of early literacy: pre-literacy discrimination, letter knowledge, phonological awareness, decoding, and vocabulary. These five constructs are explained in further detail below.

> *Pre-Literacy Discrimination*. Before children can read or even comprehend the meaning of letters, they need to be able to visually discriminate between letter shapes. For example, if a child is unable to visually distinguish a "p" from a "b", she will incorrectly identify letters and their letter sounds. Similarly, children need to be able to discriminate between the sounds of words (e.g., "cat" from "can") to facilitate listening comprehension and to match letter and word sounds with their printed versions.
>
> *Phonological Awareness*. Phonological awareness has been identified as one of the most important predictors of reading success and involves a child's facility with the sound structure of words (Phelps, 2003). Phonological skills include the ability to identify rhyming words, isolate a sound in a given word, blend individual sounds to produce a single, and detect word alliteration. We assessed the phonological awareness with two subscales from the Bader: phoneme segmentation and phoneme blending.
>
> *Letter Knowledge*. Letters are the most basic unit of reading and familiarity with the alphabet and ability to recognize letters and their corresponding sounds is a prerequisite for decoding. Letter knowledge begins with being able to identify lower and uppercase letters in a variety of fonts, but also includes understanding the representational nature of letters and connecting printed letters with their phonemic sounds. Letter knowledge is evaluated in the current study by assessing children's ability to recite the alphabet, identify lowercase letters by name, and connect lowercase letters with their sounds.
>
> *Decoding*. Decoding is the process of translating printed words into speech and is the precursor to reading fluency, the ability to read text accurately and quickly, either aloud or silent. Decoding relies on the successful acquisition of all the aforementioned reading skills, phonological awareness, letter knowledge, and pre-literacy discrimination. We measured decoding in the UPSTART study by asking children to read lists of simple pre-primer vocabulary (e.g., "and", "can",

"go", "look") and presenting them with words they might have seen in their everyday lives (e.g., "stop", "in", "out").

Vocabulary. Vocabulary has been demonstrated to be a reliable predictor of later reading scores (Snow, Burns, & Griffin, 1998) and is necessary for making meaning of written and oral language. Children's vocabulary is measured by an expressive vocabulary test where they provide names to a series of pictures.

Table 3 summarizes the alignment between the UPSTART curriculum and the literacy constructs measured by the Brigance and Bader, and also contains information about specific skills assessed by the Brigance and Bader subscales, along with possible scale ranges.

Table 3
Alignment of Outcome Measures with UPSTART Curriculum

UPSTART Curriculum	Literacy Construct	Instrument Subscale	Measured Skill	Possible Range
Language Concepts	Pre-literacy Discrimination	Auditory Discrimination	Identifies if two words sound the same	0-10
		Visual Discrimination	Identifies similarities and differences between forms, letters, and words	0-20
Comprehension/ Vocabulary	Vocabulary and Syntax	Expressive Vocabulary	Names pictures	0-27
Phonics I	Letter Knowledge	Recites Alphabet	Recites alphabet	0-26
		Lowercase Letter Knowledge	Names or recognizes lowercase letters	0-52
		Sounds of Lowercase Letters	Produces sounds of lowercase letters	0-26
Phonological Awareness	Phonological Awareness	Phonemic Blending	Blends separate word sounds into single word	0-8
		Phoneme Segmentation	Segments word into separate word sounds	0-8
Phonics II	Decoding	Survival Sight Words	Reads survival sight words that appear in public places	0-16
		Pre-Primer Vocabulary	Reads basic vocabulary words found in pre-primer reading programs	0-24

Data Collection Procedures

Data were collected for 275 treatment group children who had enrolled in UPSTART for Year 8 of the program and 556 control group children who had not enrolled in the UPSTART program. The children's parents were given an intake questionnaire during the pre-test session that collected demographic information from children, parents, and the household. The children were post-tested on the Brigance and Bader a year later before entering kindergarten.

A student data file was developed based on data collected from the intake questionnaire and from the pre-test and post-test administrations of the Brigance and Bader. The final analysis file consisted of data from 490 children, 245 treatment and 245 control, and was based on the subset of children with valid matched pre-test and post-test data, and who

had not previously used the UPSTART computerized learning program as documented through the pre-screening interview.

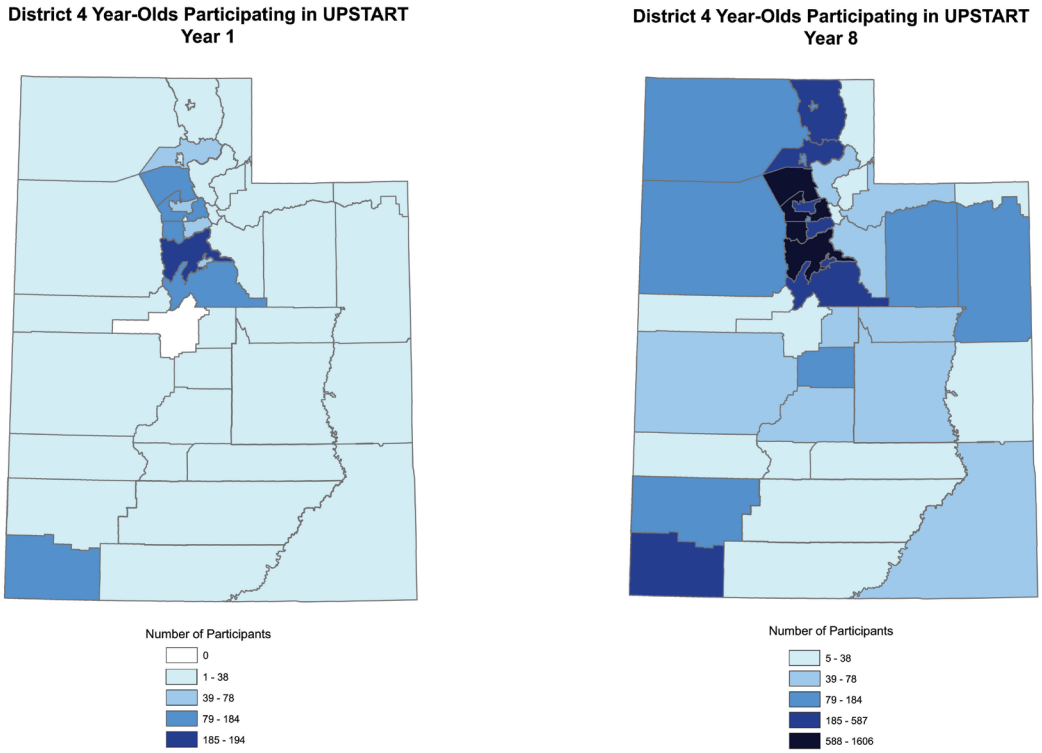
UPSTART Program Implementation

Findings reviewed in the UPSTART implementation section include eighth year enrollment, equipment provided to enrolled families by UPSTART, usage of the UPSTART curriculum in terms of instructional time logged, the proportion of UPSTART students considered to have "graduated" from the program, and the relationship between levels of UPSTART curriculum usage and literacy outcomes.

UPSTART Enrollment

The 2016-17 program year was a continued expansion of UPSTART enrollment, as the number of preschool students participating in the program in Year 8 (N=10,745) grew by 4,106 students from the previous year (Year 7, N=6,639), a 62 percent increase. Since the inception of the program, the number of students enrolled in the program rose from 1,248 children in Year 1 to 10,745 students in Year 8, an increase of over 761 percent. The maps depicted in **Figure 1** showcase UPSTART program participation by school district from the inception of the program (Year 1, N=1,248) to the most recent Year 8 (Year 8, N=10,745). As seen below in **Figure 1**, the UPSTART program has continued to further its reach over the past eight years and has increased enrollment in both urban and rural areas of the state.

Figure 1
Maps of UPSTART program participation in Year 1 and Year 8 by School District



The Waterford Institute provided a comprehensive dataset to ETI for the eighth-year UPSTART enrollment of 10,745 children, including demographic information, provisioned educational technology, UPSTART program usage, and whether or not children completed program requirements. This provisioned data was analyzed by ETI to generate the findings related to program implementation.

Some basic demographic characteristics of the C8 population are presented below in **Table 4**.

Table 4
Demographic Characteristics of C8 Population

Demographic Categories		All C8 UPSTART (N=10,745)
Child's Gender	Male	51%
	Female	49%
Child's Ethnicity	White	79%
	Hispanic	13%
	Asian/Pacific Islander	3%
	African American	1%
	Native American	<1%
	Other	3%
Child's Language	English	90%
	Spanish	9%
	Other	1%
Parent Educational Attainment	Some High School	4%
	High School Graduate	13%
	Some College	36%
	College Graduate	37%
	Advanced Degree	9%
Parent Marital Status	Married	91%
	Otherwise	9%
Household Poverty Level	Below 100%	18%
	Below 185%	42%
	Below 200%	50%

Note: Percentages may not add to 100% due to rounding. Data is from Waterford participant records.

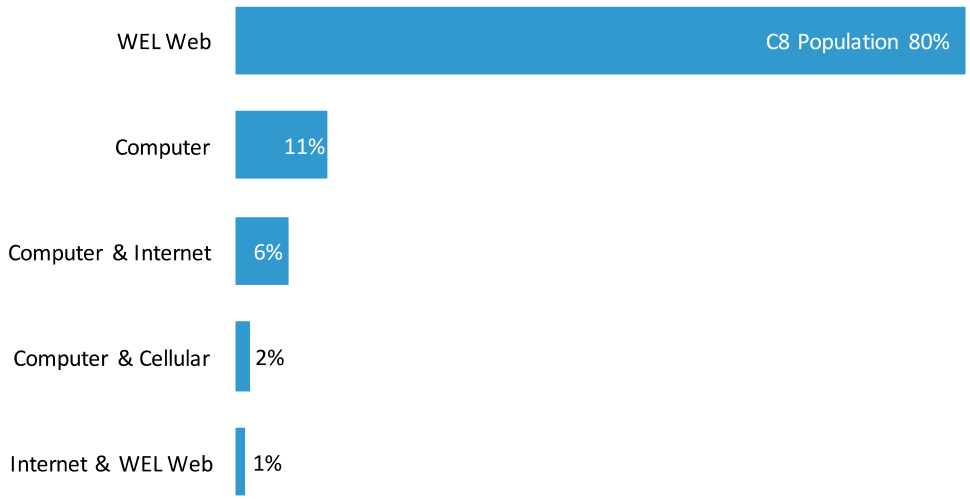
Slightly more C8 boys (51%) were enrolled than girls (49%) and in terms of ethnicity, the majority (79%) of the C8 enrollment was White, with 13% of the children being of Hispanic origin. Forty-two percent of the C8 UPSTART participants lived in families with incomes less than 185% of the federal poverty level.[6]

[6] The federal poverty definition consists of a series of thresholds based on family size. In 2017, a 100% poverty threshold for a family of four was $24,600, while a 185% threshold for a family of four was $45,510 (see U.S. Department of Health and Human Services poverty guidelines at https://aspe.hhs.gov/2017-poverty-guidelines).

Provided UPSTART Equipment

The type of education technology provided to UPSTART children in Year 8 of the program is shown in **Figure 2** for all 10,745 children enrolled in the program. The majority of UPSTART children (80%) used the Waterford website to retrieve the UPSTART program. This allowed families to access the UPSTART curriculum from their home computers.

Figure 2. Equipment provided to C8 Participants by Waterford



*Note: Percentages may not add to 100% due to rounding.

Second most frequently, UPSTART provided free personal computers to 11% of the C8 children while they participated in the program. Another 6% of the C8 program participants were provided with internet access and personal computers. The remaining 3 percent of the C8 enrollment received computers and wireless access (2%), internet and access to the Waterford website (1%), or participated in a lending library program (less than 1%) to enable them to access the UPSTART curriculum (see **Figure 2** for details).

UPSTART Usage

We reviewed program usage (time spent using the software program) for three groups: all UPSTART participants, UPSTART program graduates, and the evaluation analysis sample. The hours of instruction observed for all children documented as enrolled in the seventh year of UPSTART are summarized in **Table 5**, and are compared to program "graduates". The average level of usage for all students enrolled in the eighth year of UPSTART (N=10,745) was approximately 56 hours of instruction. The C8 academic year covered 40 weeks of instruction, beginning the week of August 29, 2016 and ending May 29, 2017.

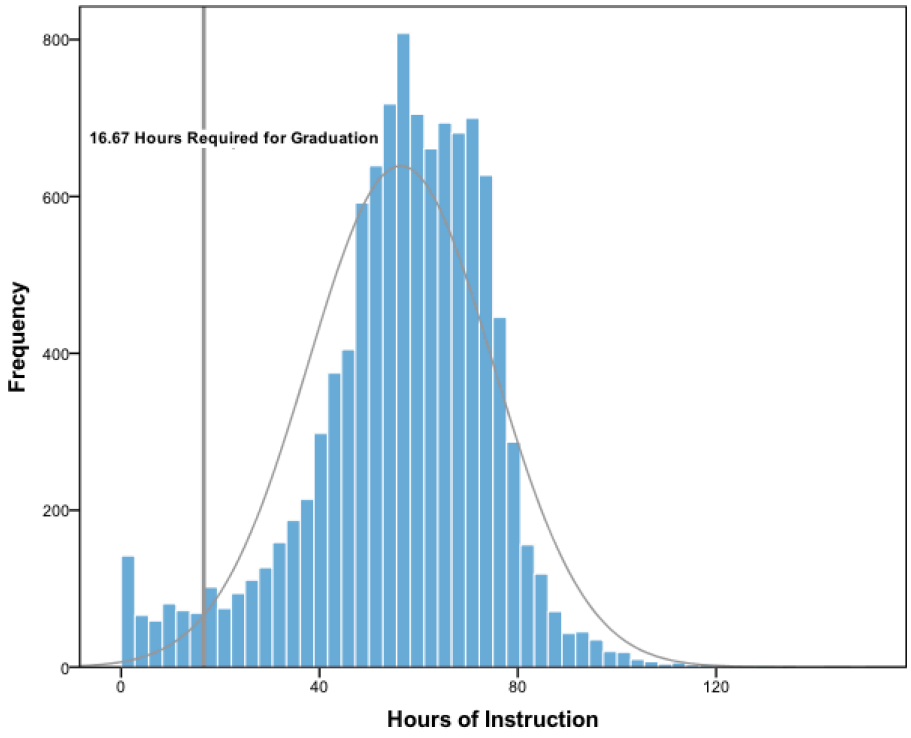
Table 5
C8 Hours of UPSTART Instruction

Group	N	Mean	SD	Range
All UPSTART	10,745	56.36	18.63	00.00 - 149.43
UPSTART Graduates	9,559	60.87	13.69	16.92 - 149.43
UPSTART Analysis Sample	245	58.98	15.79	4.88 – 98.34

Fifty-six of the 10,745 enrolled families who were provided instructional equipment (e.g., computers, an Internet subscription, and a computer drive) did not log any instructional time in the UPSTART curriculum during Year 8 of the program. For enrolled families whose children did use the curriculum, the average duration in the program was approximately 41 weeks. This usage pattern is similar to that observed in the seventh year of the program. The children in the C8 evaluation analysis sample used the UPSTART curriculum for approximately 59 hours of instruction on the average (see **Table 6**).

The histogram in **Figure 3** shows the distribution of hours of instruction for the total C8 population (N=10,745). As noted previously, fifty-six of the enrolled children logged zero hours of instruction during their time in UPSTART. At the other end of the spectrum, sixty-two children logged over 100 hours of instruction.

Figure 3. Hours of Instruction for C8 Families



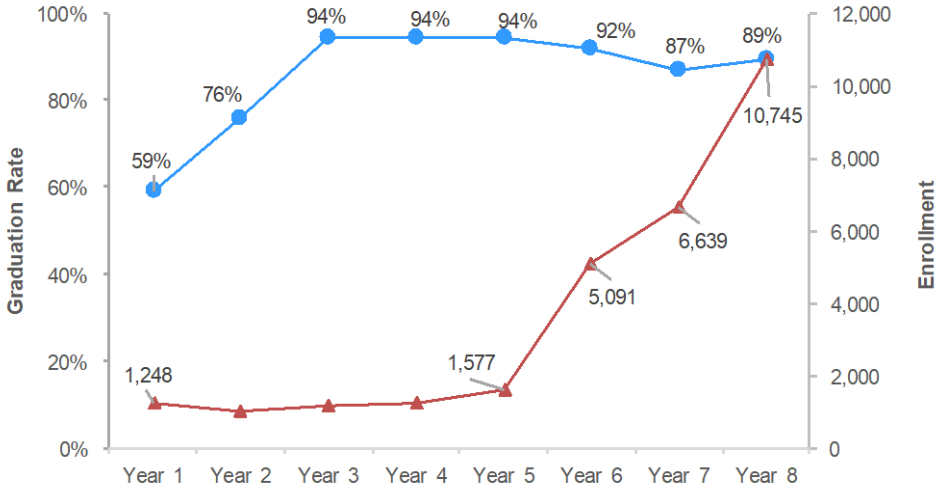
The bottom quartile of the C8 population completed 47.55 hours of instruction or less, the midpoint of the C8 distribution was 58.28 hours, and the top quartile completed in excess of 69.13 hours of instruction.

UPSTART Graduation Rate

Of the 10,745 children documented as enrolled in UPSTART in the seventh year of the program, the Waterford Institute classified 9,559 as children who had met the program's usage criteria and were thus considered to be graduates of the program. The usage criteria involved (a) logging more than 1,000 minutes (16.67 hours of instruction) with the UPSTART curriculum and (b) averaging at least one hour of instruction per week while participating in the program. UPSTART graduate status was significantly correlated with hours of instruction (r = .69) and with the number of weeks in the program (r = .56).

By these usage requirements, Cohort 8 achieved a graduation rate of 89% (i.e., 9,559/10,745 = 0.89). As seen in **Figure 4**, this is a slightly higher graduation rate than the previous year (87%) even in the face of increased enrollment, but slightly lower than the graduation rates that hovered between 92% and 94% in the initial pilot phase of the program that enrolled approximately 1,500 students in Years 3 through 5 and 5,000 students in Year 6.

Figure 4. UPSTART Graduation Rates and Enrollment



In order to further examine the features of program graduates and non-graduates, **Table 6** displays the demographic characteristics of UPSTART graduates and non-graduates in Cohort 8. Children who did not meet the program usage requirement were significantly more likely than UPSTART graduates to speak a language other than English, be a member of an underrepresented racial or ethnic minority group, have parents with lower levels of education, reside in families with parents who were not married, and have higher levels of poverty. We further explore the impact of graduation status on literacy outcomes in the **Cross-Cohort Analysis** section of this report, beginning on page 38.

Table 6
Demographic Characteristics of C8 Population

Demographic Categories		UPSTART Graduates (N=9,559)	UPSTART Non-Graduates (N=1,186)
Child's Gender	Male	51%	50%
	Female	49%	50%
Child's Ethnicity	White	80%***	69%
	Hispanic	13%	19%
	Asian/Pacific Islander	3%	4%
	African American	1%	2%
	Native American	1%	2%
	Other	2%	5%
Child's Language	English	90%**	84%
	Spanish	9%	11%
	Other	1%	5%
Parent Educational Attainment	Some High School	3%	10%
	High School Graduate	12%	22%
	Some College	36%	40%
	College Graduate	39%**	23%
	Advanced Degree	10%	6%
Parent Marital Status	Married	92%***	80%
	Otherwise	8%	20%***
Household Poverty Level	Below 100%	15%	35%***
	Below 185%	40%	62%***
	Below 200%	48%	68%***

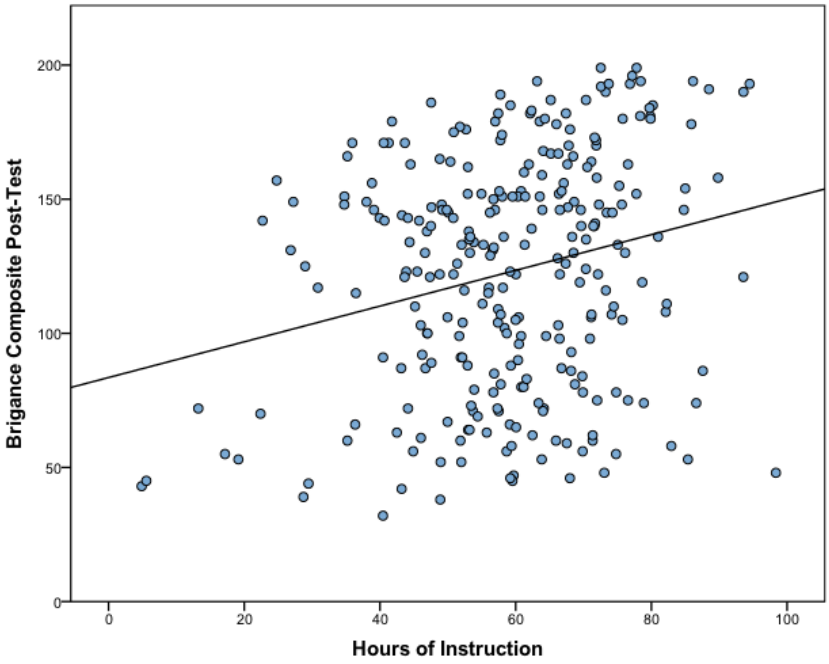
Note: Percentages may not add to 100% due to rounding. Data is from Waterford participant records.

UPSTART Usage and Literacy Outcomes

Similar to previous years, the eighth-year evaluation of UPSTART found curriculum usage to be significantly and positively related to literacy outcomes as measured by composite scores on the Brigance and Bader instruments.

The plot in **Figure 5** on the following page shows a small positive relationship between UPSTART usage (measured in hours of instruction) and Brigance post-test scores (r = .28). That is, Brigance post-test scores tend to increase with increasing hours of UPSTART usage.

Figure 5. Plot of Hours of Instruction and Brigance Post-test scores



Similarly, a correlation analysis of the relationship between hours of UPSTART instruction and the Bader composite post-test score indicates a small positive linear association between instruction time and scores on the Bader post-test (r = .30). This suggests that the acquisition of early phonological skills as measured by the Bader also tended to improve with increasing levels of exposure to UPSTART curriculum.

UPSTART Program Impacts on Literacy

This section includes results based on statistical comparisons of literacy achievement (test scores) for matched treatment and control groups during the eighth year of UPSTART implementation. The impact of the UPSTART program is shown through two lenses: effect sizes and growth scores. Both methods provide salient feedback about the impact of UPSTART. The first method helps stakeholders understand how large an impact UPSTART had on participants, while the second method shows how UPSTART students grew (compared to control students) based on two points of time.

Findings in this section were analyzed to answer the following two research questions:

> **Research Question 1.1:** *Do UPSTART students have better literacy skills at Kindergarten than control students?*
>
> **Research Question 1.2:** *Do UPSTART students show stronger literacy growth rates from preschool to Kindergarten than control students?*

The results of the matched sample are presented for each research question above, and the statistically significant ($p < .05$) findings are depicted visually[7]. We conducted a series of models that explored the impact of household income level on the outcomes of

[7] To create a concise report that highlights the most important findings for stakeholders, we did not present findings that were non-significant in figures.

interest and the results were not meaningfully different from our initial analysis. Accordingly, we chose the simplest data analytic model to test for group differences because it offered ease of interpretation for multiple audiences and more complicated models were not needed to compare differences between the treatment and control group.

Effect Sizes: An Overview

We present effect sizes throughout our reporting to provide additional context for our findings. An effect size (ES) takes the difference between two group means on an outcome variable and represents it in standard deviation units. For example, an effect size of .30 would indicate that the difference between a treatment and control group is .30 standard deviation units. Effect sizes describe the magnitude of the difference between two groups, and essentially create a standardized scale so the results are easy to interpret and have meaning. In previous reports, we have interpreted effect sizes according to Cohen's (1988) general categorization of effect sizes as small (0.2), medium (0.5), and large (0.8) as a general rule of thumb.

However, it is important to note that Cohen's broad categories were designed for a range of effect sizes across a wide spectrum of social and behavioral research and are not specifically tailored for education interventions, studies, or samples. A more appropriate and meaningful benchmark for assessing the significance of an intervention's effect size is to compare it with the effects found for similar education interventions with comparable research samples and outcome measures (Lipsey et al., 2012). If an effect is larger than those of similar interventions, it has practical significance by virtue of being larger than previously reported effect sizes. Conversely, if an effect size is lower than comparable interventions and education research studies, then the impact may not as impressive or significant.

How then, do we determine appropriate benchmarks for interventions similar to UPSTART? Researchers at the U.S. Department of Education's Institute of Education Sciences (IES) reviewed 829 effect sizes from 124 education research studies and determined that the average effect size for an evaluation that used a standardized subject outcome measure (like the Brigance/Bader) to assess a comprehensive educational intervention program that targeted individual students like UPSTART was .26 (Lipsey et. al, 2012). We provide this benchmark to contextualize the effect sizes presented in this report and to aid the reader in determining the practical significance of the effect of UPSTART – any effect size above .26 is higher than the average effect size seen in similar education evaluations. **Appendix B** provides greater detail on how the benchmark was determined. Our .26 threshold is similar to the benchmark specified by the What Works Clearinghouse (WWC), a federally funded initiative at IES that reviews educational research and interventions. The WWC considers effect sizes of .25 or larger to be "substantively important" and a qualified positive (or negative) effect, even if they do not reach statistical significance (What Works Clearinghouse, 2017).

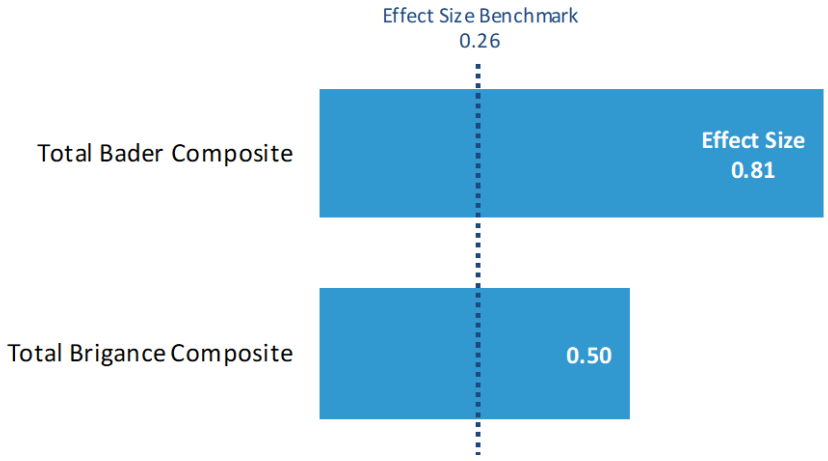
> **Research Question 1.1**: *Do UPSTART students have better literacy skills at entry to Kindergarten than control students?*

In order to demonstrate the impact of the UPSTART program, we present effect sizes that highlight the differences between UPSTART participants and a matched control group on post-test literacy measure.

Effect sizes[8] were calculated to show the magnitude of UPSTART's impact at post-test as measured by each of the 11 literacy subtests (8 Brigance subtests and 3 Bader subtests), and the Total Brigance and Bader Composites (composites include aggregated results of the subtests). **Graphs of effect sizes in this report provide a line marking the .26 benchmark to provide context and to showcase findings that have practical significance. Effect sizes with statistical significance ($p < .05$) will be presented with blue bars, while nonsignificant effect sizes will be displayed with gray bars.**

Combined post-test results showed that UPSTART participation had a medium impact on students' early literacy skill development. In the matched post-test sample[9] (N=487), UPSTART produced strong to medium effects (.81 and .50) as measured by the total Bader and Brigance composite scores that are well above the observed .26 effect size for similar interventions and evaluation studies (see **Figure 6**).

Figure 6. Brigance and Bader Posttest Analysis of Composite Scores



UPSTART children scored significantly higher than control children on seven of the eight Brigance tests and all three Bader subtests on the post-test, showing strong empirical evidence that UPSTART was successful in helping children develop key early literacy skills. The ES estimates for individual subtests on the Bader ranged from .25 (Recites Alphabet) to .60 (Pre-primer Vocabulary) and would be considered small to medium effects. The Expressive Vocabulary subtest was the only subtest on the Brigance in which the treatment and control groups' post-test scores were not significantly different. Effect sizes for the Bader were 1.00 for the Phoneme Blending subscale and .76 for the Phoneme Segmenting subscale.

[8] Effect size (Cohen's *d*) was calculated for each test as the treatment group mean minus the control group mean divided by the pooled standard deviation.
[9] Treatment Group (N = 245); Control Group (N = 242)

The effect size estimates for each statistically significant literacy subtest (9 out of 10), as measured by the Brigance and Bader instruments, are presented below in **Figure 7**. The results are organized according to the subtests' respective literacy constructs: decoding, phonological awareness, letter knowledge, and pre-literacy discrimination. Please refer to the **Outcome Measures** section beginning on page 14 for a discussion of the measurement constructs and **Table 3** for a list of all 10 subtests and their corresponding constructs.

Figure 7. Effect Size Estimates by Literacy Construct

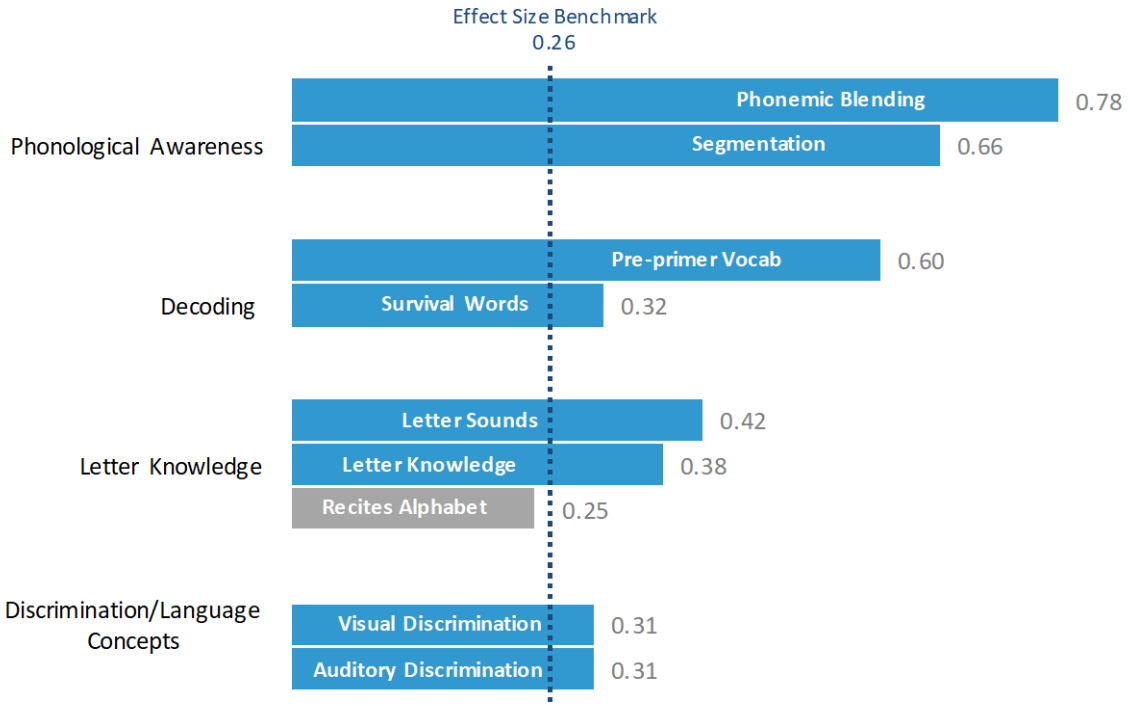
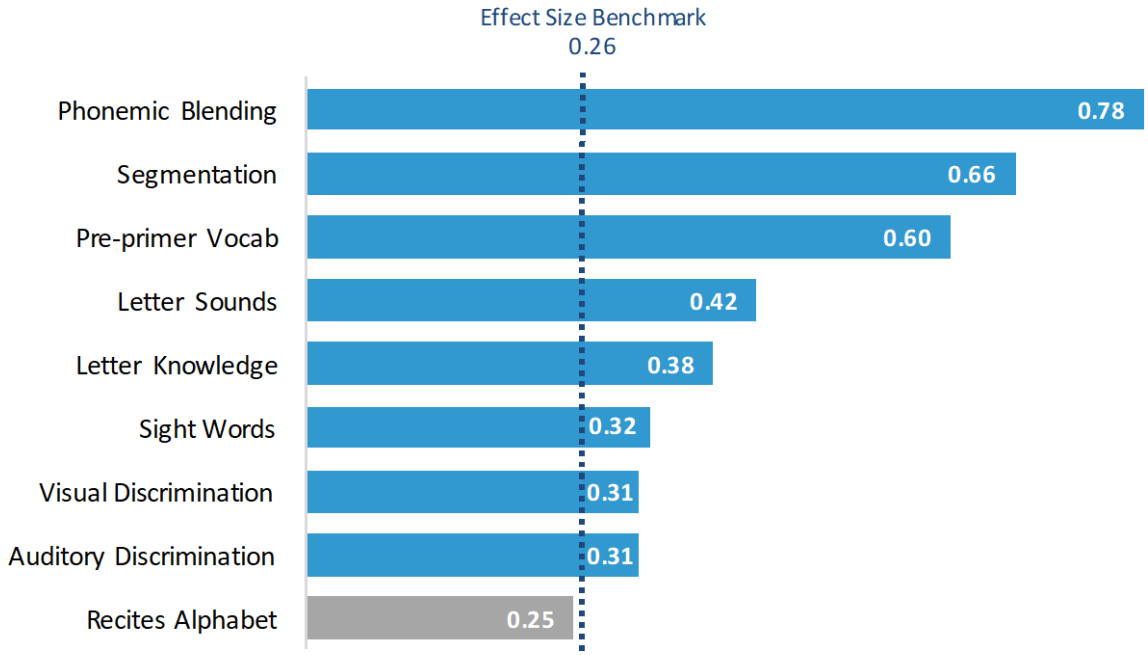


Figure 8 presents the ES of each literacy subtest by the size of their effects along with the .26 effect size benchmark from similar education intervention studies. UPSTART had the largest impact on phonemic blending (.78), phoneme segmenting (.66), and pre-primer vocabulary (.60). Effect sizes from five subtests measuring decoding, phonological awareness, and visual/auditory discrimination were above the average .26 effect size benchmark from other similar education interventions and should be considered practically significant and consequential.

Figure 8. Effect Size Estimates by Magnitude of Effect



Regression Results. In addition to computing effect sizes, we ran regression analyses to determine if pre-existing differences between the treatment and control groups on demographics and pre-test measures affected the results. Multiple linear regression analysis using the pre-test and students' gender as covariates indicated that using UPSTART produced a 4.90 increase in the Bader post-test composite score, with each point increase on the Bader pretest composite scores associated with a .94 increase on the Bader posttest. No demographic variables emerged as covariates or were significantly associated with the Bader posttest. The regression analyses did not essentially change the initial estimate of the mean overall impact on the Bader at post-test, however the linear regression analyses improved the estimate of UPSTART's overall impact on the Bader post-test from 4.69 to 5.76 points.

Using Brigance pre-test scores as covariates did not improve the estimate of UPSTART's overall impact. However, the linear combination of UPSTART participation and the Brigance composite pre-test was significantly related to performance on the Brigance post-test, R^2 = .47, adjusted R^2 = .47, F = 214.14, p < .0001. and accounted for 47% of the explained variability in posttest outcomes.

> **Research Question 1.2**: *Do UPSTART students show stronger literacy growth rates from preschool to Kindergarten than control students?*

We studied literacy growth rates while in the program as an additional way to evaluate program impacts beyond outcome score comparisons. Paired samples t-tests were performed to examine growth rates as measured by the Brigance and the Bader total test composite scores for the treatment and control group children and each subtest (Phonemic Blending, Phonemic Segmenting, Visual Discrimination, Recites Alphabet, Letter Knowledge, Letter Sounds, Auditory Discrimination, Survival Sight Words, and Pre-Primer Vocabulary). Growth rates for the treatment and control children were compared based on the observed difference scores between the post-test and the pre-test.

- The treatment group showed significantly ($p < .05$) stronger mean literacy growth rates compared to the control group on the Total Bader and Brigance Composites, with the treatment group scoring an average of 5 points higher on the Bader and 21 points higher on the Brigance.

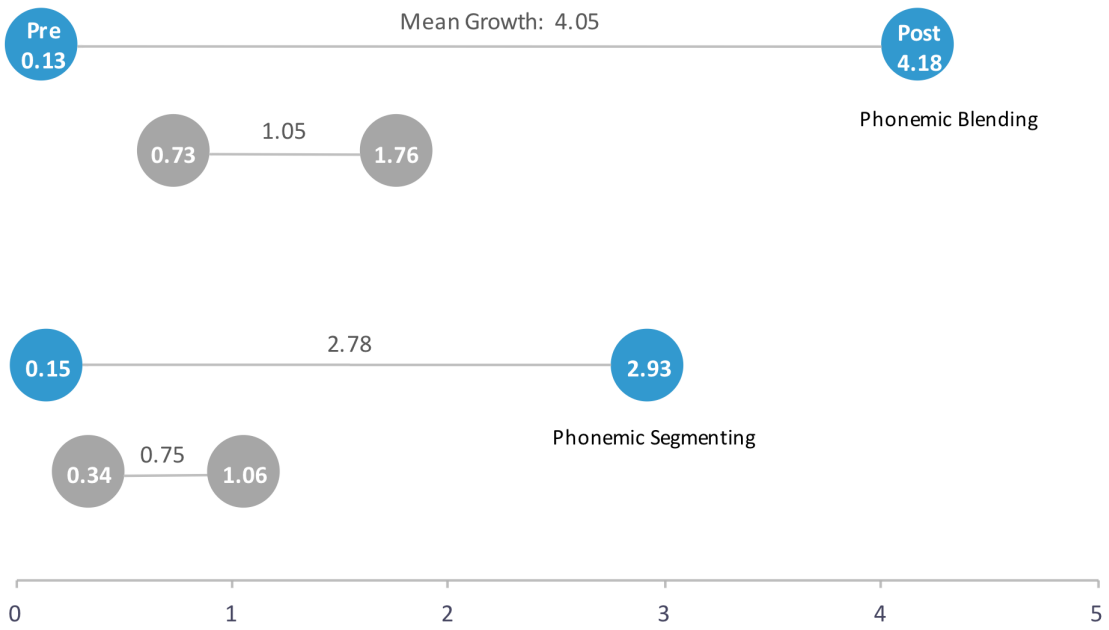
- The treatment group showed statistically stronger ($p < .05$) literacy growth rates compared to the control group on four out of eight Brigance subtests (Visual Discrimination, Letter Sounds, Survival Sight Words, and Basic Pre-Primer Vocabulary) and both Bader subtests (Phonemic Blending and Segmentation).

- There was no difference in growth rates between the treatment and control group on the following subtests: Auditory Discrimination (measures auditory discrimination between similar word pairs), Expressive Vocabulary, Reciting the Alphabet, and Letter Knowledge.

Growth rates from pre-test to post-test are shown in the figures below. Each figure categorizes the Brigance and Bader subtests that were statistically significant ($p<.05$) based on their respective literacy constructs, which include: **phonological awareness**, **decoding**, **letter knowledge, and pre-literacy discrimination**. UPSTART participants' scores are depicted in blue, while their control group counterparts are in grey.

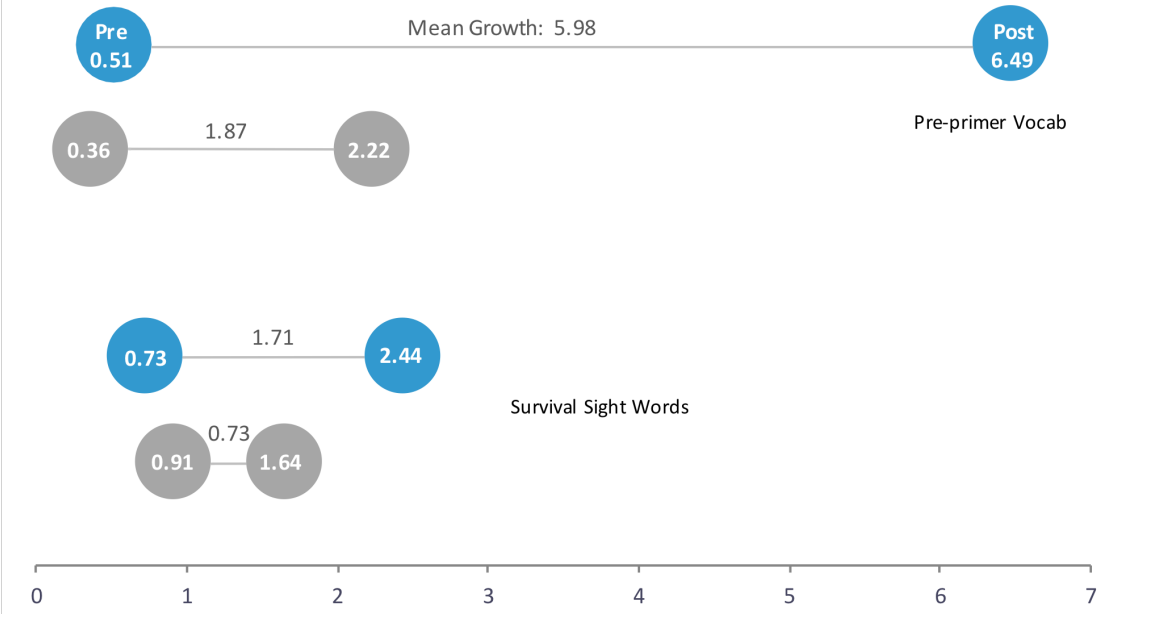
UPSTART children experienced significant, higher mean growth from pre-test to post-test compared to control children on all three subtests (rhyme recognition, phonemic blending and segmenting) that measure **Phonological Awareness**.

Figure 9. Phonological Awareness: Treatment and Control Group Pre-and-Posttest Mean Scores



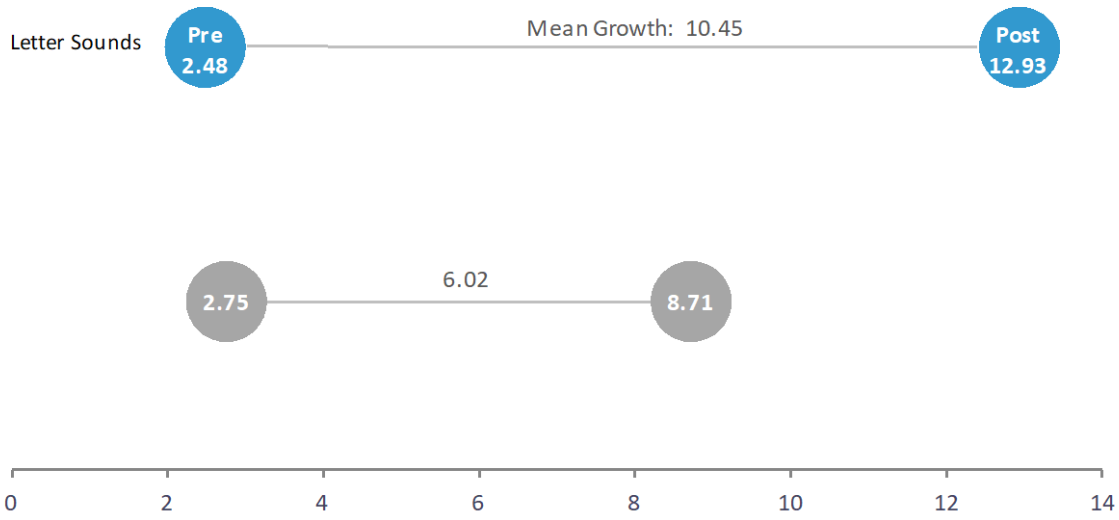
UPSTART students experienced significant, higher mean growth compared to the control group on both subtests used to measure children's **Decoding** ability, including pre-primer vocabulary and survival sight words.

Figure 10. Decoding: Treatment and Control Group Pre-and-Posttest Mean Scores



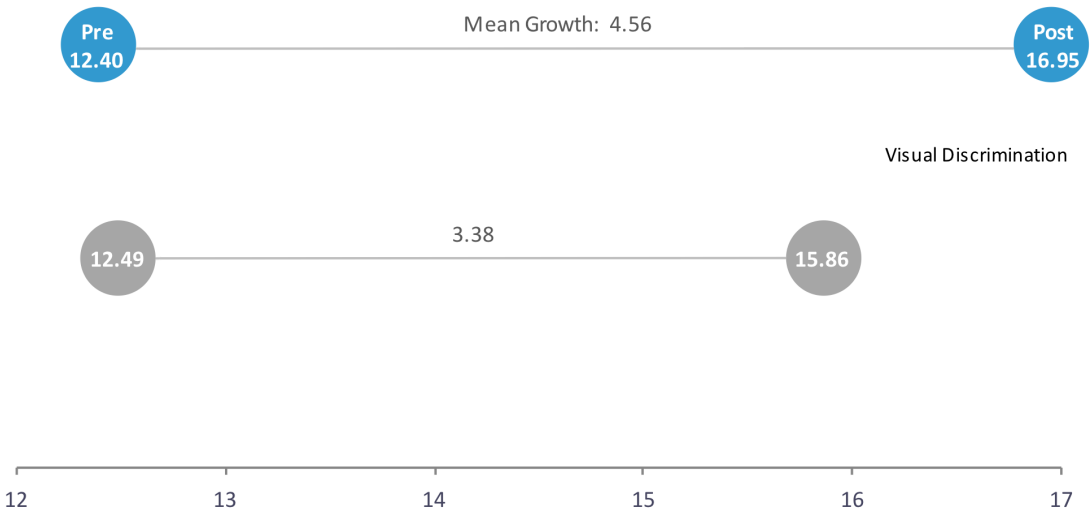
UPSTART children experienced significantly higher growth, compared to non-UPSTART children, in measuring **Letter Knowledge**. UPSTART children showed stronger growth in producing sounds of lower case letters (letter sounds). A significant difference in the growth rates of treatment and control students was not observed for the Identifying Uppercase Letters or Reciting the Alphabet subtests.

Figure 11. Letter Knowledge: Treatment and Control Group Pre-and-Posttest Mean Scores



Students who were enrolled in UPSTART had significantly higher levels of growth on one subscale measuring **Pre-Literacy Discrimination**, compared to students who did not participate in UPSTART. UPSTART children were more likely to improve on this subtest which involved visually identifying the similarities and differences between forms, letters and words. A significant difference in the growth rates between the two groups was not observed for the auditory discrimination subtest, where children identify similarities and differences between word sounds.

Figure 12. Pre-literacy Discrimination: Treatment and Control Group Pre-and-Posttest Mean Scores



Summary and Discussion

This section of the Cohort 8 (C8) evaluation report summarizes findings and trends for UPSTART implementation and impacts on early literacy skills.

Program Implementation

Based on the program enrollment demographic and usage data provided by UPSTART program officers at the Waterford Institute, the program was implemented with great success. UPSTART enrollment increased from 6,639 children in Year 7 to 10,745 children in Year 8, an increase of over 62 percent over the past year. Enrollment increased in areas across state of Utah and UPSTART has reached families in both rural and urban communities. Half of the children enrolled in Year 8 lived in families with incomes less than 200% of the federal poverty level and the majority of children were White (79%) and English speaking (80%).

Most of the C8 children accessed the UPSTART curriculum through the Waterford website (80%). Approximately 11% of the eighth-year participants received a computer loan and 6% were provided with a computer and Internet. Despite increased enrollment across the state, graduation rates at 89% were 2% higher than the previous year but slightly lower than the 92%-94% graduation rate that characterized earlier cohorts. Families with children who did not graduate from UPSTART tended to have lower levels of parental education, higher levels of poverty, and be members of underrepresented racial, ethnic, and linguistic groups.

Program Impacts on Literacy Development

While program implementation findings are important for monitoring how resources were used to enroll and graduate students, findings about literacy testing outcomes is the most important indicator of program success. UPSTART participation had a strong impact on children's emerging literacy skills based on the results from effect size and growth score analyses. The program produced statistical effects (Bader ES = .81; Brigance ES = .50) on learning compared to non-program children that are stronger, on average, than other educational evaluation studies on similar interventions with comparable outcomes and participants. The effects were seen across different measures of literacy: decoding skills, letter knowledge, pre-literacy discrimination, and phonological awareness.

We used two types of statistical comparisons to give the state multifaceted findings related to literacy achievement during the pre-kindergarten year: effect sizes and growth scores. The effect size estimates measured the differences between the treatment and control students at post-test, while the growth score analyses measured the change from pre-test to post-test for both the treatment and control groups.

We reported findings for focused literacy tests, and a majority of the results from the Brigance and Bader scales were shown to have small to large effects (effect sizes ranged from .25 to .78). Overall, the results of both analyses illustrate that UPSTART program participation had a strong impact on facilitating UPSTART students' literacy skill development in a variety of key areas. The largest impacts were found for phonemic segmentation and blending (measures phonological awareness) and pre-primer vocabulary (measures decoding skills).

UPSTART students also experienced greater growth from pre-test to post-test compared to control students in three out of five literacy constructs (phonological awareness, decoding, and letter knowledge), with the exception of the vocabulary and pre-literacy discrimination domains. Group differences in the vocabulary subtests and auditory/visual discrimination subtests were not statistically significant in the post-test analyses, indicating that these are skill areas in which UPSTART did not have a positive impact in Cohort 8.

Cross-Cohort Analysis

Preschool is a dynamic time for children, and members of the UPSTART Advisory Committee (UAC) had questions about how much real-life events effected program outcomes. Specifically, the Advisory Committee members wanted to determine how much influence program graduation status had on UPSTART outcomes and if using the program for less than what would be needed for graduation had any benefits over using the program as recommended. In addition, members of the Advisory Committee wanted to know if students who participate in both UPSTART and traditional preschool had an advantage over students who participated only in UPSTART.

While we are able to use our collected data to look at the overall effects of UPSTART within a given year when compared to a group of comparison students, we typically do not have sufficient numbers of students to look at the effect of UPSTART on subgroups of interest. However, this year after combining our assessment data across the current and past seven cohorts ("cross-cohort" data), we had a sufficient number of students for subgroup analyses.

For example, during the Cohort 7 evaluation we reported that the UPSTART graduation rate had declined from previous years and initial exploratory analysis on the small sample of program non-graduates (N = 17) indicated that program graduates had higher scores on measure of early literacy compared to non-graduates (Evaluation and Training Institute, 2017). Typically, children who fail to meet UPSTART program requirements represent 6% of our evaluation sample within a given cohort – an amount insufficient for more in-depth analysis. Conducting a cross-cohort analysis would produce a larger sample of non-graduates and allow us to determine if the positive association between graduation status and literacy outcomes is consistent when looking at a more representative group of students.

Additionally, members of the UPSTART Advisory Committee proposed looking whether attendance at a traditional preschool (either home- or center-based) in addition to UPSTART participation was associated with stronger literacy outcomes. High quality traditional preschool programs can strengthen social and emotional skills and provide children with developmentally appropriate curriculum across the learning domains of literacy, math, science, creative arts, and physical development. It is possible, as some suggested, that participating in UPSTART while being enrolled in traditional preschool would produce greater gains in achievement than with UPSTART alone. Although families who use the UPSTART program are welcome to enroll their preschooler in traditional preschool programs, typically only about one-third of families in our evaluation sample exercise that option and such small numbers prevent us from examining the impact of preschool in further detail. However, similar to the case of UPSTART graduation status as described above, by aggregating our treatment samples across the cohorts, we can create a sufficient sample of children who were enrolled in a traditional preschool program for analysis.

This cross-cohort analysis approach allows us to address questions that arose from stakeholders and from our previous evaluations about the impact of graduation status and enrollment in traditional preschool on UPSTART impacts.

Research Questions

The cross-cohort analysis is framed by research questions to assess UPSTART's impact on children who do not graduate from the program and on children who participate in UPSTART while attending traditional preschool.

The Waterford Institute has established set criteria for meeting the requirements of UPSTART to graduate from the program. We hypothesized that if UPSTART graduation has no effect on improving early literacy skills, then the preschool children who graduated from UPSTART (graduates) would perform similarly on post-test measures of early literacy development as children who did not meet the requirements for graduation (non-graduates). However, if meeting program objectives and graduating from UPSTART does have an effect on improving early literacy, then the program graduates group should perform significantly better than the non-graduates on the post-test at the beginning of kindergarten.

Additionally, to take into account the gains in performance from baseline levels, we also examine growth rates from pre-test to post-test. If being an UPSTART graduate is associated with stronger literacy growth rates, then the UPSTART graduates group would be expected to show greater gain scores (post-test score minus pre-test score) relative to the comparison group on the various literacy subtests and total test scores.

Our research questions concerning the impact of UPSTART graduation status are as follows:

> **Research Question 2.1**: *Do UPSTART students who graduate from the program in a matched sample drawn from Cohorts 2 through 8 have better literacy skills at Kindergarten than UPSTART students who do not graduate from the program?*
>
> **Research Question 2.2**: *Do UPSTART students in the matched sample drawn from cohorts 2 through 8 who graduate from the program show stronger literacy growth rates from preschool to Kindergarten than students who do not graduate from the UPSTART program?*

Participating in UPSTART does not preclude families from enrolling children in traditional preschool settings or other enrichment programs to advance their social and/or cognitive development. It is unclear from the small sample sizes present each year in previous evaluations if the beneficial effect of UPSTART was a result of the program or due to external factors such as preschool enrollment. If preschool has a unique impact on students' literacy achievement, we would predict that students who attend UPSTART and a traditional preschool outperform children who participate only in UPSTART. However, if additional preschool has no bearing on UPSTART children's performance we would expect to see similar levels of performance between children enrolled in traditional preschool and UPSTART and children who were only enrolled in the computer-based UPSTART.

Similar to previous analyses, we analyzed growth rates from pre-test to post-test in both groups to examine an alternate view of the data. If students who enrolled in a traditional preschool in addition to UPSTART show stronger literacy growth rates, then the UPSTART + Preschool group would be expected to show greater gain scores (post-test score minus pre-test score) relative to the comparison group of UPSTART only on the

various literacy subtests and total test scores. Our research questions to investigate the impact of traditional preschool and UPSTART are as follows:

> **Research Question 2.3**: *Do UPSTART students in the matched sample drawn from cohorts 2 through 8 who attend traditional pre-school in addition to UPSTART have better literacy skills at Kindergarten than students in the matched sample who did not attend traditional pre-school in addition to UPSTART?*
>
> **Research Question 2.3**: *Do UPSTART students in the matched sample drawn from cohorts 2 through 8 who attend traditional pre-school in addition to UPSTART show stronger literacy growth rates from preschool to Kindergarten than students in the matched sample who did not attend traditional pre-school in addition to UPSTART?*

Methods

Unlike the research questions for the Cohort 8 evaluation which require a comparison group of children who were not enrolled in UPSTART, the research questions for the current cross-cohort evaluation examine the differences between UPSTART participants only. Thus, historical data from students who participated in UPSTART during the last seven years of the UPSTART evaluation was used for the cross-cohort analysis (*N* = 1,149). Each UPSTART evaluation cohort used the same research design, a quasi-experimental design with pre- and post-program assessments. Moreover, a common instrument was administered across each cohort to measure early literacy outcomes, the Brigance Inventory of Early Childhood Development. (For a detailed accounting of the research method used in each evaluation cohort, please see the Methods section of the **Cohort 8 Evaluation** section of this report.)

Treatment cases from the past seven cohorts were combined into two separate data files for analysis: (1) UPSTART children who graduated from the program and those who fell short of the program requirements for graduation and (2) UPSTART children who were enrolled both UPSTART and traditional preschool (UPSTART + Preschool) and children who only participated in UPSTART during the year before kindergarten (UPSTART only). Coarsened Exact Matching (CEM) was used within each dataset to construct balanced groups for analysis (for a review of CEM as a methodological tool for creating balanced groups, please see the Methods section in the **Cohort 8 Evaluation** section of this report). UPSTART students in each data file were matched on the following characteristics: Brigance pre-test scores, ethnicity, language, and parent education.

In order to determine the impact of graduation status on early literacy, a matched dataset (N=150) was created that consisted of 75 children who graduated from UPSTART and 75 children who were enrolled in UPSTART but failed to meet program usage requirements. Demographic characteristics of the matched graduate and non-graduate samples are displayed in **Table 7**.

Table 7
Graduate vs. Non-Graduate Comparisons on Key Demographics

Demographic Categories		Graduate (N=75)	Non-Graduate (N=75)
Child Gender	Female	48%	51%
	Male	52%	49%
Child Ethnicity	Caucasian	91%	91%
	Hispanic	6%	6%
Child Language	English	96%	98%
Parent Education Level	High School Diploma	14%	12%
	Some College	50%	50%
	Bachelor's degree	30%	30%
	Graduate degree	3%	5%
Parent Marital Status	Married	88%	87%
Household Income	Under $10,000	5%	2%
	$10k-$24,999	15%	10%
	$25k-$49,999	50%	29%
	$50k-$74,999	24%	30%
	$75k-$99,999	4%	17%
	$100k or more	2%	12%

The final matched dataset (*N* = 898) used to explore the impact of additional preschool with UPSTART on literacy achievement consisted of 449 children who participated in UPSTART while being enrolled in a traditional preschool setting and 449 children who participated only in UPSTART. **Table 8** presents the demographic breakdown of the UPSTART only and UPSTART + Preschool groups.

Table 8
UPSTART + Preschool vs. UPSTART Only Comparisons on Key Demographics

Demographic Categories		Treatment (N=449)	Control (N=449)
Child Gender	Female	48%	51%
	Male	52%	49%
Child Ethnicity	Caucasian	91%	91%
	Hispanic	6%	6%
Child Language	English	96%	98%
Parent Education Level	High School Diploma	14%	12%
	Some College	50%	50%
	Bachelor's degree	30%	30%
	Graduate degree	3%	5%
Parent Marital Status	Married	88%	87%
Household Income	Under $10,000	5%	2%
	$10k-$24,999	15%	10%
	$25k-$49,999	50%	29%
	$50k-$74,999	24%	30%
	$75k-$99,999	4%	17%
	$100k or more	2%	12%

Early literacy achievement was measured with the Brigance Inventory of Early Development, which was administered by trained assessors during the summers before UPSTART began and a year later upon completion of the program and before children entered school. The Brigance consists of seven measures that function as indicators of critical skills that students must master to become proficient readers, including letter knowledge, visual and auditory discrimination, phonics, and decoding. For a complete description of the Brigance instrument, please see the **Cohort 8 Evaluation** section of this report.

Findings

We investigated two research areas: the impact of meeting UPSTART program requirements on measures of early literacy and whether or not attendance at a traditional preschool in addition to UPSTART is related to higher literacy achievement. We present the findings by research question for ease of interpretation and clarity.

Impact of Graduation Status

> **Research Question 2.1**: *Do UPSTART students who GRADUATE from the program in a matched sample drawn from Cohorts 2 through 8 have better literacy skills at Kindergarten than UPSTART students who do not graduate from the program?*

In the matched post-test sample (N = 150), students who were graduates of the UPSTART program and met all usage requirements had significantly higher post-test scores on Brigance composite (p < .05) compared to children who did not graduate the program. In addition, four of the eight Brigance post-test subtest comparisons (visual discrimination, recites alphabet, survival sight words, and basic pre-primer vocabulary) between the graduate and non-graduate groups were statistically significant (p <. 05) and favored program graduates. Effect sizes to measure the magnitude of the differences between program graduates and program non-graduates were calculated for the Brigance composite and all subscales. Mean posttest differences (Delta) and effect sizes (ES) are shown in **Table 11** for each Brigance subtest and the Total Brigance Composite score.

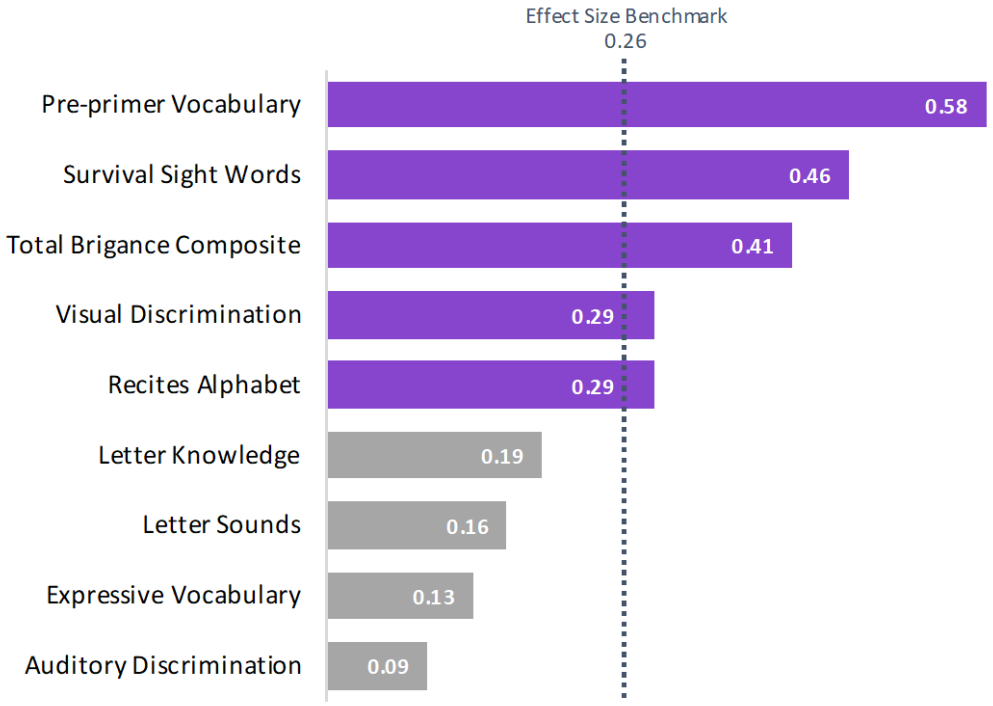
Table 9
UPSTART Graduate vs. Non-Graduate Post-Test Results

Brigance Posttest	Delta	ES
Expressive Vocabulary	0.15	0.13
Visual Discrimination	.87	.29*
Recites Alphabet	2.70	.29*
Letter Knowledge	3.41	.19
Letter Sounds	1.66	.16
Auditory Discrimination	.27	.09
Survival Sight Words	1.07	.46**
Basic Pre-primer Vocabulary	4.63	.58***
Total Brigance Composite	8.50	0.41*

*p < .05, **p ≤ .01, ***p ≤ .001

UPSTART program graduates and non-graduates effect sizes are presented graphically in **Figure 13**. Significant effect sizes are displayed in purple and nonsignificant effect sizes are shown in gray. As can be seen in **Figure 13**, all statistically significant effect sizes between UPSTART graduates and UPSTART non-graduates on post-test literacy measures surpassed the .26 benchmark for practical significance.

Figure 13. UPSTART Graduate vs. Non-Graduate Post-Test Results



The linear combination of graduation, the Brigance composite pretest, a parent with only a high school diploma, and being in the lowest income bracket (family income under $25,000) was significantly related to the Brigance posttest, R^2 = .43, adjusted R^2 = .41, F = 24.40, p < .0001. and accounted for 41% of the explained variability in posttest outcomes.

In summary, the data from our matched graduate and non-graduate groups provide evidence for a medium to large positive effect of graduation on key aspects of young children's phonics and vocabulary development as measured by the Brigance test.

> **Research Question 2.2**: *Do UPSTART students in the matched sample drawn from cohorts 2 through 8 who graduate from the program show stronger literacy growth rates from preschool to Kindergarten than students who do not graduate from the UPSTART program?*

Students who met all requirements for graduation showed statistically significant (p<.01) literacy growth rates from pre-test to post-test on all of the Brigance subtests and the Brigance composite score. Similarly, students who did not graduate from UPSTART also experienced statistically significant (p<.01) literacy growth rates from pre-test to post-test administration on the Brigance composite and on all subtests except for Auditory Discrimination.

However, as seen in **Table 10**, when looking at effect sizes that show the magnitude of differences between the two groups' literacy growth rates, we find significant differences between the UPSTART graduates and UPSTART non-graduates that favor UPSTART graduates. (Tables in **Appendix C** provide additional detail, such as group mean growth rates by subtest.) Students who met program usage requirements showed statistically stronger literacy growth rates compared to the non-graduates on the two Brigance subtests that assess decoding skills (Survival Sight Words and Basic Vocabulary), and on the overall Brigance composite score. Each of these significant growth rate effect sizes met the .26 threshold for practical significance.

Table 10
Growth Rate Effect Sizes for UPSTART Graduates and UPSTART Non-Graduates

Brigance Subscale	Effect Size
Visual Discrimination	.20
Letter Sounds	.08
Auditory Discrimination	.05
Survival Sight Words	.51**
Basic Pre-primer Vocabulary	.60***
Total Brigance Composite	.44**

*p < .05, **p ≤ .01, ***p ≤ .001

Our next set of findings focuses on the effect of enrolling in a traditional preschool setting while participating in the UPSTART program.

Impact of Traditional Preschool Enrollment

> **Research Question 2.3**: *Do UPSTART students in the matched sample drawn from cohorts 2 through 8 who attend traditional pre-school in addition to UPSTART have better literacy skills at Kindergarten than students in the matched sample who did not attend traditional pre-school in addition to UPSTART?*

In the matched posttest sample (N=898), two of the eight Brigance subtest comparisons between the traditional preschool plus UPSTART and UPSTART-only groups within treatment at posttest were statistically significant (p < .05). These two tests, Letter Knowledge and Letter Sounds, favored the UPSTART plus traditional preschool group, but only Letter Knowledge had an effect size that met the .26 benchmark for practical significance. There was no statistically significant difference between the traditional preschool plus UPSTART and UPSTART-only groups on the total Brigance Composite at post-test.

Mean posttest differences (Delta) between the UPSTART only and UPSTART with preschool groups and effect sizes (ES) are shown in **Table 11** for each Brigance subtest and the Total Brigance Composite score.

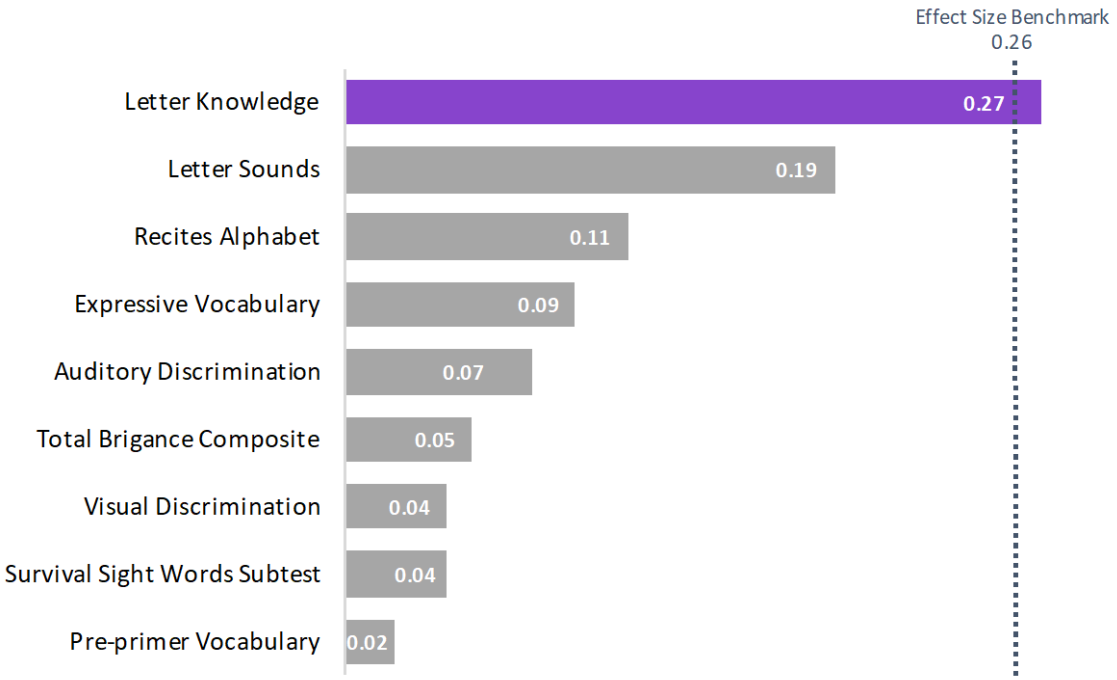
Table 11
UPSTART only and UPSTART + Preschool Post-Test Results

Brigance Posttest	Delta	ES
Expressive Vocabulary	-.12	-.09
Visual Discrimination	.10	.04
Recites Alphabet	-.98	-.11
Letter Knowledge	-4.24	-.27***
Letter Sounds	-1.82	-.19*
Auditory Discrimination	.21	.07
Survival Sight Words Subtest	.14	.04
Basic Pre-primer Vocabulary	.15	.02
Total Brigance Composite	-1.23	-.05

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

Post-test effect sizes between children who participated in UPSTART while enrolled in traditional preschool and children who participated only in UPSTART are presented graphically in **Figure 14**. Significant effect sizes are displayed in purple and nonsignificant effect sizes are shown in gray. As seen in **Figure 14**, only the effect size measuring differences on the post-test measure of letter knowledge (which favored the children enrolled in UPSTART + preschool) passed the .26 benchmark for practical significance.

Figure 14. UPSTART only vs. UPSTART + Preschool



Simple linear regression analysis using only attendance in traditional preschool plus UPSTART participation as a predictor was non-significant. The linear combination of attending a traditional preschool plus UPSTART, the Brigance composite pre-test, parents with a high school diploma only and in the lowest income bracket (family income under $25,000) explained 47% of the variability in posttest outcomes, R^2 = .48, adjusted R^2 = .44, F = 156.85, p < .0001.

In summary, the data from the matched groups UPSTART versus UPSTART plus traditional preschool sample provide evidence for a small effect of attending a traditional preschool in addition to UPSTART in terms of letter knowledge and letter sounds, but no effect in terms of other measures of young children's phonics, decoding skills, and vocabulary development as measured by the Brigance Test.

> **Research Question 2.4**: *Do UPSTART students in the matched sample drawn from cohorts 2 through 8 who attend traditional preschool in addition to UPSTART show stronger literacy growth rates from preschool to Kindergarten than UPSTART students who did not attend traditional preschool?*

While both groups (UPSTART only and UPSTART plus traditional preschool showed statistically significant (p<.01) growth rates from pre-test to post-test on the Brigance Composite and all subtests, there were no significant difference in the literacy growth rates from pre-test to post-tests *between* the two groups on the Brigance Composite or on any subtest.

Differences in growth rates from pre-test to post-test between children enrolling in preschool and UPSTART and children participating in UPSTART only are shown with effect sizes in **Table 12** for the matched sample as measured by the Brigance. No effect sizes were significant, and none met the .26 threshold for practical significance. (Tables in **Appendix C** provide additional detail, such as group mean growth rates by subtest.)

Table 12
Growth Rates for UPSTART Only and UPSTART + Preschool

Brigance Subscale	Effect Size
Visual Discrimination	0.02
Letter Sounds	0.15
Auditory Discrimination	0.02
Survival Sight Words	0.08
Basic Pre-primer Vocabulary	0.01
Total Brigance Composite	0.06

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

Summary and Discussion

Using a data file that combined UPSTART treatment students across seven cohorts, we were able to conduct additional analyses on two different groups that are typically insufficient in size within a given year: program graduates versus non-graduates and UPSTART students who enrolled in a traditional preschool versus those that participate only in UPSTART.

We found that the children who graduated from the program had higher post-test scores on the Brigance composite, as well as higher scores on the visual discrimination, recites

alphabet, survival sight words, and pre-primer vocabulary subscales when compared to children who did not meet program vendor requirements for usage. These effects ranged from .29 to .58 and can be considered to have practical significance as they were above the .26 average effect size threshold from evaluations with similar interventions, measures, and populations reported in the literature (see **Appendix B** for greater detail on how the benchmark was determined). Moreover, students who graduated from the program had significantly higher growth scores than children who did not on the Brigance composite, as well as subscales that assessed emergent decoding skills (reading survival sight words and pre-primer vocabulary).

With respect to the impact of enrolling in traditional preschool program while participating in UPSTART, we found evidence of an effect of attending a traditional preschool in addition to UPSTART in terms of letter knowledge and letter sounds, but only letter knowledge (ES = .27) had an effect size that met the .26 threshold for practical significance. There were no other effects of traditional preschool enrollment on other assessments of young children's phonics, decoding skills, and vocabulary development as measured by the Brigance Test. Additionally, there were no differences in children's scores on the overall Brigance composite – the comprehensive measure of children's letter knowledge, phonics, and decoding. The UPSTART plus preschool group showed stronger growth rates only one subtest (letter sounds).

In light of the fact that meeting program vendors' recommendations about the consistency of program use and the total amount of time spent using the software appears to have a significant impact on literacy achievement outcomes, ensuring that as many children graduate from UPSTART as possible should be a high priority for both the program vendor and the state. Students who did not graduate from the UPSTART program did not attain the full benefits of improved skills as they had lower levels of literacy achievement than students who completed the full UPSTART usage requirements. Results from the effect of additional preschool suggest that UPSTART provides high-quality literacy instruction similar to traditional preschools. While participating in a traditional preschool could convey emotional and social development benefits to children that may not be emphasized with a computer-based curriculum, there appears to be no additional advantage in the literacy outcomes of children who enroll in UPSTART and a traditional preschool.

First Grade Analysis

Evaluations of the UPSTART program have consistently shown a medium to strong impact on improving children's early literacy skills prior to entering kindergarten. For example, as reported in the most recent evaluation, students enrolled in Cohort 7 during the 2015-16 academic year experienced significant positive effects (ES = .52) compared to control children on the Brigance composite, an instrument that measures decoding skills, letter knowledge, vocabulary and syntax, and pre-literacy discrimination (Evaluation and Training Institute, 2017). Other 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, while some research points to continuing benefits of high-quality preschool experiences on cognitive outcomes into adolescence (Vandell, Belsky, Burchinal, Vandergrift, & Steinberg, 2010), 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 that 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)

As part of the UPSTART program expansion, stakeholders were interested in the long-term impact of UPSTART on students and whether program benefits present upon entry to kindergarten can be sustained once children begin elementary school. The First Grade Analysis examines whether the achievement gains from UPSTART that occurred prior to school entry were sustained through kindergarten and first grade.

Kindergarten EISP Exposure

Education initiatives such as the UPSTART program do not operate in isolation, and there are often multiple efforts or programs to foster student achievement in young learners. During the 2016-17 school year, statewide legislation through the Early Intervention Software Program (EISP) provided funding to districts to supplement kindergarten students' classroom learning with computer-based adaptive reading software programs. The goal of EISP is to provide additional individualized instruction for students in order to increase the number of students reading at grade level and to ensure students are meeting literacy achievement benchmarks. Schools interested in participating in the program submitted applications to the USBE and selected their reading software of choice from among seven vendors. Software vendors provided training and support to schools throughout the year and their programs were used in 388 schools and by 19,277 kindergarten students in 2016-17.

Consequently, it is possible that a student who was enrolled UPSTART preschool program in 2015-16, matriculated into a kindergarten classroom that was also participating in the EISP program during the 2016-17 school year. Participating in the EISP program would be major confound for the purposes of our study – both UPSTART and EISP software programs are computer-based, adaptive, and provide individualized instruction on a consistent and prescribed basis in early literacy. A student who did not participate in UPSTART but who was enrolled in a school receiving EISP services might outperform students who did not participate in either program. Additionally, because both the UPSTART preschool and EISP program involve the use of computer-based early literacy software, it is important to determine the unique impact of UPSTART preschool from participation in EISP kindergarten instruction and the possibility of potential multiple effects from participating in both programs. As the evaluators for both the UPSTART and the EISP programs, we are in a unique position to be able to determine which program (if any) a student participated in and create independent and mutually exclusive groups to ascertain the distinct impact of UPSTART on children's literacy outcomes, and the impact of the combination of UPSTART and EISP.

Research Questions

The research questions used to guide the direction of our first-grade analysis are as follows:

> **Research Question 3.1:** *Does the use of a home-based, computer-supported literacy skills training program in preschool result in stronger school-based literacy outcomes at the beginning of first grade compared to a group of peers matched in terms of demographic characteristics who did not receive the preschool program?*

We hypothesized that if UPSTART has no effect on sustaining students' literacy skills through the first grade, then the children who participated in UPSTART (the treatment group) would perform at the same level as a comparison control group (children who were not exposed to UPSTART or EISP) on measures of literacy development at the beginning of first grade. If UPSTART does have a continued impact on students' literacy achievement, then the treatment group should perform significantly better than the control group on literacy measures at the beginning of first grade.

However, in light of calls for investigation of aligned preschool-elementary school curricular approaches in sustaining preschool benefits (Jenkins et al., 2016), we conducted an explorative analysis of the impact of participating in both the UPSTART and EISP programs. Would participation in UPSTART during the preschool year, coupled with participation in EISP during the kindergarten year, lead to stronger literacy outcomes compared to students who did not participate in either program? Our second research question addresses this line of inquiry:

> **Research Question 3.2:** *Does the use of a home-based, computer-supported literacy skills training program in preschool* ***along with a computer-based kindergarten program*** *result in stronger school-based literacy outcomes at the beginning of first grade compared to a group of peers matched in terms of demographic characteristics who did not receive the preschool or kindergarten program?*

If UPSTART and participation in the EISP program has a continued impact on students' literacy achievement, then we would expect children who were enrolled in UPSTART preschool and participated in EISP to have significantly stronger performance on first grade literacy measures when compared to comparison students who did not participate in either program.

Methods

This section describes the research methods used to answer our research questions, including the research design, outcome measures, data sources, and procedures utilized to create the analytic sample.

Research Design

Due to the fact that we do not have pre-program data for the complete sample of participating students, we elected to implement the first grade evaluation of the UPSTART preschool program as a nonequivalent groups post-program only design. The evaluation design is diagrammed below in **Table 13**.

Treatment children participated in UPSTART during the seventh year of implementation (Cohort 7) the 2015-16 preschool year. While the control group remains constant (children with no UPSTART exposure or participation in EISP), the treatment group varies based on our specific analytic goals. When answering **Research Question 3.1** and exploring the unique impact of UPSTART on children's first grade literacy achievement, the treatment group consists of students who only used UPSTART. The UPSTART + EISP group is used as the treatment group to answer **Research Question 3.2** and investigate the combined effects of enrolling in UPSTART preschool program and participating in the EISP program.

Table 13
First Grade Analysis Evaluation Design

		Preschool 2015-16	**Kindergarten 2016-17**	**First Grade 2017-18**
Treatment	**UPSTART only**	UPSTART	No Program	
	UPSTART + EISP	UPSTART	EISP Program	
Control	**Control (no program use)**	No Program	No Program	
Measure				Post-Test Only DIBELS BOY 1st Grade

Because the first grade analysis necessitates a quasi-experimental design in which the treatment and control groups are not completely equivalent on factors that may influence reading achievement outcomes, 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, however, equate the groups on the basis of Beginning of the Year (BOY) Kindergarten DIBELS scores. It has been demonstrated that UPSTART students enter the school setting with higher literacy scores than comparison students, and negating that effect through statistical controls would not be an accurate representation of the short-term impact of UPSTART.

Measures

Our outcome measure consisted of the DIBELS, a standardized measure of literacy achievement for elementary school students. The DIBELS is administered to students in Grades K-3 in schools throughout the state. At the beginning of the year of kindergarten (BOY), the DIBELS measures children's competency with the alphabetic principle and basic phonic with the Letter Naming Fluency and First Sound Fluency subtests. The subtests administered at the second half of kindergarten (middle of year - MOY and end of year - EOY) and beginning of first grade (BOY) assess 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 14**).

Table 14
DIBELS Next Subscales by Administration Period

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

The DIBELS Composite score is an overall measure of children's early literacy ability and is calculated by summing the subtest scores associated with each test administration period. The DIBELS Kindergarten Composite score serves as our baseline and the First Grade Composite score serves as our outcome measure.

Data Sources

We relied on data from four different sources to create our final dataset and complete our analyses, including demographic data, literacy achievement scores, UPSTART usage information, and participation in the EISP educational software program.

- The USBE provided demographic data for students enrolled in first grade during the 2017-18 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. The USBE included a flag to identify students who were enrolled in UPSTART preschool during the 2015-16 school year.

- DIBELS Next data was provided by the USOE and consisted of Kindergarten DIBELS scores during the 2016-17 school year and beginning of year (BOY) First Grade DIBELS score from Fall 2017.
- Student-level data detailing UPSTART preschool software usage for children enrolled in Cohort 7 during 2015-16 was provided by the Waterford Institute. All students who were enrolled in UPSTART were included in our analysis, regardless of the amount of time they used the program.
- We used archival data from the EISP evaluation to identify and flag kindergarteners who participated in the EISP program during the 2016-17 program year. All students who participated in the EISP program were included in our analysis, irrespective of use.

Merged Data File

We removed instances of duplicate cases and records with missing SSIDs, baseline scores (DIBELS Kindergarten BOY) or outcome scores (DIBELS First Grade BOY) and systematically merged the data files together, using state provided identification numbers (SSIDs). Cases may have failed to merge due to students skipping or repeating grades, having incorrect SSIDs entered into the data file, or leaving the public school system (e.g., moving out of state, enrolling in home school). The complete merged data file consisted of a total of 37,446 cases, broken out into the following independent groups:

Table 15
Group Sizes for Unmatched First Grade Analysis File

Group	N=
UPSTART only	2,704
UPSTART + EISP	1,772
EISP only	12,518
Control	20,452
Total	**37,446**

One of the shortcomings of post-test only designs is selection bias, or that it is difficult to determine of any observed post-test differences between the treatment and treatment group are due to preexisting differences. In an effort to address this issue, we utilized CEM to create balanced matched samples to statistically control for significant differences between our treatment and control groups. (For a detailed discussion of CEM, please see the **Cohort 8 Evaluation**.) Our final analytic samples consisted of two data files: (1) one data file containing UPSTART only students (N = 2,701) and a matched comparison sample (N = 2,701) of students did not have UPSTART or EISP program experience and (2) a second data file containing UPSTART plus EISP students (N = 1,772) and a matched comparison sample (N = 1,772) of students who did participate in either program.

Table 16 presents key demographic characteristics for the matched analytic sample of students who only participated in the UPSTART preschool program (UPSTART Only) and their matched comparison students.

Table 16
UPSTART Only and Control Student Comparisons on Key Demographics

Demographic Categories		Treatment (N=2,701)	Control (N=2,701)
Child Gender	Female	50%	50%
	Male	50%	50%
Child Race	Caucasian	79%	79%
	Hispanic	15%	15%
Child Language	English Language Learner	9%	9%
Title 1 School	Yes	26%	26%
	Targeted for Individual Students	13%	13%
Household Income	Low Income	33%	33%

The demographic characteristics of students who participated in UPSTART as preschoolers and were enrolled in a kindergarten classroom that received EISP program software (UPSTART + EISP) and their similarly matched comparison students were displayed in **Table 17**.

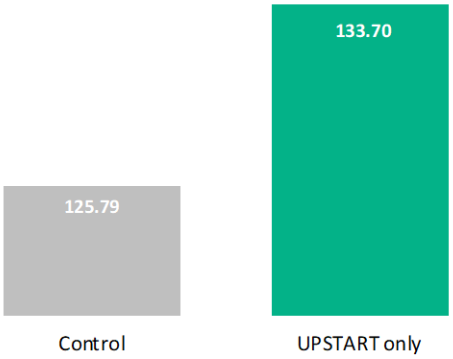
Table 17
UPSTART Only + EISP Program – Control Comparisons on Key Demographics

Demographic Categories		Treatment (N=1,772)	Control (N=1,772)
Child Gender	Female	49%	52%
	Male	51%	48%
Child Race	Caucasian	80%	80%
	Hispanic	15%	15%
Child Language	English Language Learner	9%	9%
Title 1 School	Yes	37%	37%
	Targeted for Individual Students	15%	15%
Household Income	Low Income	37%	37%

Findings

Our first set of analyses looks at the impact of enrolling only in the UPSTART preschool program on first grade literacy outcomes. When compared to a group of comparison students matched on demographic characteristics, we find evidence that first grade beginning of year (BOY) DIBELS scores are significantly higher for children who were enrolled in the UPSTART preschool program. Specifically, as seen in **Figure 15**, UPSTART students had an average BOY first grade DIBELS composite score of 133.70 compared to the average score of 125.79 for control students, a 7.91 point difference.

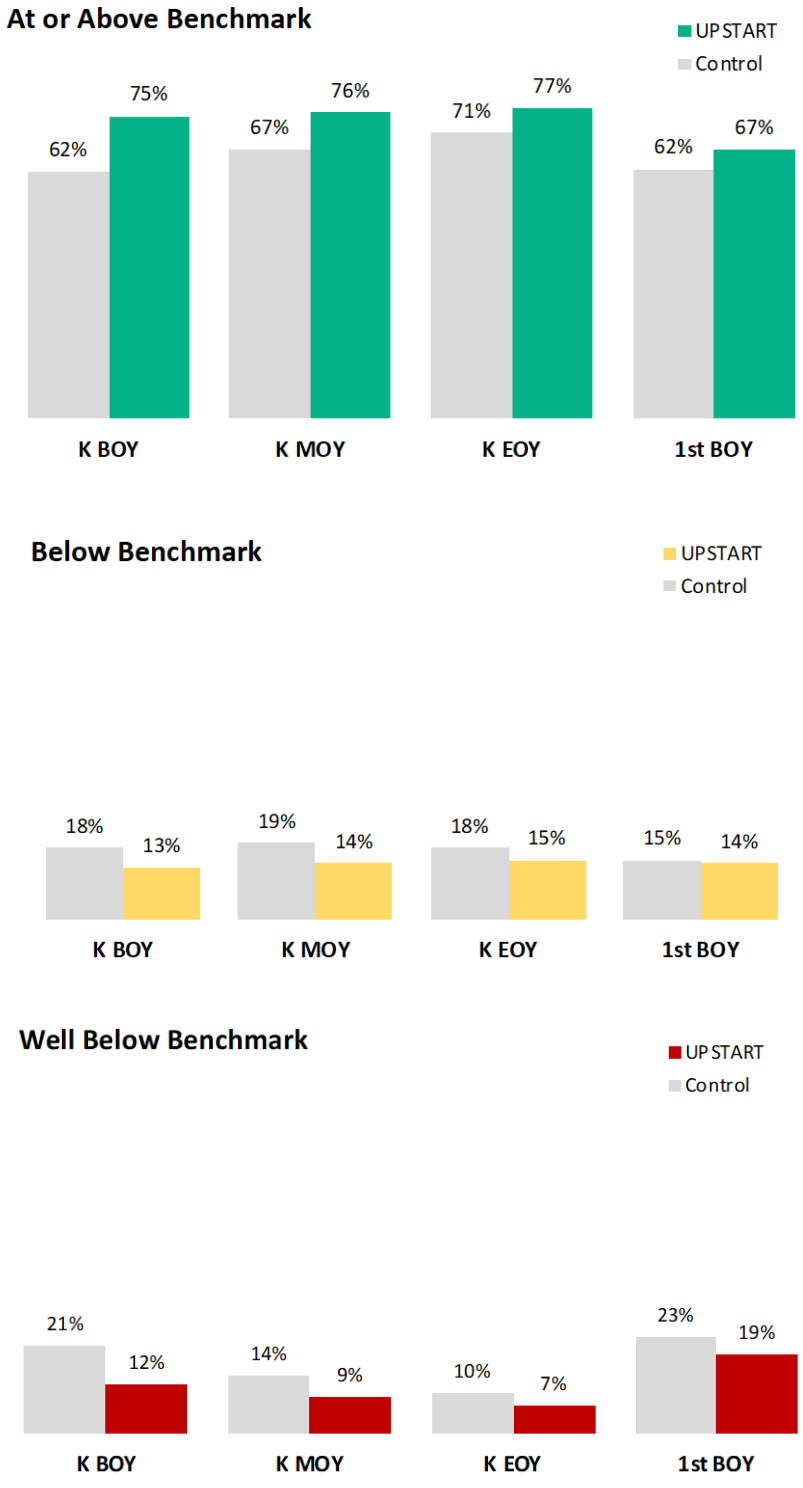
Figure 15. First Grade DIBELS Composite Scores UPSTART and Control students



The difference between UPSTART participants and their non-program comparison counterparts on the first grade BOY DIBELS composite produced an effect size of .18, which is less than the .26 effect size benchmark for similar interventions and evaluation studies. (For a more detailed discussion of effect size, please see the **Cohort 8 Evaluation**). An analysis of DIBELS composite scores at testing periods at the beginning, middle, and end of kindergarten and at the beginning of first grade using independent t-tests indicate that UPSTART children performed significantly higher on the DIBELS composite throughout kindergarten and at the beginning of first grade when compared to a group of control children who did not participate in UPSTART.

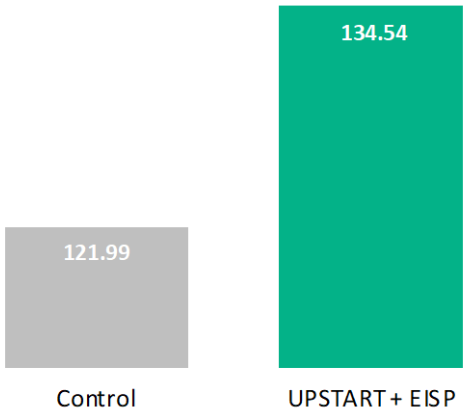
The bar graphs in **Figure 16** show the performance of children who participated in the UPSTART program with children who were not UPSTART participants on the DIBELS composite benchmark classifications that are measured at multiple time points in kindergarten and the beginning of first grade. DIBELS benchmarks are empirically derived cut points that indicate adequate reading skill for a particular grade and time of year and are categorized as 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 16** bar graphs, 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. Interestingly, both UPSTART and comparison students had lower levels of literacy achievement at the beginning of first grade (67% of UPSTART children and 62% of comparison children categorized at or above benchmark) compared to the end of kindergarten (77% of UPSTART children and 71% of comparison children categorized at or above benchmark).

Figure 16. Literacy Benchmarks Over Time: UPSTART only and Comparison Students



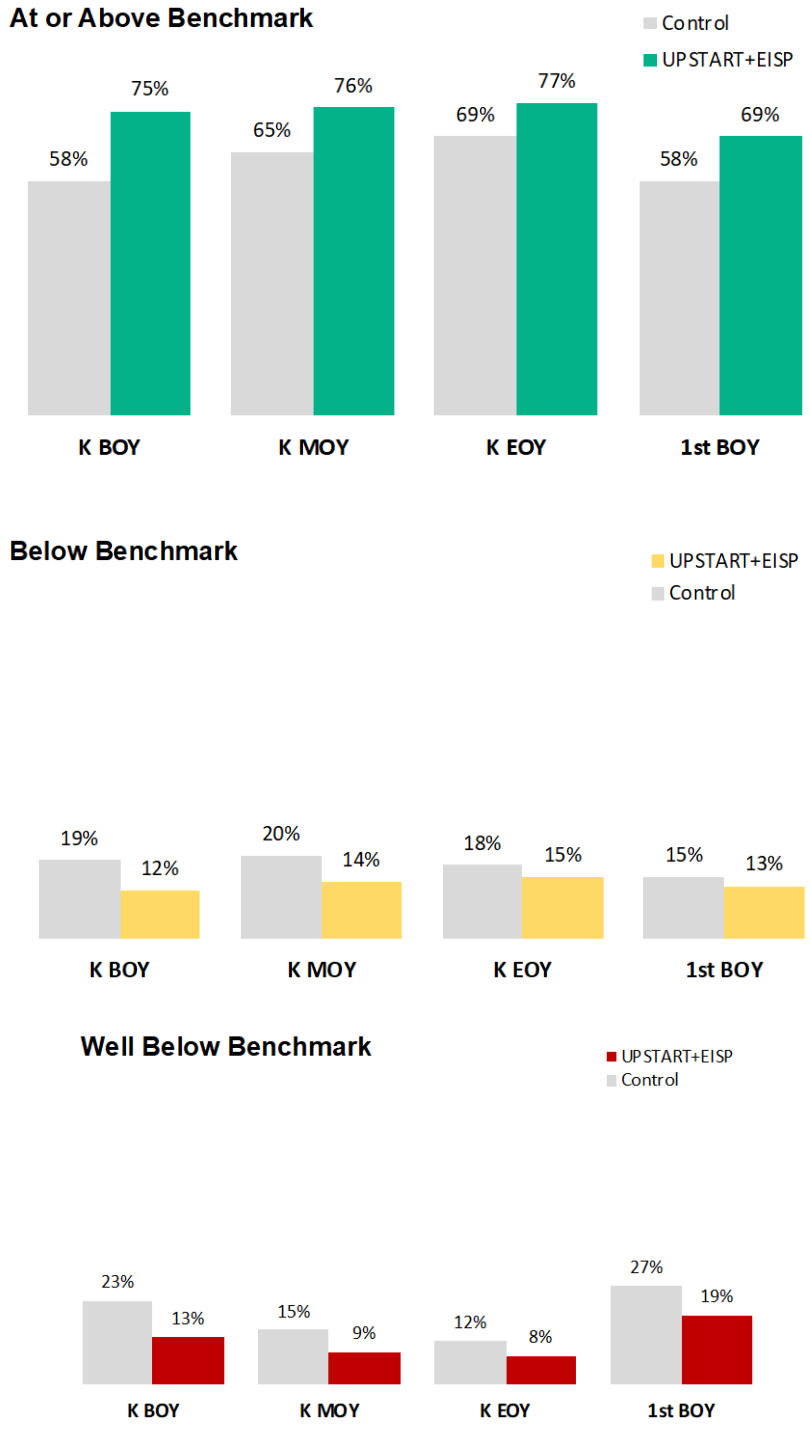
The second set of analyses takes into account the presence of the statewide EISP software program initiative and evaluates the impact of participating in UPSTART and a similar adaptive computer-based program that provides individualized literacy instruction throughout kindergarten. We found that students who participated in UPSTART during preschool and EISP during the kindergarten academic year had significantly higher scores on the first grade DIBELS composite than students who did not participate in either program. As seen in **Figure 17**, mean scores on the first grade DIBELS composite were 134.54 for the UPSTART + EISP treatment group and 121.99 for students who did not receive any literacy software, a 12.54 difference. This difference produced an effect size of .27, which is above the .26 effect size benchmark for similar studies reported in the literature and should be considered practically significant.

Figure 17. First Grade DIBELS Composite Scores UPSTART + EISP and Control students



The bar graphs in **Figure 18** show the overall performance of children who participated in the UPSTART and EISP programs with children who participated in neither on the DIBELS composite benchmark classifications measured throughout kindergarten and at the beginning of first grade. Similar to the results in the UPSTART only analysis, children who received instruction from UPSTART and EISP programs outperformed comparison students throughout kindergarten and into first grade. Children who participated in UPSTART and EISP were more likely to be classified as at or above literacy benchmarks at each assessment period, and less likely to be classified as below or well below benchmarks (see **Figure 18**). There is also an analogous pattern to the UPSTART only analysis of lower levels of literacy achievement at the beginning of first grade, with 69% of UPSTART + EISP children and 58% of comparison children categorized at or above benchmark in first grade, compared to 77% of UPSTART children and 69% of comparison children categorized at or above benchmark at the end of kindergarten.

Figure 18
Literacy Benchmarks Over Time: UPSTART/EISP and Comparison Students



Summary and Discussion

Our exploratory first grade analysis moved beyond evaluating the immediate impact of the UPSTART preschool program on preparing children for entry into traditional school environments to assess whether or not UPSTART has a sustained benefit on children's literacy achievement once children are in elementary school. Specifically, we followed Cohort 7 students through kindergarten and first grade and utilized a post-test only design to determine if UPSTART participants had higher scores on the first grade DIBELS assessment compared to students who were not enrolled in UPSTART. In an effort to isolate the effects of participating in the EISP program, a statewide computer-based literacy instruction software program for grades K-3, we excluded any student who participated in EISP as a kindergartener from our control group. We also created two treatment groups to examine potential multiplier effects from participating in both programs: students who only participated in UPSTART during their preschool year (UPSTART only) and students who participated in UPSTART as preschoolers and who participated in the EISP program as kindergarteners (UPSTART + EISP).

We found significant small effects for the sustained benefit of UPSTART. UPSTART has a positive impact on students without additional curricular support (the effect size of the UPSTART only group was .18) and an even larger impact on students who receive further individualized computer-based instruction (the effect size of the UPSTART + EISP group was .27). The effect of participating in both UPSTART and the EISP program was larger than the average effect size reported in similar evaluations with comparable interventions, measures, and students.

Because we used all students who participated in the UPSTART or EISP programs, regardless of the amount students actually used the programs, our treatment samples are considered "intent to treat" (ITT) samples. ITT samples represent the most conservative estimate of the long-term impact of UPSTART because it includes students who met vendors' requirements for program use as well as students who may have only used the program sporadically or not at all (Montori & Guyatt, 2001). However, other researchers argue that if a participant is included in the treatment group, but did not actually receive treatment, it indicates little about the treatment's efficacy (Gupta, 2011). To that end, we recommend that future analysis of the long-term effects of UPSTART include a subsample of UPSTART users who fulfilled program requirements for usage.

Summary and Recommendations

The UPSTART program shows continued success at helping preschool age children develop literacy skills and prepare for entry into kindergarten. There is also evidence UPSTART students' literacy achievement is sustained throughout kindergarten and into first grade. Given the success at improving literacy test scores, we recommend that the state continue providing the UPSTART program to children. The strong program effects support wide-scale implementation across a variety of preschool populations.

During the 2016-2017 program year, more C8 students were classified as graduates when compared to the previous cohorts (89% graduation rate in C8 versus an 87% graduate rate in C7) even in the face of a 62% enrollment increase. It is important to continually monitor program usage as children who failed to meet the program requirements for graduation had, on average, significantly lower literacy outcome scores when compared to UPSTART graduates. Moreover, our families that did not meet usage requirements were more likely to have other indicators of risk, such as lower levels of parental education, lower household incomes, and being non-native English speakers. Graduation rates need to be carefully monitored because a significant decline might erode literacy outcomes for the most at-risk students.

Program Recommendations. Although the graduation rates for C8 students were slightly higher than the previous year, as UPSTART continues its expansion it is important to continually monitor program implementation to be sure that increased enrollment does not erode graduation or usage rates, two key areas for ensuring strong student literacy achievement and future program success. Specifically, we recommend that the program vendor consider the following recommendations:

- The program vendor could develop new strategies for addressing falling usage and graduation rates among the most at-risk students (i.e. those with high levels of poverty and with English as a second language). Some potential strategies might include:
    - Establishing peer support systems among similar groups to discuss strategies for supporting children's program use.
    - Highlighting evaluation information that links graduation with higher literacy outcomes.
    - Developing targeted incentives for families with the highest risk factors for not meeting program usage requirements, such as monthly awards (extrinsic), being highlighted in UPSTART communications to social networks as "Gold Star Families" (intrinsic).

Results from the first grade analysis indicate that UPSTART children were able to maintain their advantage in literacy outcomes through the beginning of first grade, and that these effects were greater for UPSTART children who participated in the EISP program. Because the EISP program also provides students with individualized adaptive computer-based literacy instruction, it provides a logical support to build on the gains created by UPSTART and we recommend continuing the program.

Evaluation Recommendations. We recommend that the matched treatment and control group design be used for future evaluations. This research design depends on collecting sufficient data from control students to allow high matching rates to treatment students. To accomplish these high match rates, we also recommend that the state work with the evaluators to strengthen relationships with other preschool providers that serve low-income families, specifically Head Start organizations, WIC and public preschool programs to widen our ability to collect data from non-program control families. This strategy is a win-win for all involved: low-income families can help move the bar on research into early literacy (and receive financial incentives while doing it) and the state can review results across more students and have more data for evidence-based decision making about their pre-Kindergarten school readiness programs.

References

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

Brigance, A. H. (2004). Brigance Inventory of Early Development II (IED-II) (2nd ed.). N. Billerica, MA: Curriculum Associates.

Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304.

Evaluation and Training Institute. (2017, April). *Utah UPSTART program evaluation program impacts on early literacy: Year 7 Results* (Cohort 7 Technical Report). Culver City, CA: Author.

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research*, *2*(3), 109-112. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/

Guryan, J., Hurst, E., & Kearney, M. (2008). Parental education and parent time with children. *Journal of Economic Perspectives*, *22*(3), 23-46.

Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E., Clements, D., Sarama, J., Duncan, G. J. (2016). *Do high quality kindergarten and first grade classrooms mitigate preschool fadeout?* Irvine Network on Interventions in Development.

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.

Lipsey, M., Weiland, C., Yoshikawa, H., Wilson, S., & Hofer, K. (2015). Prekindergarten age cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis*, *37*, 296-313.

Mistry, R. S., Benner, A. D., Biezanz, J. C., Clark, S. L., & Howes, C. Family and social risk, and parental investments during the early childhood years as predictors of low-income children's school readiness outcomes *Early Childhood Research Quarterly*, *25*, 432-449.

Montori V.M. &, Guyatt G. H. (2001) Intention-to-treat principle. *Canadian Medical Association Journal*, *165*(10), 1339-1341. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC81628/

Neitzel, C., & Stright, A. D. (2004). Parenting behaviors during child problem solving:

The role of child temperament, mother education and personality, and the problem-solving context. *International Journal of Behavioral Development*, *28* (2), 166 - 179.

Phelps, S. (2003). *Phonological Awareness Training in a Preschool Classroom of Typically Developing Children*. Electronic Theses and Dissertations. Paper 772. http://dc.etsu.edu/etd/772

Puma, M., Bell, S., Cook, R., Heid, C. (2010). *Head Start Impact Study. Final Report*. U.S. Department of Health and Human Services, Administration for Children & Families. Washington, DC.

Shadish, Cook, and Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating math recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal*, *50*(2), 1–32.

Snow, C.E., Burns, M., S., & Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Vandell, D. L., Belsky, J., Burchinal, M., Vandergrift, N., & Steinberg, L. (2010). Do effects of early child care extend to age 15 years? Results from the NICHD Study of Early Child Care and Youth Development. Child Development, 81(3), 737-756.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills. *Child Development*, *84*(6), 2112–2130.

What Works Clearinghouse. (2017). Procedures handbook (version 4.0). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Appendix A: Comparison of C8 Evaluation Samples

The matched and unmatched samples are compared with the C8 population on key demographic characteristic reported by the program vendor in **Table A.1**. Both of the unmatched and matched samples are more homogenous than the full population of preschoolers who were enrolled in Cohort 8, with 82% and 85% of unmatched and matched children, respectively, being Caucasian and 98-99% classified as English speakers.[10]

Table A.1
Sample Comparisons on Key Waterford Demographics

Demographic Categories		C8 Population (N = 10,745)	Unmatched Sample (N=275)	Matched Sample (N=245)
Gender	Male	51%	51%	51%
	Female	49%	49%	49%
Ethnicity	Caucasian	79%	82%	85%
	Hispanic	13%	12%	9%
Child Language	English	90%	98%	99%
Parent Education Level	Some College	36%	67%	69%
	Bachelor's Degree	37%	7%	7%
Parent Marital Status	Married	91%	87%	87%
Poverty Status	Under 185%	42%	83%	82%

The C8 population had parents with higher college graduation levels and lower levels of poverty. Whereas 37% of the parents in the overall C8 population have a college degree, the modal level of parent education in the unmatched and matched sample was some college (67% and 69%, respectively). Additionally, 42% of families in the C8 sample were under the 185% federal poverty rate compared to 83% of families in the unmatched sample and 82% of families in the matched sample. As mentioned in the main body of the report, we focused on recruiting low-income families for our treatment sample to reflect the prioritization of these families by the state in the recent legislative extension of the UPSTART program.

The unmatched sample is slightly closer to representing the characteristics of the C8 population. However, the matched sample ensures that the treatment group's characteristics best mirror the control group to estimate program impact with the greatest accuracy. UPSTART outcome findings are reported in the main body of the report from the matched treatment-control sample.

[10] The testing protocol tests all children in English and requires children to understand directions in English and give verbal assent to proceed with testing. Moreover, parents need to have sufficient understanding of English to give informed consent for their participation.

Appendix B: Determining UPSTART Effect Size Benchmark

One way to assess the practical significance of an intervention is to compare its impact with effect sizes from similar evaluation studies – those that use analogous outcome measures, are evaluating a comparable intervention, or are evaluating interventions that target similar groups. Researchers at the Institute of Education Sciences (IES) reviewed 829 effect sizes from 124 education research studies conducted on K-12 students and reported an array of different effect size distributions that can provide insight into what constitutes a large or small effect relative to similar education evaluation studies (Lipsey et. al, 2012). They provide the following benchmarks to be used as normative comparisons:

- *Benchmark by outcome measure*. IES researchers looked at the type outcome measures (i.e., did researchers use a self-developed outcome measure, a general standardized outcome measure like an IQ test, or a subject-specific standardized outcome measure like a reading or math test) by grade level and found that the average effect size for education research studies evaluating elementary students with a standardized subject test (like the Brigance and Bader literacy tests) was .25. Average effect sizes were slightly higher for middle school students, but lower for high school students (.32 and .03, respectively)
- *Benchmark by intervention type*. Another metric for evaluating effect size was based on the type of intervention under investigation. Researchers sorted the interventions of reviewed studies into several broad categories (e.g., a whole school program, a teaching technique, a new instructional format, skill training, or an instructional program). The UPSTART program was closest to an instructional program, or "a relatively complete and comprehensive package for instruction in a content area like a curriculum or a more or less free standing program (e.g., science or math curriculum; reading programs for younger students; broad name brand programs like Reading Recovery; organized multisession tutoring program in a general subject area." (p. 35) The average effect size for research studies that evaluated a comprehensive instructional program such as UPSTART was .13. Larger effect sizes were found for interventions in the instructional component/skill training and teaching techniques and categories (.36 and .35, respectively).
- *Benchmark by intervention target*. A final yardstick to contextualize effect sizes focused on the targeted group of the intervention (e.g., individual students, small group, classroom, whole school, mixed.) that targeted individual students had average effect sizes of .40. Interventions that targeted individual students had the highest observed effect sizes, on average.

To determine a single benchmark, we took an average of the three different benchmarks (i.e., benchmark by outcome measure = .35; benchmark by intervention type = .13; and benchmark by intervention target = .40) and the resulting benchmark value was .26. This benchmark will be used to contextualize the effect sizes presented in this report and to aid the reader in determining the practical significance of the effect of UPSTART.

Appendix C: Cross Cohort Analysis Growth Rates

Growth Rates for UPSTART Graduates and UPSTART Non-Graduates

Brigance Test	Non-graduate (N=75) Mean Growth	Graduate (N=75) Mean Growth	Significance	Effect Size
Visual Discrimination	4.23	5.27	NS	.20
Letter Sounds	10.02	10.75	NS	.08
Auditory Discrimination	2.29	2.10	NS	.05
Survival Sight Words	.72	1.74	**	.51
Basic Vocabulary	2.92	7.47	***	.60
Total Brigance	20.18	27.33	**	.44

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

Brigance Test	UPSTART only (N=357) Mean Growth	UPSTART/PreS (N=357) Mean Growth	Significance	Effect Size
Visual Discrimination	4.40	4.33	NS	0.02
Letter Sounds	10.65	11.95	NS	-0.15
Auditory Discrimination	2.58	2.65	NS	-0.02
Survival Sight Words	2.25	2.01	NS	0.08
Basic Vocabulary	8.05	7.94	NS	0.01
Total Brigance	27.94	28.89	NS	-0.06

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