

2025

UTAH UPSTART PROGRAM EVALUATION



YEAR 16 RESULTS

Prepared by **EVALUATION AND TRAINING INSTITUTE**

Submitted to



**Utah Governor's Office
of Economic Opportunity**



EVALUATION AND TRAINING INSTITUTE *A non-profit center for research & evaluation*
12240 Venice Blvd, Suite 27, Los Angeles, California 90066 || 310-473-8367



12240 Venice Blvd. Suite 27
Los Angeles, CA 90066
www.eticonsulting.org

Founded in 1974, the Evaluation & Training Institute (ETI) is a non-profit consulting firm, headquartered in Los Angeles, dedicated to working with schools, post-secondary institutions, public agencies, private foundations, community-based organizations, and professional organizations. We specialize in third-party program evaluations covering many fields, including education, literacy, STEM, social services, health, and prevention. Many of our evaluations have been instrumental in the development of public policy as well as state and federal legislation. Throughout, our focus is on helping clients improve their programs as well as maintain accountability to funders and oversight committees. For more information on the Evaluation and Training Institute, please visit our website or contact ETI:

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

Allison Coordt, Ph.D., Senior Project Manager
acoordt@eticonsulting.org

Emma Rees, MPP, Junior Project Manager
erees@eticonsulting.org

Citation: Rees, E., Coordt, A. & Hobbs, J. (2025). *Utah UPSTART Program Evaluation: Year 16 Program Results*. Evaluation and Training Institute

ACKNOWLEDGEMENTS

The Evaluation and Training Institute (ETI) thanks Anju Thapaliya (Program Manager, Strategic Initiatives) from the Governor's Office of Economic Opportunity for the collaboration on this evaluation project. We also thank Katy Challis (Director of Privacy), Aaron Brough (Director of Data and Statistics), and Jonathan Collins (Research Consultant) from the Utah State Board of Education for their efforts in sharing state PEEP and Acadience data with us in order to conduct our analysis.

Additionally, we extend our thanks to Jim Lund, Haya Shamir, Ph.D., Melinda Dalton, and Erik Yoder from Waterford for continuing to provide ETI with the necessary UPSTART data used to complete the evaluation each program year.

Table of Contents

<u>EXECUTIVE SUMMARY</u>	<u>1</u>
<u>INTRODUCTION.....</u>	<u>3</u>
<i>EVALUATION OBJECTIVES</i>	<i>4</i>
<i>UPSTART PROGRAM DESCRIPTION</i>	<i>5</i>
<u>PRE-KINDERGARTEN PROGRAM IMPLEMENTATION STUDY</u>	<u>6</u>
<i>RESEARCH QUESTIONS</i>	<i>7</i>
<i>PROCEDURE</i>	<i>7</i>
<i>UPSTART ENROLLMENT</i>	<i>7</i>
<i>UPSTART USAGE</i>	<i>9</i>
<i>UPSTART GRADUATION RATE</i>	<i>11</i>
<u>PRE-KINDERGARTEN ACADEMIC ACHIEVEMENT STUDY</u>	<u>13</u>
<i>RESEARCH DESIGN</i>	<i>13</i>
<i>RESEARCH QUESTIONS</i>	<i>14</i>
<i>PROCEDURE</i>	<i>14</i>
<i>CONTROL STUDENT MATCHING PROCESS</i>	<i>15</i>
<i>LINEAR REGRESSION MODEL</i>	<i>15</i>
<i>OUTCOME MEASURES.....</i>	<i>16</i>
<i>INTENT TO TREAT ANALYSIS RESULTS</i>	<i>18</i>
<i>PROGRAM USE ANALYSIS RESULTS</i>	<i>19</i>
<i>LOW INCOME ANALYSIS RESULTS</i>	<i>22</i>
<u>KINDERGARTEN ACADEMIC ACHIEVEMENT STUDY.....</u>	<u>24</u>
<i>RESEARCH DESIGN.....</i>	<i>25</i>
<i>RESEARCH QUESTIONS</i>	<i>25</i>
<i>KINDERGARTEN DATA COLLECTION</i>	<i>26</i>
<i>KINDERGARTEN MEASURES.....</i>	<i>26</i>
<i>KINDERGARTEN INTENT TO TREAT ANALYSIS RESULTS.....</i>	<i>27</i>
<i>KINDERGARTEN LOW INCOME ANALYSIS RESULTS</i>	<i>28</i>
<i>KINDERGARTEN LOW INCOME GRADUATION ANALYSIS</i>	<i>29</i>
<u>RECOMMENDATIONS FOR PROGRAM IMPROVEMENT.....</u>	<u>30</u>

STUDY LIMITATIONS AND RESEARCH RECOMMENDATIONS.....	31
REFERENCES.....	34
APPENDIX A. PEEP PERFORMANCE LEVELS.....	35
APPENDIX B. ACADIENCE READING AND MATH MEASURES	37
APPENDIX C. ACADIENCE READING AND MATH BEGINNING OF YEAR BENCHMARK LEVELS.....	38
APPENDIX D. PEEP ENTRY AND EXIT DISTRIBUTIONS	40

List of Figures

Figure 1. UPSTART Program Enrollment Over Time.....	8
Figure 2. Distribution of Hours of Reading Instruction for C16 Families	10
Figure 3. Average Weekly Minutes of Reading and Math Use	11
Figure 4. End of Year Benchmark Levels in Reading and Math among Program and Non-Program Students ...	27
Figure 5. Benchmark Levels in Reading and Math among Low-Income Program and Non-Program Students at End of Year.....	29
Figure 6. Benchmark Levels in Reading and Math among Low-Income Graduates, Non-Graduates, and Non-Program Students at Exit.....	30

List of Tables

Table 1. UPSTART Reading and Math Curriculum	6
Table 2. Demographic Characteristics of C16 Population	8
Table 3. C16 Hours of UPSTART Instruction	10
Table 4. Waterford Graduation Criteria.....	12
Table 5. PEEP Exit Profile	17
Table 6. Literacy Regression Summary.....	18
Table 7. Numeracy Regression Summary	19
Table 8. Literacy Quartiles Predicted Means	20
Table 9. Literacy Predicted Means for High Reading Usage Students	21
Table 10. Numeracy Predicted Means for High Math Usage Students.....	22
Table 11. Literacy Predicted Means for Low-Income Students	23
Table 12. Numeracy Predicted Means for Low-Income Students.....	24
Table 13. Acadience Learning Subscales	26

EXECUTIVE SUMMARY

The Evaluation and Training Institute (ETI) developed this report for the Utah Governor's Office of Economic Opportunity (GOEO) to evaluate the performance of the UPSTART program in its sixteenth year. This evaluation focused on students from Cohort 16, who were enrolled during the 2024-2025 program year. The report includes an analysis of pre-k program implementation data (the "Pre-K Implementation Study") and academic achievement testing data (the "Pre-K Academic Achievement Study"), as well as an analysis of kindergarten academic achievement testing data for Cohort 15 (the "Kindergarten Academic Achievement Study;" students who were enrolled in Kindergarten during this report cycle).

Key Pre-K Program Implementation Study Highlights:

- **Graduation Rates Declined:** During the 2024-2025 program year, 63% of UPSTART students (N= 13,220) successfully graduated from the program, down from the previous year's graduation rate of 70%.
- **Consistent Engagement Drives Success:** All participants, regardless of graduation status, engaged with the UPSTART reading program for an average of 29 hours throughout the year. Graduates, however, logged around 39 hours on average. Consistent engagement was crucial for achieving the program's time-centered objectives.

Pre-K Academic Achievement Notable Findings:

- **UPSTART strengthens early literacy:** Pre-k participants scored significantly higher than non-participants on measures of literacy at the end of the program year.
- **Dosage matters:** Students who used UPSTART most intensively (highest use quartile) demonstrated the largest literacy gains, with a moderate effect size compared to controls.



- **Numeracy results were less conclusive:** Overall, UPSTART participants did not show significant advantages over their matched peers in numeracy test scores. We believe that ceiling effects in the PEEP assessment likely limited the ability to detect differences between program and non-program students at the end of their pre-kindergarten year.
- **High-usage benefits in numeracy:** Despite overall nonsignificant results, students with the highest program usage achieved modest but statistically significant numeracy gains.
- **Impacts on low-income students are not conclusive:** Results for low-income students showed positive trends, but findings were not statistically significant due to small sample size. We recommend moving to low-income study that combines cohorts of students to build larger sample sizes.

Kindergarten Academic Achievement Study Highlights:

- **Higher Kindergarten Achievement:** UPSTART participants demonstrated stronger kindergarten reading and math outcomes compared to their peers who did not participate in the UPSTART program during preschool.
- **Stronger Outcomes for Low Income Students:** A higher percentage of low-income UPSTART students reached or surpassed benchmark by the end of the year, compared to those who did not use the software during their pre-k year.

Program Recommendations

The UPSTART program continues to boost student literacy and kindergarten achievement. Program use is strongly correlated with a positive impact on literacy test scores and students who graduate consistently show higher academic achievement outcomes, particularly in delayed achievement measures at the end of kindergarten. We recommend that Waterford strengthen its strategies to improve consistent student engagement and graduation rates, which have declined steadily in recent years.



INTRODUCTION

The Evaluation and Training Institute (ETI), a non-profit research and consulting firm, spearheaded a multi-year evaluation of the UPSTART program in Utah. This technical report presents detailed findings from the program's sixteenth year, covering the 2024-2025 implementation period. We have also developed a non-technical companion report, “UPSTART Evaluation Highlights,” which presents our key insights in a more accessible and concise format. The primary aim of the evaluation was to measure the impact the UPSTART program has on academic achievement, aiding the state and stakeholders in understanding the program’s impact.

This report marks an important transition in the UPSTART evaluation research methods: we employed a more rigorous research design than in recent years, comparing the early literacy and numeracy skills of UPSTART participants with those of non-participating peers. The quasi-experimental research design was based on analyzing student test scores using Utah’s current early education assessments. Several years ago, the Utah State Board of Education introduced the Preschool Entry and Exit Profile (PEEP) for preschool and the Kindergarten Entry and Exit Profile (KEEP) for kindergarten. Since then, the KEEP has been phased out and replaced by Acadience Learning as the statewide kindergarten assessment, while the PEEP remains the standard measure for preschool. These shifts in assessment tools have influenced how analyses are conducted across evaluation cycles, shaping the availability of comparison groups, statewide testing coverage, and access to control data.

The Year 16 evaluation reflects these changes, marking the first use of the PEEP within a treatment–control design for pre-k and Acadience to measure subsequent kindergarten outcomes. Accordingly, this report focuses on two groups of children: Year 16 pre-k participants (Cohort



16) and Year 15 kindergarten participants (Cohort 15), who completed kindergarten in 2025 after participating in UPSTART the prior year.

Report Roadmap

This report begins with an overview of the evaluation objectives and a description of the UPSTART program. Next, we present findings from three related studies: (1) our *pre-k implementation study* examining students' participation in UPSTART, (2) our *pre-k academic achievement study*, examining literacy and numeracy achievement outcomes, and (3) our *kindergarten academic achievement study*, examining the achievement outcomes for students who participated in UPSTART during pre-k and completed kindergarten in 2025. Finally, the report concludes with a discussion of key findings from the 2024–2025 program year and considerations to guide future evaluation efforts.

Evaluation Objectives

UPSTART's curriculum places primary emphasis on developing early literacy skills while also incorporating numeracy to strengthen children's overall school readiness skills. Since school readiness requires progress in both domains, the current evaluation was designed to capture the program impact on literacy and numeracy, while also studying measures of program implementation. For the 2024–2025 program year, the evaluation pursued the following objectives:

- *Pre-K Implementation Study:*
 - Document the extent to which Cohort 16 pre-kindergarteners engaged with the online curriculum as intended.
- *Pre-K Academic Achievement Study:*
 - Assess the impact of the program on early literacy and numeracy skills by comparing outcomes for Cohort 16 UPSTART participants (treatment group) and



a group of matched students who did not use the program (control group) at the end of the pre-k year.

- Explore the pre-kindergarten achievement outcomes for students from low-income families.
- *Kindergarten Academic Achievement Study:*
 - Evaluate kindergarten reading and math outcomes by comparing former Cohort 15 UPSTART participants with a control group at the end of kindergarten.
 - Analyze the kindergarten achievement outcomes specifically for students from low-income families.

UPSTART Program Description

Waterford recognizes that nearly half of the nation’s four-year-olds face socioeconomic barriers that hinder their access to early education. In response, the Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART) program was developed—a home-based computer readiness initiative designed to equip young children for school entry and long-term academic success. This program provides Utah families with pre-k children, including those who are underserved or from lower-income backgrounds, with essential resources such as technology, internet access, and support. Families engage with an adaptive curriculum that offers online instruction in early literacy and math, effectively preparing their children for kindergarten. In the 2024-2025 school year, a total of 13,220 pre-k children participated in the UPSTART program¹, marking its sixteenth year of operation.

The UPSTART software used adaptive lessons, digital books, animated songs, and activities to deliver individualized early literacy and numeracy content. Once children met the

¹ The program evaluation numbers are based on the in-home implementation and children who have at least 1 minute of UPSTART usage.

required literacy usage, they were able to additionally engage in the math/science curriculum. The skills taught by the UPSTART program are outlined in **Table 1**.

Table 1. UPSTART Reading and Math Curriculum

Early Reading Curriculum	Early Math and Science Curriculum
Phonological Awareness	Numbers and Operations
Phonics	Measurement and Data
Comprehension and Vocabulary	Geometry
Language Concepts	Science Concepts

Children were encouraged to use the UPSTART reading program for a minimum of 1500 minutes across the program year (or 15 minutes a day, 5 days a week). No minimum use recommendation was suggested for the math curriculum. All enrolled families were provided with parental resources and technical support from Waterford customer service representatives.

PRE-KINDERGARTEN PROGRAM IMPLEMENTATION STUDY

This section presents the Cohort 16 implementation study, outlining the guiding research questions, data collection procedures, and detailed findings. This part of the evaluation centers on understanding how UPSTART was implemented during the preschool year.

Summary of Key Findings

Students demonstrated strong engagement with the UPSTART reading software, averaging about 29 hours of instruction throughout the year, down slightly from last year but still above recommended levels. Overall, about 63% of students were classified as graduates of the reading software, with graduates logging around 39 hours on average. This pattern underscores that consistent engagement was key to meeting the program’s graduation requirements.

Research Questions

The pre-k implementation study addressed the following research questions:

Research Question 1.1: To what extent did children use the UPSTART program as defined by Waterford (measured in minutes of instruction per week/total minutes for duration of program)?

Research Question 1.2: What proportion of the participants successfully met the requirement for program completion (i.e., “graduated” as defined by Waterford)?

Procedure

ETI analyzed program usage data provided by Waterford for all enrolled children, to examine how the UPSTART program was implemented during the 2024–2025 program year. These records included detailed usage variables such as average weekly minutes, average number of days used per week, average session length, and cumulative time spent in the program. In addition, parents supplied demographic information about their child and household during the registration process. Together, these datasets were analyzed to describe program implementation and identify patterns of student program use across the Year 16 participants.

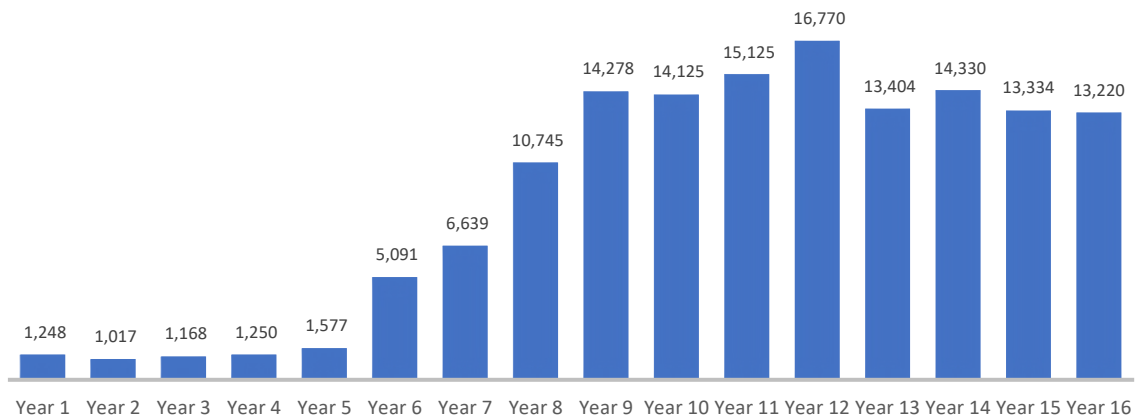
The following section outlines our implementation findings from Year 16 based on the performance of all children statewide who participated in the program during the 2024-2025 year.

UPSTART Enrollment

During the 2024-2025 program year, 13,220 pre-k students used the program as part of C16, similar to the previous year’s cohort (N = 13,334). **Figure 1** illustrates the growth of the program’s reach over time from (Year 1, N=1,248) to the most recent Year 16 (N=13,220).



Figure 1. UPSTART Program Enrollment Over Time



The demographic makeup of C16 was similar to previous program years, with most of the cohort classified as Caucasian/White (72%) with 11% from Latino/a origin (**Table 4**). Approximately one third (35%) of the families who reported income in Year 16 were under 185% poverty².

Table 2. Demographic Characteristics of C16 Population

Demographic Categories		C16 UPSTART
Child's Gender		(n=13,220)
	Male	51%
	Female	49%
Child's Ethnicity		(n=13,220)
	Caucasian/White	72%
	Latino/a	11%
	Asian/Pacific Islander	3%
	African American/Black	1%
	Other	11%
	Decline to Answer	3%
Child's Language		(n=13,220)

² Not all families provided responses to every demographic question; therefore, the percentages in **Table 2** represent the proportion among those who provided information. This case-wise deletion method is a conventional, traditional, and simple technique used to manage missing data, particularly when reporting univariate statistics (Schafer, J. L., & Graham, J. W. 2002).

Demographic Categories		C16 UPSTART
	English	92%
	Spanish	7%
	Other	2%
Parent Educational Attainment		(n=13,220)
	Did Not Complete High School	2%
	High school graduate	11%
	Some College	23%
	College Graduate	43%
	Advanced Degree	20%
	Decline to Answer	2%
Parent Marital Status		(n=12,791)
	Married	89%
	Otherwise	11%
Poverty Level		(n=10,504)
	Under 185% Poverty	35%
	Above 185% Poverty	65%
<i>Percentages in the table are based on those providing a response in Waterford's participant records and may not add to 100% due to rounding.</i>		

UPSTART Usage

Research Question 1.1: To what extent did children use the UPSTART program as defined by Waterford (measured in minutes of instruction per week/total minutes for duration of program)?

The C16 academic year spanned 37 weeks of instruction, starting the week of September 9, 2024, and concluding on May 19, 2025. We analyzed program usage data to assess how families engaged with the program. On average, enrolled families participated for approximately 30 weeks. Waterford recommends a total of 1,500 minutes (or 25 hours) of usage for the reading curriculum throughout the program. This year, the average usage for all students (N=13,220) was around 1,736 minutes (or 29 hours) of reading instruction, a slight decrease from the previous year but still significantly above the recommended amount. UPSTART reading graduates used

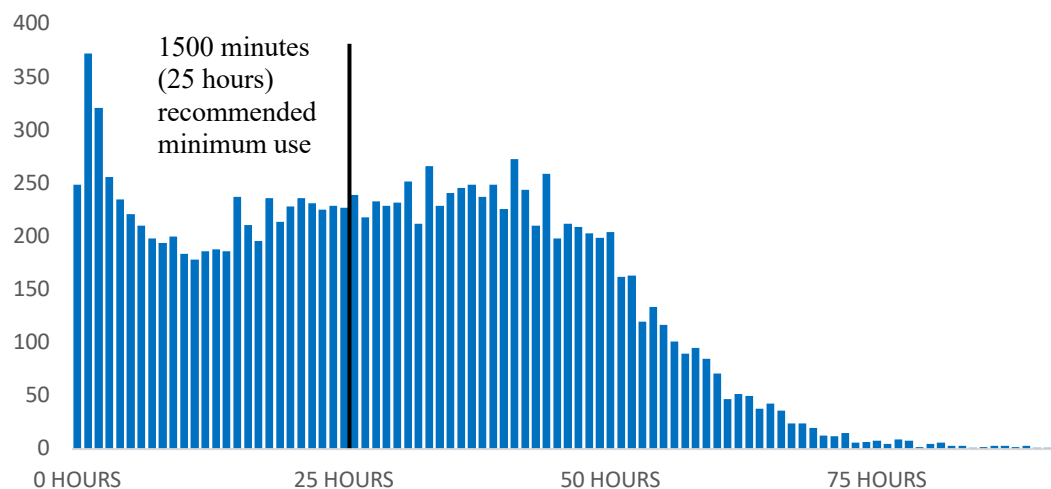
the software for approximately 2,360 minutes (or 39 hours) of instruction. Approximately 95% of students (n=12,621) who participated in the reading program also utilized the optional math software throughout the school year. On average, students engaged with the math software for 719 minutes over the course of 26 weeks, averaging about 24 minutes per week. A summary of the Year 16 data for instructional reading and math hours are presented in **Table 3**.

Table 3. C16 Hours of UPSTART Instruction

Group	N	Mean	SD	Range
UPSTART Reading	13,220	28.94	18.15	0.02-167.53
UPSTART Reading Graduates	8,330	39.33	14.01	0.02-167.53
UPSTART Math	12,621	11.98	13.89	0.02-141.17

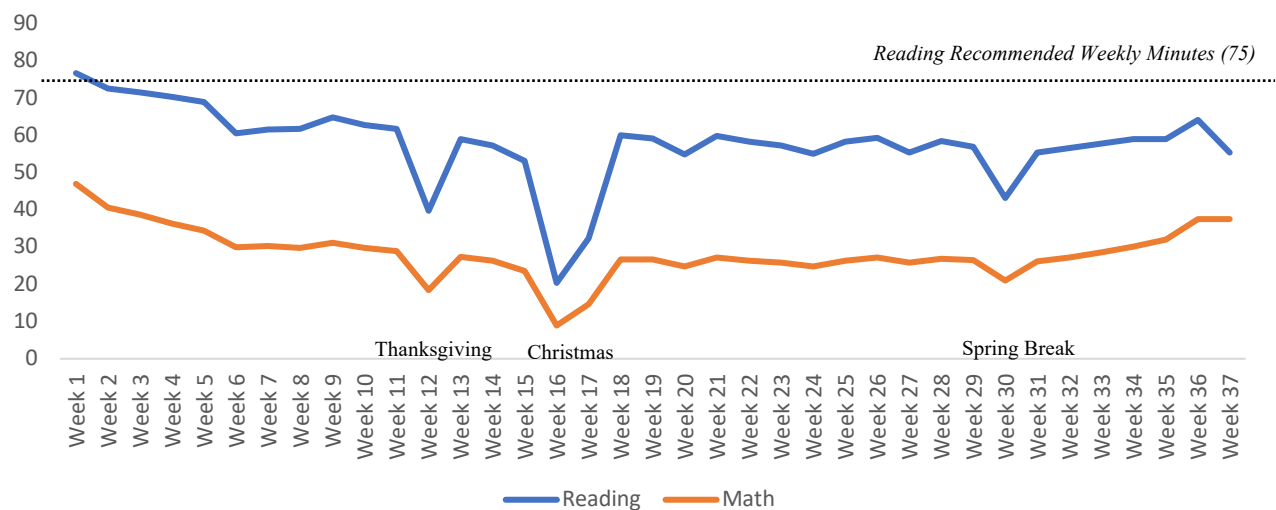
Figure 2 illustrates the distribution of reading instruction hours for the total C16 population (N=13,220). As shown, most children surpassed the program's minimum requirement of 25 hours; however, approximately 43% of participants recorded less than the recommended criteria for program completion.

Figure 2. Distribution of Hours of Reading Instruction for C16 Families



The bottom quartile of the C16 population completed 14.08 hours of reading instruction or less, the midpoint of the C16 distribution was 28.81 hours, and the top quartile completed 42.47 hours or more of instruction. Thirteen students used the program for 100 hours or more. The pattern of weekly usage throughout the program mirrored that of previous years, with declines in program engagement corresponding to holidays and school breaks (see **Figure 3**).

Figure 3. Average Weekly Minutes of Reading and Math Use



On average, the UPSTART children for the 2024-2025 program year logged below the recommended weekly reading duration of 75 minutes. Engagement with the math curriculum was lower, as anticipated for an optional component of the program.

UPSTART Graduation Rate

Research Question 1.2: What proportion of the participants successfully met the requirement for program completion (i.e., “graduated” as defined by Waterford)?

In Year 16, Waterford classified 8,330 children as graduates out of the 13,220 students, resulting in a graduation rate of 63% ($8,330/13,220 = 0.63$). Waterford students can meet the

graduation requirements in a couple of ways: (a) by logging at least 1,500 minutes (25 hours) of instruction with the UPSTART reading curriculum, and/or (b) achieving a specified level on the Waterford Assessment of Core Skills (WACS) adaptive assessment³.

Students who achieve high scores on the WACS—indicating readiness for kindergarten—are considered graduates of the program, regardless of their level of program usage. Similarly, students who meet the usage requirement (even if their WACS scores are lower) also qualify for graduation. As seen in **Table 4**, approximately 26% of students demonstrated both high achievement on the WACS and high usage, while only 6% achieved high scores with low usage. About 37% of students were unable to engage with the program as recommended and recorded low or no WACS scores, thus not meeting the graduation criteria for the UPSTART program.

Table 4. Waterford Graduation Criteria

	Low Program Usage (Less than 1500 minutes)	High Program Usage (1500 minutes or more)
High WACS achievement (>= Kindergarten Beginning AND >= 1 Level of Growth) OR (>= Kindergarten Intermediate AND No Pretest)	6% High achievement Low Usage	26% High Achievement High Usage
Low WACS Achievement (< Kindergarten Beginning OR <1 Level of Growth) OR (< Kindergarten Intermediate AND No Pretest)	7% Low Achievement Low Usage	14% Low achievement High Usage
No WACS	30% No EOY Assessment Data Low Usage	17% No EOY Assessment Data High Usage
<i>Graduation rate and categories provided by Waterford; bold text denotes graduate</i>		

³ This adaptive assessment provides precise data for both high and low-scoring students by evaluating key literacy skills such as phonological awareness, vocabulary, listening comprehension, phonics, and reading comprehension.

PRE-KINDERGARTEN ACADEMIC ACHIEVEMENT STUDY

Year 16 marked an important transitional point for evaluating the UPSTART program, as it was the first year in which achievement outcomes were examined using the PEEP to compare pre-kindergarten participants (treatment) with a matched sample of students who did not use the UPSTART program (control). We've provided a summarized snapshot of the findings followed by a detailed description of the research methods and analyses.

Summary of Key Findings

Results indicate that the UPSTART program has a positive impact on early literacy, with pre-k participants demonstrating significantly stronger literacy skills at the end of the program year than non-participants, although the strength of these results are mediated by program usage. Program use plays an important role: students who engaged most intensively with UPSTART experienced the largest literacy score gains, reflecting a moderate effect size compared to control groups. While numeracy outcome scores were less conclusive overall, likely due to ceiling effects in the assessment, students in the highest usage quartile showed modest improvements in numeracy scores. For low-income participants, results suggested positive trends across both literacy and numeracy measures, although small sample sizes limited our ability to determine statistical significance, underscoring the need for larger or combined cohorts to strengthen evidence for this subgroup.

Research Design

ETI implemented a quasi-experimental matched treatment–control group research design to compare literacy and numeracy outcomes between pre-k children who participated in the UPSTART program and those students who did not use the software. Both treatment and control children completed a PEEP Entry assessment prior to the start of the 2024–2025 pre-k year and a



PEEP Exit assessment at its conclusion. The timeline below outlines the sequence of assessments used to evaluate program impacts on early literacy and numeracy development.

	Summer 2024		Summer 2025	
Registered UPSTART Students	PEEP Entry	UPSTART	PEEP Exit	Kindergarten
Control Students	PEEP Entry		PEEP Exit	Kindergarten

Research Questions

The current impact evaluation was guided by several research questions focused on UPSTART’S impact on students’ literacy and numeracy achievement:

Research Question 2.1: Does participation in the UPSTART program produce an overall treatment effect? Specifically, do UPSTART students demonstrate significantly higher literacy and numeracy scores at the end of the pre-k year compared to non-participants (control group)?

Research Question 2.2: Do literacy and numeracy outcomes vary by program usage? In other words, do students with higher levels of UPSTART participation achieve stronger literacy and numeracy outcomes than students with lower usage or no participation?

Research Question 2.3: What is the impact of UPSTART on students from low-income backgrounds? Specifically, do low-income UPSTART participants outperform low-income non-participants on measures of literacy and numeracy at the end of the pre-k year?

Procedure

ETI collected literacy and numeracy assessment data at the beginning (“PEEP Entry”) and end (“PEEP Exit”) of the pre-k school year from a sample of students participating in the UPSTART program. These children comprised the treatment group. The Utah State Board of Education (USBE) provided PEEP Entry and Exit assessment data for a sample of students who did not participate in UPSTART, which served as the control group. ETI analyzed PEEP Entry and Exit data from both groups to assess differences in literacy and numeracy outcomes. All assessments were conducted in one-on-one sessions by trained test administrators using



standardized testing procedures. UPSTART parents provided consent for their child to participate in the evaluation and completed an online survey.

Control Student Matching Process

ETI matched a group of students who did not participate in UPSTART (control group) to a group of UPSTART students (treatment group) using Coarsened Exact Matching (CEM; Iacus, King, and Porro, 2008). Students were matched on data from the beginning of the preschool year (PEEP Entry) and across several important characteristics, including gender and race. For our literacy analysis, we matched on the PEEP Entry Literacy score and for numeracy we used the PEEP Entry Numeracy score.

The CEM procedure was designed to maximize retention of treatment cases. Given the larger pool of control students, slight pretest imbalances remained between the matched groups. These differences were statistically addressed through weighting, which adjusted for variation in the mean values of covariates. Despite these differences, our approach produced well-balanced analytic samples. Balance on characteristics used to match students is measured by the L1 statistic. The L1 Statistic is a comprehensive measure of global imbalance (Iacus et al., 2008). It is based on the L1 difference between the multidimensional histogram of all pretreatment covariates in the treated group and that in the control group. Because lower L1 values indicate less imbalance, this score demonstrates that the treatment and control groups were highly comparable across covariates (Literacy L1 statistic= 0.07758146; Numeracy L1 statistic=0.07758146).

Linear Regression Model

We used an ordinary least squares (OLS) regression model to estimate differences in outcome scores between the treatment and control group while controlling for other predictors of



posttest achievement variables. Our independent variable was treatment group status (1/0), and we included other predictor variables to control for their effects in our models, including: PEEP Entry test scores, gender, and ethnicity to adjust for their influence on exit literacy or numeracy scores. By controlling for these covariates, we increased our ability to show a causal link between program use and outcomes while holding other factors unrelated to the program constant. In cases where the regression model was not statistically significant, we did not calculate effect sizes.

Outcome Measures

The study focused primarily on foundational literacy outcomes, including phonological awareness, letter knowledge, and vocabulary, while also examining numeracy skills such as numeral recognition and quantity discrimination. The Preschool Entry and Exit Profile (PEEP) served as the primary outcome measure to assess achievement outcomes for both treatment and control groups. The PEEP captures performance in early literacy and numeracy skills and was developed collaboratively by Utah educators, higher education faculty, and members of the USBE. It was designed as a resource for stakeholders—including parents, teachers, and educational leaders—by providing information about the academic and lifelong learning practices essential for children entering and exiting pre-k. (USBE, 2024;2025). Specifically, the PEEP assessment⁴ is described as being used to:

- Determine levels of performance upon entry and exit of pre-k.
- Identify students who may need early intervention instruction and promote differentiated instruction for all students.
- Analyze the effectiveness of programs.
- Provide opportunities for data-informed decision-making and cost-benefit analysis of

⁴ Additional information on the PEEP is provided in **Appendix A**.

early learning initiatives.

- Identify effective instructional practices or strategies for improving student achievement outcomes in a targeted manner.
- Understand the influence and impact of pre-k in both the short- and long-term

PEEP Exit subscales measured literacy constructs such as *vocabulary* and *letter knowledge*, while the numeracy constructs included skills such as *counting* and *numeral discrimination*. The PEEP Exit consists of two subscales and produces a total score for literacy and a total score for numeracy. All question areas are presented in **Table 5**.

Table 5. PEEP Exit Profile

Literacy (Score range 0-108)	Numeracy (Score range 0-36)
Oral language	Oral counting
Alphabet knowledge: uppercase	Numeral identification
Alphabet knowledge: lowercase	Number sense: 1-1 correspondence, cardinality, and quality to numeral
Writing letters	Number sense: numeral to quantity
Phonological awareness	Discrimination: quantity
Alphabetic principle	Discrimination: shape creation
Concept of print: directionality	
Concept of print: letter and word	
<i>PEEP Exit Profile updated June 2024</i>	

It is important to note that the PEEP is a relatively new measure, developed by the state to use a common assessment for pre-k students. Psychometric properties of the PEEP, such as reliability and validity of the measure, have not yet been published. The evaluators used untransformed (raw) PEEP scores to conduct the analysis.

Detailed Results and Analysis

Results from a series of OLS regression models were used to evaluate the impact of the UPSTART program on early school readiness skills. Statistical models were used to determine three types of program impacts: 1) Intent to Treat Analysis (all students who used the program),

2) Program Use Analysis, and 3) Low Income Analysis. Within each sample, we present our results for literacy and numeracy separately.

Intent to Treat Analysis Results

We first examined the overall treatment effect by comparing outcomes for all students who used the program⁵ (Intent To Treat, ITT) and control samples. Our guiding research question was:

Research Question 2.1: Does participation in the UPSTART program produce an overall treatment effect? Specifically, do UPSTART students demonstrate significantly higher literacy and numeracy scores at the end of the pre-k year compared to non-participants (control group)?

ITT Literacy

As seen in **Table 6**, by the end of the program year, UPSTART students outperformed their matched control counterparts on measures of early literacy. The predicted mean PEEP exit score for the treatment group was 85.75, compared to 82.39 for the control group, representing a statistically significant difference of 3.36 points ($p = .021$). The effect size of 0.15 indicates a moderate but meaningful advantage for UPSTART participants relative to non-participants. While the difference is modest in magnitude, it provides evidence that UPSTART contributed positively to literacy development during the pre-k year.

Table 6. Literacy Regression Summary

	Condition	P-value	Predicted Mean	Diff.	ES
Literacy	Treatment	0.021	85.75	3.36	0.15
	Control		82.39		
<i>Note.</i> Data source: Matched sample size ctrl= 6906, tr= 264; L1 statistic= 0.07758146; Hedges’ g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater. “--”: effect size not calculated, model not statistically significant					

⁵ Students included in our ITT sample used the software for one minute or more throughout the program.

ITT Numeracy

Analysis of the PEEP Numeracy outcomes showed no significant overall treatment effect for UPSTART participants within our intent to treat sample (See **Table 7**). The treatment group had a predicted mean score of 33.37, compared to 33.56 for the control group, resulting in a nonsignificant difference of -0.19 points ($p = .371$).

Table 7. Numeracy Regression Summary

	Condition	P-value	Predicted Mean	Diff.	ES
Numeracy	Treatment	0.371	33.37	-0.19	--
	Control		33.56		
<i>Note.</i> Data source: Matched sample size ctrl= 4229, tr= 266; Numeracy L1 stat=0.07758146; Hedges’ g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater, "--": effect size not calculated, model not statistically significant					

Although these findings suggest no measurable advantage for UPSTART students in numeracy, interpretation is complicated by the characteristics of the PEEP Numeracy scale distribution. The distribution of PEEP Exit Numeracy scores demonstrated ceiling effects, also known as a negatively skewed score distribution, indicating that many students in both treatment and control groups scored at the top of the measure's range of scores, restricting the range of possible outcome scores. The score range restriction reduced the variance in the data, making it difficult to compare groups. Reduced variance lowers statistical power and makes treatment effects harder to detect (Kanyongo et. al, 2007). It is possible that the nonsignificant PEEP Numeracy Scale results may reflect limitations of the measure rather than the absence of true program effects in numeracy (for more information about PEEP score distributions, please see **Appendix D**).

Program Use Analysis Results

Next, we investigated how varying levels of program usage influenced outcomes by dividing the treatment sample into quartiles ranging from lowest to highest use.

Research Question 2.2: Do literacy and numeracy outcomes vary by program usage? In other words, do students with higher levels of UPSTART participation achieve stronger literacy and numeracy outcomes than students with lower usage or no participation?

Program Use Literacy

An analysis of program literacy achievement by usage quartile revealed important differences in literacy outcomes for UPSTART participants (See **Table 8**). Control students had a predicted mean score of 82.39 on the PEEP exit assessment. Among UPSTART participants, outcomes varied depending on program usage. Students in the lowest usage quartile scored slightly below controls ($M = 78.75$), though this difference was not statistically significant ($p = .233$). Students in the second and third quartiles scored above controls ($M = 85.63$ and 84.41 , respectively), but these differences were also nonsignificant.

Table 8. Literacy Quartiles Predicted Means

Condition	Minutes of Use	P-value	Predicted Mean	Effect Size Compared to Controls
Control			82.39	
Usage Quartile 1	0-1,685	0.233	78.75	--
Usage Quartile 2	1686-2469	0.257	85.63	--
Usage Quartile 3	2470-3010	0.469	84.41	--
Usage Quartile 4	3011-10052	0.000	92.90	<u>0.48</u>
<i>Note.</i> Data source: Matched sample size ctrl= 4229, tr= 266; Hedges' g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater, "--": effect size not calculated, model not statistically significant				

By contrast, students in the highest usage quartile demonstrated a large and statistically significant advantage ($M = 92.90$, $p < .001$, $ES = 0.48$). This effect size approaches the threshold for a moderate impact, indicating that high levels of UPSTART participation were strongly

associated with improved early literacy outcomes. Taken together, the results suggest that while modest usage does not reliably differentiate outcomes from controls, intensive engagement with the UPSTART program is linked to substantial literacy gains.

When isolating and studying only the highest usage group of UPSTART participants (Usage Quartile 4) and matching them to a comparable control sample, the results indicate a clear and meaningful program impact. As seen in **Table 9**, high-usage students (N = 69) achieved an average predicted PEEP Exit score of 92.19, compared to 82.11 for the control group (N = 5,128)⁶. This 10.08-point difference was statistically significant ($p < .001$) and represents a large effect size (ES = 0.45). These findings suggest that high levels of use with UPSTART is strongly associated with higher literacy outcomes by the end of the pre-k year. Overall, this analysis highlights that UPSTART participants who engage most consistently with the program experience the greatest literacy gains relative to their peers.

Table 9. Literacy Predicted Means for High Reading Usage Students

	Condition	P-value	Predicted Mean	Diff.	ES
High Usage Students	Treatment	0.00	92.19	10.08	<u>0.45</u>
	Control		82.11		
<i>Note.</i> Data source: Matched sample size ctrl= 5128, tr= 69; Hedges’ g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater, “--”: effect size not calculated, model not statistically significant, “--”: effect size not calculated, model not statistically significant					

Program Use Numeracy

We conducted a quartile analysis of program usage for students who used the math software but found no statistically significant differences across quartiles. However, when we

⁶ Although the L1 statistic (0.193) indicates somewhat less balance between the treatment and control groups compared to the full matched sample, the observed impact remains robust and meaningful.

focused specifically on students with the highest levels of math usage (Quartile 4), results revealed a modest but statistically significant advantage in numeracy outcomes compared to the control sample (See **Table 10**). High-usage students (N = 64) achieved a predicted mean PEEP Numeracy score of 34.17, compared to 33.15 for control students (N = 3,629). This 1.02-point difference was statistically significant ($p = .03$) and corresponded to an effect size of 0.28, reflecting a medium impact.

Table 10. Numeracy Predicted Means for High Math Usage Students

	Condition	P-value	Predicted Mean	Diff.	ES
High Usage Students	Treatment	0.03	34.17	1.02	0.28
	Control		33.15		
<i>Note.</i> Data source: Matched sample size ctrl= 3629, tr= 64; L1= 0.17673523; Hedges’ g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater, “--”: effect size not calculated, model not statistically significant					

Although the effect size was smaller than that observed for our literacy high usage students, the findings suggest that dedicated participation in UPSTART may yield measurable gains in numeracy. However, interpretation should remain cautious given the limited sample size of high-usage treatment students and the ceiling effects observed in the PEEP Numeracy scale, which constrain the ability to capture the full extent of possible gains.

Low Income Analysis Results

Finally, we conducted an exploratory analysis of low-income families, to better understand the program's impact on students who may be most in need of additional educational support.

Research Question 2.3: What is the impact of UPSTART on students from low-income backgrounds? Specifically, do low-income UPSTART participants outperform low-income non-participants on measures of literacy and numeracy at the end of the pre-k year?

Low Income Literacy Analysis

Using Coarsened Exact Matching (CEM), we matched 60 low-income⁷ treatment students to a subset of 1,415 low-income control students. As seen in **Table 11**, the predicted literacy marginal mean for treatment students was 83.96, compared to 78.75 for controls—a difference of 5.21 points in favor of UPSTART participants. The relatively small base size for the sample of low-income treatment students limited statistical power, making it difficult to estimate the magnitude of the treatment effect with confidence.

Table 11. Literacy Predicted Means for Low-Income Students

	Condition	P-value	Predicted Mean	Diff.	ES
Low Income	Treatment	0.122	83.96	5.21	--
	Control		78.75		
<i>Note.</i> Data source: Matched sample size ctrl= 1415, tr= 60; L1 Statistic= 0.15230396; Hedges’ g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater, “--”: effect size not calculated, model not statistically significant					

Low Income Numeracy Analysis

Our analysis among low-income students on numeracy outcomes showed no statistically significant differences between UPSTART participants and control students. As seen in **Table 12**, the predicted numeracy marginal mean for treatment students was 30.36, compared to 29.94 for controls—a difference of less than one points in favor of UPSTART participants. Again, the relatively small base size for the sample of low-income treatment students limited statistical power, making it difficult to detect moderate effects with confidence.

⁷ For control students, low income is defined as 185% of poverty or under. For treatment students, low income is defined as 200% of poverty or under.

Table 12. Numeracy Predicted Means for Low-Income Students

	Condition	P-value	Predicted Mean	Diff.	ES
Low Income	Treatment	0.603	30.36	0.41	--
	Control		29.94		
<i>Note.</i> Data source: Matched sample size ctrl= 2,217, tr= 60; L1 Statistic= 0.00000000000000661; Hedges' g effect size (ES) benchmark is indicated in the table as follows if significant: Small: 0 to < .10; <i>Medium</i> , italicized text: .10 < .30, <u>Large</u> : bold and underlined text: .30 or greater, "--": effect size not calculated, model not statistically significant					

Overall, these findings suggest a potential positive impact of UPSTART for low-income families, but the evidence is not conclusive given the limited sample. Further research with larger samples of low-income students, or through combining cohorts across multiple years, will be needed to more accurately estimate program effects within this subgroup.

KINDERGARTEN ACADEMIC ACHIEVEMENT STUDY

As part of the Year 16 UPSTART evaluation, we examined a sample of UPSTART students from the 2023-2024 program year (Cohort 15 students) who completed kindergarten in 2025. We had access to kindergarten assessment data for over 42,000 students, enabling us to conduct a comprehensive analysis. This included comparing outcomes between program graduates and non-graduates, as well as a more detailed exploration of achievement outcomes among low-income students.

Summary of Key Findings

Overall, the findings indicate that students who used the UPSTART program demonstrated stronger kindergarten reading and math outcomes compared to their peers who did not participate in the UPSTART program during preschool. Low-income students and program graduates had a higher percentage of students at or above benchmark in reading and math by the

end of the year, compared to students who did not use the software. Our findings suggest that UPSTART participation in pre-k, and especially program graduation, was associated with improved learning achievement at the end of kindergarten.

Research Design

ETI conducted a pretest/posttest observational research design using assessment data from the USBE. We identified a substantial sample of Year 15 UPSTART students from the state's dataset and analyzed their kindergarten reading and math⁸ in achievement outcomes, measured by the Acadience Learning assessment. This approach also enabled us to establish a state-wide comparison group by excluding the UPSTART participants. The timeline below illustrates the evaluation design used to assess kindergarten outcomes.

2023-2024	Fall 2024		Spring 2025
Pre-K UPSTART Program Year	Acadience Beginning of Year	Kindergarten	Acadience End of Year
No UPSTART			

Research Questions

The kindergarten analysis was designed to answer the following research questions:

Research Question 3.1: How did reading and math achievement for kindergarteners who participated in the UPSTART program at pre-k, compare to the pattern of performance for Non-UPSTART students?

Research Question 3.2: How did the reading and math achievement for low-income kindergarteners who participated in the UPSTART program at pre-k, compare to pattern of performance for low-income Non-UPSTART students?

Research Question 3.3: How did reading and math achievement for low-income kindergarteners compare across three groups, (1) those who graduated from the UPSTART program at pre-k, (2) those who participated but did not graduate, or (3) those who did not participate at all?

⁸ In our pre-k study, reading was referred to as early literacy and math was referred to as numeracy.

Kindergarten Data Collection

We utilized Utah state Acadience Learning testing data, transferred from the USBE, to evaluate the reading and math outcomes of Cohort 15 UPSTART students compared to their non-program peers. Our analysis focused on Acadience reading and math data for approximately 11,000 UPSTART students and 30,000 non-program students, collected at both the beginning and end of the kindergarten year⁹.

Kindergarten Measures

Our kindergarten analysis used Acadience Learning data to measure the achievement outcomes among all kindergarten students. Both the Acadience Learning Reading and Math assessments are used statewide to measure student's acquisition of early literacy and math skills at the beginning, middle, and end of the academic year. The measures for each Acadience Learning assessment are presented in **Table 13**¹⁰.

Table 13. Acadience Learning Subscales

Acadience Reading	Acadience Math
Reading Composite Score	Math Composite Score
First Sound Fluency (FSF)	Beginning Quantity Discrimination (BQD)
Phoneme Segmentation Fluency (PSF)	Number Identification Fluency (NIF)
Nonsense Word Fluency (NWF)	Next Number Fluency (NNF)

We primarily focused on beginning of year and end of year Reading Composite and Math Composite scores, which are calculated using a combination of reading or math scales.

Composite scores provide the best overall estimate of student's reading or math proficiency for the student's grade level and time of year.

⁹ Cohort 15 sample sizes are based on those identified in both the state SIS dataset and in the Acadience dataset for the 2024-2025 school year. Not all students had complete Acadience scores.

¹⁰ A more detailed explanation of the Acadience assessment is available in **Appendix B**.

Detailed Results and Analysis

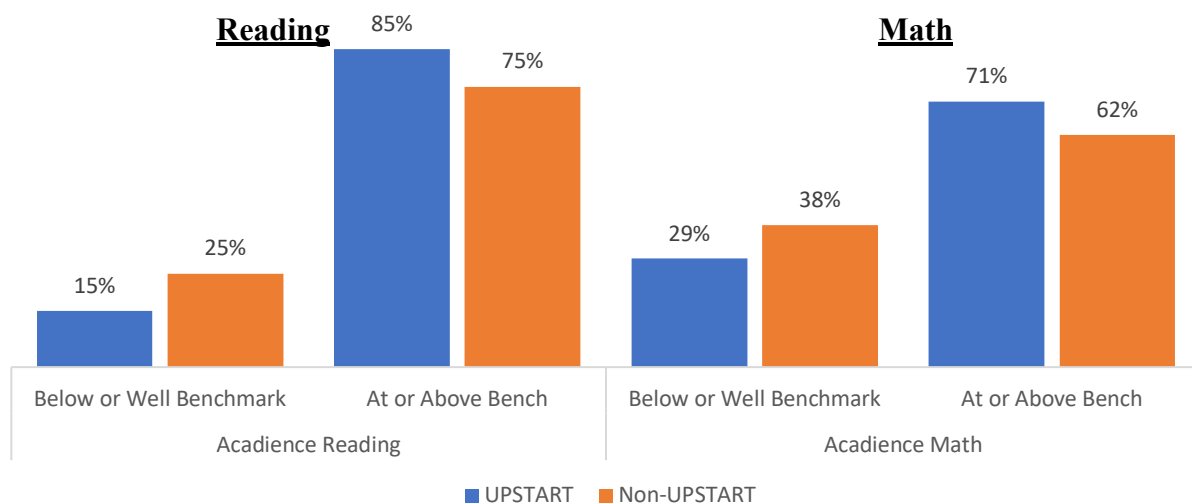
An observational pretest/posttest design was used to examine three types of program impacts: 1) Intent to Treat Analysis (all students who used the program), 2) Low-Income Analysis, and 3) Low Income Graduation Analysis.

Research Question 3.1: How did reading and math achievement for kindergarteners who participated in the UPSTART program at pre-k, compare to the pattern of performance for Non-UPSTART students?

Kindergarten Intent To Treat Analysis Results

UPSTART students were associated with higher scores at the end of kindergarten, compared to students who did not use the UPSTART software (**Figure 4**). For example, on Acadience Reading measures, 85% of the Cohort 15 UPSTART students (blue bars) finished kindergarten at or above benchmark, compared to 75% of students who did not use the software. Likewise, on the Acadience math assessment, 71% of Cohort 15 students finished at or above benchmark, compared to 62% of non-UPSTART students.

Figure 4. End of Year Benchmark Levels in Reading and Math among Program and Non-Program Students



Note: Reading EOY: UPSTART n=10,633, Non-UPSTART n=29,158; Math EOY: UPSTART n=10,498, Non-UPSTART n=28,951

It should be noted that we did not control for differences in student characteristics and thus this analysis was descriptive in nature. For more information about Acadience Beginning of Year composite scores for both groups, see **Appendix C**.

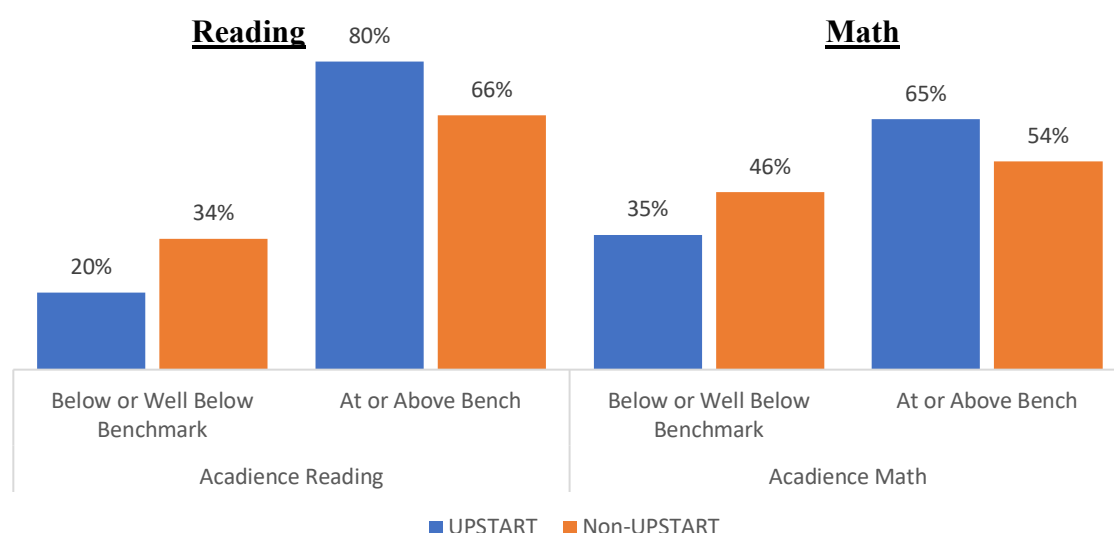
Kindergarten Low Income Analysis Results

Research Question 3.2: How did literacy and numeracy achievement outcomes for low-income students compare among those who used the UPSTART program at pre-k and those who did not participate at all?

We conducted a targeted analysis of kindergarten achievement outcomes for students identified as low-income based on demographic data from the Student Information System (SIS) at the start of the school year¹¹. At the end of the year, we found that 80% of pre-k UPSTART students were at or above benchmark, compared to only 66% of their non-program peers. In terms of math, 65% of pre-k UPSTART students were at or above benchmark, while 54% of non-UPSTART met or exceeded the benchmark (**Figure 5**). These findings suggest that low-income and at-risk students participating in UPSTART perform notably better than their non-program counterparts by the end of the kindergarten year.

¹¹ Student information system (SIS) data was provided by Utah State Board of Education (USBE) and collected at the beginning of the year. Students identified as low income are considered to be economically disadvantaged and include students who are eligible for free and reduced priced lunch.

Figure 5. Benchmark Levels in Reading and Math among Low-Income Program and Non-Program Students at End of Year



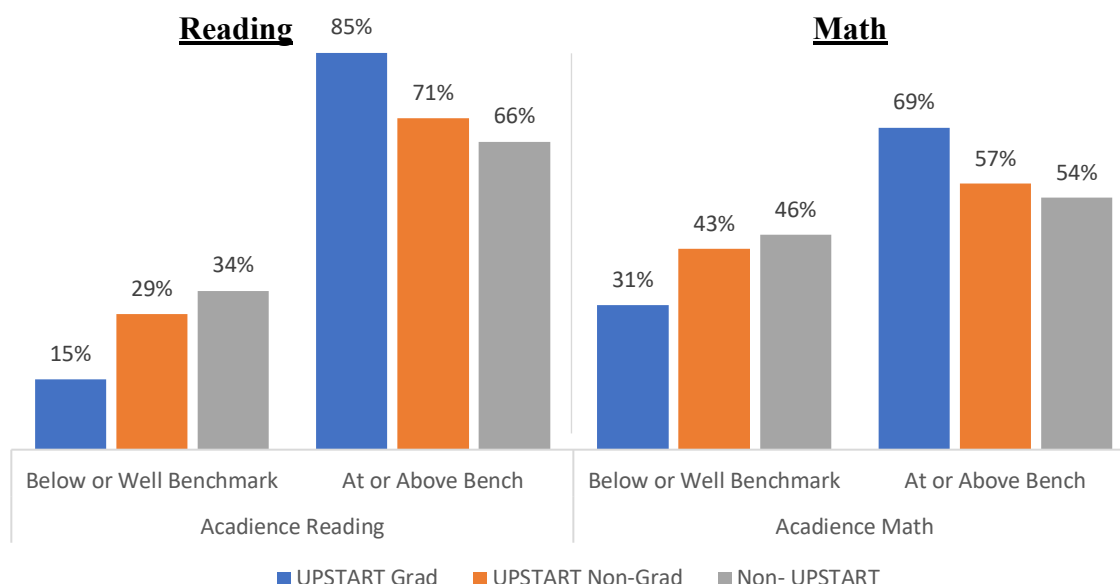
Note: Reading- UPSTART n= 2,266, Non-UPSTART n=10,130; Math- UPSTART n= 2,248, Non-UPSTART n=10,049

Kindergarten Low Income Graduation Analysis

Research Question 3.3: How did literacy and numeracy achievement outcomes for low-income kindergarteners compare among the following three groups, (1) those who graduated from the UPSTART program at pre-k, (2) those who participated but did not graduate, or (3) those who did not participate at all?

In **Figure 6**, we show our findings on kindergarten achievement among three groups of low-income students. A larger percentage of low-income students who participated in UPSTART at pre-k—both those who met graduation criteria (blue bars) and those who did not (orange bars)—were at or above benchmark by the end of the year, compared to students who did not participate at all (gray bars). Specifically, 85% of low-income UPSTART graduates and 71% of non-graduates finished kindergarten at or above the reading benchmark, while only 66% of non-program students reached this level.

Figure 6. Benchmark Levels in Reading and Math among Low-Income Graduates, Non-Graduates, and Non-Program Students at Exit



Note: Reading- UPSTART Grad n=1421, UPSTART Non-Grad n=845, Non-UPSTART n=10,130; Math- UPSTART Grad n= 1,412, UPSTART Non-Grad n= 836, Non-UPSTART n=10,049

Additionally, approximately 69% of graduates demonstrated adequate math skills by the end of the year. In contrast, about 57% of UPSTART non-graduates and 54% of non-UPSTART students were at or above benchmark in Acadience math outcomes by the end of the year. Although all students showed skill development by the end of kindergarten, pre-k UPSTART students—particularly graduates—exhibited a clear advantage in both reading and math skills compared to their non-graduate and non-program peers.

RECOMMENDATIONS FOR PROGRAM IMPROVEMENT

In its sixteenth year of implementation, the UPSTART program continues to demonstrate a strong positive impact on student achievement. Program use is strongly correlated with a positive impact on literacy test scores, and a standard metric for sufficient program use is

graduation status. Students who graduate UPSTART consistently show higher academic outcomes, including in delayed achievement measures at the end of kindergarten; however, graduation rates have declined in recent years, a trend that may impact students' success in the program. We recommend that Waterford strengthen its approach to improving program engagement and graduation rates, particularly for students experiencing risk factors that affect school readiness, including economic disadvantage or restricted access to early educational resources.

STUDY LIMITATIONS AND RESEARCH RECOMMENDATIONS

The UPSTART evaluation findings should be interpreted within the context of the study limitations, specifically, regarding the available data and measurement methods used. In this section, we outline these limitations and also provide our recommendations for improving future research and evaluation.

Income Data: While analyses of low-income participants indicated positive trends for UPSTART students in both literacy and numeracy, the small sample size restricted statistical power and limited our ability to detect significant effects. The low-income treatment sample size was limited because income reporting was optional for parents and therefore only available for a subset of UPSTART participants. As such, these results should be interpreted as exploratory, rather than conclusive.

Pre-kindergarten Academic Achievement Measurement: Students' PEEP Exit data exhibited ceiling effects, indicating that scores were negatively skewed by the end of the pre-kindergarten year. While test scores were normally distributed at the beginning of the year, many students—across both treatment and control groups—scored at or near the maximum range on

the post-assessment, particularly on the numeracy measure. This pattern reduces the ability to detect meaningful differences between groups and increases the likelihood of a Type II error (Ho and Yu, 2015)—that is, concluding that a treatment had no effect when it did. Consequently, the nonsignificant results in numeracy may reflect limitations of the PEEP assessment rather than an actual lack of program impact in that domain.

Based on insights gained from recent evaluation years, we propose several recommendations for enhancing future evaluations:

- ***Address Type II Error from Pre-K Exit Data:*** The PEEP Exit Numeracy data exhibits ceiling effects that limit the ability to detect program impacts. Incorporating kindergarten beginning of year (BOY) Acadience scores, which are normally distributed, alongside PEEP Exit data would enable a multi-measure pre-k analysis, that may help address limitations due to non-normal data distributions.
- ***Continue Using Matched Comparison Design:*** Assistance from the USBE (through a data sharing agreement) allowed us to conduct a quasi-experimental study, which is a strong evaluation research design that supports causal interpretation of findings. Future studies should build on this approach to strengthen our understanding of the impact of UPSTART.
- ***Connect Pre-k and Kindergarten Data to Strengthen Kindergarten Study:*** Future research should link baseline pre-k data with kindergarten Acadience outcome data. By matching the state-issued student IDs assigned in preschool to those used in kindergarten, we can analyze children's progress from pre-k into elementary school. This would enhance the kindergarten study by shifting it from a purely observational approach to a quasi-experimental research design.

- ***Consider Combining Cohorts Across Time:*** To increase statistical power and enable more detailed subgroup analyses, such as low-income impact analyses, future evaluations could combine multiple program cohorts across years to build larger subgroup sample sizes. This approach would allow for more robust subgroup comparisons and deeper insights into the program’s impact on student learning.
- ***Expand Longitudinal Analyses:*** Results indicate that there are benefits from the program beyond the pre-kindergarten year. Future evaluations should explore opportunities tracking UPSTART students into the early elementary years to assess sustained impacts on academic achievement and other relevant indicators of student success as they navigate K-12 education.

REFERENCES

- Dynamic Measurement Group, Inc. (2016, September). *Acadience Reading Benchmark Goals and Composite Score*. <https://Acadience.org/papers/AcadienceNextBenchmarkGoals.pdf>.
- Ho, Andrew D., & Yu, Carol C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, 75(3), 365-388.
- Iacus, Stefano M., Gary King, and Giuseppe Porro. 2008. *Matching for Causal Inference without Balance Checking*. <http://gking.harvard.edu/files/abs/cem-abs.shtml>.
- Kanyongo, Gibbs Y.; Brook, Gordon P.; Kyei-Blankson, Lydia; and Gocmen, Gulsah (2007). Reliability and Statistical Power: How Measurement Fallibility Affects Power and Required Sample Sizes for Several Parametric and Nonparametric Statistics," *Journal of Modern Applied Statistical Methods*: Vol. 6: Iss. 1, Article 9. DOI: 10.22237/jmasm/1177992480
- Powell-Smith, K., Good, R.H., III, & Dewey, E.N., & Latimer, R.J. (2014). *Assessing the Readability of Acadience AD Oral Reading Fluency and Daze*. (Technical Report No.16). Eugene, OR: Dynamic Measurement Group.
- Utah State Board of Education. (2021). *PEEP Cut Scores*. Utah State Board of Education. https://schools.utah.gov/assessment/assessment/assessments/pre_k/PreKPEEPCutScores.pdf
- Utah State Board of Education. (2025, August). *PEEP: School Year 2024–2025*. Utah State Board of Education. https://schools.utah.gov/assessment/assessment/assessments/pre_k/24PEEPAnnualReport.pdf
- Utah State Board of Education. (2024, June). *Utah's Pre-Kindergarten Entry and Exit Profile (PEEP): Entry Test Administration Manual*. Utah State Board of Education. https://schools.utah.gov/assessment/assessment/assessments/pre_k/24PreKPEEPEntryTAM.pdf
- Utah State Board of Education. (2024, June). *Utah's Pre-Kindergarten Entry and Exit Profile (PEEP): Exit Test Administration Manual*. Utah State Board of Education. https://schools.utah.gov/assessment/assessment/assessments/pre_k/24PreKPEEPExitTAM.pdf

APPENDIX A. PEEP Performance Levels

The USBE created PEEP performance levels to help further identify student needs. (USBE, 2021). At the PEEP Entry, Literacy and Numeracy scores fall into two performance levels. For both subscales, students who score within the performance level 1 range (0-23 points) are classified as “support needed” and demonstrate limited prerequisite knowledge and skills in literacy and numeracy. Students who score within the performance level 2 category are considered to be at benchmark. **Table A1** provides more information on the skills associated with each performance level.

Table A1. PEEP Entry Performance Levels

Subscale	Performance Level	PEEP Score Range	Level of Support	Associated Skills
Literacy	Performance Level 1	0-20 Points	Support Needed	Identifies few to no correct objects in the picture, Demonstrates few to no concepts of print, Identifies few to no letters of the alphabet, Draws few to no lines or shapes correctly, Can produce the first part of few to no compound words, Associates few to no sounds to the letters of the alphabet, Responds to a question in an incomplete sentence or off- topic or does not respond
	Performance Level 2	21-31 points	At Benchmark	Identifies most correct objects in the picture, Demonstrates concepts of print, Identifies some letters of the alphabet, Draws most lines or shapes correctly, Can produce the first part of most compound words, Associates some sounds to the letters of the alphabet, Responds to a question on-topic and speaks in complete sentences
Numeracy	Performance Level 1	0-23 Points	Support needed	Identifies few to no shapes, Counts few to no numbers from 1 to 10 Recognizes few to no numbers from 0 to 5, Identifies few to no groups that have more objects
	Performance Level 2	24-32 points	At Benchmark	Identifies all or most shapes, Counts most numbers from 1 to 10, Recognizes most numbers from 0 to 5, Counts objects using one-to-one correspondence up to 5, Demonstrates emerging cardinality in counting, Sorts objects correctly by given category, Duplicates and/or extends simple patterns, Identifies all or most groups that have more objects

At the PEEP Exit, Literacy and Numeracy scores are divided into three performance levels. For both subscales, students who score within the performance level 1 range at the exit have not demonstrated prerequisite knowledge and may require significant intervention. Students who are within the performance level 2 range (Literacy - 26-46 points, Numeracy- 18-28 points) have minimal prerequisite knowledge and may require some intervention in order to succeed. Students with the highest literacy and numeracy scores (literacy- 47-108 points and numeracy- 29-36 points) have demonstrated sufficient knowledge and skills and may require minimal intervention to succeed. **Table A2** provides more information on the skills associated with each performance level.

Table A2. PEEP Exit Performance Levels

Subscale	Performance Level	PEEP Score Range	Level of Support	Associated Skills
Literacy	Performance Level 1	0-25 Points	Significant Intervention	Names few to no objects/actions in a picture, identifies no letters of the alphabet, writes few to no letters, isolates few to no first sounds, associates few to no sounds to the letters of the alphabet, demonstrates few to no concepts of print
	Performance Level 2	26-46 points	Some Intervention	Names some objects/actions in a picture and describes it with a phrase(s), identifies few letters of the alphabet, writes some letters, which may include those in their name, isolates some first sounds, associates some sounds to the letters of the alphabet, demonstrates some concepts of print
	Performance Level 3	47-108 Points	Minimal Intervention	Names most objects/actions in a picture and describes it with a sentence(s), identifies some letters of the alphabet, writes his or her name and many letters, isolates many first sounds, associates many sounds to the letters of the alphabet, demonstrates concepts of print
Numeracy	Performance Level 1	0-17 Points	Significant Intervention	Counts few to no numbers from 1 to 20, identifies few to no numbers between 0 and 10, identifies one to no groups that have more objects, draws one to no shapes
	Performance Level 2	18-28 points	Some Intervention	Counts some numbers from 1 to 20, identifies some numbers between 0 and 10 counts some objects using 1-1 correspondence, associates numbers to small quantities less than 5, identifies some groups that have more objects draws some shapes
	Performance Level 3	29-36 points	Minimal Intervention	Counts many numbers from 1 to 20, identifies many numbers between 0 and 10, counts objects using 1-1 correspondence, associates numbers to large quantities less than 10, identifies groups that have more objects, draws most shapes

APPENDIX B. Acadience Reading and Math Measures

Acadience Reading is a statewide assessment used to measure students' acquisition of early literacy skills at the beginning, middle, and end of the academic year. According to a technical report produced by the Dynamic Measurement Group (Powell-Smith, et al., 2014), "*The Acadience measures map on to the critical early reading skills identified by the National Reading Panel (2002) and include indicators of phonemic awareness, Alphabetic principle, vocabulary and oral language development, accuracy and fluency with connected text, and comprehension.*" **Tables B1 and B2** provides a summary of the Acadience subscales used in our Kindergarten analyses.

Table B1. Acadience Reading Scales

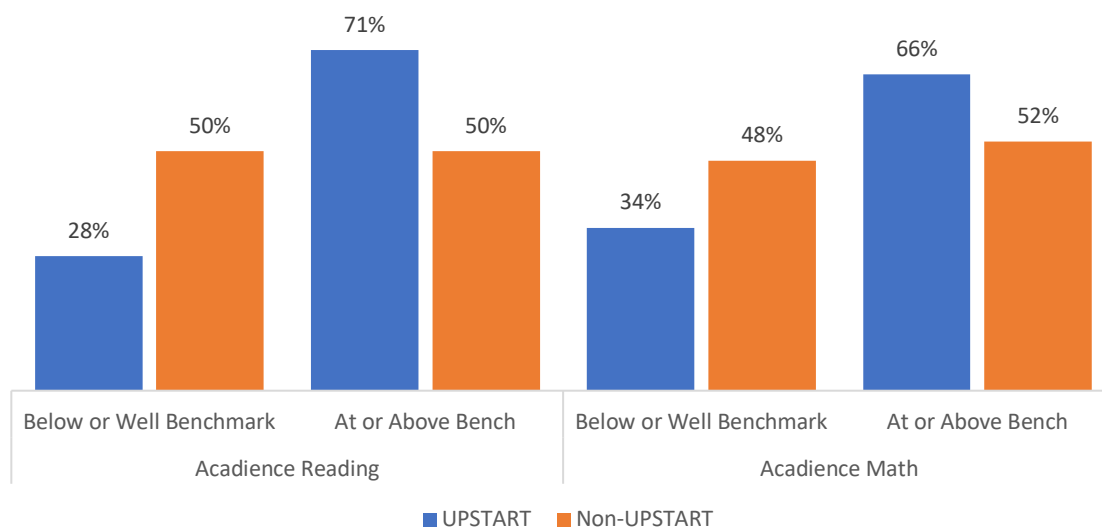
Acadience Reading Scale	Description	Early Reading Construct	Time Point
Composite Score	Acadience Composite Score is a combination of multiple Acadience scores	Overall estimate of reading proficiency	BOY, MOY, EOY
First Sound Fluency (FSF)	A brief direct measure of a student's fluency in identifying initial sounds in words.	Phonemic Awareness	BOY, MOY
Phoneme Segmentation Fluency (PSF)	Assesses the student's fluency in segmenting a spoken word into its component parts of sound segments.	Phonemic Awareness	MOY, EOY
Nonsense Word Fluency (NWF)	Assesses knowledge of basic letter sound correspondences and the ability to blend letter sounds into consonant-vowel-consonant and vowel-consonant words. Designed to measure alphabetic principle and basic phonics.	Alphabetic Principle and Basic Phonics	MOY, EOY

Table B2. Acadience Math Scales

Acadience Math Scale	Description	Early Math Construct	Grade
Composite Score	Acadience Composite Score is a combination of multiple Acadience scores	Overall estimate of math proficiency	BOY, MOY, EOY
Beginning Quantity Discrimination (BQD)	Assesses a student's ability to discriminate between two quantities	Magnitude Comparison and Subsidization	BOY, MOY, EOY
Number Identification Fluency (NIF)	Assesses a student's ability to orally identify the numerals 1 through 99	Number Identification	BOY, MOY, EOY
Next Number Fluency (NNF)	Assesses a student's ability to extend the county sequence	Counting	BOY, MOY, EOY

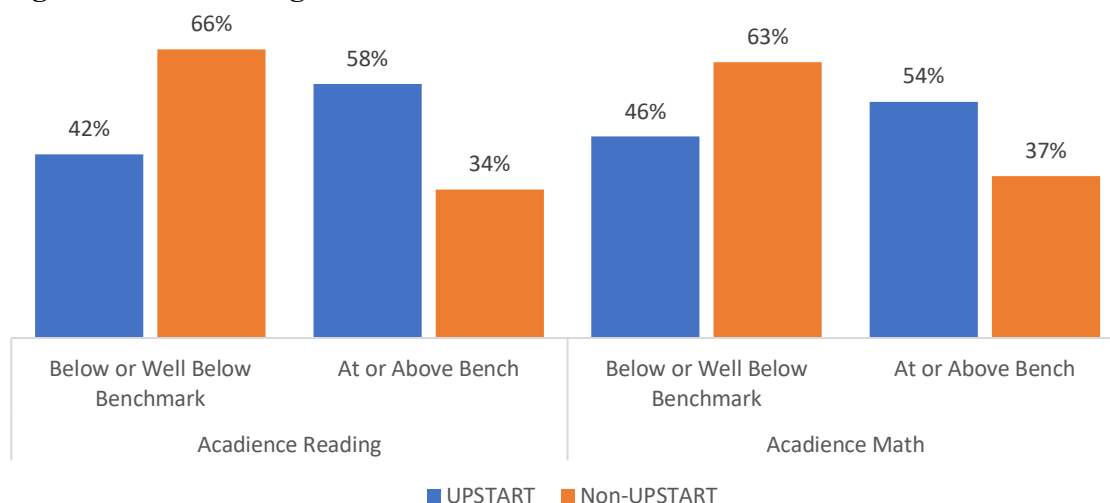
APPENDIX C. Acadience Reading and Math Beginning of Year Benchmark Levels

Figure C1. Beginning of Year Benchmark Levels in Reading and Math among Program and Non-Program Students



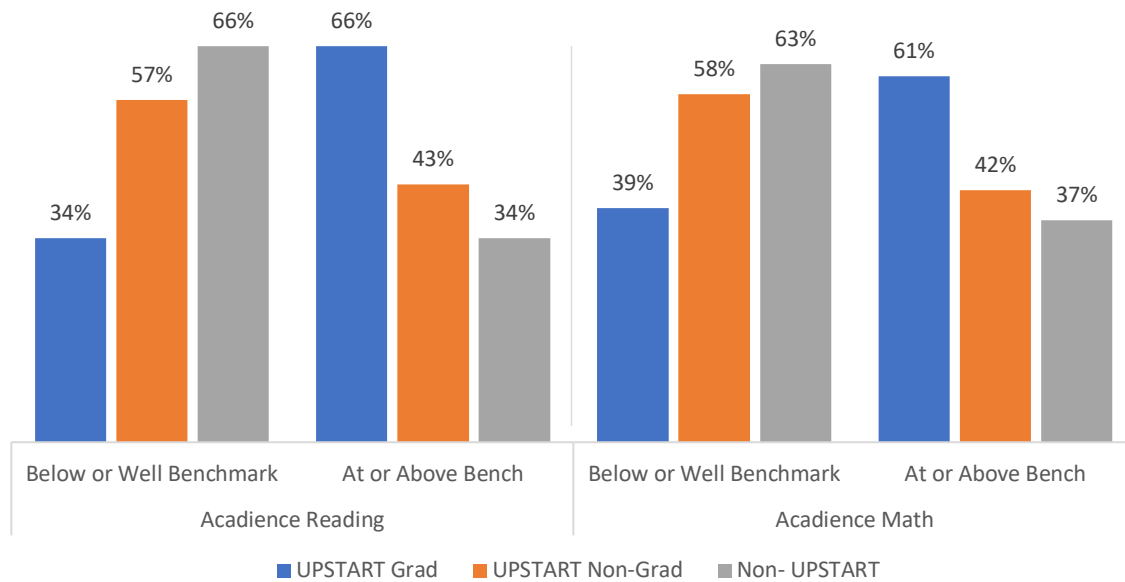
Note: Reading EOY: UPSTART n=10,633, Non-UPSTART n=29,158; Math EOY: UPSTART n=10,498, Non-UPSTART n=28,951

Figure C2. Beginning of Year Benchmark Levels in Reading and Math among Low-Income Program and Non-Program Students at End of Year



Note: Reading- UPSTART n= 2,2273, Non-UPSTART n=10,234; Math- UPSTART n= 2,232, Non-UPSTART n=10,013

Figure C3. Beginning of Year Benchmark Levels in Reading and Math among Low-Income Graduates, Non-Graduates, and Non-Program Students at Exit



Note: Reading- UPSTART Grad n=1,423, UPSTART Non-Grad= 850, Non-UPSTART n=10,234; Math UPSTART Grad n=1,398, UPSTART Non-Grad= 834, Non-UPSTART n=10,298

APPENDIX D. PEEP Entry and Exit Distributions

Figure D1 displays the distribution of the PEEP Literacy scores for students who did not use the UPSTART software in pre-k (control group). PEEP Entry Literacy scores are normally distributed (left/orange bar chart). In contrast, PEEP Exit Literacy scores are skewed to the right. The distribution of literacy scores for the ETI treatment sample follows a similar pattern, suggesting comparable overall trends in score distribution between the groups.

Figure D1. PEEP Literacy Score Distribution

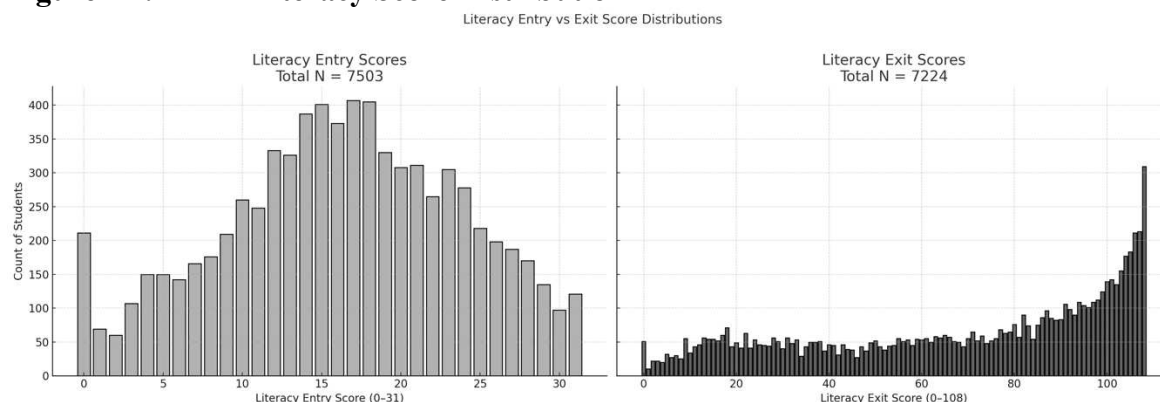
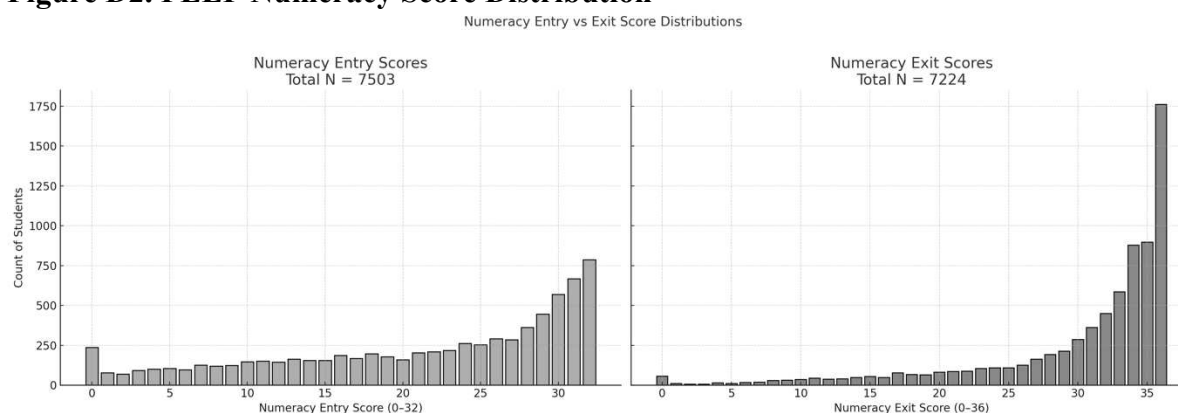


Figure D2 displays the distribution of PEEP Numeracy scores for control students. PEEP Entry Numeracy scores are negatively skewed, with most students scoring toward the upper end of the scale. The negative skew becomes more pronounced at exit (right), indicating that many students achieved high numeracy scores by the end of the school year. The ETI treatment group exhibits a similar distribution pattern at the Entry and Exit.

Figure D2. PEEP Numeracy Score Distribution





Evaluation and Training Institute
12240 Venice Blvd. Suite 27
Los Angeles, CA 90066
www.eticonsulting.org

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

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

