

Utah UPSTART Education Program Evaluation Kindergarten Outcomes: Program Impacts on Reading Proficiency

Cohort 1/Year 1 Results Technical Report

August 2011

Submitted by:



Evaluation and Training Institute
11500 W. Olympic Blvd.
Suite 626
Los Angeles, CA 90064

All correspondence should be sent to:

Jon Hobbs, Ph.D.
jhobbs@eticonsulting.org

Contents

EXECUTIVE SUMMARY	1
INTRODUCTION	3
BACKGROUND	3
EVALUATION DESIGN	4
Research Questions	4
Outcome Measures	6
Test Data Collected	6
KINDERGARTEN ANALYSIS	8
Exploratory Analyses	9
Variable Transformations	9
OLS Regression Analysis	9
Diagnosing Heteroskedasticity	10
Weighted Least Squares Analysis	10
Analysis of Implementation Time	10
UPSTART IMPLEMENTATION	12
UPSTART Enrollment	12
UPSTART Equipment Provided	12
UPSTART Graduates	13
UPSTART Usage	14
How UPSTART Usage Relates to Reading Proficiency	17
UPSTART OUTCOMES	21
Does UPSTART improve reading readiness?	21
Does UPSTART sustain improvements in reading readiness?	22
SUMMARY AND CONCLUSIONS	25
Test Data Collected and Analyzed	25
The Analysis	25
Findings: UPSTART Implementation	26
Findings: UPSTART Impact on Reading Proficiency in Kindergarten	27

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. A majority (56%) of the 1,248 preschool children enrolled in the first year of UPSTART were from low-income families. The evaluation of UPSTART's first year of implementation was specifically designed to assess the program's impact on developing the children's reading proficiency once they enrolled in kindergarten. Other objectives included documenting the extent to which participants used the computerized curriculum; establishing the relationship between curriculum usage and reading proficiency outcomes; and documenting the program's completion or "graduation" rate.

The Kindergarten Analysis

A weighted least squares regression approach was used to estimate reading proficiency at the beginning of kindergarten as measured by the DIBLES Next (DN) beginning kindergarten composite. Low-income status was used as a weight variable to correct for non-constant variance in estimating the regression coefficients. English Language Learner (ELL) status, special education status, and gender were used as control variables in the final model to estimate reading proficiency for children who had participated in UPSTART the previous school year compared to children from the general kindergarten population who had not participated in UPSTART.

An ordinary least squares regression approach was used to estimate reading proficiency in middle kindergarten as measured by the DIBLES Next (DN) middle kindergarten composite. A hierarchical block design was used in the final model, which controlled for race, ELL status, low-income status, and special education status in examining differences in reading proficiency between the UPSTART treatment group and the kindergarten control group.

The effect of UPSTART usage on reading proficiency 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 kindergarten analysis sample. The final model statistically controlled for ELL status and special education status in kindergarten in comparing usage at each quartile with usage at the fourth quartile in estimating the effect of UPSTART on reading proficiency at the beginning and middle of kindergarten using the DN Composites.

Findings: UPSTART Implementation

Most of the first year UPSTART participants (70%) received a computer drive with the UPSTART curriculum loaded on it. Approximately 11% of the first year participants were loaned a computer and given free Internet access to help them access the UPSTART curriculum. Another 7% of the first year participants were loaned a personal computer to use at home while

participating in UPSTART. The remaining 10% to 12% of the first year participants were provided with various combinations of educational technology to enable them to access the UPSTART curriculum, including wireless and cellular devices.

- The kindergarten test sample had a mean of 68 hours of participation in the UPSTART curriculum over the first year of the project. This compares with an average of 76 hours of instruction for program “graduates” and an average of 53 hours of instruction for all students enrolled in UPSTART in the first year.
- Length of participation in the UPSTART curriculum was significantly and positively correlated with reading proficiency at the beginning ($r=.30$) and middle ($r=.29$) of kindergarten.
- Reading proficiency was shown to improve with increasing levels of UPSTART curriculum usage for both beginning and middle kindergarten children who had participated in UPSTART in preschool. These analyses controlled for ELL and special education status in kindergarten, both of which negatively impact reading proficiency outcomes.
- The UPSTART graduation rate in the first year of the program was 59%. UPSTART graduation status was not significantly correlated with reading proficiency at either the beginning ($r=.13$) or middle ($r=.10$) of kindergarten.

Findings: UPSTART Impact on Reading Proficiency in Kindergarten

- UPSTART participation had a moderately strong impact on improving the reading proficiency of UPSTART participants compared to nonparticipants at the beginning of kindergarten, as measured by the DIBELs Next Beginning Kindergarten Composite. The final model controlled for the significant influence of ELL status, special education status, and gender on reading proficiency outcomes at the beginning of kindergarten.
- UPSTART participants maintained their achievement gains through the middle of kindergarten as measured by the DIBELs Next Middle Kindergarten Composite. The final model controlled for the significant influence of ELL status, special education status, low income status and race on reading proficiency outcomes in middle kindergarten.

Based on the first year results, the evidence suggests that UPSTART’s use of education technology in a home based approach has merit for facilitating the development of school readiness in young preschool children.

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 first year of operation during the 2009-10 school year, the project's implementation contractor – the Waterford Institute – enrolled 1,248 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 evaluation of UPSTART's first year of implementation was designed specifically to assess the program's impact on developing the children's reading proficiency once they enrolled in kindergarten. Other objectives included documenting the extent to which participants used the computerized curriculum; establishing the relationship between curriculum usage and reading proficiency outcomes; and documenting the program's completion or "graduation" rate.

A majority (56%) of the 1,248 preschool children that enrolled in the first year of UPSTART were from low income families, according to data provided by the Waterford Institute. Slightly more boys (52%) were enrolled than girls (48%). In terms of ethnicity, the vast majority (81%) of the enrollment was Caucasian, 13% were Hispanic, 2% were of Asian descent, 1% were Black, and 1% were Native American. Ethnicity was unknown for 1% of the first year enrollment.

Background

The Utah State Department of Education's (USOE) quasi-experimental framework for the evaluation involved a variation of the Posttest-Only Design with Nonequivalent Groups. The USOE's scope of work suggested the need for repeated posttests in kindergarten and grade one. This design is sketched below where NR means "not randomly assigned," X stands for the implementation of the UPSTART treatment and O1, O2, and O3 stand for repeated posttests at the beginning and end of kindergarten and at the beginning of grade one.

NR	X	O1	O2	O3

NR		O1	O2	O3

In this research design, the UPSTART children received the Waterford Early Learning Program through the use of interactive personal computers at home in the year before kindergarten and are then compared with a group of children that did not participate in the UPSTART preschool program. The two main problems with this design are:

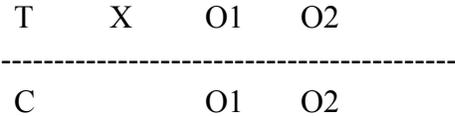
- **Selection bias:** We don't know if any observed posttest differences between the treatment and control group are due to pre-existing differences.
- **Mortality:** Even if the groups had been the same at the beginning, we don't know if any observed posttest differences are due to differential dropout rates of families from the study groups that changed the composition of the groups over time.

Thus, design limitations include not knowing whether selection effects or differential dropout rates account for any between group differences that might be observed after the preschool program. We proposed two things that could be done to improve this design: (1) use a pretest in fall 2009 with both treatment and control groups, or (2) use multiple control groups that bracket the expected effects. The USOE declined to permit the use of pretests and delays in obtaining data from the Waterford Institute and the Utah public schools derailed our plans to select a "criterion" (or All Star) control group.

Lacking the desired degree of control through experimental design, our fallback option in the first year evaluation was to exercise statistical control methods through the use of linear multiple regression analysis and analysis of covariance.

Evaluation Design

The kindergarten evaluation design that was implemented is a posttest-only repeated measures design with nonequivalent groups and is diagramed below, where T stands for children who received the UPSTART preschool program, and C stands for the general population Control group of students that 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 measurements taken at the beginning of kindergarten, and O2 indicates measurements taken in the middle of kindergarten.



Research Questions

We hypothesized that if UPSTART has no effect on improving reading readiness, then the kindergarten children who participated in UPSTART – the treatment group – would be expected to perform at the same level as the general population control group on kindergarten measures of reading proficiency at the beginning of kindergarten. If UPSTART does have an effect on improving reading readiness, then the treatment group should perform significantly better than

the control group when first measured at the beginning of kindergarten. If UPSTART sustains its gains over time, then the treatment group would be expected to continue to perform better than the control group when retested in the middle of kindergarten.

Our research questions for the school readiness component of the evaluation study are as follows:

RQ1: *Does UPSTART improve reading readiness?*

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

$$T > C @ O1$$

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

$$T = C @ O1$$

RQ2: *Does UPSTART sustain improvements in reading readiness?*

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

$$T > C @ O1 \text{ and } O2$$

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

$$T = C @ O2$$

In the kindergarten analysis, the outcomes of interest are measures of early literacy skills relevant to emerging readers such as early phonemic awareness, letter recognition, awareness of concepts of print and oral language comprehension.

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 minutes of exposure per participant per week?*

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

RQ5: *How does 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 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 outcomes of interest.

Outcome Measures

The reading skills taught by the Waterford Early Learning Program at Level 1 of the curriculum¹ include:

- Phonological Awareness: rhyming, initial sound recognition, phonemic segmenting and blending.
- Phonics: letter name knowledge, sound knowledge, and word reading.
- Comprehension and Vocabulary: vocabulary knowledge and comprehension strategy development.
- Language Concepts: print concepts and basic oral language skills.

The Utah public schools in which the UPSTART participants enrolled for kindergarten in the fall of 2010 mostly used the latest (7th) edition of *The Dynamic Indicators of Basic Early Literacy Skills* (DIBELS), called the DIBELS Next, to measure student reading proficiency. We were able to obtain DIBELS Next reading proficiency data from seven Utah public school districts for 137 kindergarten students who had participated in UPSTART. The cooperating school districts also provided us with DIBELS Next data for over 9,000 kindergarten students who had not participated in UPSTART.

At the kindergarten level, the DIBELS Next (DN) primarily measures phonemic awareness and early phonics skills. We used DN Composite Scores to measure student reading proficiency. The DN Composite for beginning kindergarten is the total score obtained by summing the scores of the First Sound Fluency (FSF) subtest and the Letter Naming Fluency (LNF) subtest. The DN Composite for middle kindergarten is the total score obtained from summing the scores of the FSF subtest, LNF subtest, Phoneme Segmentation Fluency (PSF) subtest, and the Nonsense Word Fluency Subtest. The latter measures a child's competency with the alphabetic principle and with early phonics skills.

Test Data Collected

Ten Utah school districts supplied test data for 258 children enrolled in public kindergartens in 2010 who had participated in UPSTART the previous school year. One school district submitted Developmental Reading Assessment (DRA) data for 62 students. Two school districts submitted DIBELS test data for 59 students. Seven school districts supplied DIBELS Next (DN) test data for 137 students.

¹ 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.

Only the DN data were sufficient for the desired statistical analysis. Consequently, the kindergarten analysis was carried out with DN data using a sample of 137 kindergarten students who had participated in the UPSTART preschool program the previous year and a control group of 247 kindergarten students who had not participated in UPSTART. Control students were randomly selected from the pool of DN test data at each of the seven participating school districts (Box Elder, Granite, Iron, Ogden, Salt Lake, Tooele, and Weber) using a 2:1 ratio of control to treatment group students in order to optimize sample size for the analysis.

See Appendix A for a summary of the statistical power analysis that was conducted to inform the sampling plan. See also Appendix B for further details on the samples.

Kindergarten Analysis

A kindergarten student data file was developed based on data collected from the Waterford Institute and from ten Utah public school districts that agreed to supply data for the first year evaluation. A summary of the variables examined in the DIBELS Next analysis is shown in Table 1. The DIBELS Next composite scores for the beginning and middle of kindergarten were chosen as the dependent variables for the analysis since they provide the best overall estimate of the student’s reading proficiency, according to the DIBELS Next Technical Manual.

*Table 1
Variables in the Kindergarten Analysis of Reading Readiness*

Predictor Variables	Data Values
Group	Treatment vs. Control
District	School District
Gender	Male vs. Female
Ethnicity	Asian Black/African American Caucasian/White Hispanic American Indian/Alaska Native Hawaiian/Pacific Islander Multi-Ethnic Other Unknown
Low Income (Free or Reduced Price Lunch)	Yes/No
English Language Learner (ELL)	Yes/No
Primary Language Spoken	English Spanish Other
Attended Any Preschool	Yes/No
Title 1 Compensatory Education	Yes/No
Special Education	Yes/No
Birth Date	Mm/dd/yyyy format
UPSTART “Graduate”	Yes/No
Usage	UPSTART Minutes Logged
UPSTART Provided Equipment	9 categories of equipment
Entry Date	Date enrolled in UPSTART
UPSTART Usage by week	Minutes per week
DIBELS Next Outcome Variables	Data Values
Composite Score Beginning K	0 to High
Composite Score Middle K	0 to High

Exploratory Analyses

The treatment group and control group data were examined separately using descriptive statistics and the SPSS Explore procedure to describe the two groups on the variables assessed and to check the assumptions underlying use of the general linear model (e.g., normally distributed data and homogeneity of variance). This helped determine how the data were distributed on the variables assessed, identified the extent to which there was missing data, and the extent to which the assumptions of the general linear model were met. Correlations were also run to assess the degree of relationship between the independent and dependent variables and to examine the independence of the predictor variables. Preliminary regression analyses also provided diagnostic data for checking assumptions, particularly homogeneity of variance with respect to how well the models functioned in predicting outcomes over different levels of the predictors.

Because of substantial missing data, the following variables could not be used in the subsequent analyses: primary language spoken, preschool attendance, and Title I status. The preliminary analyses also identified problems with the assumption of homogeneity of variance, particularly with the control student data. This led to the use of transformed response scores for some analyses.

Variable Transformations

Based on the distribution of the ethnicity variable, it was necessary to re-code it as a dummy variable measured as White vs. other (coded 1 for White, 0 otherwise). Gender was also re-coded as a dummy to create a continuous scaled variable called Male (1 if male and 0 if female).

As noted above, non-constant variance was diagnosed in preliminary regression runs and the response variables were subsequently transformed in an attempt to resolve this problem. Variance stabilizing transformations using square root and log transformations were carried out for the two DIBLES Next composite variables to correct for violations of the assumption of homogeneous residuals (i.e., the difference between predicted and observed reading proficiency scores).

OLS Regression Analysis

The initial kindergarten analysis used an ordinary least squares (OLS) multiple regression approach to estimate the effect of UPSTART participation on reading proficiency at the beginning and middle of kindergarten. Reading proficiency was measured by DIBLES Next composite scores for the beginning and middle of kindergarten. A 2-block hierarchical regression design was employed in which a set of five covariates was entered first to control for differences between the treatment and control group, followed by the treatment-control group comparison. The covariate control set included dummy variables for gender (Male vs. Female), ethnicity (White vs. Other), low income (Yes vs. No), English Language Learner (Yes vs. No),

and whether the child received special education services in kindergarten (SPED vs. Otherwise). The treatment and control group samples are described in Appendix B by school district and compared (unadjusted) on the five covariate measures.

Separate OLS regressions were run in estimating reading proficiency effects for the beginning and middle of kindergarten. The OLS regression procedure was successful in estimating the impact of UPSTART as measured by the DN Composite for the middle of kindergarten. However, difficulties were encountered in estimating impacts for beginning kindergarten students as discussed immediately below.

Diagnosing Heteroskedasticity

In the beginning kindergarten analysis, scatterplots of the residuals in both the initial OLS analyses and in the subsequent analyses using transformed composite scores (square root and log transformations) showed that the assumption of homogeneity of variance had been violated. The diagnosis of heteroskedasticity brought into question the accuracy of the effects estimates for beginning kindergarten using the OLS regression procedure. The solution was to determine what was causing the error variances to differ over levels of the covariates and to shift to a weighted least squares analysis. Exploratory analyses revealed that there was significantly greater variability on the DN Composite for the low income and special education control group students relative to non-low income and regular education control students or treatment group students within the beginning kindergarten sample.

Weighted Least Squares Analysis

The principle source of heteroskedasticity turned out to be the low income variable as opposed to the special education variable. The low income dummy was then used as a weight variable in a weighted least squares (WLS) analysis to correct for heterogeneous variances in the beginning kindergarten analysis. Covariates for special education status, ELL status, ethnicity (White vs. Other) and gender (percent male) were entered as control variables in the WLS analysis in estimating reading proficiency differences between treatment and control group children at the beginning of kindergarten. The initial WLS analysis showed that the ethnicity covariate was not statistically significant in the beginning kindergarten analysis, so it was dropped from the final analysis. The final results reported for the effects of UPSTART participation on reading proficiency at the beginning of kindergarten are based on the WLS analyses and include covariates for ELL, SPED, and gender, with low income status as the weight variable.

Analysis of Implementation Time

An analysis of covariance was used to determine the relationship between the amount of instruction received by UPSTART participants and reading proficiency outcomes. The usage variable (UPSTART Minutes Logged) was co-linear (redundant) with the treatment-control

group dummy variable since its value was zero for the control group students (who never used the UPSTART curriculum) and a positive value for the UPSTART treatment students (who did). For that reason, the usage variable could not be incorporated into the multiple regression analysis with the Group variable. The alternative was to run an analysis of covariance subset for the UPSTART treatment group using an ordinal version (ordered categories) of the usage variable to see what the impact of time in the program was on reading proficiency. This was accomplished by creating a new variable called Usage Group in which usage was factored into four levels corresponding to quartiles of usage for the DN test samples. The analysis of covariance that was run on Usage Group incorporated the five covariates used in the regression analysis as control measures.

Results

Findings are reported first in answering research questions 3, 4, and 5 about the implementation of the UPSTART program. We also report data describing the education technology equipment provided to enrolled children by UPSTART. Next, we report findings on the impact of UPSTART to answer research questions 1 and 2 about the extent to which UPSTART may have facilitated the development of reading proficiency in kindergarten for children who participated in the home-based education technology preschool program compared to children who did not participate in the program.

UPSTART Implementation

Findings reviewed under UPSTART implementation include enrollment in the first year, equipment provided to enrolled families by UPSTART, usage of the UPSTART curriculum, UPSTART graduates, and the relationship between levels of UPSTART curriculum usage and reading proficiency outcomes.

UPSTART Enrollment

The Waterford Institute provided documentation for a first-year UPSTART enrollment of 1,248 children. The Waterford Institute enrolled most (almost 80%) of the first-year UPSTART students in April and May of 2009. Waterford continued to enroll families over the summer and into the fall of 2009, concluding enrollment in January 2010.

UPSTART Equipment Provided

The kind of education technology supports provided to children enrolled in UPSTART is shown in Table 2 for all 1,248 students enrolled in the first year and for the DIBLES Next kindergarten sample. The majority of the first year UPSTART students (approximately 70%) 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 majority of students in the kindergarten test sample (approximately 64%) also received a computer drive with the curriculum loaded on it.

Next most often, UPSTART loaned personal computers to almost 11% of the enrolled students and gave them free access to the Internet while they used the equipment. A slightly higher percentage of the kindergarten test sample (about 17%) received a free computer loan with free Internet access in the first year of the program. Another 7% of the program participants were given access to a home computer for free while they participated in the program. Similarly, 8% of the kindergarten test sample was given access to a home computer for free while they participated in the program. The remaining 10 to 12% of the enrolled children received various combinations of computer technology to enable them to access the UPSTART curriculum.

Table 2
Percent of Students Provided Equipment by UPSTART

Equipment Provided	All UPSTART	Kindergarten Sample
Drive	70.5	63.5
Computer & Internet	10.7	16.8
Computer	6.8	8.0
Computer & Cellular	5.9	6.6
Computer & Wireless	2.4	2.2
Internet & Drive	1.9	0.7
Computer with Wireless & Internet	0.8	1.5
Cellular & Drive	0.6	0.0
Other	0.4	0.7
Sample Size	N = 1,248	N = 137

UPSTART Graduates

The Waterford Institute defined an UPSTART "graduate" as a participant who:

- Was actively engaged in the program for at least nine months; and
- Had at least 1,000 minutes of usage (16.67 hours) while in the program.

Children who enrolled in the spring of 2009 and only used the program for four months (e.g., through the summer of 2009 only) were not considered “graduates” by the Waterford Institute.

Of the 1,248 children documented as enrolled in UPSTART in the first year of the program, the Waterford Institute classified 741 as graduates of the program. This converts to a *graduation rate of 59%* ($741/1248 = .59$, or 59%).

While the Waterford Institute’s definition of an UPSTART graduate could be considered somewhat arbitrary, it does focus the program design on the theoretical desirability of enrolling preschool children for a full year and not just the summer. It also sets a desired level of minimum exposure to the curriculum, which is 16.7 hours or approximately 13.3 weeks at 75 minutes per week. The logic of the Waterford Institute’s definition of a graduate, however, is somewhat elusive.

The arbitrariness of the definition of what constitutes an UPSTART graduate is reinforced by its lack of an empirical grounding in being related to a valued literacy outcome like reading

proficiency. We found that UPSTART graduation status was not significantly correlated with reading proficiency at either the beginning ($r=.13$) or middle ($r=.10$) of kindergarten based on the test samples in the kindergarten analysis. However, we did find that increasing exposure to the curriculum is a significant predictor of reading proficiency.

In order for UPSTART “graduate” status to be more meaningful, an empirical cut point needs to be established that has some reliable bearing on being predictive of reading proficiency. That will require further research using a criterion-referenced test of reading proficiency. However, we do examine the relationship between level of instruction and reading proficiency in the final section on implementation below.

UPSTART Usage

UPSTART instruction was initiated in mid-April 2009 and continued through the first week in July 2010 for a total possible of 65 weeks of instruction in the “first year” of the program. The program design called for 15 minutes of instruction per day, 5 days a week for a total of 75 minutes of instruction per week. A student enrolled for 65 weeks in the first year would be expected to have engaged in at least 4,875 minutes of instruction, or 81.25 hours of instruction. The average level of usage for all students enrolled in the first year was 53 hours of instruction, which converts to 42 weeks of UPSTART implementation on the average. The students in the DIBELS Next test samples used the UPSTART curriculum for 68 hours of instruction on the average, or approximately 54 weeks of instruction over the course of the first program year.

Students considered to be UPSTART graduates by the Waterford Institute used the UPSTART curriculum for 76 hours of instruction on the average, or approximately 61 weeks over the course of the first program year. Based on these calculations, an UPSTART “graduate” would have to have been engaged for approximately 94% of the total possible instructional time available during the first year of the program (i.e., $61/65 = .94$, or 94%).

The hours of instruction observed for all students documented to be enrolled in the first year of UPSTART are summarized in Table 3 compared to “graduates” and the students in the kindergarten analysis sample. On the average, UPSTART students received 53 hours of instruction whereas those classified by the Waterford Institutes as graduates of the program received almost 76 hours of instruction on the average. The kindergarten analysis sample was in between at almost 68 hours of instruction.

Table 3
Hours of UPSTART Instruction

Group	N	Mean	SD	Range
All UPSTART	1,248	53.16	40.92	<1 – 313.57
UPSTART “Graduates”	741	75.95	37.31	16.86 - 313.57

Kindergarten Sample	137	67.73	44.04	<1 – 237.92
---------------------	-----	-------	-------	-------------

Inspection of the histograms in Figures 1-3 showing the distributions of hours of instruction for the three groups suggests that the kindergarten analysis sample is more representative of the first year UPSTART program population than the “graduates” group.

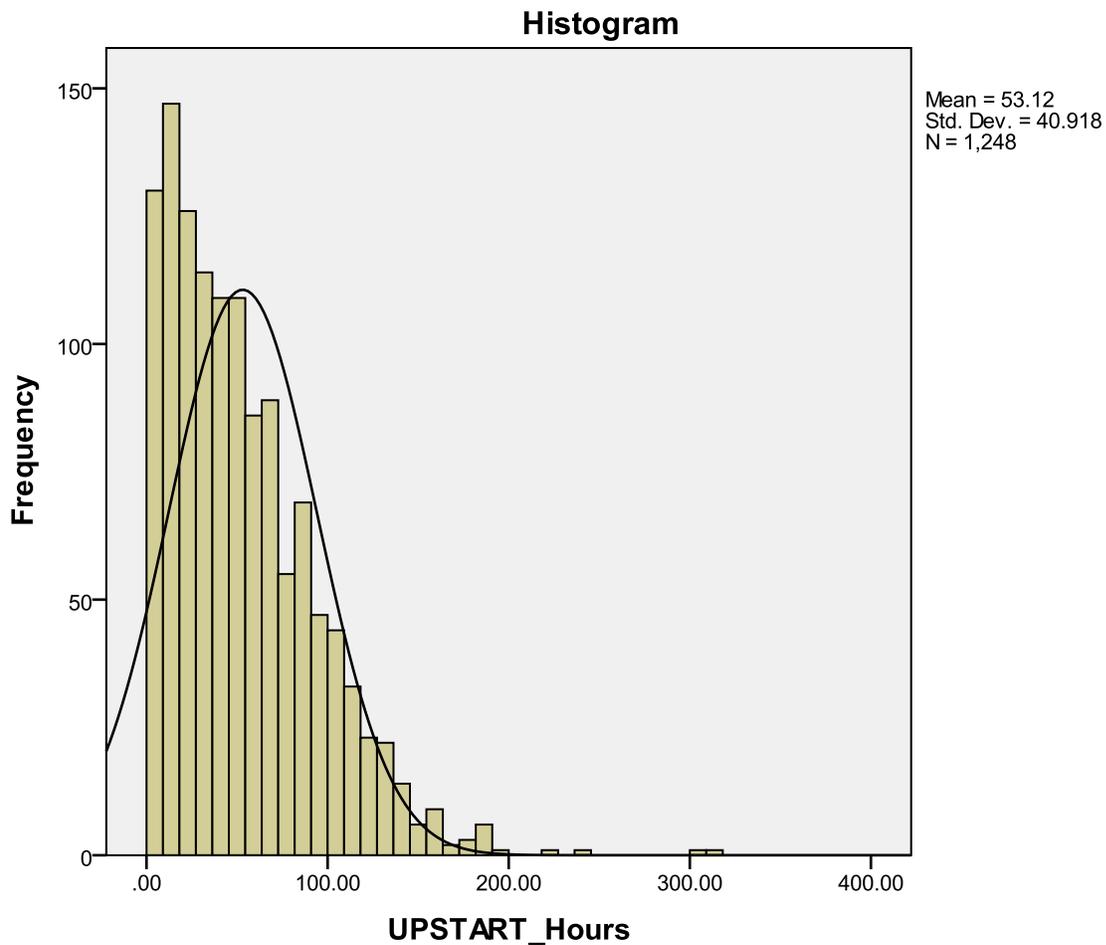


Figure 1. Hours of Instruction for All Students Enrolled in UPSTART in Year 1

Usage of the UPSTART curriculum for all students enrolled in the first year of the program (see Figure 1) is right-skewed with a mean of approximately 53 hours of instruction and a standard deviation of 41 hours. Because of the right-skewed nature of the distribution, the median – 45 hours of instruction -- is a more accurate representation of the average usage of the UPSTART curriculum. Approximately 1.5% of the enrollment completed less than one hour of instruction. At the other end of the distribution, approximately 1% of the enrollment completed 180 or more hours of instruction.

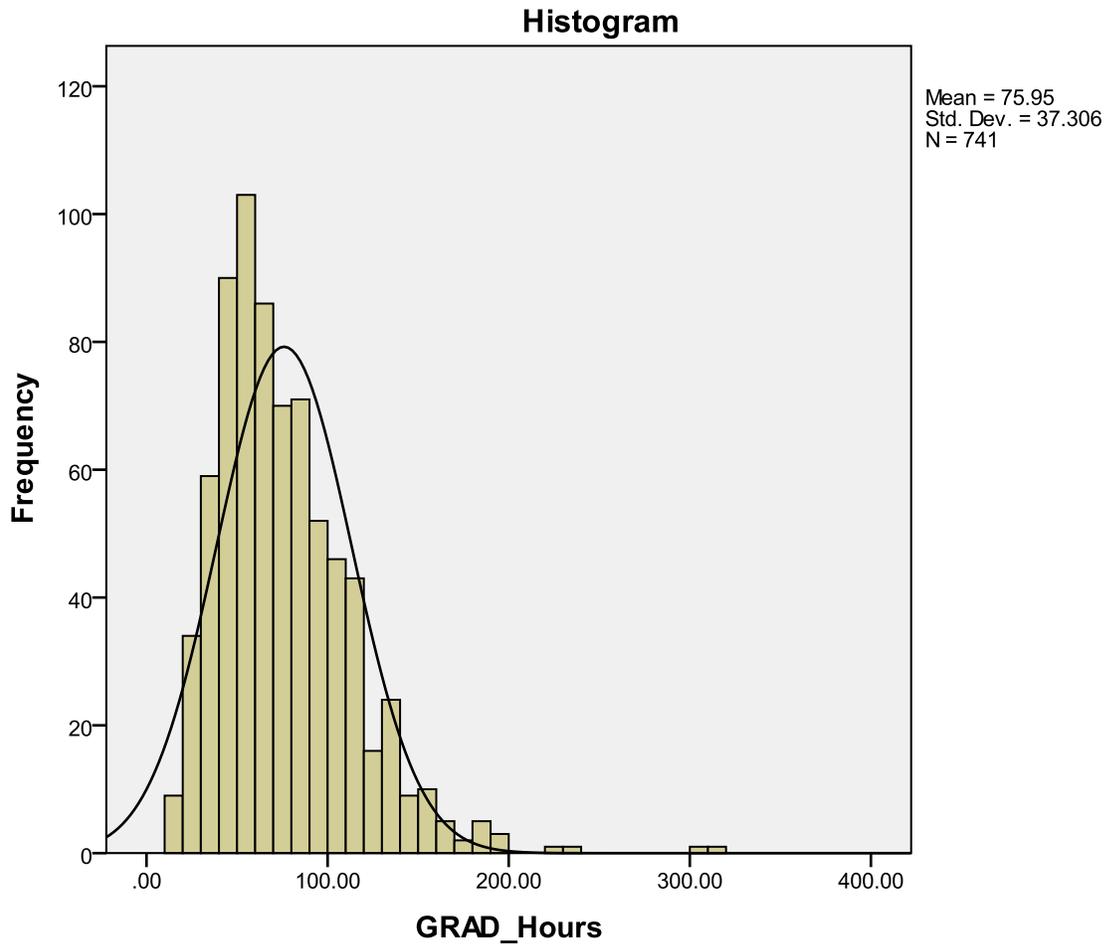


Figure 2. Hours of Instruction for UPSTART Graduates in Year 1

Lop off the bottom quintile from Figure 1 and you have Figure 2. The usage distribution for the “graduates” subgroup starts with 16.9 hours of instruction and runs to 313.6 hours of instruction, as with the full program group. The graduates’ distribution is right-skewed as well, which makes its median value of approximately 69 hours of instruction the more accurate representation of central tendency for this group. Because the bottom 20% of the graduates’ distribution of usage hours has been removed, its skewness is actually more pronounced than that of the total program enrollment (i.e., a skew statistic of 1.407 compared to 1.237).

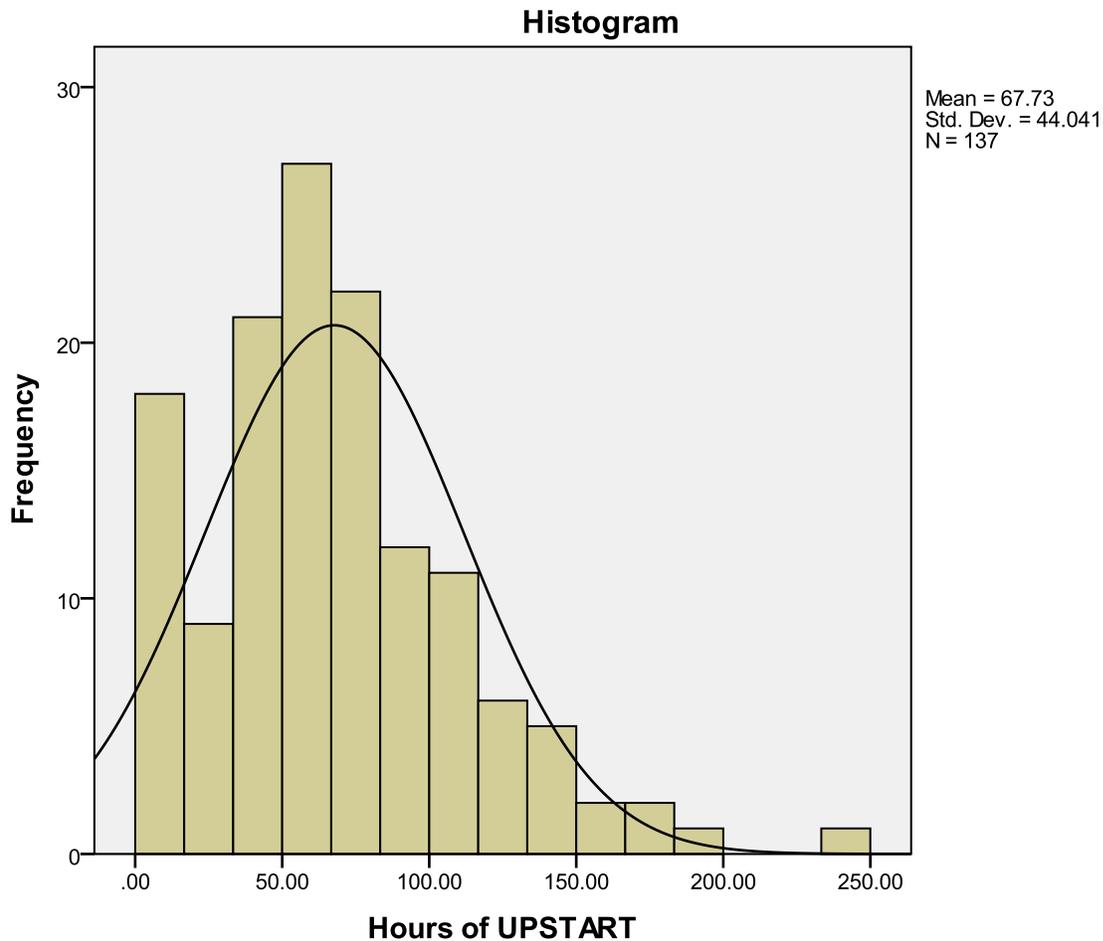


Figure 3. Hours of Instruction for Year 1 UPSTART Kindergarten Analysis Sample

UPSTART usage for the kindergarten analysis sample (see Figure 3) is right-skewed (skew statistic = .845) with a mean of approximately 68 hours of instruction and a standard deviation of 44 hours. The analysis sample’s median is 60 hours of instruction. At the low end, less than 4% of the analysis sample completed less than one hour of instruction. At the high end of the distribution, approximately 2% of the sample completed 180 or more hours of instruction.

How UPSTART Usage Relates to Reading Proficiency

As previously noted, we did establish in the kindergarten analysis that UPSTART curriculum usage is positively and significantly correlated with reading proficiency. This moderately strong relationship was observed at both the beginning ($r=.30$) and middle ($r=.29$) of kindergarten. We also found that increases in reading proficiency were also related to increasing levels of UPSTART curriculum usage. Table 4 shows that UPSTART usage is significantly related to

reading proficiency, statistically controlling for the effects of ELL and special education status (income level, race, and gender are not significant predictors in this model).²

Table 4
Tests of Between Subjects Effects: DN Composite Score - Beginning K-

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	21768.801 ^a	5	4353.760	8.629	.000	.257
Intercept	274324.319	1	274324.319	543.678	.000	.813
ELL	12687.081	1	12687.081	25.144	.000	.167
SPED	3185.005	1	3185.005	6.312	.013	.048
UsageGroup	8436.421	3	2812.140	5.573	.001	.118
Error	63071.366	125	504.571			
Total	358277.000	131				
Corrected Total	84840.168	130				

Adjusted R Squared = .23

In Table 5 below, each usage group is identified by its quartile value, 1 through 4. The covariance model compares the effects of each level of usage with the fourth quartile level of usage for the kindergarten analysis sample and displays the difference in DN Composite scores in the column labeled B – expressed as a regression coefficient. The parameter estimates for Usage Group indicate that participants in the first quartile of usage (under 40 hours in the kindergarten sample) score approximately 23 points lower on the DIBLES Next at the beginning of kindergarten than participants in the fourth quartile of usage (90 or more hours in the kindergarten sample). The parameter estimate for the second quartile of usage indicates that those who used the UPSTART curriculum between 40 and 60 hours scored almost 17 points lower on the average than fourth quartile UPSTART users (90 or more hours). Finally, the parameter estimate for the third quartile of usage indicates that those who used the UPSTART curriculum between 60 and 90 hours scored approximately 13 points lower on the average than fourth quartile UPSTART users (90 or more hours). These estimates clearly indicate a linear trend such that reading proficiency in kindergarten increases with increasing levels of UPSTART curriculum usage in preschool.

² The Partial Eta Square statistic for Usage Group in Table 4 indicates that increasing exposure to the UPSTART curriculum accounts for about 12% of the reading proficiency measured by the DN composite at the beginning of kindergarten. This indicates a moderately strong effect of UPSTART at the beginning of kindergarten.

Table 5
Parameter Estimates: DN Composite Score - Beginning K-

Parameter	B	Std. Error	t	Sig.	Partial Eta Squared
Intercept	64.053	4.244	15.094	.000	.646
ELL	-27.772	5.538	-5.014	.000	.167
SPED	-26.229	10.440	-2.512	.013	.048
[UsageGroup=1.00]	-22.852	5.808	-3.935	.000	.110
[UsageGroup=2.00]	-16.564	5.559	-2.980	.003	.066
[UsageGroup=3.00]	-12.681	5.596	-2.266	.025	.039
[UsageGroup=4.00]	0 ^b

a. Parameter set to zero because it is redundant

This positive linear trend in reading proficiency associated with usage quartile is evident in Figure 4 below

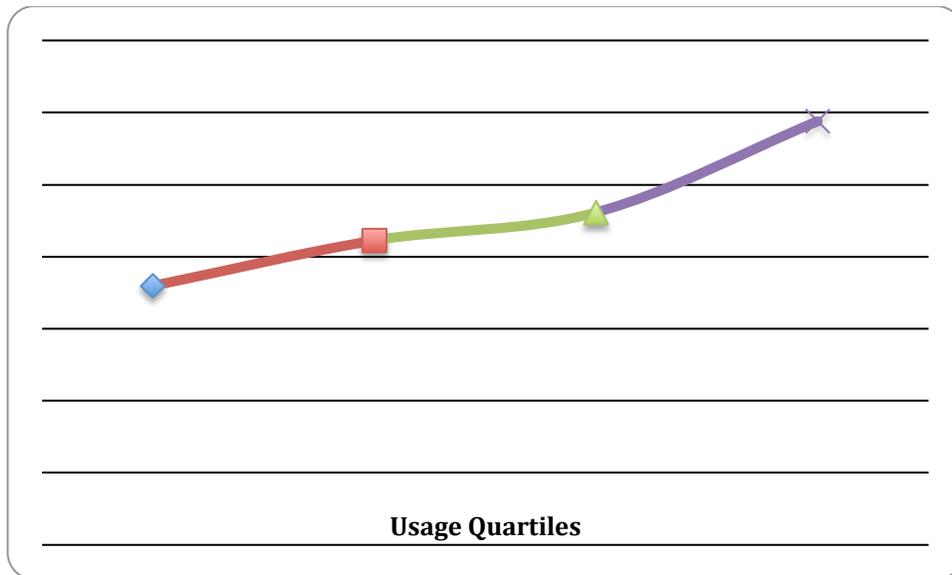


Figure 4. Mean DN Composite B Scores by UPSTART Usage Quartile

The results are much the same for the UPSTART children in middle kindergarten – see Tables 6 and 7 -- and show that increases in reading proficiency correspond with increasing usage of the UPSTART curriculum. In middle kindergarten, UPSTART usage accounts for about 19% of the children’s reading proficiency as measured by the DN Composite at that level.

Table 6
Tests of Between Subjects Effects: DN Composite Score - Middle K-

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	95390.609 ^a	5	19078.122	7.346	.000	.223
Intercept	3081911.705	1	3081911.705	1186.759	.000	.903
ELL	57905.486	1	57905.486	22.298	.000	.148
SPED	15451.424	1	15451.424	5.950	.016	.044
UsageGroup	31826.081	3	10608.694	4.085	.008	.087
Error	332404.943	128	2596.914			
Total	3759082.000	134				
Corrected Total	427795.552	133				

Adjusted R Squared = .193

In Table 7, the parameter estimates for Usage Group indicate that participants in the first quartile of usage (under 40 hours) score almost 40 points lower on the DIBLES Next in the middle of kindergarten compared to participants in the fourth quartile of usage (90 or more hours). The parameter estimate for the second quartile of usage indicates that those who used the UPSTART curriculum between 40 and 60 hours scored almost 38 points lower on the average than fourth quartile UPSTART users (90 or more hours). The parameter estimate for the third quartile of usage (60-90 hours) indicates that these participants scored about 24 points lower on the average than fourth quartile UPSTART users.

Table 7
Parameter Estimates: Composite Score - Middle K- DIBELS Next

Parameter	B	Std. Error	t	Sig.	Partial Eta Squared
Intercept	194.622	9.454	20.585	.000	.768
ELL	-59.227	12.543	-4.722	.000	.148
SPED	-53.328	21.862	-2.439	.016	.044
[UsageGroup=1.00]	-39.811	12.929	-3.079	.003	.069
[UsageGroup=2.00]	-37.523	12.501	-3.002	.003	.066
[UsageGroup=3.00]	-24.112	12.583	-1.916	.058	.028
[UsageGroup=4.00]	0 ^b

Again, these estimates suggest a linear trend, with reading proficiency increasing with increasing levels of UPSTART curriculum usage. The data for this trend are presented in Figure 5 below. The shallow slope in Figure 5 is consistent with the with partial eta squared statistic for Usage Group in Table 6 which suggests a smaller effect for UPSTART at the middle kindergarten level.

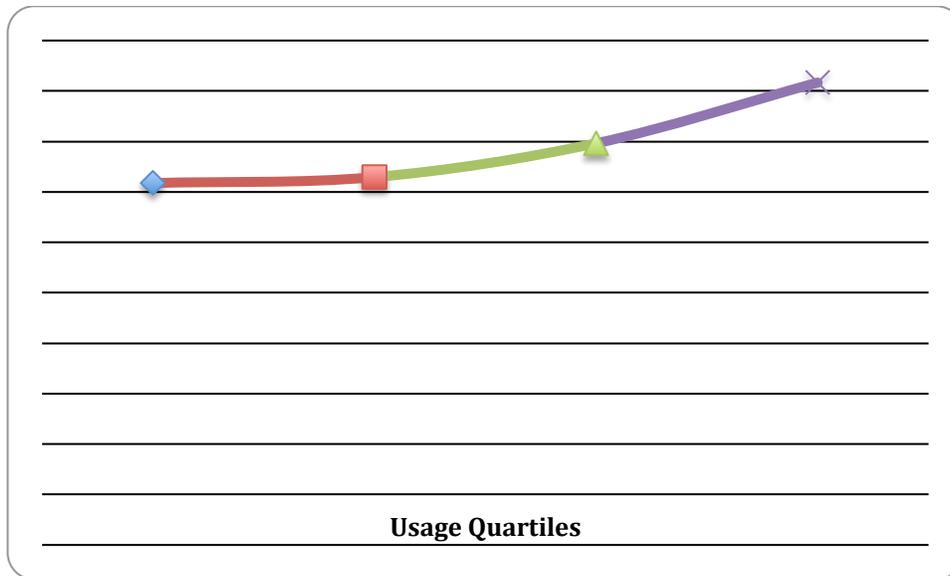


Figure 4. Mean DN Composite M Scores by UPSTART Usage Quartile

UPSTART Outcomes

A weighted least squares regression approach was used to estimate reading proficiency at the beginning of kindergarten as measured by the DIBLES Next (DN) beginning kindergarten composite. Low income status was used as a weight variable to correct for non-constant variance in estimating the regression coefficients. ELL status, special education status and gender were used as control variables in the final model to estimate reading proficiency for children who had participated in UPSTART the previous school year compared to children from the general kindergarten population who had not participated in UPSTART.

Does UPSTART improve reading readiness?

The WLS ANOVA summary table shows that the four-factor regression model is statistically significant. The results of interest are shown in Table 9.

Table 8
WLS ANOVA Summary Table for DN Composite – Beginning of Kindergarten

	Sum of Squares	Df	Mean Square	F	Sig.
Regression	19573.703	4	4893.426	11.422	.000
Residual	54837.500	128	428.418		
Total	74411.203	132			

Table 9 shows that ELL status, special education status and being male all tend to depress beginning kindergarten reading proficiency (DN Composite) scores by approximately 8 to 23 points on the average. Special education students score 23 points lower on the DIBLES Next composite at the beginning of kindergarten compared to regular education students on the average. ELL students score 15 points lower than English proficient students on the average at the beginning of kindergarten. And boys score almost 8 points lower than girls on the DN Composite at the beginning of kindergarten. Statistically controlling for these three factors and weighting the data on the basis of income status, we find that children who had participated in UPSTART during preschool scored almost 18 points higher in reading proficiency as measured by the DN Composite compared to beginning kindergarten children who did not participate in UPSTART prior to enrolling in public school. *Based on these results, the evidence is that UPSTART appears to significantly improve reading readiness for beginning kindergarten students.*

Table 9
WLS Regression Coefficients for DN Composite – Beginning of Kindergarten

	Unstandardized Coefficients		Standardized Coefficients		t	Sig.
	B	Std. Error	Beta	Std. Error		
(Constant)	30.122	3.074			9.797	.000
ELL	-15.278	3.909	-.304	.078	-3.909	.000
SPED	-23.303	6.668	-.271	.078	-3.495	.001
MALE	-7.664	3.625	-.162	.077	-2.114	.036
GROUP	17.772	3.946	.345	.077	4.503	.000

Adjusted R Square = .24

Considering the amount of variance accounted for by the beginning kindergarten model (an adjusted R Square of .24) and judging from the Beta statistic (the standardized regression coefficient value of .35), the UPSTART effect size would be considered to represent a *moderately strong impact*.³ Thus, the available evidence from the first year of the program is that UPSTART does appear to have a moderately strong impact on helping young children develop beginning reading skills to a significantly greater extent than they would have without participating in the program.

Does UPSTART sustain improvements in reading readiness?

The DN Composite for middle kindergarten covers more reading skills than the beginning reading composite. At the middle kindergarten level, phonemic awareness continues to be measured on the DIBLES Next through the First Sound Fluency and Letter Naming Fluency subtests. However, the Phoneme Segmentation Fluency subtest is added along with measures of

³ See Chapter 9 in Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

the alphabetic principle using the Nonsense Word Fluency Subtest and whole word reading is scored on this subtest as well.

The OLS regression model summary for middle kindergarten is displayed in Table 10 and shows that the four measured covariates (Model 1) account for 15% of the variance in DIBLES Next reading proficiency in the middle of kindergarten. The model summary also shows that adding the GROUP variable (Model 2) makes a statistically significant contribution in accounting for variation in DIBLES Next test scores, meaning that UPSTART participation continued to make a difference in kindergarten reading proficiency through the middle of kindergarten. The overall model accounts for 17 percent of middle kindergarten reading proficiency as measured by the DIBLES Next and indicates that UPSTART students were able to sustain their gains in reading proficiency through the middle of kindergarten.

Table 10
OLS Regression Model Summary
DN Composite – Middle of Kindergarten

Model	R	R Square	Adj R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.402 ^a	.162	.153	54.544	.162	18.126	4	376	.000
2	.430 ^b	.185	.174	53.866	.023	10.524	1	375	.001

Table 11 shows ELL status, low income status, and special education status all tend to depress middle kindergarten reading proficiency by approximately 14 to 46 points. Ethnicity also has an effect on reading proficiency in the middle kindergarten model. Specifically, Caucasian students tend to have higher reading proficiency scores – almost 17 points higher on the average -- than non-White students by the middle of kindergarten.

Statistically controlling for these four factors, we find that middle kindergarten children who had participated in the UPSTART preschool program scored approximately 19 points higher in reading proficiency as measured by the DN Composite for middle of kindergarten compared to middle kindergarten children who did not participate in UPSTART prior to enrolling in public school. Based on this evidence, we can say that the *UPSTART participants appear to have sustained their gains in reading proficiency through the middle of kindergarten.*

Table 11
 OLS Regression Coefficients for DN Composite – Middle of Kindergarten

	Unstandardized Coefficients		Standardized Coefficients		t	Sig.
	B	Std. Error	Beta	Partial r		
(Constant)	136.378	6.991			19.508	.000
ELL	-27.881	8.438	-.184	-.168	-3.304	.001
LOW INCOME	-13.724	6.105	-.113	-.115	-2.248	.025
SPED	-46.108	11.741	-.186	-.199	-3.927	.000
WHITE	16.592	6.953	.133	.122	2.386	.018
GROUP	19.089	5.884	.154	.165	3.244	.001

Adjusted R Square = .17

Considering the amount of variance accounted for by the middle kindergarten model (an adjusted R Square of .17) and judging from the Beta statistic for the Group variable (the standardized regression coefficient value of .15), UPSTART’s *sustained effect size* would continue to be considered a *moderately strong impact*. Thus, the available evidence indicates that the UPSTART effect has been sustained through the middle of kindergarten.

In our follow-up, we will see if the UPSTART effect continues to be sustained when we check the children’s reading proficiency skills again in first grade.

Summary and Conclusions

This final section of the first year evaluation report summarizes:

- The data that were collected and analyzed;
- The analysis methods employed;
- Findings regarding UPSTART implementation; and
- Findings on UPSTART's impact on reading proficiency as measured in kindergarten.

Test Data Collected and Analyzed

Test data were obtained from 10 Utah public school districts for 258 children who had participated in UPSTART during its first year of operation during the 2009-2010 school year. The test data came from three assessments: the DIBELS Next, the DIBELS, and the DRA. Control group data were provided by the school districts for all nonparticipating students tested. DRA test data were supplied for 62 participants from one school district; DIBELS data were supplied by two districts for 62 participants; and DIBELS Next (DN) data were supplied by seven districts for 137 participants.

We needed a sample size of at least 90 UPSTART participants in order to use multiple regression analysis to test a model with six factors, which was our intent. The model to be estimated involved five covariates (gender, race, ELL status, low income status, and special education status) plus the independent variable of interest: the UPSTART treatment group compared to nonparticipant controls. Test data were sufficient only for an analysis using the DIBELS Next data supplied by seven school districts for 137 former UPSTART students.

The response variables selected for the kindergarten analysis included the DN Composites for the beginning and middle of kindergarten. Thus, the final analysis of UPSTART's impact in kindergarten was based on DN Composite scores for 137 children enrolled in seven Utah school districts.

The Analysis

A weighted least squares (WLS) regression approach was used to estimate reading proficiency at the beginning of kindergarten as measured by the DIBELS Next (DN) beginning kindergarten composite. Low income status was used as a weight variable to correct for non-constant variance in estimating the regression coefficients. ELL status, special education status, and gender were used as control variables in the final WLS model to estimate reading proficiency for children who had participated in UPSTART the previous school year compared to children from the general kindergarten population who had not participated in UPSTART.

An ordinary least squares (OLS) regression approach was used to estimate reading proficiency in middle kindergarten as measured by the DIBLES Next (DN) middle kindergarten composite. The final OLS model employed a hierarchical block design in which a control set of four covariates (dummy variables for race, ELL status, low income status, and special education status) were entered first followed by the set of interest (treatment vs. control).

The dependent variables modeled were:

- The DN beginning kindergarten composite, which included the First Sound Fluency subtest and the Letter Naming Fluency subtest; and
- The DN middle kindergarten composite, which in addition to higher levels of the first two subtests also included the Phoneme Segmentation and Nonsense Word Fluency subtests.

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 kindergarten analysis sample. The final ANCOVA models statistically controlled for ELL status and special education status in estimating the effect of UPSTART usage on reading proficiency at the beginning and middle of kindergarten using the DN composites. In the ANCOVA analyses, the effects of usage at quartiles one through three were compared with usage at the fourth quartile.

Findings: UPSTART Implementation

The Waterford Institute provided documentation for a first-year UPSTART enrollment of 1,248 children. A majority (56%) of the 1,248 preschool children that enrolled in the first year of UPSTART were from low income families, according to data provided by the Waterford Institute. Slightly more boys (52%) were enrolled than girls (48%). In terms of ethnicity, the vast majority (81%) of the enrollment was Caucasian, 13% were Hispanic, 2% were of Asian descent, 1% were Black, and 1% were Native American. Ethnicity was unknown for 1% of the first year enrollment.

Most of the first year participants (70%) received a computer drive with the UPSTART curriculum loaded on it. Approximately 11% of the first year participants received a computer loan and free Internet access to help them access the UPSTART curriculum. Another 7% of the first year participants were loaned a personal computer to use at home while participating in UPSTART. The remaining 10 to 12% of the first 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 kindergarten test sample had a mean of 68 hours of UPSTART curriculum usage over the first year of the project. This compares with an average of 76 hours of instruction for program “graduates” and an average of 53 hours of instruction for all students enrolled in UPSTART in the first year.
- Length of participation in the UPSTART curriculum was significantly and positively correlated with reading proficiency at the beginning ($r=.30$) and middle ($r=.29$) of kindergarten.
- Reading proficiency improved with increasing levels of UPSTART curriculum usage for both beginning and middle kindergarten children.
- The UPSTART graduation rate in the first year of the program was 59%. UPSTART graduation status was not significantly correlated with reading proficiency at either the beginning ($r=.13$) or middle ($r=.10$) of kindergarten.

Findings: UPSTART Impact on Reading Proficiency in Kindergarten

- UPSTART participation had a moderately strong impact on improving the reading proficiency of UPSTART participants compared to nonparticipants at the beginning of kindergarten, as measured by the DIBELs Next beginning kindergarten composite. This analysis controlled for the significant influence ELL status, special education status, and gender on reading proficiency outcomes at the beginning of kindergarten.
- UPSTART participants maintained their achievement gains through the middle of kindergarten as measured by the DIBELs Next middle kindergarten composite. This analysis controlled for the significant influence of ELL status, special education status, low income status and race on reading proficiency outcomes in middle kindergarten.

Appendix A

Statistical Power Analysis

Two power analyses were conducted in planning the sample sizes needed for the first year kindergarten analysis of UPSTART. The power analyses differed in terms of assumptions about the minimum effect size sizes to detect for UPSTART in terms of the R-Squared increment that would be detected over and above that of the control variable set. Both analyses assumed a hierarchical regression analysis in which a set of 5 covariates would be entered in a control block that would yield an R-squared of .20, followed by the UPSTART treatment vs. control group variable of interest. Alpha was set at .05 and power was set at .80.

In Model 1, the minimum UPSTART effect was specified as an R-squared increment of .02. This was the absolute smallest effect that could be meaningfully detected. The power analysis determined that a sample size of 310 cases per group would be required to detect such a small treatment effect.

In Model 2, the minimum UPSTART effect was specified as an R-squared increment of .07. This was also a small effect but was considered more meaningful as a realistic estimate of the UPSTART effect. The power analysis determined that a sample size of 90 cases per group would be required to detect a treatment effect of this size.

In planning the samples, it was determined that the analysis required a treatment group of at least 90 cases. However, it was recognized that 300 cases per group would be more optimal. The 137 treatment group cases met the minimum sample size criterion. Since we had a pool of over 9,000 control cases, we decided to optimize the control group sample by using a 2:1 ratio of control cases to treatment group cases and consequently selected a random sample of 274 controls stratified by school district.

Appendix B Samples

*Table B.1
Treatment-Control Group Sample Sizes by School District*

School Districts	Sample Sizes	
	Treatment Group	Control Group
Box Elder	16	32
Granite	58	116
Iron	10	20
Ogden	6	12
Salt Lake	27	54
Tooele	10	20
Weber	10	20
Total	137	274

*Table B.2
Treatment-Control Sample Demographics*

Demographic	Group	N	Mean	SEM	t	Significance
% Male	Treatment	137	59	4	2.18	**
	Control	274	48	3		
% White	Treatment	137	77	4	3.20	**
	Control	274	62	3		
% ELL	Treatment	137	15	3	-1.27	NS
	Control	274	20	2		
% Low Income	Treatment	137	30	4	-2.87	**
	Control	274	44	3		
% SPED	Treatment	137	5	2	-0.44	NS
	Control	274	6	2		

**p<.01

The reader should note that the two critical covariates in most of the first year impact analyses are the ELL and SPED variables and that the treatment and control groups are both statistically equivalent on these two covariates.