Utah UPSTART Program Evaluation
Program Impacts on Early Literacy

Year 5 Results

Cohort 5 Technical Report

March 2015

Submitted by:

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

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

Contents

Executive Summary

Established as a pilot demonstration project by the Utah State legislature, UPSTART uses educational technology in a home-based approach to develop the school readiness skills of preschool children. The Waterford Institute implemented the project and provided documentation for a fifth-year UPSTART enrollment of 1,577 children. The vast majority (70%) of the 1,577 preschool children who enrolled in the fifth year of UPSTART were from low income families, according to data provided by the Waterford Institute. Slightly more girls (50.5%) were enrolled than boys (49.5%). In terms of ethnicity, the vast majority of the Cohort 5 (C5) enrollment was Caucasian (74%) with 20% being of Hispanic origin. The ethnicity of the remaining 6% of the C5 enrollment was of African American?, Asian, Native American, Pacific Islander, or unknown origin.

Evaluation Design

The evaluation of UPSTART's fifth year of implementation used a pretest-posttest control group design to *assess the program's impact on developing the children's early literacy skills in preschool*. Other objectives included (a) documenting the extent to which participants used the computerized curriculum; (b) establishing the relationship between curriculum usage and literacy outcomes; and (c) documenting the degree to which the C5 participants met the program's curriculum usage criteria, otherwise referred to as the "graduation" rate.

An ordinary least squares (OLS) regression approach was used to estimate posttest differences in the development of literacy skills between a sample of UPSTART participants (the treatment group) and a group of similar nonparticipants (the control group) in the year prior to enrollment in Kindergarten. The children were measured on two tests of early literacy skills: the Brigance Inventory of Early Development and the Bader Reading and Language Assessment. Covariates used in the analyses to adjust for initial between group differences included pretest scores on the respective tests and selected demographic characteristics that independently influenced posttest outcomes. Additionally, differences between the treatment and control groups in their early literacy growth rates as measured by the two tests were examined.

The effect of UPSTART usage on literacy skill development was examined for UPSTART participants using an analysis of covariance in which usage levels were split into quartiles based on the usage distribution of the preschool analysis sample. The statistical models controlled for the children's initial level of literacy development, as measured by the pretest score on each of the two respective tests (i.e., the Brigance and the Bader). The effect of UPSTART usage on literacy skill development was assessed by comparing the adjusted mean posttest performance on the Brigance and the Bader at each usage quartile with the fourth quartile of usage. The purpose

of the analysis was to determine whether literacy development increased with increasing usage of the UPSTART curriculum.

Descriptive statistics were also computed to describe the population of students that enrolled in the fifth year of UPSTART (i.e., Cohort 5, hereafter abbreviated as C5). The C5 UPSTART population descriptors included student demographics, the equipment that C5 participants received, hours of UPSTART curriculum usage, and the graduation status of C5 students. Graduates were participants who met the UPSTART program's curriculum usage criteria.

Data Collection

271 preschool children were recruited for the C5 evaluation study; 109 treatment group children who had enrolled in UPSTART for Year 5 of the program and 162 control group children who had not enrolled in the UPSTART program. The 109 treatment group children came from an initial random sample of 150 C5 enrollees whose families were contacted about participating in the C5 evaluation, and recruited for testing. The control children were recruited from preschools, daycare centers, and other childcare organizations in Utah. The children's parents were given an intake questionnaire (see Appendix A) at the time their children were pretested on the Brigance and Bader during the summer of 2013. The children were subsequently posttested on the Brigance and Bader a year later during the summer of 2014.

The final analysis sample for the C5 evaluation used the data collected from 194 children. The UPSTART treatment group was composed of 94 children who passed the screening interview and were able to provide valid and matched pretest and posttest data on the Brigance (89 of whom also had valid and matched Bader data). The non-UPSTART control group was composed of a random sample of 100 control children who were selected from the 141 cases with valid and matched Brigance and Bader test data. These two files were then merged to form the final analysis sample in which the treatment and control group samples were reasonably balanced in size.

Data Analysis

To determine whether UPSTART children had developed better literacy skills at Kindergarten entry compared to control group children, group equivalence on the pretests was first examined using independent sample t-tests. Relationships between the demographics and posttest scores were then examined using correlation analyses. Next, posttest differences between the treatment and control groups were examined for both the Brigance and Bader. Finally, posttest differences were re-examined by adjusting for initial differences between the treatment and control groups with the use of multiple regression analysis. The regressions used a hierarchical block design in which the pretest and a set of demographic covariates were entered first, followed by the treatment-control group indicator.

The magnitude of UPSTART's impact was also estimated using effect size estimates as measured by standardized treatment and control group differences on the Brigance and Bader posttests. Effect sizes for each of the two tests and their subtest were interpreted as demonstrating a small, medium or large effect on the children's early literacy development.

To determine whether UPSTART students exhibited stronger literacy growth rates relative to control students from preschool to kindergarten, paired sample t-tests were run to obtain gain scores (i.e., posttest minus pretest) for the matched Brigance and Bader treatment groups on the total test and each of the subtests. Statistically significant growth rates were determined by examining confidence intervals for the treatment and control groups for each test measure at the 99% confidence interval.

The effect of UPSTART usage on reading proficiency was examined for UPSTART participants using an analysis of covariance (ANCOVA) in which usage levels were split into quartiles based on the usage distribution of the preschool analysis sample. The final ANCOVA models statistically controlled for initial literacy skills as measured by the pretest on each respective measure. The ANCOVA analyses estimated the effects of usage at quartiles one through three compared with usage at the fourth quartile, controlling for initial levels of literacy development.

UPSTART Implementation Results

As in previous years, most of the C5 participants (74%) received a computer drive with the UPSTART curriculum loaded on it. Approximately 8% of the fifth year participants received a computer loan and a free Internet subscription to help them access the UPSTART curriculum. Another 7% of the C5 participants were loaned a personal computer to use at home while participating in UPSTART. The remaining 11% of the fifth year participants were provided with various combinations of educational technology to enable them to access the UPSTART curriculum, including wireless and cellular devices.

Findings about UPSTART curriculum usage are summarized below.

- The C5 preschool population averaged approximately 71 hours of instruction over the 2013-2014 school year. This is the same level of average usage as found in the C4 program year.
    - By comparison, C5 program "graduates" averaged approximately 75 hours of instruction.
    - The C5 analysis samples had a mean of approximately 72-73 hours of usage.
- The UPSTART graduation rate continued to hold at 94% in Year 5. This is the same level of program attainment as realized in Program Years 3 and 4.

- UPSTART graduate status in the fifth year of the program was significantly correlated with hours of instruction (r=.58) as well as with the duration of program participation as measured by weeks of attendance (r=.74).

- UPSTART graduate status in the fifth year of the program was significantly correlated with early literacy outcomes as measured by the Brigance (r=.24, p<.05) and the Bader (r=.22, p<.05) at the beginning of Kindergarten.

- UPSTART curriculum usage among C5 children was significantly correlated with literacy skills at the beginning of Kindergarten as measured by the Brigance (r = .45, p<.01) and the Bader (r = .41, p<.01). Controlling for initial literacy skills, the correlation of UPSTART usage with Kindergarten outcomes was somewhat lower: r=.34 (p<.01) for the Brigance sample and r=.40 (p<.01) for the Bader sample.

- UPSTART usage accounted for 16 to 19% of the variance in literacy skills developed by C5 children as measured by the Bader and Brigance posttests respectively at Kindergarten entry.

UPSTART Impact Results

In the fifth year of the program, UPSTART was observed to have a small overall impact on the development of participating children's early phonics skills as measured by the *Brigance* assessment. Adjusting for pre-existing differences, UPSTART participants on the average scored almost 12 points higher on the overall Brigance relative to control group children. Significant program impacts (relative to control group performance) were found in six of the ten areas assessed by the Brigance. Compared to similar nonparticipants, UPSTART produced:

- Small effects in helping young children learn how to recite the alphabet, name and recognize lower case letters, as well as produce the sounds of lower case letters;
- Medium size effects in helping children learn how to hear and see differences in letters and words; and
- Large effects in helping children learn how to read basic vocabulary words found in pre-primer reading programs.

- Relative to controls, UPSTART children showed significantly stronger growth rates in (a) learning how to pronounce letter sounds, (b) learning how to tell the difference between letter sounds, and (c) in developing their vocabulary.

- As noted above, the largest phonics-related impact observed in UPSTART's fifth year of operation was in the development of *vocabulary*. This finding replicates results found in the third and fourth year evaluations of UPSTART.

- Overall, UPSTART achieved medium size effects on improving the phonological awareness skills of participants in Year 5 of the program as measured by the Total Bader assessment. These results replicate the evaluation findings from Year 4 of the program.

- On average, UPSTART treatment group children scored an average of almost six points higher on the Bader posttest (regression adjusted) relative to control group children. Compared to similar nonparticipants, UPSTART produced:

    - Small effects in helping young children recognize pairs of words that rhymed; and
    - Medium size effects on helping young children learn how to blend and segment phonemes.

- Relative to controls, the UPSTART treatment group showed stronger growth rates from pretest to posttest in Phoneme Blending and Phoneme Segmenting skills as well as on the overall Bader assessment.

UPSTART strives to improve preschool children's knowledge of letter sounds, develop their early reading vocabulary, and help young children learn how to pronounce words and begin to read. In its fifth year of operation, UPSTART made substantial progress in helping young children reach these early literacy development goals in preparing the children for Kindergarten entry.

Introduction

UPSTART is a pilot 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 fifth year of operation during the 2013-14 school year, the project's implementation contractor – the Waterford Institute – enrolled 1,577 preschool children and provided them a game formatted program of early literacy instruction delivered by personal computers and the Internet, designed to prepare them academically for kindergarten. The 1,577 children enrolled in the fifth year cohort, hereafter referred to as C5, participated in UPSTART from September 2013 through June 2014.

The evaluation of UPSTART's fifth year of implementation used a pretest-posttest control group design to *assess the program's impact on developing the children's early literacy skills in preschool*. Other objectives included (a) documenting the extent to which participants used the computerized curriculum; (b) establishing the relationship between curriculum usage and literacy outcomes; and (c) documenting the program's completion or "graduation" rate as measured by the proportion of the enrollment that met the criteria established for usage of the program's curriculum.

Slightly more C5 girls (50.5%) were enrolled than boys (49.5%). In terms of ethnicity, the vast majority (74%) of the C5 enrollment was Caucasian, with 20% of the children being of Hispanic origin. The ethnicity of the remaining 6% of the C5 enrollment was composed of children from African American, Asian, Native American, Pacific Islander and unknown backgrounds. The primary language spoken by the vast majority of the C5 children was English (84%). Approximately 15% of the C5 children spoke Spanish and 1% spoke other languages. Twelve percent of the C5 children had a diagnosed disability, most often speech impairments.

A majority (70%) of the 1,557 preschool children who enrolled in the fifth year of UPSTART were from low income families, according to data provided by the Waterford Institute. Most commonly, the C5 parents had some college (35%) or had achieved a bachelor's degree (37%). The vast majority of the C5 parents were married (89%).

C5 Evaluation Methods

The Cohort 5 evaluation continued to use the quasi-experimental research design variant of the nonequivalent comparison group design described previously. Specifically, the design implemented in the C5 evaluation (as in the previous three years) used a treatment group and an untreated comparison group, with both pretest and posttest data collected on the same children over a 12 month interval during the year prior to enrollment in Kindergarten. The design is

diagramed below. NR indicates that the evaluation was a quasi-experiment since the children were not randomly assigned to groups.

The C5 study recruited 271 preschool children: 109 treatment group children who had enrolled in UPSTART for Year 5 of the program (the 2013-2014 school year) and 162 nonparticipating control group children. The 109 treatment group children came from an initial random sample of 150 C5 UPSTART enrollees whose families were contacted about participating in the C5 evaluation. Approximately 130 of these UPSTART families agreed to participate in the C5 evaluation. Of these 130 families, 21 children were lost from the study for reasons related to testing feasibility. That is, they either could not be tested due to a diagnosed disability, they spoke a language other than English, or they failed to show up for pretesting. The remaining 109 UPSTART children subsequently participated in pretesting prior to entering the program over the summer of 2013. The control children were recruited from preschools, daycare centers, and other childcare organizations in Utah and pretested over the summer of 2013. Both the treatment and control children were posttested over the summer of 2014.

In the diagram below, T stands for the children who received the UPSTART preschool program during its fifth year of operation, and C stands for the comparison group children who did not participate in UPSTART. The "X" indicates that the UPSTART children received the Waterford Early Learning Program prior to Kindergarten and that the children from the control group did not. O1 indicates pretest measurements taken in the summer of 2013 and O2 indicates posttest measurements taken in the summer of 2014 for the C5 treatment and control group children.

NR	T	01	X	O2
NR	C	O1		02

The use of both a pretest and a comparison group facilitates our ability to examine potential threats to validity, which could jeopardize a clear interpretation of the results.[1] Because the study is not a randomized control trial, the groups are nonequivalent by definition, and consequently selection bias can be assumed to operate to some degree in some manner. The pretest allows 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). The pretest also allows us to examine the nature and degree of attrition in the study and whether it differentially affects one group more than the other.

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

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 the comparison group on posttest 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 comparison group on the posttest 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 pretest to posttest. If UPSTART shows stronger literacy growth rates, then the treatment group would be expected to show greater gain scores (posttest score minus pretest score) relative to the comparison group on the various subtests and total test scores.

With respect to concerns for school readiness, our research questions for the C5 evaluation study were as follows:

RQ1: D*o UPSTART students have better early literacy skills at kindergarten compared to control group students?*

If the answer is yes, then we would expect to see:

T > C @ O2 (controlling for differences at O1)

If the answer is no, then we would expect to see:

T = C @ O2 (controlling for differences at O1)

RQ2: D*o UPSTART students show stronger literacy growth rates from preschool to kindergarten compared to control group students?*

If the answer is yes, then we would expect to see:

T > C @ O2- O1 (growth)

If the answer is no, then we would expect to see:

T = C @ O2-O1 (growth)

In the preschool analysis, the outcomes of interest were measures of early literacy skills relevant to emerging readers such as phonological awareness, letter recognition, and letter sound knowledge and vocabulary development.

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

RQ3: *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?*

RQ4: *What percent of the participants completed the full implementation program (i.e., "graduated" as defined by the Waterford Institute)?*

RQ5: *How does the level of UPSTART curriculum usage relate to reading readiness outcomes?*

Data for research questions 3 and 4 were obtained from records maintained by the Waterford Institute and are answered in this report by descriptive statistics. The answer to Research Question 5 was derived from statistical analyses of 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.

Outcome Measures

The reading skills taught by the Waterford Early Learning Program at Level 1 of the curriculum[2] include:

- Phonological Awareness: phonemic segmenting and blending.
- Phonics: letter name knowledge, sound knowledge, and word reading.
- Comprehension and Vocabulary: vocabulary knowledge.
- Language Concepts: oral reading fluency.

The Brigance. The Brigance Inventory of Educational Development was selected as an early literacy measure of phonics and vocabulary knowledge and as a measure of pre-Kindergarten academic and cognitive skills. Ten of the Brigance scales were administered from the *language development and academic/cognitive domains*, as described below.

The Brigance language development scales included the:

- *Expressive Objects subtest*: the child is asked to name pictures shown by an assessor. (Total possible subtest score = 27)
- *Receptive Objects subtest*: the child is asked to point to pictures named by an assessor. (Total possible subtest score = 27)
- *Expressive Grammar subtest*: the child is assessed on the ability to use plural *s*, *ing*, prepositions, and interpret and talk about an illustration. (Total possible subtest score = 12)

The Brigance academic and cognitive literacy scales included the:

[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.

- *Visual Discrimination* subtest: the assessment focuses on the child's ability to identify similarities and differences between forms, uppercase letters, lowercase letters, and words. (Total possible subtest score = 20)
- *Recites Alphabet* subtest: the child is asked to recite the alphabet. (Total possible subtest score = 26)
- *Lowercase Letter Knowledge* subtest: the child is asked to name and recognize (point to) lower case letters presented by an assessor. (Total possible subtest score = 52)
- *Sounds of Lowercase Letters* subtest: the child is assessed on the ability to produce sounds of lowercase letters. (Total possible subtest score = 26)
- *Auditory Discrimination* subtest: the assessment focuses on the child's ability to identify if two words sound the same or different. (Total possible subtest score = 10)
- *Survival Sight Words* subtest: the assessment focuses on the child's ability to read survival sight words that appear on signs in public places. (Total possible subtest score = 16)
- *Basic Pre-Primer Vocabulary* subtest: the assessment focuses on the child's ability to read basic vocabulary words found in pre-primer reading programs. (Total possible subtest score = 24)
- *Total Brigance:* sum of the language and cognitive subtest scores. (Total possible score = 240)

As shown above, the version of the Brigance used in the C5 evaluation is a fairly comprehensive early literacy assessment comprised of 10 subtests in which the total test ranges from a score of zero to a score of 240. The Brigance is weighted toward the academic/cognitive literacy domain which accounts for approximately 73% of the total test score.

The Bader. The Bader was selected as a measure of *phonological awareness*. The Bader is comprised of three subtests, as follows:

- *Rhyme Recognition*: the child is asked to say yes if a pair of words presented orally by the assessor end the same way or to say no if the word pair do not end the same. (Total possible subtest score = 10)
- *Phonemic Blending*: the child is presented with a sequence of phonemes and is asked to say the word they constitute. (Total possible subtest score = 8)
- *Phoneme Segmentation*: the child is presented with a word and is asked to say the word sounds that make up the word in correct sequence. (Total possible subtest score = 8)
- Total Bader: sum of the Bader subscale scores. (Total possible Bader score = 26)

As revealed above, the Bader employs a relatively narrow test scale and measures the child's phonological awareness, considered an important predictor of later reading ability. Phonological awareness involves the child's ability to detect the sound structure of spoken words at three levels: rhyming, syllables, and phonemes.

Data Collection

As previously stated, 271 four year-old children were recruited for the C5 study: 109 treatment group children who had enrolled in UPSTART for Year 5 of the program and 162 control group children who had not enrolled in the UPSTART program. The children's parents were given an intake questionnaire (see Appendix A) at the time their children were pretested on the Brigance and Bader during the summer of 2013. The children were posttested on the Brigance and Bader a year later in the summer of 2014.

A preschool student data file was developed based on data collected from the intake questionnaire and from the pretest and posttest administrations of the Brigance and Bader. The final analysis file was based on the subset of children with valid matched pretest and posttest data, and who had not previously used the Upstart computerized learning program as documented through the intake interview.

Measurement Attrition and Creation of the Final Analysis Sample

Of the 271 preschool children recruited for the C5 evaluation, it was determined through the intake interview that 15 of the treatment group children had previously used the UPSTART online learning program. The screening interview resulted in these 15 treatment group cases being removed from the final analysis sample. All 94 of the treatment group children provided complete and valid Brigance testing data, but five of these children found the Bader pretest too difficult and could not produce a valid, nonzero test score. This resulted in a final Bader test sample of 89 treatment group children.

In conducting a quality control review of the data, it was determined that two of the control group cases had multiple missing demographic data fields that could not be reasonably corrected through imputation procedures. Consequently, these two control cases were removed from the final analysis file. Further review of the control group data showed that 10 children did not take the Brigance posttest and that test data problems were evident with 21 control children who initially took the Bader pretest. These problems involved a combination of not being able to score on the Bader pretest or posttest or not showing up at all for the Bader posttest. In the end, it was determined that valid and matched Brigance and Bader test data were available for 141 control group children.

The C5 attrition results are summarized in Table 1.

Table 1
C5 Attrition Summary

Attrition Indicator	*Treatment*	*Control Group*	*Total Sample*
Recruited & Accepted		162	271
Passed Screening Interview	94	162	256
Passed Data QC Review	94	160	254

Attrition Indicator	*Treatment*	*Control Group*	*Total Sample*
Matched & Valid Brigance Pre/Post	94	152	246
Matched & Valid Bader Pre/Post	89	141	230

Prior exposure to the UPSTART curriculum was the most common cause of attrition in the treatment group sample. The other common cause of attrition for both treatment and control group children was not being able to score on the Bader. This occurred on both the pretest and the posttest, but most commonly on the Bader pretest. Based on the data made available, not showing up for the posttest was a problem specific to the C5 control families.

The C5 evaluation's overall measurement attrition rate was 10%. As previously stated, the most common cause of measurement attrition was not being able to score on the Bader. The C5 measurement attrition rate of 10% is an improvement over previous years in which measurement attrition ranged from 15% to 24%.

The final analysis sample for the C5 evaluation used the data collected from 194 children. The UPSTART treatment group was composed of 94 children who passed the screening interview and were able to provide valid and matched pretest and posttest data on the Brigance (89 of whom also had valid and matched Bader data). The non-UPSTART control group was composed of a random sample of 100 control children who were selected from the 141 cases with valid and matched Brigance and Bader test data. These two files were then merged to form the final analysis sample in which the treatment and control group samples were reasonably balanced in size. Table 2 shows the composition of the final analysis sample.

Table 2
C5 Final Analysis Sample

Dataset	*Treatment* *N*	*Control Group* *N*	*Total Sample* *N*
Bader	89	100	189
Brigance	94	100	194
Demographics	94	100	194

Data Analysis: Do UPSTART children have better literacy skills at Kindergarten than control group children?

The general strategy for determining whether there was an impact of the UPSTART preschool experience on young children's literacy skills was to compare a sample of program participants with a similar group of nonparticipants on Brigance and Bader posttest scores collected at the beginning of Kindergarten. This strategy assumes that the two groups are initially similar on

factors that influence the literacy skills measured at Kindergarten. These factors could include initial differences between the groups on measured literacy skills (e.g., pretest scores) as well as demographic factors that differentiate the treatment and control groups (e.g., the child's ethnicity) if they are significantly related to posttest performance.

If the treatment and control groups are essentially similar at the beginning of UPSTART on factors affecting posttest literacy outcomes of interest, then any observed differences on the posttest can be reasonably attributed to participation in UPSTART. Alternatively, if there is significant initial nonequivalence between the groups, then statistical adjustments to the posttest outcomes using regression analysis will be necessary in leveling the playing field so that more accurate and fair comparisons can be made.

The equivalence of the treatment and control groups in the final analysis samples were examined on the basis of the Brigance and Bader pretest scores and on the basis of those demographic characteristics that were significantly related to the posttests. Group equivalence on the pretests was examined using independent sample t-tests. Relationships between the demographics and the posttest scores were examined using correlation analyses.[3]

While initial between-group differences were not found on pretest measures of early literacy, the pretest measures were found to exert a strong influence on posttest scores; see the correlations between the pretest and posttest measures in Appendix C. Additionally, there were some differences between the treatment and control group on demographics that influenced total posttest scores on the Brigance and Bader (i.e., child ethnicity and whether the child had experience with a computer in preschool). This necessitated a final set of analyses using multiple regression analysis to adjust posttest scores due to the influence of the pretests and some of the between-group differences in demographics that affected posttest performance.

Posttest differences between the treatment and control groups were first examined for both the Brigance and Bader using independent sample t-tests. Ultimately, posttest differences were re-examined by adjusting for initial differences between the treatment and control groups with the use of multiple regression analysis. The regressions used a hierarchical block design in which the pretest and a set of demographic covariates were entered first, followed by the treatment-control group comparison. Effect size estimates are also graphically presented for all posttest differences between the treatment and control groups on the Brigance and Bader.

[3] It was necessary to transform a number of the demographic measures from nominal measures to scale measures by creating "dummy variables" on the basis of the dominant characteristics of the sample. For example, parent's marital status was transformed into whether the parent was married or not, or percent married.

Data Analysis: Do UPSTART children show stronger literacy growth rates from preschool to Kindergarten than control group children?

To determine whether UPSTART students show stronger literacy growth rates from preschool to Kindergarten compared to control students, paired sample t-tests were run to compare pretest and posttest gain scores for the matched Brigance and Bader *treatment groups* on the total test and each of the subtests. The same analysis was performed with the Brigance and Bader matched *control groups*. Statistically significant growth rates were determined by examining confidence intervals for the treatment and control group gain scores for each test measure at the 99% confidence interval.[4] Bar charts are displayed for each set of gain score comparisons.

Analysis of Implementation Time

A simple regression analysis and an analysis of covariance (ANCOVA) were used to determine the relationship between curriculum usage (measured by the amount of instruction received by UPSTART participants) and literacy outcomes. An ordinal version (ordered categories) of UPSTART usage (transformed to hours of instruction) was used to see what the impact of instructional time in the program was on literacy outcomes as measured by total scores on the Brigance and Bader posttests. This was accomplished by creating a new variable called Usage Group in which hours of instruction was factored into four levels corresponding to quartiles. The ANCOVA was run separately for the Brigance and the Bader with Usage Group as the independent variable and the respective pretests as a covariate.

[4] To guard against Type I error (falsely rejecting the null hypothesis) in conducting tests of statistical inference (e.g., t-tests and multiple regression analysis) the criterion for statistical significance was based on the error rate for the collection of comparison data required by the Brigance and the Bader. For example, comparisons among the ten Brigance subtest means and the total test at the .05 level can result in at least half of the statistical tests being significant by chance: 11 (.05) = .55. Increasing the rigor of the significance criterion by moving to the .01 level for the Brigance still leaves the possibility that the effective significance level for the collection of comparisons is .11, not .01: 11 (.01) = .11). This consideration resulted in a decision rule to set the confidence level at 99% and $p<.01$ for the collection of comparisons across the Bader and Brigance. For further detail, see Kirk R.E. (1968). *Experimental Design: Procedures for the Behavioral Sciences*. Belmont, CA: Wadsworth Publishing Company.

Findings on UPSTART Implementation

Findings reviewed under UPSTART implementation include enrollment in the fifth year, 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 Waterford Institute provided documentation for a fifth-year UPSTART enrollment of 1,577 children. Some basic demographic characteristics of the C5 population are presented below in Table 3 compared to the treatment group in the C5 analysis sample (N=94).

Table 3
Basic Demographic Characteristics: C5 population vs. C5 Analysis Sample

Demographic Categories		***All UPSTART*** (N=1,577)	***Analysis Sample*** (N=94)
Child's Gender	Boys	50%	42%
	Girls	50%	58%
Child's Ethnicity	Caucasian	74%	88%
	Hispanic	20%	5%
	Asian/Pacific Islander	2%	--
	Black	1%	1%
	Native American	1%	--
	Multiracial	--	5%
	Other	2%	--
Child's Primary Language	English	84%	96%
	Spanish	15%	3%
	Other	1%	1%
Parent Educational Attainment	Some High School	6%	2%
	High School Graduate	14%	55%
	Some College	35%	40%
	College Graduate	37%	2%
	Advanced Degree	8%	--
	Unknown	2%	--
Parent Marital Status	Married	89%	90%
	Otherwise	11%	10%

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

As shown in Table 3, the analysis sample was somewhat of a more advantaged subgroup compared to the C5 population from the standpoint of the child's ethnicity and primary language.

That is, the C5 population is under-represented in the analysis sample in terms of Hispanics and Spanish speakers. However, this is somewhat of an artifact of the sample selection process which required the selected UPSTART children to be proficient English language speakers in order to participate in the testing. A substantially larger recruitment effort would have been required to find additional Hispanic families with bilingual children who would agree to participate in the evaluation.

UPSTART Equipment Provided

The kind of education technology provided to UPSTART children in Year 5 of the program is shown in Table 4 for all 1,577 children enrolled and for the C5 analysis sample. As in past years, the vast majority of UPSTART children (74%) received a computer drive with the UPSTART curriculum loaded on it. This allowed families to access the UPSTART curriculum from their home computers. Similarly, the students in the C5 analysis sample most often (84%) also received a computer drive with the curriculum loaded on it.

Second most frequently, UPSTART provided personal computers and free Internet subscriptions to 8% of the C5 children; 5% of the C5 analysis sample were also provided with personal computers and free Internet subscriptions. Another 7% of the C5 program participants were given access to a home computer for free while they participated in the program. Similarly, 5% of the C5 analysis sample was given access to a home computer for free while they participated in the program. The remaining 11% of the C5 enrollment received various combinations of computer technology to enable them to access the UPSTART curriculum. See Table 4 for details.

Table 4
Equipment provided to C5 Participants

Equipment Provided	***All UPSTART*** (N=1,577)	***Analysis Sample*** (N=94)
Drive	74%	84%
Computer & Internet	8%	5%
Computer	7%	5%
Computer & Wireless	4%	4%
Computer & Cellular	4%	--
Internet & Drive	2%	--
Cellular & Drive	1%	1%
Computer with Wireless & Internet	<1%	--

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

UPSTART Graduates

Of the 1,577 children documented as enrolled in UPSTART in the fifth year of the program, the Waterford Institute classified 1,484 as children who had met the program's usage criteria and were thereby considered to be graduates of the program. The usage criterion 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. By this definition, Cohort 5 achieved a *graduation rate of 94%* (i.e., 1,484/1,577 = 0.94). This is the same level of UPSTART criterion usage as achieved in the previous two years.

UPSTART graduate status was significantly correlated with hours of instruction (r = .58) and with the number of weeks in the program (r = .74). Additionally, UPSTART graduate status in the fifth year of the program was significantly correlated with early literacy outcomes as measured by the Brigance (r=.24, p<.05) and Bader (r=.22, p<.05) posttests.

UPSTART Usage

The hours of instruction observed for all children documented asenrolled in the fifth year of UPSTART are summarized in Table 5 compared to "graduates" and the children in the C5 preschool analysis samples. The average level of usage for all students enrolled in the fifth year of UPSTART was approximately 71 hours of instruction; this is the same average level of usage as documented in the fourth year of the program. The C5 academic year covered 44 weeks of instruction, beginning the week of September 2, 2013 and ending June 30, 2014.

Six of the 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 5 of the program. These families dropped out of the program within six weeks of enrollment. For enrolled families whose children did use the curriculum, the average duration in the program was approximately 40 weeks. This usage pattern is similar to that observed in the fourth year of the program.

The children in the C5 analysis samples used the UPSTART curriculum for approximately 72-73 hours of instruction on the average (see Table 5).

Table 5
C5 Hours of UPSTART Instruction

Group	*N*	*Mean*	*SD*	*Range*
All UPSTART	1,577	71.34	24.79	00.00 – 168.79
UPSTART Graduates	1,484	74.94	20.58	16.76 - 168.79
Brigance Analysis Sample	94	72.44	25.63	5.52 – 168.79
Bader Analysis Sample	89	73.17	25.29	5.52 – 168.79

The histograms in Figures 1-4 show the distribution of hours of instruction for the total C5 population (Figure 1), the C5 graduates (Figure 2), and the C5 analysis samples (Figure 3 for the Brigance sample and Figure 4 for the Bader sample). All four histograms show hours of instruction to be essentially normally distributed.

C5 Population Usage. In the C5 population (see Figure 1), UPSTART curriculum usage was normally distributed with an average usage level of approximately 71 hours. As previously noted, six of the enrolled children logged zero hours of instruction during their time in UPSTART (i.e., up to six weeks). At the other end of the spectrum, five children logged over 150 hours of instruction.

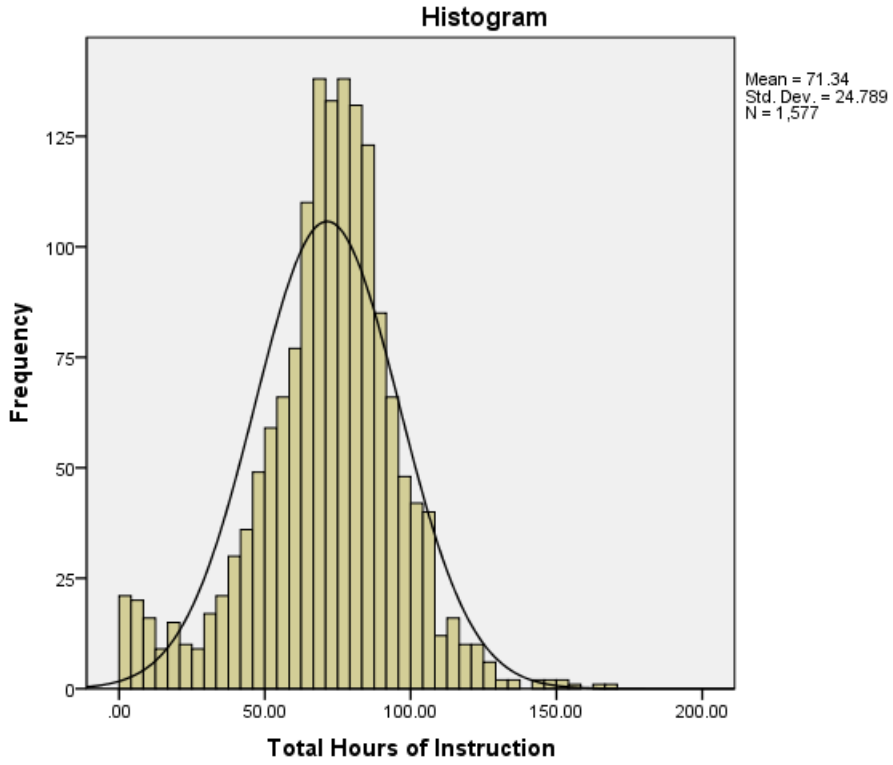

Figure 1
Hours of Instruction for All Students Enrolled in UPSTART in Year 5

The bottom quartile of the C5 population completed 59.23 hours of instruction or less. The midpoint of the C5 population distribution (the median) was 73.83 hours of instruction. The top quartile of the C5 population completed in excess of 85.85 hours of instruction.

C5 Graduate Usage. UPSTART curriculum usage for the subset of graduates (N=1,484; see Figure 2) was normally distributed with an average usage level of 74.94 hours of instruction. As noted previously, five of the C5 participants – all graduates -- logged over 150 hours of

instruction. The bottom quartile of the C5 graduates ranged from 16.76 hours to 62.81 hours of usage. The midpoint of the C5 graduate distribution (the median) was 74.97 hours of instruction. The top quartile of the C5 graduates completed in excess of 86.74 hours of instruction.

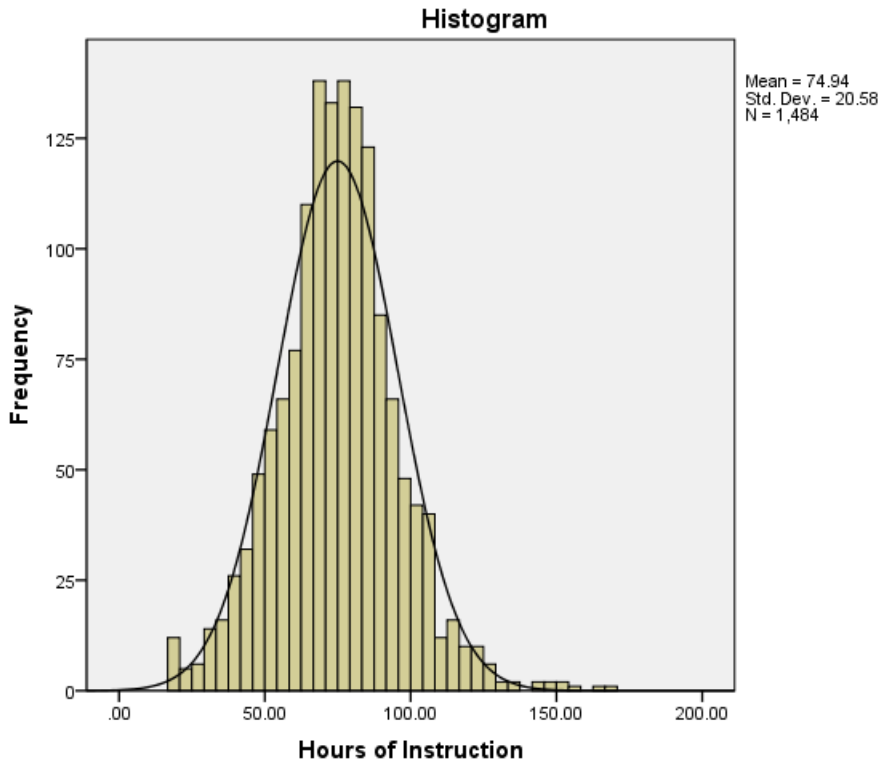

Figure 2
Hours of Instruction for UPSTART Graduates in Year 5

C5 Analysis Sample Usage. UPSTART curriculum usage for the C5 analysis samples was essentially normally distributed with an average level of usage of approximately 72-73 hours of instruction, depending on the test. Curriculum usage for children in the analysis samples ranged from a low of 5.52 hours of instruction to a high of 168.79 hours of instruction.

The mean level of usage for the C5 Brigance analysis sample (N=94, see Figure 3) was 72.44 hours of instruction with a standard deviation of 25.63 hours. The Brigance analysis sample's median is 75.88 hours of instruction. For the usage analysis with the Brigance sample, hours of instruction are distributed as follows by quartile of usage:

- 1st Quartile: 5.52 hours to 61.00 hours
- 2nd Quartile: 64.56 hours to 75.78 hours
- 3rd Quartile: 75.98 hours to 83.87 hours

- 4th Quartile: 85.67 hours to 168.79 hours

The four quartiles divide the set of observations of usage into four portions that are approximately equal in terms of the proportion of observations in each portion (i.e., 23 or 24 cases per quartile in the Brigance sample).

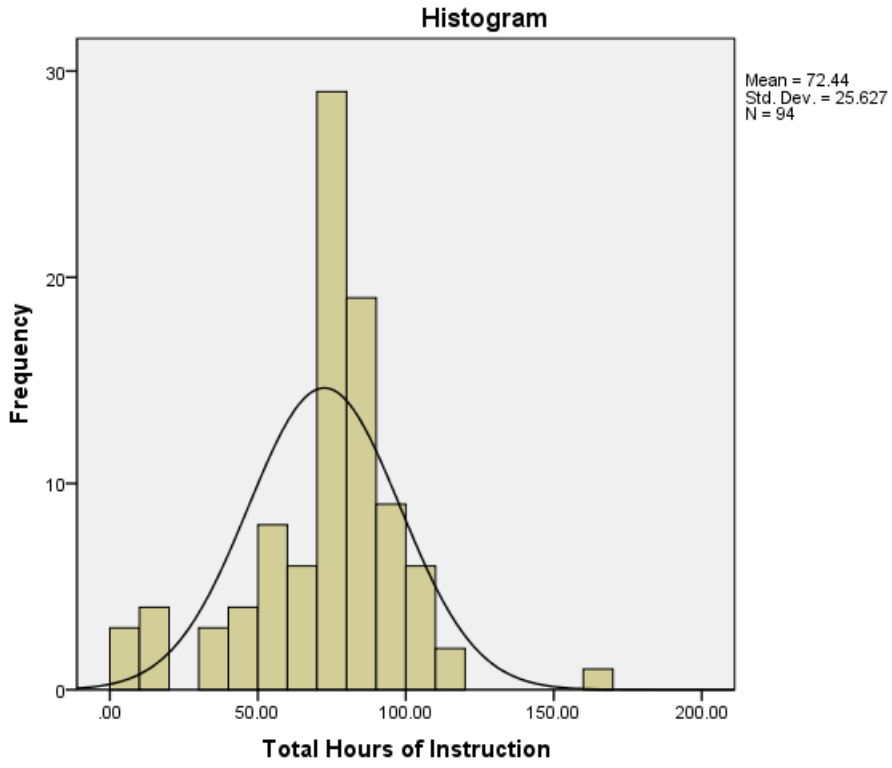

Figure 3
Hours of Instruction for C5 Brigance Analysis Sample

The mean level of usage for the C5 Bader analysis sample (N=89, see Figure 4) was 73.17 hours of instruction with a standard deviation of 25.29 hours. The Bader analysis sample's median is 76.51 hours of instruction. For the usage analysis with the Bader sample, hours of instruction are distributed as follows by quartile of usage:

- 1st Quartile: 5.52 hours to 64.56 hours
- 2nd Quartile: 68.16 hours to 76.51 hours
- 3rd Quartile: 77.05 hours to 83.87 hours
- 4th Quartile: 85.67 hours to 168.79 hours

The four quartiles divide the set of observations of usage into four portions that are approximately equal in terms of the proportion of observations in each portion (i.e., 22 or 23 cases per quartile in the Bader sample).

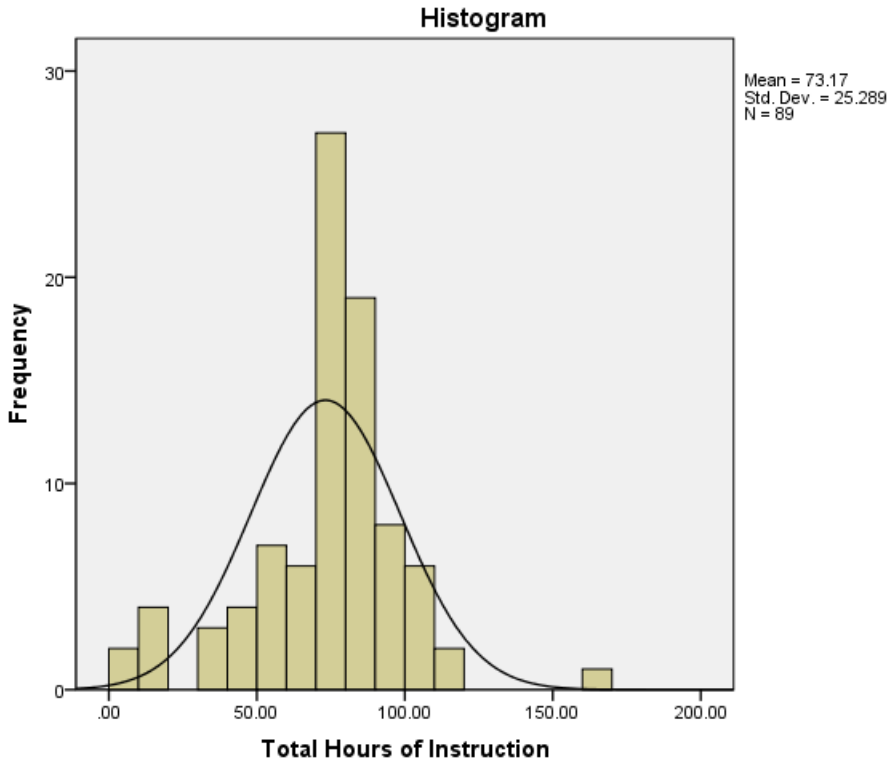

Figure 4
Hours of Instruction for C5 Bader Analysis Sample

UPSTART Usage and Literacy Outcomes

Like previous years, the fifth year evaluation of UPSTART found curriculum usage to be significantly and positively related to literacy outcomes as measured by total posttest scores on both the Brigance and the Bader. The correlation between UPSTART usage and literacy outcomes measured by the Brigance Total Posttest was positive and statistically significant ($r=.45$, $p<.01$, $n=94$). Similarly, the correlation between UPSTART usage and literacy outcomes measured by the Bader Total Posttest was also positive and statistically significant ($r= .41$, $p<.01$, $n=89$). The relationship between usage and literacy outcomes was examined further as discussed below.

UPSTART Usage as a Predictor of Brigance Outcomes. A simple regression of usage on Brigance posttest scores showed that usage (measured in hours) accounted for 19% of the variance in literacy outcomes measured by the Brigance (adjusted $R^2 = .19$). See Table 6 for the regression model summary and Table 7 for the Analysis of Variance (ANOVA) summary.

Table 6
Usage Model Summary for Brigance Data

R	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
.446	.199	.190	36.053

The independent variable is Total Hours of Instruction.

Table 7
ANOVA for Brigance Usage Outcomes

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	29711.538	1	29711.538	22.858	.000
Residual	119585.015	92	1299.837		
Total	149296.553	93			

The independent variable is Total Hours of Instruction.

The relationship between usage and Brigance posttest scores was moderately strong as indicated by the statistically significant standardized regression coefficient *Beta* shown in Table 8 (Beta = .45, p<.01).

Table 8
Usage Coefficients for Brigance Outcomes

	Unstandardized Coefficients		*Standardized Coefficients*		
	B	**Std. Error**	**Beta**	**t**	**Sig.**
Total Hours of Instruction	.697	.146	.446	4.781	.000
(Constant)	118.666	11.203		10.592	.000

The curve fitting plot presented in Figure 5 shows *a moderately strong linear relationship* (adjusted $R^2 = .19$) between UPSTART usage (measured in hours of instruction) and Brigance posttest scores. That is, Brigance posttest scores tend to increase with increasing hours of

UPSTART usage.[5] This suggests that the acquisition of early phonics skills measured by the Brigance tend to improve with increasing levels of exposure to the UPSTART curriculum.

Figure 5
Plot of Hours of Instruction and Brigance Posttest Scores

Because usage effects are confounded by pretest scores[6] to some extent, it was desirable to re-examine the relationship between UPSTART usage and literacy outcomes at Kindergarten entry by controlling for initial levels of literacy development. This was accomplished with an analysis of covariance (ANCOVA) in which literacy outcomes at Kindergarten entry were examined at each usage quartile (measured in hours of instruction as described previously for the analysis samples). This analysis used pretest scores to control for initial levels of literacy development.

Brigance ANCOVA. The Brigance pretest was used as a control covariate in an ANCOVA with the Brigance analysis sample. The ANCOVA results for the Brigance test sample shown in Table 9 reveal that the usage factor is statistically significant ($p<.05$). The

[5] Confidence in this conclusion – that the relationship is linear – is strong since the statistical power for estimating this relationship is strong (Power = .99, using Cohen's [1988] power table 9.3.2).

[6] When pretest scores are controlled for, the correlation between usage and Brigance posttest scores drops from r = .44 (simple correlation) to r = .34 (partial correlation). Similarly but less so, when pretest scores are controlled for, the correlation between usage and Bader posttest scores drops from r = .41 (simple correlation) to r = .40 (partial correlation).

amount of variance in Brigance Posttest scores accounted for by usage, controlling for prior literacy achievement, is approximately 9% (see the Partial Eta Squared statistic column).

Table 9
UPSTART Usage as a predictor of Brigance Total Posttest Scores - Beginning K

Tests of Between-Subjects Effects

Dependent Variable: Brigance Posttest

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Partial Eta Squared*	*Noncent. Parameter*	*Observed Power*[b]
Corrected Model	74360.839	4	18590.210	22.079	.000	.498	88.317	1.000
Intercept	57912.867	1	57912.867	68.782	.000	.436	68.782	1.000
Brigance_Pre	51930.947	1	51930.947	61.678	.000	.409	61.678	1.000
Brigance_Usage	7592.835	3	2530.945	3.006	.035	.092	9.018	.691
Error	74935.714	89	841.974					
Total	2840118.000	94						
Corrected Total	149296.553	93						

Adjusted R Squared Model = .48 Computed using alpha = .05

Each quartile usage group is identified by its quartile value (1-4) in Table 10. The covariance model shown in Table 10 compares the literacy achievement effects of each level of usage with the fourth quartile level of usage for the C5 analysis sample, controlling for the influence of initial literacy skills, and displays the difference in Brigance total posttest scores in the column labeled B – expressed as a regression coefficient. The parameter estimates in Table 10 *suggest a curvilinear trend* in early literacy development (as measured by total Brigance posttest scores). That is, *early literacy achievement measured by the Brigance tends to peak at the third quartile of UPSTART usage and then tends to drop off. In terms of usage levels, achievement on the Brigance tends to be highest for children whose UPSTART usage is between 75 and 85 hours*.

The statistical power of the Brigance ANCOVA is less than optimal (power = .69) because of the relatively small quartile usage samples (i.e., 23-24 children per quartile usage group). Consequently, we are not as confident as we would like to be in drawing the conclusion that the usage trend for the C5 population is curvilinear rather than simply linear. Nevertheless, that is what the C5 usage data suggest with respect to the Brigance sample.

Table 10
Parameter Estimates: Brigance Total Posttest Scores - Beginning K

Parameter	*B*	*Std. Error*	*t*	*Sig.*	*Partial Eta Squared*	*Observed Power*
Intercept	93.382	13.117	7.119	.000	.363	1.000
Brigance_Pre	.685	.087	7.854	.000	.409	1.000
Usage_Group=1.00	-20.275	8.868	-2.286	.025	.055	.619
Usage_Group=2.00	-1.111	8.646	-.128	.898	.000	.052
Usage_Group=3.00	2.791	8.615	.324	.747	.001	.062
Usage_Group=4.00	0					

The covariance-adjusted Brigance posttest means can be seen more clearly by usage quartile in Table 11. These data suggest that literacy achievement differs significantly for UPSTART children in usage quartile 1 compared to UPSTART children in usage quartiles 2 through 4. From a descriptive standpoint, *the apparent curvilinear trend in the data shown in Figure 7 suggests that literacy achievement increases with usage level up to quartile 3 and then declines*.

Table 11
Parameter Estimates: Brigance Total Posttest Scores by Usage Quartile

Dependent Variable: Brigance Posttest

Brigance Usage Group	*Mean*	*Std. Error*	*95% Confidence Interval*	
			Lower Bound	**Upper Bound**
Quartile 1	153.652[a]	6.055	141.622	165.682
Quartile 2	172.816[a]	6.052	160.791	184.842
Quartile 3	176.719[a]	5.927	164.942	188.495
Quartile 4	173.927[a]	6.206	161.596	186.259

Covariates appearing in the model are evaluated at the following values:

Brigance Pretest = 117.511.

Table 12 presents data showing how UPSTART early literacy development varies according to differences between usage quartiles. These pairwise comparisons reveal that:

- UPSTART children with quartile 3 level usage have significantly greater literacy achievement than children with quartile 1 level usage (i.e., a difference of 23 points on the Brigance posttest, statistically significant at the 95% confidence interval); and that

- Literacy achievement for UPSTART children with quartile 3 level usage is not significantly different from that of UPSTART children with quartile 2 or 4 level usage.

Table 12
Pairwise Comparisons of Usage Quartiles on Brigance Literacy Achievement

Dependent Variable: Total Brigance Posttest

(I) Brigance Quartiles	*(J) Brigance Quartiles*	*Mean Difference (I-J)*	*Std. Error*	*Sig.*[a]	*95% Confidence Interval for Difference*	
					Lower Bound	**Upper Bound**
Quartile 1	Quartile 2	-19.164	8.582	.168	-42.323	3.994
	Quartile 3	*-23.067*[*]	8.442	*.045*	-45.846	-.287
	Quartile 4	-20.275	8.868	.148	-44.207	3.656
Quartile 2	Quartile 1	19.164	8.582	.168	-3.994	42.323
	Quartile 3	-3.902	8.474	1.000	-26.770	18.966
	Quartile 4	-1.111	8.646	1.000	-24.441	22.219
Quartile 3	Quartile 1	*23.067*[*]	8.442	*.045*	.287	45.846
	Quartile 2	3.902	8.474	1.000	-18.966	26.770
	Quartile 4	2.791	8.615	1.000	-20.457	26.039
Quartile 4	Quartile 1	20.275	8.868	.148	-3.656	44.207
	Quartile 2	1.111	8.646	*1.000*	-22.219	24.441
	Quartile 3	-2.791	8.615	*1.000*	-26.039	20.457

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

*. The mean difference is significant at the .05 level.

From a substantive perspective, UPSTART children in the Brigance sample who attained at least quartile 2 level usage – at least *65 hours, or more* – tended to develop the early literacy skills measured by the Brigance to a significantly greater degree than children who used the UPSTART curriculum less so.

The apparent curvilinear trend in Brigance literacy achievement levels across usage quartiles is shown in Figure 7. The observed UPSTART usage effect on the development of literacy skills measured by the Brigance assessment is *moderately strong*.

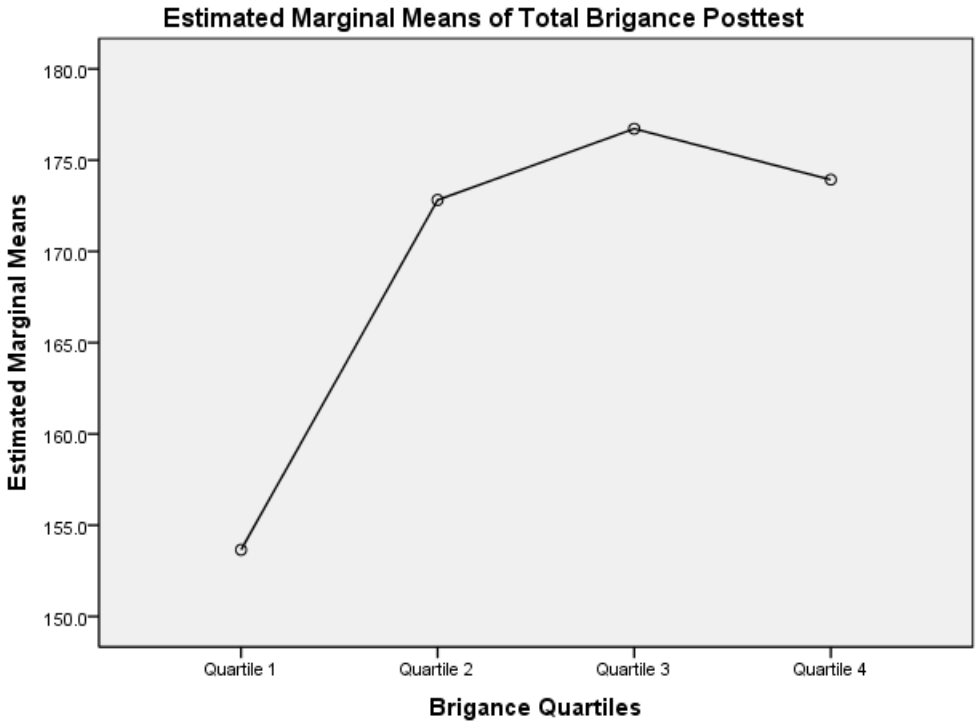

Figure 7
Brigance Literacy Development by UPSTART Usage Quartiles

UPSTART Usage as a Predictor of Bader Outcomes. The same procedures were performed with the C5 UPSTART usage data and the Bader posttest scores. Similar results were found: 16% of the variance in Bader posttest outcomes could be accounted for by UPSTART usage (adjusted R^2 = .16) and the relationship between usage and Bader posttest scores was statistically significant ($p<.01$) and moderately strong (Beta = .41, adjusted R^2 = .16). See Table 13 for the regression model summary, Table 14 for the Analysis of Variance (ANOVA) summary, and Table 15 for the regression coefficients showing the impact of usage on Bader posttest scores.

Table 13
Usage Model Summary for Bader Data

R	*R Square*	*Adjusted R Square*	*Std. Error of the Estimate*
.411	.169	.160	6.117

The independent variable is Total Hours of Instruction.

Table 14
ANOVA for Bader Usage Outcomes

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	662.720	1	662.720	17.714	.000
Residual	3254.831	87	37.412		
Total	3917.551	88			

The independent variable is Total Hours of Instruction.

Table 15
Usage Coefficients for Bader Outcomes

	Unstandardized Coefficients		*Standardized Coefficients*		
	B	**Std. Error**	**Beta**	t	Sig.
Total Hours of Instruction	.109	.026	.411	4.209	.000
(Constant)	9.734	1.995		4.880	.000

The curve fitting plot presented in Figure 6 shows a *moderately strong linear relationship* between UPSTART usage and the development of phonological awareness skills as measured by the Bader assessment (standardized regression coefficient for usage = 0.41, $p<.01$). This suggests that young children's phonological awareness skills (e.g., rhyming and sounding out words) tend to improve with increasing levels of exposure to the UPSTART curriculum.[7]

[7]Confidence in this conclusion is also strong – that the relationship is linear – since the statistical power for estimating this relationship is strong (Power = .92, using Cohen's [1988] power table 9.3.2).

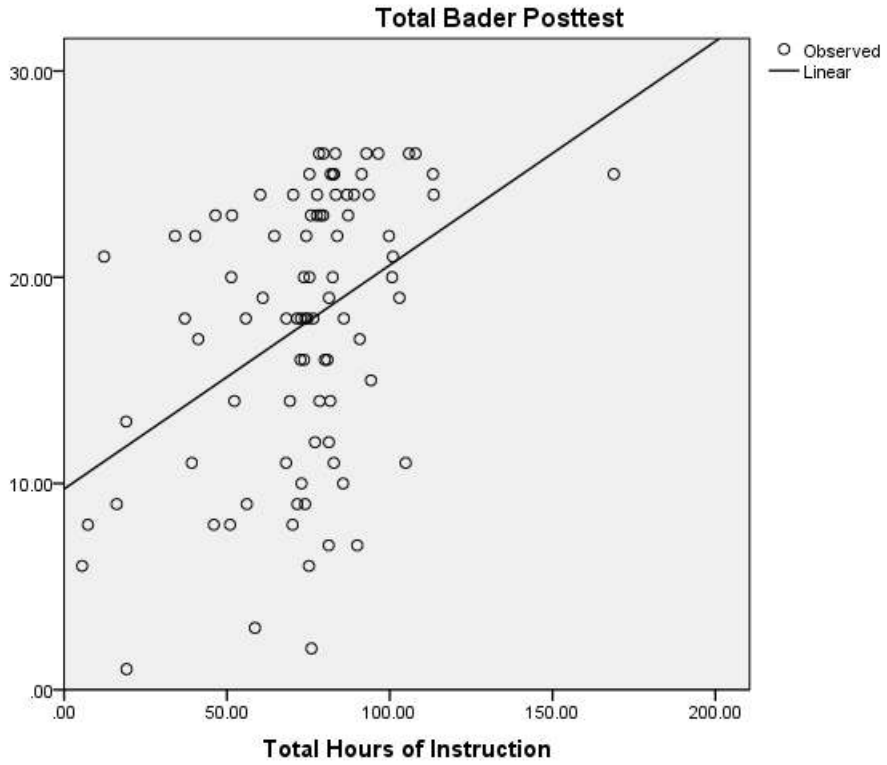

Figure 6
Plot of Hours of Instruction and Bader Posttest Scores.

Bader ANCOVA. The Bader pretest was used as a control covariate in an ANCOVA with the C5 Bader test sample. The ANCOVA results for the Bader test sample shown in Table 16 reveal that the usage factor is statistically significant ($p<.01$). The amount of variance in Bader Posttest scores accounted for by usage, controlling for prior literacy achievement, is approximately 14% (see the Partial Eta Squared statistic column).

Table 16
Tests of Between Subjects Effects: Bader Total Posttest - Beginning K

Dependent Variable: Bader Postttest

Source	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Partial Eta Squared*	*Noncent. Parameter*	*Observed Power*[b]
Corrected Model	1051.141	4	262.785	7.701	.000	.268	30.804	.996
Intercept	4260.048	1	4260.048	124.841	.000	.598	124.841	1.000
Bader_Pre	451.389	1	451.389	13.228	.000	.136	13.228	.949
Bader_Usage	478.805	3	159.602	4.677	.005	.143	14.031	.881
Error	2866.409	84	34.124					
Total	31719.000	89						
Corrected Total	3917.551	88						

Adjusted R Squared Model = .23 Computed using alpha = .05

In Table 17, each usage group from the Bader sample is identified by its quartile value (1-4). The covariance model compares the effects of each level of usage with the fourth quartile level of usage for the C5 analysis sample, controlling for the influence of initial literacy skills, and displays the difference in Bader total posttest scores in the column labeled B – expressed as a regression coefficient. The parameter estimates in Table 17 *suggest an S-shaped curvilinear trend* in which literacy achievement (as measured by total Bader posttest scores) is significantly greater for UPSTART children in the *4th quartile of usage compared to UPSTART children in the 1st and 2nd quartiles of usage*.[8]

[8] The analysis of quartile effects at the upper end of the usage spectrum is somewhat constrained by limitations in statistical power associated with the relatively small sample sizes of the Bader quartile usage groups (22-23 cases per group).

Table 17
Parameter Estimates: Bader Total Posttest Score - Beginning K

Parameter	*B*	*Std. Error*	*t*	*Sig.*	*Partial Eta Squared*	*Noncent. Parameter*	*Observed Power*[a]
Intercept	16.485	1.724	9.564	.000	.521	9.564	1.000
Bader_Pre	.499	.137	3.637	.000	.136	3.637	.949
Usage_Group =1.00	-5.131	1.761	-2.913	.005	.092	2.913	.821
Usage_Group=2.00	-4.842	1.764	-2.744	.007	.082	2.744	.774
Usage_Group=3.00	-0.705	1.765	-0.400	.691	.002	.400	.068
Usage_Group=4.00	0						

The covariance-adjusted Brigance posttest means can be seen by usage quartile in Table 18. From a descriptive standpoint, these data *suggest* that early literacy achievement tends to differ for UPSTART children in usage quartiles 1 and 2 compared to UPSTART children in usage quartiles 3 and 4.

Table 18
Parameter Estimates: Bader Total Posttest Scores by Usage Quartile

Dependent Variable: Bader Posttest

Bader Usage Group	*Mean*	*Std. Error*	*95% Confidence Interval*	
			Lower Bound	**Upper Bound**
Quartile 1	15.240	1.226	12.803	17.678
Quartile 2	15.530	1.246	13.053	18.007
Quartile 3	19.666	1.245	17.189	22.143
Quartile 4	20.371	1.251	17.883	22.860

Covariates appearing in the model are evaluated at the following values: Bader Pretest = 7.787.

The pairwise comparisons in Table 19 identify the quartile differences that are statistically significant. The Table 19 pairwise comparisons show that:

- UPSTART children with *quartile 4 level usage* have developed significantly greater phonological awareness than children with quartile 1 or 2 level usage. These differences are statistically significant at the 95% confidence interval.

- The phonological awareness skills of UPSTART children with quartile 3 level usage are not significantly different from that of UPSTART children with quartile 4 level usage. (This is a difference of less than one point on the Bader posttest).

From a *substantive perspective*, it appears that UPSTART children who attain at least quartile 3 level usage – at least 77 hours of curriculum usage or more – tend to benefit the most from UPSTART in terms of the development of phonological awareness skills.

Table 19

Pairwise Comparisons of Usage Quartiles on Brigance Literacy Achievement

Dependent Variable:Total Bader Posttest

(I) Bader Quartiles	*(J) Bader Quartiles*	*Mean Difference (I-J)*	*Std. Error*	*Sig.*[a]	*95% Confidence Interval for Difference*[a]	
					Lower Bound	**Upper Bound**
Quartile 1	Quartile 2	-.290	1.749	1.000	-5.015	4.436
	Quartile 3	-4.426	1.748	.079	-9.149	.297
	Quartile 4	*-5.131*[*]	1.761	*.028*	-9.891	-.371
Quartile 2	Quartile 1	.290	1.749	1.000	-4.436	5.015
	Quartile 3	-4.136	1.761	.127	-8.896	.623
	Quartile 4	*-4.842*[*]	1.764	*.045*	-9.610	-.074
Quartile 3	Quartile 1	4.426	1.748	.079	-.297	9.149
	Quartile 2	4.136	1.761	.127	-.623	8.896
	Quartile 4	-.705	1.765	1.000	-5.475	4.065
Quartile 4	Quartile 1	*5.131*[*]	1.761	*.028*	.371	9.891
	Quartile 2	*4.842*[*]	1.764	*.045*	.074	9.610
	Quartile 3	.705	1.765	1.000	-4.065	5.475

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

*. The mean difference is significant at the .05 level.

Figure 8 confirms the results of the pairwise comparisons and shows that UPSTART children with usage levels at or above quartile 3 tend to develop phonological awareness skills (as measured by the Bader Posttest at Kindergarten entry) to a substantially greater degree than UPSTART children with less usage.

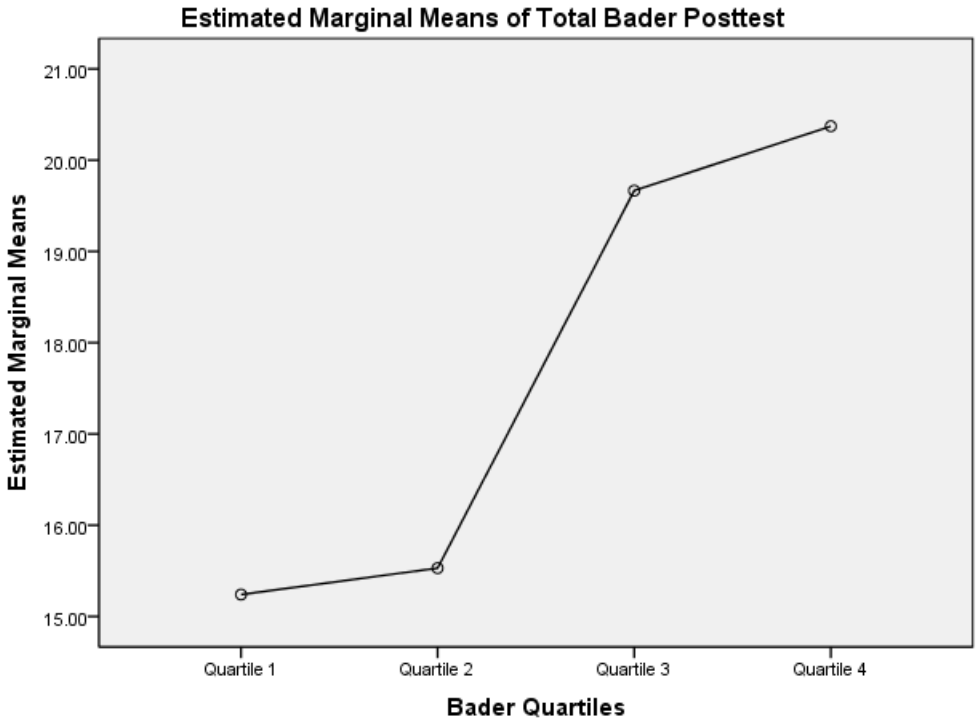

Figure 8
Bader Literacy Development by UPSTART Usage Quartiles

Findings on UPSTART Outcomes

In this section, evaluation findings are reviewed that address the C5 study questions on UPSTART outcomes. These questions include:

Research Question 1: *Do UPSTART students have better literacy skills at Kindergarten than control students?*

Research Question 2: *Do UPSTART students show stronger literacy growth rates from Preschool to Kindergarten than control students?*

For each of these two questions, results for the Brigance are reviewed first, followed by the results for the Bader.

Do UPSTART students have better literacy skills at Kindergarten than control students?

As discussed previously, the analytic strategy for answering Research Question 1 proceeded through the following phases:

- Pretest Analysis
- Covariate Analysis
  - Identifying significant pre-existing differences between the treatment and control groups
  - Identifying significant posttest predictors that differentiate the treatment and control groups
- Posttest Analysis
- Multiple Regression Analysis

Brigance Pretest Results. There were no statistically significant differences between the UPSTART treatment group children and the control group children on the Brigance at pretest. Table 20 shows the Brigance pretest results.

Table 20
Brigance Pretest Analysis of Treatment-Control Group Differences

Brigance Pretest	*Group*	*N*	*Mean*	*SD*	*t*	*Significance*
Expressive Objects	Treatment	94	24.266	1.647	1.137	NS
	Control	100	24.960	2.064		
Receptive Objects	Treatment	94	26.628	0.950	0.038	NS
	Control	100	26.620	1.716		
Expressive Grammar	Treatment	94	8.936	1.412	0.496	NS
	Control	100	8.840	1.284		
Visual Discrimination	Treatment	94	13.745	4.823	0.760	NS
	Control	100	13.230	4.607		
Recites Alphabet	Treatment	94	9.970	8.751	-1.063	NS
	Control	100	11.370	9.441		
Lowercase Letter Knowledge	Treatment	94	19.755	19.133	-0.696	NS
	Control	100	21.710	19.946		
Sounds of Lowercase Letters	Treatment	94	5.649	8.156	-0.492	NS
	Control	100	6.260	9.092		
Auditory Discrimination	Treatment	94	6.638	3.382	-0.358	NS
	Control	100	5.820	3.663		
Survival Sight Words	Treatment	94	1.309	1.606	-1.148	NS
	Control	100	1.660	2.527		
Basic Pre-primer Vocabulary	Treatment	94	0.606	2.574	-0.970	NS
	Control	100	1.100	4.253		
Total Brigance	Treatment	94	117.511	36.140	-0.731	NS
	Control	100	121.570	40.872		

There were two statistically and substantively significant differences in demographic characteristics between the treatment and control children in the Brigance analysis sample that were related to posttest outcomes. One of these related to the child's ethnicity. The UPSTART treatment group children were more likely to be Caucasian (88% vs. 77%) whereas the control group children were more likely to be Hispanic (18% vs. 5%). The other difference was that the control group children were more likely to have had experience with computers in daycare or a preschool setting (22 vs. 6%).[9] See Appendix B and Appendix C for further details. Child ethnicity (Caucasian vs. Otherwise) and the Brigance pretest were used as covariates in preliminary regression analyses to adjust posttest outcomes for pre-existing between group differences. The Brigance pretest was retained as a statistical control variable in the final regression analysis.

Brigance Posttest Results. Posttest results showed that the UPSTART treatment group children performed significantly better than the control group children on six of the 10 Brigance subtests*: Visual Discrimination, Recites the Alphabet, Lowercase Letter Knowledge, Letter*

[9] This variable was actually more of an indicator of control group membership since the control children were mostly recruited from preschools. (The correlation between "Used PC in Daycare" and "Study Group" was r = -.22, $p < .01$) .

Sounds, Auditory Discrimination, and Vocabulary. While the UPSTART treatment group children scored 8.7 points higher than the control group children on the Total Brigance posttest, this difference was not statistically significant at the 95% confidence interval (p=.14). Treatment-control group differences on the Brigance at posttest are presented in Table 21.

Table 21
Brigance Posttest Analysis of Treatment-Control Group Differences

Brigance Posttest	*Group*	*N*	*Mean*	*SD*	*t*	*Significance*
Expressive Objects	Treatment	94	25.989	1.406	0.568	NS
	Control	100	25.900	1.049		
Receptive Objects	Treatment	94	26.968	0.176	-0.077	NS
	Control	100	26.970	0.171		
Expressive Grammar	Treatment	94	9.787	1.860	-0.384	NS
	Control	100	9.880	1.492		
Visual Discrimination	Treatment	94	18.617	1.814	4.478	**
	Control	100	17.090	2.850		
Recites Alphabet	Treatment	94	19.809	8.425	2.252	*
	Control	100	16.850	9.853		
Lowercase Letter Knowledge	Treatment	94	42.809	13.876	2.387	*
	Control	100	37.430	17.403		
Sounds of Lowercase Letters	Treatment	94	17.585	9.367	3.194	**
	Control	100	13.040	10.445		
Auditory Discrimination	Treatment	94	8.309	2.813	3.788	**
	Control	100	6.640	3.313		
Survival Sight Words	Treatment	94	3.670	3.900	1.544	NS
	Control	100	2.890	3.058		
Basic Pre-primer Vocabulary	Treatment	94	11.000	9.727	5.037	**
	Control	100	4.730	7.368		
Total Brigance	Treatment	94	169.190	40.066	1.482	NS
	Control	100	160.490	41.592		

*p≤.05 **p≤.01

Using the data from Table 21, effect sizes[10] were calculated to show the magnitude of UPSTART's impact at posttest as measured by each of the 10 Brigance subtests and the Total Brigance. The effect size (ES) estimates are presented in Table 22 and show the magnitude of the average performance difference in standard deviation units between the treatment group and the control group on each of the Brigance assessments administered at posttest in the C5 evaluation.

UPSTART children scored significantly higher on six of the ten Brigance subtests at the posttest follow-up. The ES estimates for Recites Alphabet, Letter Knowledge and Letter Sounds ranged from .30 to .43 and would be considered small effects by Cohen.[11] The ES estimates for visual

[10] An effect size was calculated for each test as the treatment group mean minus the control group mean divided by the control group standard deviation.
[11] See Chapter 2 in Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

and auditory discrimination ranged from .50 to .53 and would be considered medium size effects by Cohen. The UPSTART children's performance on the Basic Vocabulary subtest would be considered a large impact (ES = .85). On average, treatment group children scored almost 9 points higher on the total Brigance at posttest compared to control group children. This overall difference approached statistical significance (p=.14) and substantively would be considered a small effect.

Table 22
Brigance Effect Size Estimates

Brigance Posttest	***Effect Size***	***Significance***	***Magnitude of Effect***
Expressive Objects	0.085	NS	
Receptive Objects	-0.111	NS	
Expressive Grammar	-0.062	NS	
Visual Discrimination	0.535	**	Medium
Recites Alphabet	0.300	*	Small
Letter Knowledge	0.309	*	Small
Letter Sounds	0.435	**	Small
Auditory Discrimination	0.504	**	Medium
Survival Sight Words	0.255	NS	(Small)
Basic Pre-Primer Vocabulary	0.850	**	Large
Total Brigance	0.209	NS	(Small)

*p≤.05 **p≤.01

Figure 9 (on the next page) shows the Brigance effect size estimates for significant C5 results by total test and subtest in bar chart format.

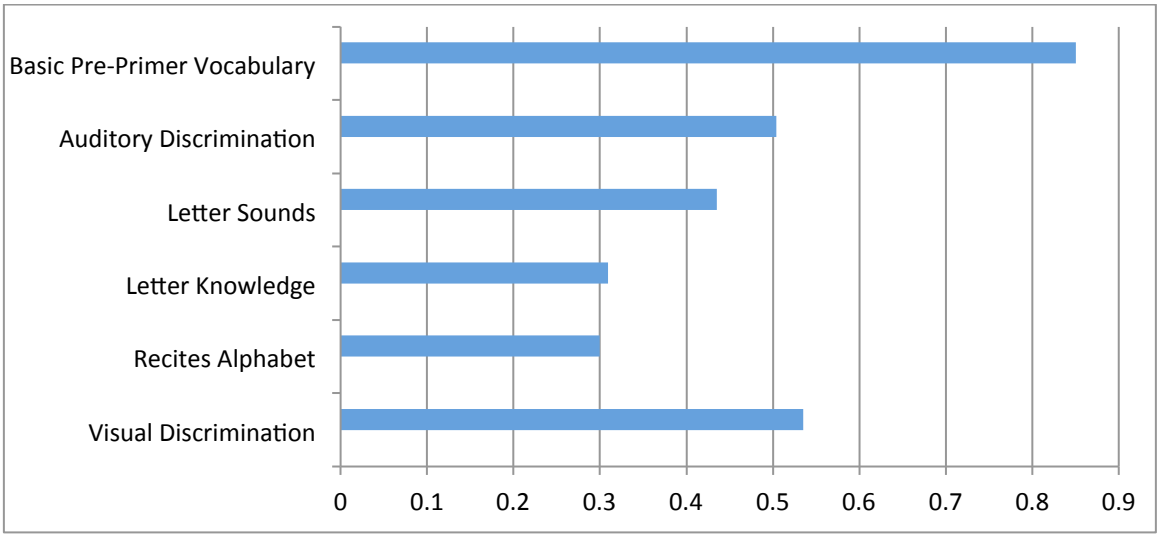

Figure 9
UPSTART Impact as Measured by the Brigance in Effect Size Units

Brigance Posttest Regression Results. Adjusting for initial differences in literacy skills between the treatment and control groups through the use of multiple regression analysis, it was found that the treatment group children outscored the control group children on the *overall Brigance* posttest by 11.84 points. The regression-adjusted posttest difference on the Total Brigance is approximately 3 points larger than the raw t-test results reviewed earlier (i.e., an average difference of approximately 9 points versus 12 points).

The final Brigance regression model[12] is shown in Tables 23 and 24.

Table 23
OLS ANOVA Summary Table for Total Brigance Posttest – Beginning of Kindergarten

Model	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Regression	174731.446	2	87365.723	111.622	.000
Residual	149494.806	191	782.695		
Total	324226.253	193			

Predictors: Total Brigance Pretest, Study Group (Treatment vs. Control)

[12] The preliminary regression model showed that the child ethnicity covariate (Caucasian vs. Otherwise) was statistically non-significant when entered into the regression equation with the Brigance pretest.

Table 24
OLS Regression Coefficients for Total Brigance Posttest – Beginning of Kindergarten

Model	*Unstandardized Coefficients*		*Standardized Coefficients*			
	B	Std. Error	Beta	t	*Sig.*	*Partial Correlation*
(Constant)	66.611	6.939		9.599	. 000	
Brigance Pretest	.772	.052	.727	14.784	.000	.731
Treatment Group	11.836	4.025	.145	2.941	.004	.208

Adjusted R Square = .53 Partial R Square = .04

The observed effect size for the Brigance treatment group using regression adjusted estimates is found by computing the partial R square statistic for Treatment Group, which in this case has a value of .04 (i.e., .208 squared = .043). A partial R square value of .04 is in the small effects size range (see Cohen, 1988; Chapter 9). The regression-based estimate of UPSTART's overall impact on early literacy growth as measured by the Brigance – an average difference of almost 12 points on the Brigance compared to the control group -- is similar to the un-adjusted estimate obtained from the t-test procedure. Both estimates would be classified as small effects.

Bader Pretest Results. There were no statistically significant differences between the UPSTART treatment group children and the control group children on the Bader at pretest. Table 25 shows the Bader pretest results.

Table 25
Bader Pretest Analysis of Treatment-Control Group Differences

Bader Pretest	*Group*	*N*	*Mean*	*SD*	*t*	*Significance*
Rhyme Recognition	Treatment	89	5.99	3.07	0.37	NS
	Control	100	5.83	2.89		
Phoneme Blending	Treatment	89	1.28	2.31	-1.49	NS
	Control	100	1.86	2.99		
Phoneme Segmenting	Treatment	89	0.52	1.57	-1.62	NS
	Control	100	0.96	2.16		
Total Bader	Treatment	89	7.79	4.59	-1.09	NS
	Control	100	8.65	6.21		

The same two demographic differences between the treatment and control children were evident in the Bader analysis sample. The UPSTART treatment group children in the Bader analysis sample were more likely to be Caucasian (89% vs. 77%) whereas the control group children were more likely to be Hispanic (18% vs. 5%). Additionally, control group children were more likely to have had experience with a computer in daycare (22% vs. 7%). See Appendix B and

Appendix C for further details. Child ethnicity (Caucasian vs. Otherwise) and whether the child used a computer in daycare were used as covariates along with the Bader pretest in preliminary regression analyses to adjust posttest outcomes for pre-existing between group differences. The Bader pretest was retained as a statistical control variable in the final regression analysis.

Bader Posttest Results. Table 26 presents the Bader posttest results and shows a statistically significant treatment group effect for all subtests as well as for the Total Bader. The mean observed (unadjusted) difference between the treatment and control group on the Total Bader posttest – which favored the UPSTART treatment group – was 5.33 points.

Table 26
Bader Posttest Analysis of Treatment-Control Group Differences

Bader Posttest	***Group***	***N***	***Mean***	***SD***	***t***	***Significance***
	Treatment	89	7.96	2.66		
Rhyme Recognition	Control	100	7.07	3.22	2.07	*
	Treatment	89	5.21	2.93		
Phoneme Blending	Control	100	3.15	3.27	4.58	**
	Treatment	89	4.51	3.24		
Phoneme Segmenting	Control	100	2.12	3.02	5.24	**
	Treatment	89	17.67	6.67		
Total Bader	Control	100	12.34	7.33	5.21	**

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

Effect size estimates for the Bader posttest results are presented in Table 27 using Cohens d. These data show the magnitude of the average performance difference in standard deviation units between the C5 treatment group and the control group on each of the Bader assessments administered at posttest in the C5 evaluation. The size of UPSTART's impact on early literacy skills measured by the Bader assessment ranged from small to medium. Overall, UPSTART produced medium-size impacts on enhancing preschool children's phonological awareness in year 5 of the program.

Table 27
Bader Effect Size Estimates

Bader Posttest	***Effect Size***	***Significance***	***Magnitude of Effect***
Rhyme Recognition	0.27	*	Small
Phonemic Blending	0.63	**	Medium
Phoneme Segmenting	0.79	**	Medium
Total Bader	0.73	**	Medium

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

Figure 10 shows the Bader effect size estimates by total test and subtest in bar chart format.

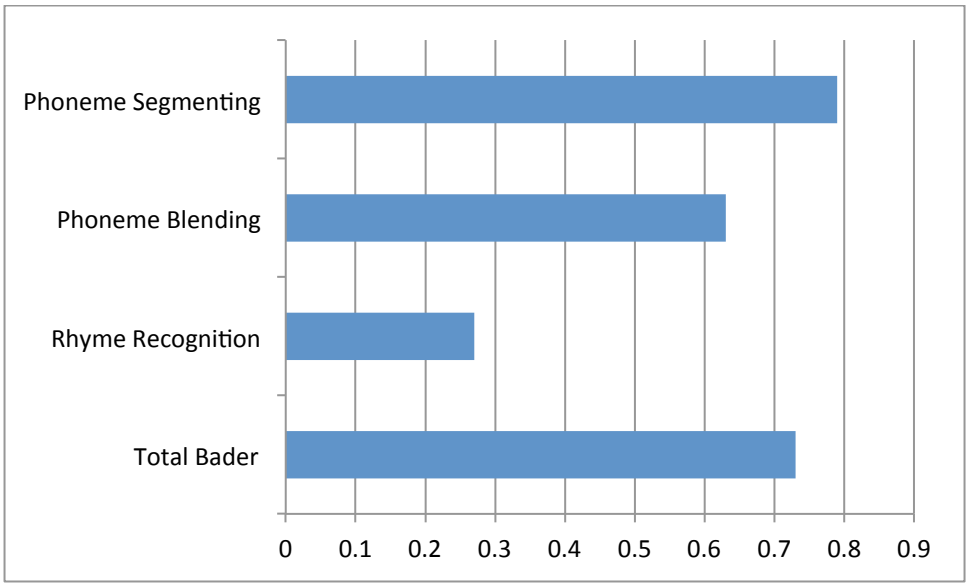

Figure 10
UPSTART's Impact as measured by the Bader in Effect Size Units

Bader Posttest Regression Results. Statistically adjusting for the initial between group differences using multiple regression analysis, it was found that the treatment group outscored the control group on the Total Bader by 5.93 points on the average. This is almost exactly the same result as found previously in the C4 evaluation. The final Bader regression model[13] for the C5 program is shown in Tables 28 and 29.

Table 28
OLS ANOVA Summary Table for Total Bader Posttest – Beginning of Kindergarten

Model	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Regression	4038.251	2	2019.125	57.376	.000
Residual	6545.601	186	35.191		
Total	10583.852	188			

Predictors: Bader Pretest and Study Group (Treatment vs. Control)

[13] Preliminary regression models showed that the effect of the demographic covariates (child's ethnicity if Caucasian and whether the child used a computer in daycare) were statistically non-significant when entered into the regression equation with the total Bader pretest.

Table 29
OLS Regression Coefficients for Total Bader Posttest – Beginning of Kindergarten

	Unstandardized Coefficients		*Standardized Coefficients*			*Partial*
	B	Std. Error	Beta	t	*Sig.*	*Correlation*
(Constant)	6.371	.904		7.051		
Total Pretest	.690	.079	.506	8.757	.000	.540
Study Group	*5.930*	.867	.396	6.838	.000	.448

Adjusted R Square = .37 Partial R Square = .19

The partial R square for Study Group in the Bader regression analysis is .20[14], which suggests a medium overall effect size (see Cohen, 1988; Chapter 9) for the C5 UPSTART program in helping to develop children's phonological awareness as measured by the overall Bader. This interpretation is consistent with the effect size estimate obtained for the total Bader posttest (see Table 27). It should also be noted that the statistically adjusted overall treatment effect of 5.93 is similar to, but somewhat larger than the raw difference score reported previously in the discussion of the t-test findings for the Bader posttest (i.e., 5.93 vs. 5.33 points).

Do UPSTART students show stronger literacy growth rates from preschool to Kindergarten than control students?

Paired samples t-tests were performed to examine growth rates as measured by the Brigance and the Bader total test batteries and subtests for the treatment and control group children. Growth rates for the treatment and control children were compared based on the observed difference scores between the posttest and the pretest. Significant differences in growth rates were estimated on the basis of whether or not the confidence intervals of the treatment and control groups overlapped at the 99% Confidence Interval of the Mean Growth Rate.

Brigance Growth Score Results. There was statistically significant growth from pretest to posttest for the matched Brigance treatment group sample (N=94) on the Total Brigance and on all ten subtests. Similar results were observed for the matched Brigance control group (N=100) except for the Expressive Grammar, Receptive Objects, and Expressive Objects subtests, which did not show significant growth from pretest to posttest (using the 99% confidence interval).

Growth rates were significantly different at the 99% CI between the treatment and control group on three of the Brigance subtests: *Letter Sounds, Auditory Discrimination, and Vocabulary.* On these three subtests, the confidence intervals do not overlap. Differences in growth rates in these

[14] Partial r squared = $.448^2 = .200$

three areas all favored the UPSTART treatment group. That is, *UPSTART children showed significantly stronger growth rates in (a) learning how to pronounce letter sounds, (b) learning how to tell the difference between letter sounds, and (c) in developing their vocabulary*. The Brigance growth rate results are presented in Table 30.

Table 30
Treatment-Control Group Differences in Growth Rates on the Brigance

Brigance Test	***Control Group***		***Treatment Group***		***Significance p≤.01***
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Expressive Objects	.940	.461 – 1.419	.723	.404 – 1.043	NS
Receptive Objects	.350	-.105 – .805	.340	.103 - .578	NS
Expressive Grammar	1.040	.572 - 1.508	.851	.340 – 1.362	NS
Visual Discrimination	3.860	2.647 – 5.073	4.872	3.674 – 6.070	NS
Recites Alphabet	5.480	2.863 - 8.097	9.830	6.871 – 12.789	NS
Letter Knowledge	15.720	11.282 - 20.158	23.053	18.313 – 27.793	NS
Letter Sounds	6.780	4.655 – 8.905	11.936	9.526 – 14.346	**
Auditory Discrimination	.820	-.407 – 2.407	2.670	1.645 – 3.695	**
Survival Sight Words	1.230	.685 - 1.775	2.362	1.429 – 3.294	NS
Basic Vocabulary	3.630	2.075 - 5.185	10.394	7.850 – 12.937	**
Total Brigance	38.920	31.753 - 46.087	51.681	43.213 - 60.149	NS

The differences in growth rates between the treatment and control groups are also shown in bar chart format in Figure 11.

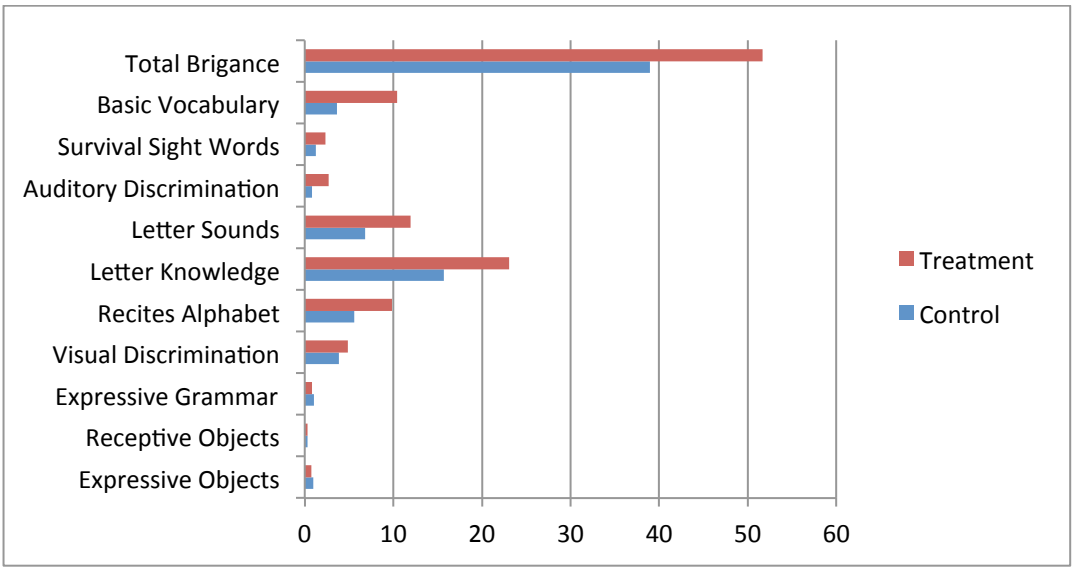

Figure 11

Growth Rate Comparisons on the Brigance

Bader Growth Score Results. There was statistically significant growth from pretest to posttest for the matched Bader treatment group sample (N=89) and for the matched Bader control group sample (N=100) on the Total Bader and all of the Bader subtests. Additionally, the UPSTART treatment group showed significantly stronger growth rates (statistically significant at the 99% CI) relative to the control group on the Total *Bader* and on two of the three Bader subtests. Specifically, *the UPSTART treatment group showed stronger growth rates from pretest to posttest for Phoneme Blending and Phoneme Segmenting skills as well as for the overall Bader assessment*. Differences in growth rates between the treatment and control groups were not statistically significant at the 99% CI for phonological awareness skills measured by the Rhyme Recognition subtest of the Bader. These results are shown in Table 31. The Bader growth rate results for the C5 group are essentially the same as found last year for the C4 group.

Table 31
Treatment-Control Group Differences in Growth Rates on the Bader

Bader Test	***Control Group***		***Treatment Group***		***Significance p≤.01***
	Mean Growth	99% CI Growth Rate	Mean Growth	99% CI Growth Rate	
Rhyme Recognition	1.240	.276 – 2.204	1.966	.901 – 3.031	NS
Phoneme Blending	1.290	.597 – 1.983	3.933	3.093 - 4.771	**
Phoneme Segmenting	1.160	.427 - 1.893	3.989	3.064 – 4.914	**
Total Bader	3.690	2.156 - 5.224	9.889	8.075 - 1.670	**

Figure 12 uses bar charts to compare the growth rates of the treatment and control group as measured by the Total Bader and each of its subtests from pretest to posttest for the C5 matched samples.

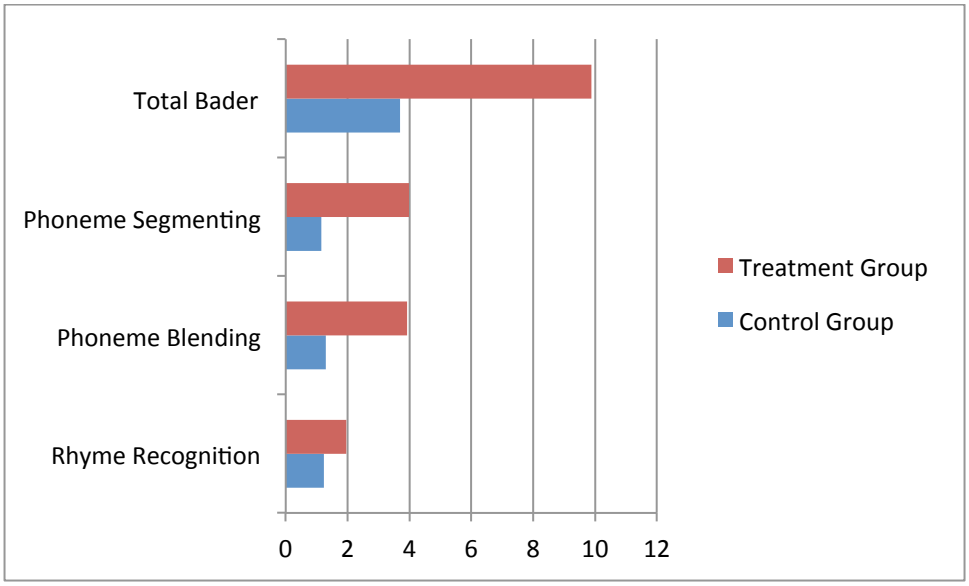

Figure12
Growth Rate Comparisons on the Bader

Summary and Conclusions

This final section of the Cohort 5 (C5) evaluation report summarizes;

- The data that were collected and analyzed;
- The analysis methods employed;
- C5 findings and trends in UPSTART implementation and usage;
- C5 findings and trends in UPSTART's impact on the development of early literacy skills;
- Implications for UPSTART practice; and
- Study limitations and implications for future UPSTART evaluation.

Data Collection

271 preschool children were recruited for the C5 evaluation whose parents had agreed to participate in the study. This number included 109 treatment group children who had enrolled in UPSTART for Year 5 of the program and 162 control group children who had not enrolled in the UPSTART program. The children's parents were administered an intake questionnaire (see Appendix A) at the time their children were pretested on the Brigance and Bader assessments over the summer of 2013. The treatment and control group children were subsequently posttested on the Brigance and Bader in the summer of 2014.

The intake interview established that 15 of the 109 children recruited for the treatment group had had prior exposure to the UPSTART online curriculum, most likely through siblings who had previously participated in the program. These 15 cases were removed from the final analysis sample. All 94 of the remaining treatment group children provided complete and valid pretest and posttest data from the Brigance; 89 of the 94 treatment group children provided complete and valid pretest and posttest data from the Bader assessment.

A quality control review determined that complete, valid and matched Brigance and Bader test data were available for 141 of the 162 control group children. In order to ensure that the treatment and control group samples were reasonably balanced and of approximately equal size, a random sample of 100 control group children were selected from the 141 remaining control cases to form the final control group sample.

The final analysis sample consisted of 100 control group children with matched Bader and Brigance test data. The final treatment group sample was composed of 89 children with matched Bader data and 94 children with matched Brigance data.

The C5 study's overall measurement attrition rate was 10%, caused primarily by children not being able to score on the Bader pretest. The C5 study's measurement attrition rate of 10% is an improvement over previous years which ranged from 15% to 24%.

Data Analysis

To determine whether UPSTART children have better literacy skills at Kindergarten compared to control group children, group equivalence on the pretests was first examined using independent sample t-tests. Relationships between the demographics and the posttest scores were then examined using correlation analyses.[15] Next, posttest differences between the treatment and control group were examined for both the Brigance and Bader. Finally, posttest differences were re-examined by adjusting for initial differences between the treatment and control groups with the use of multiple regression analysis. The regressions used a hierarchical block design in which the pretest and a set of demographic covariates were entered first, followed by the treatment-control group comparison.

The magnitude of UPSTART's impact was also estimated using effect size estimates as measured by standardized treatment and control group differences on the Brigance and Bader posttests. Effect size estimates for each of the two tests and their subtest were interpreted as demonstrating either a small, medium or large effect on the children's early literacy development.

To determine whether UPSTART students show stronger literacy growth rates from preschool to kindergarten relative to control students, paired sample t-tests were run to compare pretest and posttest scores for the matched Brigance and Bader treatment groups on the total test and each of the subtests. Statistically significant growth rates were determined by examining confidence intervals for the treatment and control groups for each test measure at the 99% confidence interval.

The effect of UPSTART usage on early reading proficiency was examined for UPSTART participants using a simple regression analysis and an analysis of covariance (ANCOVA) in which usage levels were split into quartiles based on the usage distribution of the preschool analysis samples. The final ANCOVA models statistically controlled for initial literacy skills as measured by the pretest on each respective measure. The ANCOVA analyses estimated the effects of usage at quartiles one through three compared with usage at the fourth quartile, controlling for the children's initial level of literacy development.

Findings on UPSTART Implementation

The UPSTART implementation contractor – the Waterford Institute – provided documentation for a fifth-year project enrollment of 1,577 preschool children. A majority (70%) of the C5 enrollment were from low income families. Slightly more girls (50.5%) were enrolled than boys (49.5%). In terms of ethnicity, the vast majority of the C5 enrollment was Caucasian (74%), 20%

[15] It was necessary to transform a number of the demographic measures from nominal measures to scale measures by creating "dummy variables" on the basis of the dominant characteristics of the sample. For example, parent's marital status was transformed into whether the parent was married or not, or percent married.

were Hispanic, and the remaining 6% were of African American, Asian, Native American, Pacific Islander, or unknown origin. The primary language spoken by the vast majority of the C5 children was English (84%). Approximately 15% of the C5 children spoke Spanish and 1% spoke other languages. Twelve percent of the C5 children had a diagnosed disability, most often involving speech impairments.

As in previous years, most of the C5 participants (74%) received a computer drive with the UPSTART curriculum loaded on it. Approximately 8% of the fifth year participants received a computer loan and a free Internet subscription to help them access the UPSTART curriculum. Another 7% of the C5 participants were loaned a personal computer to use at home while participating in UPSTART. The remaining 11% of the fifth year participants were provided with various combinations of educational technology to enable them to access the UPSTART curriculum, including wireless and cellular devices.

Findings about UPSTART usage are summarized below.

- The average level of UPSTART curriculum usage in Year 5 was 71 hours. This was the same level of average usage as in Year 4 and four hours higher than in Year 3.

- The C5 preschool analysis samples had a mean of approximately 72-73 hours of UPSTART curriculum usage over the fifth year of the project. This compares with an average of approximately 75 hours of instruction for program "graduates" and approximately 71 hours of instruction for all students enrolled in UPSTART in the fifth year of the program.

- The UPSTART graduation rate continued to hold at 94% in Year 5. This is the same level of program attainment as realized in Years 3 and 4.

- UPSTART graduate status in the fifth year of the program was significantly correlated with hours of instruction ($r=.58$) as well as with the duration of program participation as measured by weeks of attendance ($r=.74$).

- UPSTART graduate status in the fifth year of the program was significantly correlated with early literacy outcomes at the beginning of Kindergarten as measured by the Brigance posttest ($r=.24$, $p<.05$) and the Bader posttest ($r=.22$, $p<.05$).

- UPSTART curriculum usage was significantly correlated with literacy skills at the beginning of Kindergarten as measured by the Brigance posttest ($r = .45$, $p<.01$) and the Bader posttest ($r = .41$, $p<.01$). Controlling for initial levels of literacy skills, the correlation of UPSTART usage with Kindergarten outcomes was somewhat lower: $r=.34$ ($p<.01$) for the Brigance sample and $r=.40$ ($p<.01$) for the Bader sample.

- UPSTART usage accounted for 16 to 19% of the variance in literacy skills developed by C5 children at Kindergarten entry as measured by the Bader and Brigance posttests respectively.

- There appears to be a moderately strong *curvilinear relationship* between UPSTART curriculum usage and preschool children's early literacy development. Early literacy skills measured by the Brigance tended to peak and level off with the attainment of at least 65 hours of curriculum usage. The development of phonological awareness skills measured by the Bader tended to peak and level off after about 77 hours of usage.

Findings on UPSTART Impact

In the fifth year of the program, UPSTART was observed to have a small overall impact on the development of participating children's early phonics skills as measured by the Brigance assessment. Adjusting for pre-existing differences between the treatment and the control group, UPSTART participants on the average scored almost 12 points higher on the overall Brigance posttest compared to similar preschool children not enrolled in the program. Significant program impacts (relative to control group performance) were found in six of the ten areas assessed by the Brigance. Compared to similar nonparticipants, UPSTART produced:

  - Small effects in helping children learn how to recite the alphabet, name and recognize lower case letters, as well as produce the sounds of lower case letters;
  - Medium size effects in helping children learn how to hear and see differences in words and in the letters of the alphabet; and
  - Large effects in helping children learn how to read basic vocabulary words found in pre-primer reading programs.

- UPSTART children showed significantly stronger growth rates in (a) learning how to pronounce letter sounds, (b) learning how to tell the difference between letter sounds, and (c) in developing their vocabulary.

- As noted above, the largest phonics related impact observed in UPSTART's fifth year of operation was in the development of *vocabulary*. *This finding replicates results found in the third and fourth year evaluations of UPSTART*.

- Overall, UPSTART achieved medium size effects on improving the phonological awareness skills of participants in Year 5 of the program as measured by the Total Bader assessment. These results replicate the evaluation findings from Year 4 of the program.

- On average, UPSTART treatment group children scored an average of almost six points higher on the Bader posttest (regression adjusted) relative to control group children. Compared to similar nonparticipants, UPSTART produced:

- Small effects in helping young children recognize pairs of words that rhymed; and
- Medium size effects in helping children learn how to blend and segment phonemes.

- The UPSTART treatment group showed stronger growth rates from pretest to posttest for Phoneme Blending and Phoneme Segmenting skills as well as on the overall Bader assessment.

  Two of the early reading skills that UPSTART focuses on include helping young children improve their knowledge of letter sounds and helping them develop their vocabulary. Similarly, the program also focuses on helping young children learn how to read by developing their phoneme blending and segmenting skills. In Year 5 of the program, UPSTART made substantial progress in these areas toward achieving its goals for helping to develop young children's early literacy.

Implications for Practice

The results of the usage analysis probably have the greatest implications for UPSTART instructional practice. The C5 children who benefitted the most from UPSTART in terms of developing their early phonics-related skills – as measured by the Brigance assessment – were those children who used the online curriculum for at least 65 hours. Alternatively, those children who benefited the most in terms of developing their phonological awareness skills – as measured by the Bader assessment -- were those children who used the online curriculum for at least 77 hours. These results suggest that young children may need more instructional time for developing phonological awareness skills relative to basic phonics skills.

Study Limitations and Implications for UPSTART Evaluation

The implications for practice just discussed were derived from an analysis of covariance (ANCOVA) procedure conducted as part of the evaluation's usage analysis. The reliability of the ANCOVA estimates of usage effects were adversely affected by the study's sample size. Our confidence in the Bader's estimates of usage effects are greater than those obtained from the Brigance, in part, because the usage effects detected by the Bader were stronger than those detected by the Brigance. Nevertheless, the statistical power of the ANCOVA usage analysis -- with both the Bader and the Brigance samples -- was less than optimal because of the sample size requirements for this particular statistical procedure. In short, the ANCOVA allows us to obtain the kind of information we seek about the effects of differing levels of usage, but it requires larger sample sizes than we have been able to obtain.

We used the ANCOVA procedure to estimate the effects of four usage groups, defined by quartiles along the usage distribution. In the Brigance sample (N = 94), there were 23-24 children in each of the quartile usage groups. The observed power for the Brigance ANCOVA was .69 whereas the desired level of statistical power for an inferential analysis is .80 or greater. To achieve the desired level of statistical power, we would need quartile usage subgroups of

approximately 40 children each.[16] This would require obtaining a final analysis sample of approximately 160 UPSTART children.

The UPSTART recruitment pool was composed of 150 families. The C5 evaluation was able to utilize approximately 63% of this recruitment base in forming the final treatment group analysis sample (i.e., 94/150 = .626). Based on a recruitment success rate of 63%, the evaluation would need to begin recruitment with a pool of approximately 255 families (i.e., x=160/.63=254) in order to achieve the desired sample size (N=160). Thus, in order to produce findings from a usage analysis employing the ANCOVA procedure with sufficient confidence, future UPSTART evaluations would need to start with a larger recruitment pool in the neighborhood of 255 families. We suggest that recruitment be expanded for the Cohort 6 evaluation.

[16] Based on the need to detect moderate size usage effects, which we know from the C5 analyses, and using a criterion of 95% confidence.

Appendix A

UPSTART Evaluation
Parent Intake Form

Please check the appropriate response with an "**X**". Choose only ONE response for each question.

1. Have any of your children participated in the UPSTART program in the past?
   □$_1$ Yes □$_2$ No

   1a. If yes, did your 4-year-old also use the program?
   □$_1$ Yes □$_2$ No

Child Information

2. What is your child's birthday? _____

3. What is your child's gender?
   □$_1$ Male □$_2$ Female

4. What year will your child be entering Kindergarten? _____

5. What is your child's ethnicity?
   □$_1$ Hispanic □$_2$ Native American/Alaskan Native □$_3$ Asian/Pacific Islander
   □$_4$ Caucasian □$_5$ African American □$_6$ Other:_____

6. What is your child's primary language?
   □$_1$ English □$_2$ Spanish □$_3$ Portuguese □$_4$ Chinese
   □$_5$ German □$_6$ Japanese □$_7$ Other:_____

7. Is your child currently attending a daycare/preschool?
   □$_1$ Yes □$_2$ No

   7a. **If yes**, approximately how many hours a week does your child attend a daycare/preschool?
   □$_1$ less than 10 hours □$_2$ 10-19 hours □$_4$ 20-24 hours
   □$_5$ 25-29 hours □$_6$ 30-34 hours □$_7$ 35 or more hours

8. Does your child have access to a computer in your house?
   □$_1$ Yes □$_2$ No

9. Does your child use a computer in her/his day care or preschool?
   □$_1$ Yes □$_2$ No □$_3$ Not Applicable (not in day care or preschool)

10. How comfortable is your child using a computer?
    □$_1$ Very comfortable □$_2$ Somewhat comfortable □$_3$ Somewhat uncomfortable
    □$_4$ Not comfortable □$_5$ Very uncomfortable

Caregiver Information

11. What is your relation to the participating child?
    □1 Mother □2 Father □3 Grandmother □4 Grandfather
    □5 Step Father □6 Step Mother □7 Other:_____

12. What is your ethnicity?
    □1 Hispanic □2 Native American/Alaskan Native □3 Asian/Pacific Islander
    □4 Caucasian □5 African American □6 Other:_____

13. What is your primary language?
    □1 English □2 Spanish □3 Portuguese □4 Chinese
    □5 German □6 Japanese □7 Other:_____

14. What is the highest level of education you have completed?
    □1 Did not complete high school □2 High school diploma/GED □3 High school
    □4 Some college □5 Bachelor's degree □6 Masters degree □7 Doctorate

15. What is your paid employment status:
    □1 Full time □2 Part time □3 Not working

16. What is your spouse's paid employment status:
    □1 Full time □2 Part time □3 Not working
    □4 Not Applicable (single parent)

17. What is your marital status?
    □1 Married □2 Separated □3 Divorced □4 Unmarried

18. How many people live in your home (including you and all your children)?
    □1 One □2 Two □3 Three □4 Four □5 Five □6 Six or more

19. What is your total household annual income?
    □1 under $10,000 □2 $10,000-$24,999 □3 $25,000-$49,999
    □4 $50,000-$74,999 □5 $75,000-$99,000 □6 $100,000 or more

Thank you for participating in the Utah UPSTART Evaluation!

Appendix B
Brigance Sample: Treatment – Control Group Differences on Demographics

Covariate	*Group*	*N*	*Mean*	*SD*	*t*	*Significance*
Child is female	Treatment	94	.58	.50	.072	NS
	Control	100	.58	.50		
Child is Caucasian	Treatment	94	.88	.32	2.098	*
	Control	100	.77	.42		
Child is Hispanic	Treatment	94	.05	.23	-2.813	**
	Control	100	.18	.39		
Child's primary language is English	Treatment	94	.96	.20	.245	NS
	Control	100	.95	.22		
Child attended preschool or daycare	Treatment	94	.35	.48	-5.342	**
	Control	100	.71	.47		
Child used computer in preschool or daycare	Treatment	94	.06	.25	-3.156	**
	Control	100	.22	.42		
Child has access to a computer at home	Treatment	94	.94	.25	.913	NS
	Control	100	.90	.30		
Child comfort level with computers	Treatment	94	1.83	.79	-1.482	NS
	Control	100	2.02	.99		
Parent is Caucasian	Treatment	94	.91	.28	1.210	NS
	Control	100	.86	.35		
Parent's primary language is English	Treatment	94	.96	.20	1.565	NS
	Control	100	.90	.30		
Parent Educational Attainment (recoded)[17]	Treatment	94	2.43	.58	-7.818	**
	Control	100	3.22	.82		
Parent is married	Treatment	94	.90	.30	.950	NS
	Control	100	.86	.35		
Parent is working	Treatment	94	.43	.50	-.024	NS
	Control	100	.45	.50		
Household size	Treatment	94	5.05	.92	1.304	NS
	Control	100	4.86	1.14		
Household annual income category	Treatment	94	3.55	1.33	-.834	NS
	Control	100	3.70	1.12		

**p≤.01
*p≤.05

[17] 1 = High School Dropout; 2 = High School Graduate, 3 = Some College; 4 = College Graduate

Appendix C
Pearson Correlations with Total Posttest Scores

Variable	***Brigance*** (N=194)	***Bader*** (N=189)
Study Group[18]	.11	.36**
Pretest	.72**	.48**
Child is Female	.09	.05
Child is Caucasian	.33**	.22**
Child is Hispanic	-.31**	-.21**
Child's primary language is English	.26**	.22**
Child attended daycare/preschool	.02	-.06
Child has computer access at home	.11	. 08
Child used computer in preschool/daycare	-.26**	-.16*
Child's computer comfort	.01	-.04
Parent is Caucasian	.29**	.22**
Parent's primary language is English	.26**	.21**
Parent Educational Attainment	.09	-.10
Parent is married	.26**	.19**
Parent is employed	-.08	.03
Household size	.09	.10
Household income	.28**	.29**

**p≤.01 *p≤.05

[18] Coded 1 if Treatment Group and 0 if Control Group

Appendix D
Bader Sample: Treatment – Control Differences on Demographics

Covariate	*Group*	*N*	*Mean*	*SD*	*t*	*Significance*
Child is female	Treatment	89	.58	.50	.072	NS
	Control	100	.58	.50		
Child is Caucasian	Treatment	89	.89	.32	2.176	*
	Control	100	.77	.42		
Child is Hispanic	Treatment	89	.05	.21	-3.036	**
	Control	100	.18	.39		
Child's primary language is English	Treatment	89	.97	.18	.553	NS
	Control	100	.95	.22		
Child attended preschool or daycare	Treatment	89	.35	.48	-5.314	**
	Control	100	.71	.47		
Child used computer in preschool or daycare	Treatment	89	.07	.25	-3.084	**
	Control	100	.22	.42		
Child has access to a computer at home	Treatment	89	.94	.23	1.127	NS
	Control	100	.90	.30		
Child comfort level with computers	Treatment	89	1.85	.79	-1.268	NS
	Control	100	2.02	.99		
Parent is Caucasian	Treatment	89	.92	.27	1.358	NS
	Control	100	.86	.35		
Parent's primary language is English	Treatment	89	.97	.18	1.853	NS
	Control	100	.90	.30		
Parent Educational Attainment (recoded)[19]	Treatment	89	2.44	.58	-7.592	**
	Control	100	3.22	.82		
Parent is married	Treatment	89	.91	.29	1.082	NS
	Control	100	.86	.35		
Parent is working	Treatment	89	.43	.50	-.317	NS
	Control	100	.45	.50		
Household size	Treatment	89	5.06	.92	1.309	NS
	Control	100	4.86	1.14		
Household annual income category	Treatment	89	3.61	1.31	-.528	NS
	Control	100	3.70	1.12		

**p≤.01
*p≤.05

[19] 1 = High School Dropout; 2 = High School Graduate, 3 = Some College; 4 = College Graduate