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

2017-2018 Findings

Submitted to the Utah State Board of Education
September 2018

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

Table of Contents

Executive Summary

EVALUATION PURPOSE AND RESEARCH QUESTIONS

The Evaluation and Training Institute (ETI) is conducting a three-year evaluation of the High-Quality School Readiness Expansion (HQSR-E) program to determine how participation in high-quality preschool programs impact children's school readiness. In 2017-2018, the second year of the evaluation, we focused on answering the following three overarching research questions:

1. Did high-quality preschool children have better early literacy skills at entry to kindergarten compared to children who were not enrolled in a high-quality preschool program?
2. Did high-quality preschool children have better early math skills at entry to kindergarten compared to children who were not enrolled in a high-quality preschool program?
3. What were the social-emotional development skills of program children at the conclusion of the preschool year?

EVALUATION APPROACH

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

In an effort to control for pre-existing differences between our treatment and comparison samples, we matched comparison students to treatment students on literacy and math achievement before the program started. We combined the Cohort 1 (2016-2017) and Cohort 2 (2017-2018) students to increase the sample sizes of our treatment groups and detect small program effects. The matched groups created equal baseline levels of literacy and math achievement, and allowed us to have groups balanced across these skill levels prior to conducting any statistical analysis. We used different types of analyses, including ordinary least squares (OLS) regression analysis and descriptive analysis of outcome data.

FINDINGS

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

Literacy Skill Development:

- ✓ Using descriptive analysis, the majority of high-quality preschool students across all three program models entered kindergarten with literacy quotient levels rated as average or above average when compared to a national normed sample of their peers: 77% of UPSTART children had average or above average literacy quotients at the end of preschool, followed by 71% of high-quality public preschoolers, and 69% of high-quality private preschoolers.
- ✓ Children using UPSTART had higher overall literacy test scores at post-test than the comparison children. Specifically, UPSTART students had significantly higher scores on subtests that measured letter knowledge, listening comprehension, and phonological awareness.
- ✓ Participation in the high-quality public and private preschools did not result in significantly better literacy outcomes when compared to a group of similar children not attending high-quality preschools.
- ✓ Children enrolled in high-quality private preschools had significantly lower phonological awareness scores when compared to a group of similar comparison children.

Social Emotional Development (SED):

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

Math Skill Development:

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

RECOMMENDATIONS

Our recommendations were developed in response to the study findings, theoretical considerations and limitations discussed in the body of this report. Our primary recommendation is to continue the program and evaluation to develop more evidence about its impacts over time, particularly its long-term impacts. As the program enters its third year, we will now be able to study its impacts from preschool into first grade, which is crucial information for policy decisions.

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

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

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

HQSR-E 2017-2018 Report

Evaluation Purpose & Research Questions

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

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

In this report we include findings for a combined preschool sample, which includes data collected during the first and second school year of implementation. Combining Year 1 and Year 2 program student cohorts was necessary for increasing our statistical power and identifying small impacts from the program.

Program Background

In March of 2016, the Utah State Legislature provided grant funding for a multi-year project to expand access to high-quality preschool programs to economically disadvantaged children. To receive grant funding and qualify for high-quality status, Local Education Agencies (LEAs) are scored on a rubric to ensure their programs meet certain standards for high-quality programs. The process involves the submission of a grant application, an interview in which the LEA reviews supplementary materials submitted as evidence of high-quality elements, and classroom observations. In the grant applications, LEAs describe their need for funding, plans for recruiting students, and how their program meets each high-quality element listed in the legislation, including, staff with at least a bachelor's degree in an early childhood education related

field or a child development associate certification, evidence-based curriculum that aligns to the Utah Early Childhood Standards adopted by the State Board of Education, and class sizes that do not exceed 20 students, among other criteria. Classroom observations are conducted using the Early Childhood Environment Rating Scale (ECERS-3 and ECERS-E), an observational tool that rates the quality of early education programs on the following six areas: space and furnishings, personal care routines, language and literacy, learning activities, interaction, and program structure. Local Education Agencies (LEAs) must achieve a minimum overall ECERS score of three or four, depending on the program length, as well as meet the additional rubric criteria.

In 2017-2018, there were fourteen high-quality public-school districts[1], six private high-quality preschool providers, and one computer/home-based high-quality provider, Waterford's UPSTART program. In addition to expanding access to high-quality preschool programs, the legislation provided funding to increase the quality of early childcare professionals through a professional development program that helps educators earn their Child Development Certificate.

Evaluation Methods

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

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

Figure 1: Preschool Study Yearly Testing and Data Collection

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

[1] HQSR-E districts included: Cache, Davis Community, Davis HOPE, Duchesne, Granite, Iron, Jordan, Logan, Murray, Salt Lake City, Sevier, Washington, and Weber.
[2] "Program students" were operationally defined as economically disadvantaged students who attended a high-quality preschool.

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

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

We measured math and social and emotional development with the Brigance IED III. Selected scales from the Brigance Math measured children's ability to count by route, read numerals, and identify missing numerals in a sequence. The social and emotional development scale was a parent survey and assessed parental perceptions of children's interpersonal and self-regulatory skills. Based on recommendations outlined in the Cohort 1 report (Evaluation and Training Institute, 2017), we modified the socio-emotional development measure from a dichotomous response scale to a five-point Likert response scale to increase variability and reduce measurement ceiling effects. Lastly, we collected background information on participants' home environment through a parent intake survey developed by the ETI project team. See **Appendix A** for more details about the measures used in the evaluation.

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

---

[3] High-quality public districts tested in Cohort 1 were Davis HOPE; Salt Lake City; Weber; tested Cohort 2 high-quality public districts were SLC; Weber; Granite; Davis Community; Davis HOPE; Jordan.
[4] Cohort 1 high-quality private sites were Progressive Preschool, Head Start (SSL), Head Start (CCH); Cohort 2 high-quality sites were CAP Head Start, Hilltop Christian, Mountainland Head Start, YMCA New Bridge.

Following the recommendations in the Cohort 1 report, we combined data from Cohorts 1 and 2 to ensure a larger analytic sample and greater confidence in the stability of our statistical findings (Evaluation and Training Institute, 2017). Before we combined the data across cohorts, we conducted analyses and reviewed all outcomes to ensure that there were no widespread differences between the two cohorts. **Table 1** displays the number of children tested in Cohort 1 and 2.

Table 1. HQSR-E Cohort 1 and 2 Number Tested

Experimental group	Number Tested Cohort 1	Number Tested Cohort 2
Public	75	99
Private	58	65
UPSTART	93	101
Non-HQ comparison	134	112

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

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

After removing cases with incomplete or missing data, cases in which the program or control condition changed (e.g. control children who enrolled in a high-quality preschool after the pre-test), and creating statistically matched groups with CEM, the final analytic samples for each measure and treatment group is listed in **Table 2** (see **Appendix C** for sample demographic characteristics).

Table 2. HQSR-E Analytic Sample

Group	Literacy (Brigance)	Letter Knowledge (Brigance)	Phonological Awareness (Bader)	Listening Comprehension (PELI)	Math (Brigance)
HQ Public	108	102	131	98	97
Non-HQ	108	102	131	98	97
HQ Private	81	83	84	77	78
Non-HQ	81	83	84	77	78
UPSTART	125	115	145	122	83
Non-HQ	125	115	144	122	83

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

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

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

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

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

Findings

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

Literacy

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

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

We used an OLS regression model to determine the impact of the different implementation models on children's overall literacy development while controlling for prior achievement, preschool duration, and family socioeconomic factors. **Table 3** presents the effect size estimates for the Brigance Global Literacy composite, an overall measure of alphabet knowledge, vocabulary, phonics, and language concepts. An asterisk denotes a statistically significant finding and negative effect sizes indicate the control group performed better than the treatment group.

Table 3. Post-test Analysis of Literacy Composite Effect Sizes, OLS Regression Model

Construct	UPSTART	Public	Private
Global Literacy	**.78***	NS	NS

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

As seen in **Table 3**, the UPSTART group was the only program group to produce a statistically significant positive effect on the Brigance Global Literacy composite. The UPSTART group produced a meaningful impact on the development of early literacy skills when compared to children who were enrolled in traditional non-HQ preschools, with an effect size of .78 that exceed our .26 benchmark for practical significance. There were no significant differences between children enrolled in high-quality public or private preschools and their comparison counterparts on the Brigance Literacy composite.

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

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

Table 4 presents the effect size estimates for skills measuring letter knowledge, oral comprehension, and phonological awareness for each high-quality preschool model.

As seen in **Table 4**, there were few statistically significant positive differences between program and control group children in literacy domain areas, with the exception of UPSTART, which had medium to strong effects in areas measuring phonological awareness (ES = .84) and letter knowledge (ES = .60). UPSTART had a small though statistically significant positive impact on students' listening comprehension skills (ES = .29). Comparison children recruited from non-high-quality preschools had significantly stronger outcomes compared to the high-quality public preschool group on phonological awareness (ES = -.26), however the effect size was small and did not exceed the benchmark for practical significance. There were no significant or meaningful differences between children enrolled in high-quality public or private preschools and children enrolled in traditional preschools on scales measuring listening comprehension or letter knowledge.

Table 4. Effect Size Estimates by Literacy Domain

Literacy Domain	Skills Tested	UPSTART	Public	Private	Instrument
Letter Knowledge	Letter Sounds	**.60****	NS	NS	Brigance
	Recites Alphabet				
	Letter Knowledge				
Listening Comprehension	Inference/Prediction	**.29***	NS	NS	PELI
	Recollection				
	Cloze				
Phonological Awareness	Rhyme Recognition	**.84****	**-.26***	NS	Bader
	Phonemic Blending				
	Phonemic Segmenting				

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

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

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

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

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

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

Table 5 presents the percentage of children in each program group who had literacy quotient levels that fell into the categories of below average, average, and above average at the conclusion of the HQSR-E Program.

Table 5. Post-Test Literacy Quotient Levels by HQSR-E Program Group

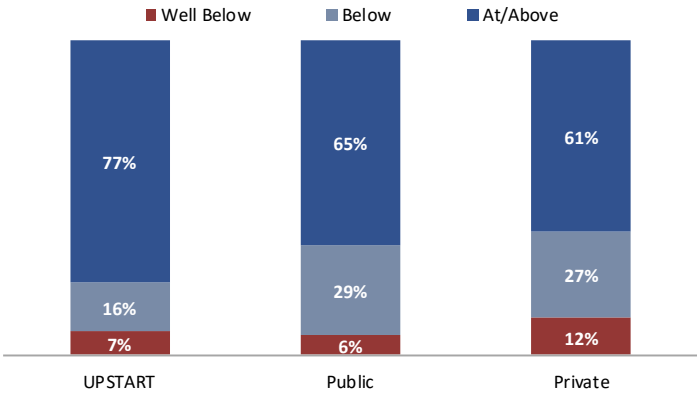
Quotient Level	UPSTART (N = 125)	Public (N = 108)	Private (N = 81)
Below average	23%	30%	31%
Average	22%	30%	36%
Above average	55%	41%	33%

Overall, regardless of HQSR-E program model, the majority of students are entering kindergarten with literacy quotient levels of average or above, indicating that they are on target and ready to learn. Over three-fourths of children (77%) who participated in UPSTART had quotient scores rated as average or above at the end of the preschool year, with 55% of children scoring literacy quotients of above average. Moreover, approximately 70% of students enrolled in high-quality public and private preschool settings scored average or above on the Brigance literacy measure before entering kindergarten, with 41% and 33%, respectively, scoring above average. Slightly fewer students attending high-quality private preschools scored above average on the Brigance literacy (33%), compared to students enrolling in high-quality public preschools (41%) or UPSTART (55%).

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

Figure 2 presents post-test PELI listening comprehension benchmark levels for children enrolled in the three high-quality preschool models. The majority of children had listening comprehension scores that were at or above benchmark goals, with the UPSTART program having the highest percentage of on-target children (77%), followed by high-quality public preschools (65%) and high-quality private preschools (61%).

Figure 2. Listening Comprehension Benchmark Status at Post-test



One important caveat to the Brigance Literacy quotient and PELI comprehension benchmark analysis is that they do not take pre-existing differences into account and presents the literacy achievement of children at one point in time (before kindergarten entry). The next research question addresses the growth in literacy skills for children between the pre-test and post-test assessments.

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

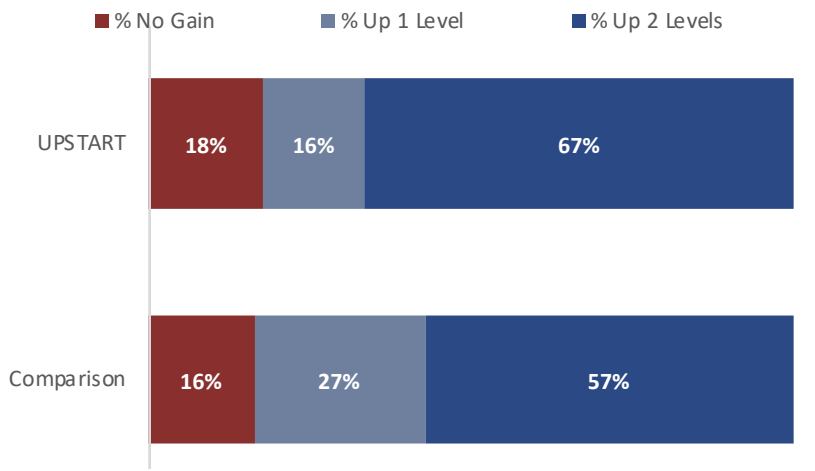
In addition to literacy quotients, the Brigance literacy scale is associated with age equivalents based on a child's literacy knowledge compared to other children at a particular age. If a child scores below the average knowledge age for his/her biological age, he/she might be at risk of struggling in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

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

High-quality preschool interventions that use age-appropriate curricula can contribute to improvements in academic areas, particularly for children who lack the skills that predict readiness for kindergarten (Phillips et al., 2017). Focusing on the children most at risk prior to beginning the HQSR-E program, or children whose literacy content knowledge age was at or below 80% of their biological age at pre-test, we measured the extent to which children mastered enough literacy concepts during the course of the preschool year to improve their knowledge age at post-test. If children move from a knowledge age at or below 80% of their biological age to a knowledge age close to their actual biological age (90% or above), they are on target with similarly aged peers and are more prepared to enter formal school settings.

Figure 3 follows the 49 UPSTART children and their 45 non-high-quality comparison counterparts who were classified as 80% or below their biological age at pre-test and examines their performance at post-test. Findings are based on descriptive analysis and comparisons to norm groups.

Figure 3. Change in Knowledge Age, UPSTART and Comparison Preschoolers

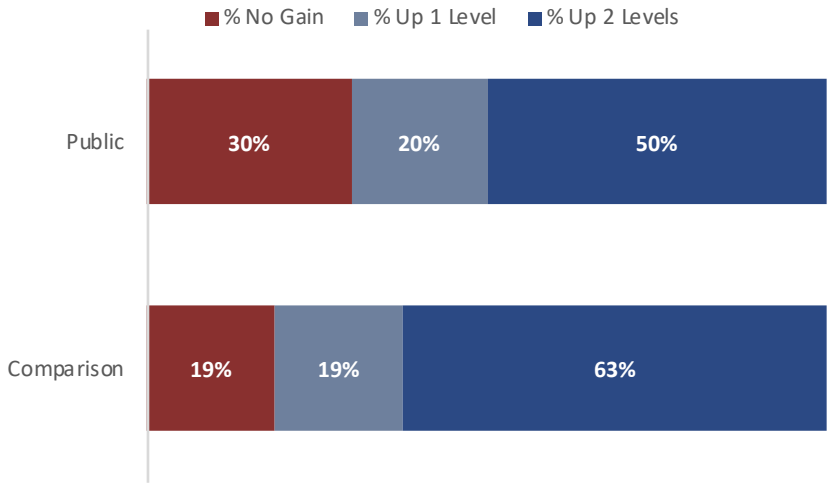


N= UPSTART (45); Comparison (49)
Students had a content knowledge age at or below 80% of their biological age (BA) at pre-test.
% No gain: Stayed at or below 80% BA; % Up 1 Level: 81-90% below BA; Up 2 levels: 91% to or at/above BA

As seen in **Figure 3**, more than two-thirds (67%) of at-risk UPSTART children increased their knowledge age two levels, raising their knowledge age from 80% or below their biological age (i.e., at-risk) to 91% or above their biological age (i.e., on target) during the course of the program, compared to 57% of children enrolled in traditional non-HQSR-E programs who achieved the same result. Eighteen percent of UPSTART children and 16% of non-high-quality comparison children remained at 80% or below their biological age at the end of the preschool year.

Figure 4 displays the change in knowledge age for the 48 high-quality public preschoolers and 50 comparison children that were classified as at-risk at pre-test (i.e., had a literacy knowledge age of 80% or below their biological age).

Figure 4. Change in Knowledge Age, high-quality Public and Comparison Preschoolers

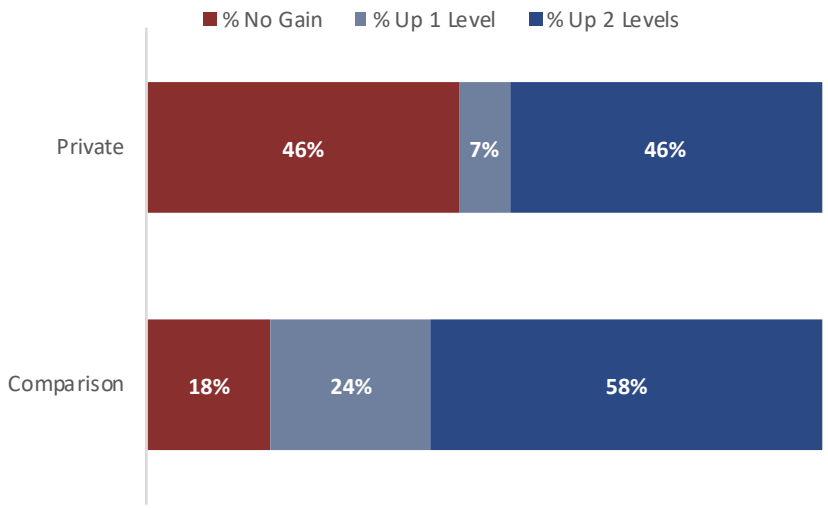


N= Public (48); Comparison (50)
Students had a content knowledge age at or below 80% of their biological age (BA) at pre-test.
% No gain: Stayed at or below 80% BA; % Up 1 Level: 81-90% below BA; Up 2 levels: 91% to or at/above BA

Half of the high-quality public preschoolers classified as at risk moved up two levels to their target knowledge age, compared to 63% of at-risk children who were enrolled in non-high-quality preschools (see **Figure 4**). Thirty percent of high-quality public at-risk children had a literacy knowledge age that remained at 80% or below their biological age after a year or preschool, while 19% of at-risk children in traditional preschools similarly saw no change in their knowledge age.

Preschoolers attending high-quality private sites with literacy knowledge ages of 80% or below their biological age are presented in **Figure 5**, along with their non-high-quality counterparts.

Figure 5. Change in Knowledge Age, High-quality Private & Comparison Preschoolers



N= Private (41); Comparison (38)
Students had a content knowledge age at or below 80% of their biological age (BA) at pre-test.
% No gain: Stayed at or below 80% BA; % Up 1 Level: 81-90% below BA; Up 2 levels: 91% to or at/above BA

Forty-six percent of high-quality private preschoolers with a literacy knowledge age of 80% or below their chronological age increased their knowledge age two levels to 90% and above their chronological age, compared to fifty-eight percent of comparison non-high-quality children. Forty-six percent of high-quality private preschoolers, compared to 16% of non-HQ comparison children, remained at 80% or below their biological age at the end of the preschool year.

Social-emotional Development

Social-emotional development includes the child's experience, expression, and management of emotions and the ability to establish positive and rewarding relationships with others (Early Education and Support Division, 2017). Children's social and emotional development are important skillsets for determining if children are emotionally ready to learn and interact with their peers in a school setting. For example, Denham (2006) explains that "compared with less socially competent peers, children who enter school with greater competence show a number of benefits related to their relationships with teachers and classmates, readiness to learn, school engagement,

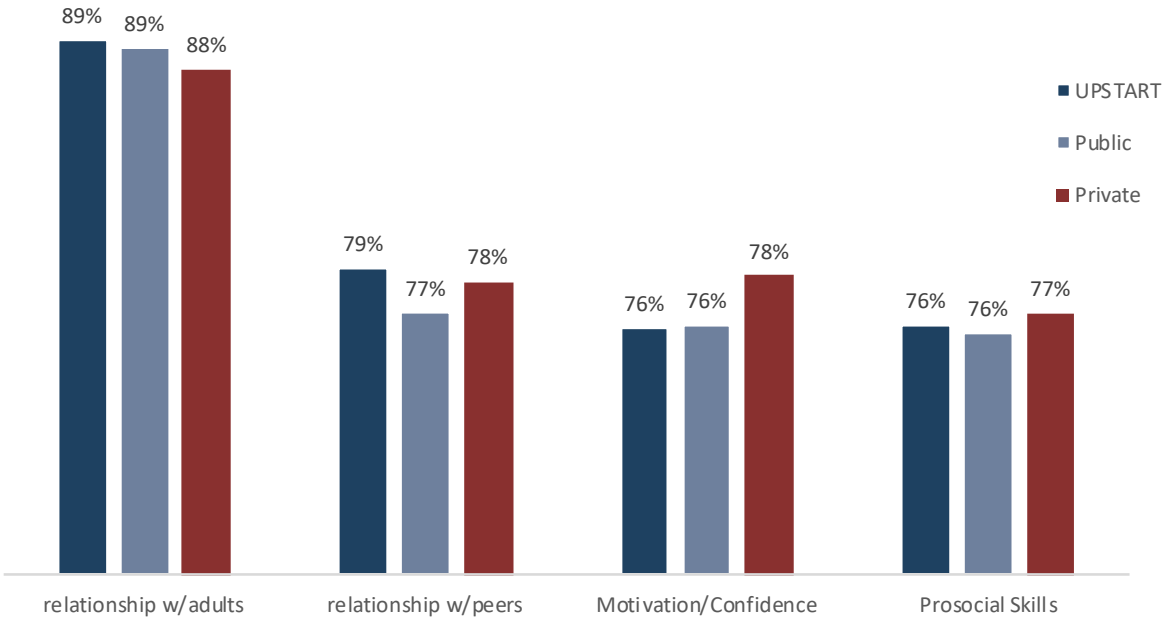
and overall academic adjustment." We administered a survey to parents that measured children's social-emotional development in four areas, including areas measuring their interpersonal skills (relationships with adults and peers) and self-regulatory skills (prosocial skills and confidence and motivation). SED analyses findings are reported for Cohort 2 participants and not the combined C1/C2 analyses sample due to changes made to the survey scale from C1 to C2. Our findings are presented below.

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

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

Figure 6 presents the percent of possible skills developed in these areas, on average, for each of these four social-emotional development constructs by HQSR-E program group. In general, parents rated the SED skills of their children similarly, with little variation on SED subscales between children enrolled in the different high-quality program models.

Figure 6. Post-program Social-emotional Development skills by Program Group



N= UPSTART (98); Public (91); Private (51)
Percentages based on the total count of skills divided by total possible in scale. Scores reported by parents on an adapted SED instrument developed by Curriculum Associates.

Parents indicated that children had the strongest SED skills in behaviors related to relationships with adults (88-89%), followed by relationships with peers (77-79%), motivation and self-confidence (76-78%), and prosocial skills (76-77%). Most children across HQSR-E program groups developed the social-emotional development skills appropriate to their age, indicating they were ready for kindergarten in this area.

Math

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

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

Table 6 depicts the differences between treatment and control groups on a composite measure of math that assesses specific early numeracy skills. An asterisk denotes a statistically significant finding and negative effect sizes indicate the control group performed better than the treatment group.

Table 6. Posttest Analysis of Math Effect Sizes, OLS Regression

Scale	Skills Tested	UPSTART	Public	Private	Instrument
Math Composite	Counts by rote	NS	NS	**-.36***	Brigance
	Reads numerals				
	Identifies missing numbers				

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

As depicted in **Table 6**, UPSTART[5] or high-quality public program participation did not have a statistically significant positive impact on children's math skill development when compared to similar children who did not participate in the HQSR-E program. In the case of the high-quality private preschoolers, comparison children from non-HQ preschools had significantly stronger outcomes in numeracy skills on the Brigance Math composite (ES = -.36).

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

Discussion, Limitations & Recommendations

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

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

When considering that our evaluation was centrally focused on early academic achievement as a predictor of school success, the discussion of unmeasured aspects of children's development is important. These unmeasured factors may be impacting outcomes. Executive function, for example, is a wide-ranging umbrella construct that pertains to the cognitive and behavioral processes that "organize and direct all cognitive activity, emotional response, and overt behavior" (Isquith, Crawford, Esby, & Gioia, 2005). Executive functioning qualities include the ability to selectively focus on a particular task (attentional control), switch focus (cognitive flexibility), hold and use front of mind information (working memory), and control impulses (inhibitory control) (Ackerman & Friedman-Krauss, 2017). It is possible that the presence or lack of executive functioning skills may be influencing outcomes as opposed to the participation in high-quality preschools. However, if we measure executive functioning in future cohorts, we will be able to control for these skills in our statistical models and increase our confidence that they are not responsible for program effects (or lack thereof).

Evaluation Limitations

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

Timing of Data Collection. In many pre-posttest research designs, the pretest occurs prior to any exposure to the intervention, or program, being studied. However, this timeline is not always practical in its application, especially when it comes to measuring school-related outcomes. For example, the timing between receiving a list of program participants and when testing may begin can be affected by late school registration, staffing constraints and other challenges. Similar to our experience with Cohort 1, we did not receive contact lists of Cohort 2 program participants from public and private high-quality sites until after school started. It was our priority to test children as quickly as possible before they received too much exposure to the program, and we completed most of our testing by October 2017. This reduces the window of time between pretest and posttest, and could possibly diminish growth if pretest (baseline) scores are already being influenced by the program.

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

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

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

Recommendations

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

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

Our final recommendation is to consider broadening the evaluation measures to use data from the newly implemented Utah Kindergarten Entry and Exit Profile (KEEP). The KEEP was not available when this evaluation was designed, and we believe it may offer

additional information about HQSR-E program impacts beyond the measures of literacy, math and social-emotional development that we currently use. As the foundational measure of school readiness in the state, we believe that using KEEP scores alongside our established measures of early literacy and math would add another view and help to better understand the relationship between HQSR-E models and their impacts on students.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendices

Appendix A. Study Measures

Participating children were administered a battery of assessments in the beginning of preschool and were re-tested at the end of preschool. We measured early reading, numeracy and social and emotional development[6] using subscales from the Brigance Inventory for Early Child Development (IED II/III) (Brigance 2010, 2013), and the Preschool Early Literacy Indicator (PELI). **Table A1** details the instrument subscales used and maps them against their respective outcome measures.

Table A1
Preschool Outcome Measures

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

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

Appendix B. Data Collection and Analyses Methods

Scheduling and Recruitment. For both the Cohort 1 and 2 samples, ETI worked with the site contacts to transfer a contact list of treatment (e.g. study participants) families who may be eligible to participate in the study. ETI's team of schedulers contacted parents about study participation, which consisted of introducing the study, its purpose and benefits, determining eligibility to participate, and scheduling families once the potential participant was deemed eligible and willing. To be eligible, treatment families were screened based on the following criteria:

- Parent must have a preschool aged child who is not entering kindergarten until 2018;
- Child must be English proficient and at least one parent must be English proficient;
- Child must not be diagnosed with a disability that could affect the testing;
- Child must not be participating in UPSTART (unless they are in the UPSTART treatment sample).

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

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

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

Cohort 1 Testing Timeline. ETI quickly scheduled and tested families after receiving contact lists from participating sites. The computer/homebased preschool program, UPSTART, provided a contact list early in the summer and pre-test data collection was completed by August of 2016. ETI received the final list of public and private high-quality preschool sites in August of 2016 and began data collection at those sites on September 16th, 2016. Sites provided ETI with contact lists from September 2nd through

October 4th, 2016[7], and sites with earlier preschool start dates were given priority. Much of our pre-test data had been collected by October 10th, 2016 (90% of all our data). However, data were collected through mid-to-late October to accommodate families who had missed their earlier appointments and to target difficult to reach families in private high-quality preschools and increase the size of our smallest sample.

Cohort 2 Testing Timeline. Immediately after receiving contact lists from sites participating in the 2017-2018 program year, ETI began scheduling and testing families. The computer/homebased preschool program, UPSTART, provided a contact list early in the summer and pre-test data collection was completed by August of 2017. ETI received the final list of public and private high-quality preschool sites in August of 2017 and began data collection at the first of those sites on September 9th, 2017. Sites provided ETI with contact lists from August 15th through October 3rd, 2017, and sites with earlier preschool start dates were given priority. Much of our pre-test data had been collected by November 6th (90% of all our data).

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

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

   Domains Studied: Literacy; Math

   Research Questions:

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

Description: We used ordinary least squares (OLS) regression models to compare differences between each of the three treatment conditions and a control group on outcomes measuring early literacy and math, while controlling for differences in baseline scores, age in months at post-test and family characteristics (e.g. parent education level, poverty status, months in preschool). The control group of students were used in each regression model. We conducted significance[8] testing to identify findings in which we could not reject the null hypotheses: that outcomes were not a result of the treatment condition. In other words, we used this method to determine if findings were more likely to be due to the treatment effects, or due to chance occurrence (e.g. Type I

[7] Weber data collection activities occurred later than other sites as a result of replacing Granite in the sample.

[8] Asterisks are used to indicate statistically significant findings based on P-values within our results: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$.

error). We reported the predicted means for each group based on OLS regression results as well as effect sizes in the body of the report and/or Appendices. **Effect sizes above the .26 threshold that are statistically significant are presented in bold throughout our findings.** This threshold was created as a way to benchmark the strength of our findings against those found in similar intervention programs. The purpose and use of effect size is described in more detail in **Appendix E** to assist with interpreting the results.

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

   Domains Studied: Literacy

   Research Questions:

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

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

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

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

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

Children's content knowledge scores can also be used to determine to what extent a child is ready to learn as they matriculate into kindergarten. From norm group referenced data, we can calculate if a child is performing at, above or below their normative counterparts at the age of entry into kindergarten. If a child scores below the average content knowledge for their biological age, they might be at risk of struggling in

kindergarten. The risk goes up for students who performed far below their age equivalent norm group comparisons. For example, if a child is performing at 90% of his/her normed achievement level, he/she is closer to the content knowledge target than a child performing only at 60 or 70%. We examined the differences in growth from pre-to-posttest among the four experimental groups and for children who we identified as at-risk (e.g. content knowledge age was 80% or below their biological age ranges).

3. **Method:** Descriptive analyses

   Domains Studied: Interpersonal and Self-Regulatory Skills

   Research Questions:

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

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

Appendix C. Cohort Sampling

Cohort 1 High-Quality Preschool Site Sampling
ETI received a list of High-Quality Preschool Program sites (public and private) and a separate list of UPSTART participants. ETI reviewed the public and private high-quality list and selected three preschool providers from each to include in our Year One sample. ETI selected three of the larger districts to participate in 2016-2017: Salt Lake City, Granite, and Davis. However, the final sample consisted of Salt Lake City, Weber, and Davis school districts after it was determined that logistical barriers in Granite would prevent us from completing data collection within a feasible timeline.

Table C1. Cohort 1 2016-2017 HQSR-E LEA's

LEA	Program Type
Davis Hope*	Public
Granite	Public
Iron	Public
Jordan	Public
Murray	Public
Salt Lake City*	Public
Sevier	Public
Washington SPED	Public
Weber*	Public
Progressive Preschool*	Private
Head Start (SSL)*	Private
Head Start (CCH)*	Private
Centro de la Familia	Private

*Sites participating in the 2016-2017 evaluation

Cohort 2 High-Quality Preschool Site Sampling
ETI received a list of High-Quality Preschool Program sites (public and private) and a separate list of UPSTART participants. ETI reviewed the public and private high-quality list and selected five public preschool providers and four private preschool providers to include in our Year Two sample. ETI selected the five public preschool providers based on the large number of participating preschoolers available to test, their proximity to each other and those that were not using UPSTART.

Table C2. Cohort 2 2017-2018 HQSR-E LEA's

LEA	Program Type
Davis Community*	Public
Davis HOPE*	Public
Granite*	Public
Jordan*	Public
Murray	Public
Salt Lake City*	Public
Sevier	Public
Washington SPED	Public
Weber*	Public
CAP Head Start*	Private
Centro de la Familia	Private
Hilltop Christian*	Private
Mountainland HS*	Private
YMCA New Bridge*	Private

*Sites participating in the 2017-2018 evaluation

Student Sample

At the student level, ETI focused on treatment preschool aged students who were enrolled in high quality preschool sites, who were English proficient[9], and who were not identified with special needs that ETI could not accommodate during testing. Demographics for each sample are presented in **Tables C3 through C5**.

Table C3. Treatment-Control Sample Demographics, high-quality Public Pre-K

Category	Indicator	Public (N=108)		Control (N=108)	
		n	%	n	%
Race/ Ethnicity	% White	94	87%	90	83%
	% Hispanic	20	19%	19	18%
Gender	% Female	59	55%	61	57%
Preschool	% Attended preschool	108	100%	85	79%
Household	% Married	82	76%	86	80%
	% Under 200% poverty	80	74%	77	71%
	Average household size	5.01		5.48	
	Average household income	$46,731		$46,680	

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

Table C4. Treatment-Control Sample Demographics, UPSTART

Category	Indicator	UPSTART (N=125)		Control (N=125)	
		n	%	n	%
Race/ Ethnicity	% White	109	87%	105	84%
	% Hispanic	11	9%	22	18%
Gender	% Female	63	50%	64	51%
Preschool	% Attended preschool	83	66%	107	86%
Household	% Married	108	86%	100	80%
	% Under 200% poverty	99	79%	93	74%
	Average household size	5.97		5.61	
	Average household income	$46,280		$45,574	

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

Category	Indicator	Private (N=81)		Control (N=81)	
		n	%	n	%
Race/ Ethnicity	% White	57	70%	60	74%
	% Hispanic	22	27%	24	30%
Gender	% Female	45	56%	47	58%
Preschool	% Attended preschool	81	100%	73	90%
Household	% Married	55	68%	60	74%
	% Under 200% poverty	62	77%	48	59%
	Average household size	4.88		5.30	
	Average household income*	$33,302		$51,547	

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

Appendix D. Using and Interpreting Effect Sizes

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

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

Appendix E. HQSR-E Detailed Findings

High-quality Public Preschool Results

Table E1. Literacy Analysis of Treatment-Control Differences, High-quality Public Preschool

Literacy	Group	*n*	Predicted Mean	Δ	*ES*	*Statistical Significance*
Brigance Composite	Treatment (Public)	108	96.84	-.92	-.03	NS
	Control	108	97.77			
Brigance Letter Knowledge	Treatment	102	42.33	-.08	.00	NS
	Control	102	42.41			
PELI Oral Comprehension	Treatment	98	17.48	-.39	-.14	NS
	Control	98	17.86			
Bader Phonological Awareness	Treatment	131	1.79	-.94	-.26	*
	Control	131	2.73			

*Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; NS: Non-significant finding and indicates no meaningful difference between Treatment and Control groups and should not be interpreted as such.

Table E2. Math Analysis of Treatment-Control Differences, High-quality Public Preschool

Math	Group	*n*	Predicted Mean	Δ	*ES*	*Statistical Significance*
Composite	Treatment (Public)	97	22.47	-.12	-1.17	NS
	Control	97	23.63			

*Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; NS: Non-significant finding and indicates no meaningful difference between Treatment and Control groups and should not be interpreted as such.

UPSTART Preschool Results

Table E3. Brigance Literature Analysis of Treatment-Control Differences, UPSTART

Literacy	Group	*n*	Predicted Mean	Δ	*ES*	*Statistical Significance*
Brigance Composite	Treatment (UPSTART)	125	117.75	26.34	.78	***
	Control	125	91.40			
Brigance Letter Knowledge	Treatment	115	58.42	17.71	.60	***
	Control	115	40.70			
PELI Oral Comprehension	Treatment	122	18.07	.84	.29	*
	Control	122	17.23			
Bader Phonological Awareness	Treatment	145	6.51	4.20	.84	***
	Control	144	2.32			

*Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; NS: Non-significant finding and indicates no meaningful difference between Treatment and Control groups and should not be interpreted as such.

Table E4. Math Analysis of Treatment-Control Differences, UPSTART

Math	Group	*n*	Predicted Mean	Δ	*ES*	*Statistical Significance*
Composite	Treatment (UPSTART)	83	25.47	.37	.04	NS
	Control	83	25.10			

*Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; NS: Non-significant finding and indicates no meaningful difference between Treatment and Control groups and should not be interpreted as such.

High-Quality Private Preschool Results

Table E5. Brigance Literature Analysis of Treatment-Control Differences, High-quality Private Preschool

Literacy	Group	*n*	Predicted Mean	Δ	*ES*	*Statistical Significance*
Brigance Composite	Treatment (Private)	81	88.70	-3.98	-.12	NS
	Control	81	92.68			
Brigance Letter Knowledge	Treatment	83	42.90	4.84	.19	NS
	Control	83	38.05			
PELI Oral Comprehension	Treatment	77	16.84	-.38	-.14	NS
	Control	77	17.22			
Bader Phonological Awareness	Treatment	84	1.16	-.90	-.26	NS
	Control	84	2.06			

*Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; NS: Non-significant finding and indicates no meaningful difference between Treatment and Control groups and should not be interpreted as such.

Table E6. Math Analysis of Treatment-Control Differences, High-quality Private Preschool

Math	Group	*n*	Predicted Mean	Δ	*ES*	*Statistical Significance*
Composite	Treatment (Private)	78	19.54	-3.03	-.36	*
	Control	78	22.57			

*Note: $^{*}p<.05$, $^{**}p<.01$, $^{***}p<.001$ denotes statistical significance; NS: Non-significant finding and indicates no meaningful difference between Treatment and Control groups and should not be interpreted as such.

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

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

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