Rural UPSTART Preschool Study:

Preliminary Evaluation Results for Investing in Innovation (i3) Grant U411B130020

June 2016

Primary Authors:
Melanie Overby
L. Jon Hobbs

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

Contact for correspondence
Jon Hobbs, Ph.D.
jhobbs@eticonsulting.org

Suggested citation
Evaluation and Training Institute. (2016). *Rural UPSTART Preschool Study: Preliminary Evaluation Results.* Retrieved from http://eticonsulting.org/i3

Abstract

In 2013 the Waterford Institute was awarded a validation grant from the Investing in Innovation Fund (i3) to fund the validation of “Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART)”, a pilot project that uses a computer-based preschool program to develop the school readiness skills of preschool children in rural Utah. Researchers used a randomized control trial design to evaluate the impact of the UPSTART program in advancing children’s early literacy skills. Preschoolers in the experimental group were randomly assigned to the UPSTART Reading software program with a recommended use of 15 minutes per day, five days a week, while students assigned to the control group participated in the UPSTART Math software program with identical usage requirements. Two standardized tests, the Brigance Inventory for Early Development and the Preschool Early Literacy Indicators (PELI) were administered at baseline and a year later to assess early literacy skills: letter knowledge, phonological awareness, decoding, oral comprehension, visual and auditory discrimination, and oral language and vocabulary. Results from independent sample t-tests revealed that children in UPSTART Reading had higher scores on letter recognition, phoneme manipulation, phonological awareness, and ability to decode common words and pre-primer vocabulary compared to their counterparts enrolled in UPSTART Math. There were no differences between the two groups on subtests that measured visual or auditory discrimination, oral comprehension, vocabulary, and oral language. This report previews information to be submitted to the National Evaluation of i3 (NEi3) and provides the foundation for future impact analyses of the UPSTART preschool program.

Table of Contents

Introduction

The Investing in Innovation Fund (i3) provides funding to support local educational agencies and nonprofit organizations in their efforts to expand the implementation of innovative practices that have been demonstrated to have a positive impact on a variety of student outcomes, such as improving achievement, closing achievement gaps, and increasing high school graduation rates. The goal of i3 grants is to develop and expand promising practices that can be shared as best practices and can be scaled up to other communities based on proven success.

In 2013 the Waterford Institute was awarded an i3 grant to fund the validation of "Utah Preparing Students Today for a Rewarding Tomorrow (UPSTART)", a pilot project that uses a home-based education technology approach to develop the school readiness skills of preschool children in rural Utah. The grant is designed to validate the effectiveness of UPSTART as a promising educational practice for possible scale-up in other environments at a later time.

One of the requirements of the i3 program is that each grantee must conduct a rigorous and independent evaluation in order to (1) determine the impact of the intervention on educational outcomes, and (2) to assess the extent to which the intervention was implemented as intended. The Evaluation and Training Institute (ETI) was contracted to conduct a systematic evaluation of UPSTART's effectiveness in rural communities in order to help expand our understanding of what works in education and for whom and what contexts specific interventions are effective.

Report Purpose

This report presents early results from the randomized preschool study conducted over the 2014-15 school year. As they are part of an ongoing evaluation of the Rural UPSTART project, the findings presented here are preliminary and provide the foundation for future investigation of how the UPSTART program impacts children's emerging literacy in preparation for entry into kindergarten. Future reports will examine differential treatment effects, model additional predictor variables and other more nuanced comparisons of these data.

The subsequent sections of this report are organized as follows. First, we present a brief overview of research concerning literacy and school readiness and the needs of students in rural education settings. Next, we describe our evaluation research design and methods in evaluating the rural UPSTART preschool program. Lastly, we present baseline and post-test scores of treatment and control groups on literacy outcome measures and discuss the implications of our findings for future work on the Rural UPSTART evaluation project.

Review of Literature

School readiness is a multi-faceted concept that conveys important advantages to children. Student who enter kindergarten ready to learn get better grades, are more likely to graduate high school, and have a greater chance of entering successful careers as adults than children who are less ready when they begin kindergarten. Conversely, those children that start school developmentally behind other children tend to perform at lower levels later in school. Children's outcomes during preschool and early childhood predict their later academic, economic, and professional achievements (Currie, 2001; Heckman & Masterov, 2004).

Research has consistently linked emergent early literacy skills with later reading achievement. The National Early Literacy Panel conducted a large scale meta-analysis of early literacy research and found that in addition to conventional literacy skills such as decoding, oral reading fluency, and comprehension that have a clear and consistently strong relationship with later reading skills, there are also emergent literacy skills such as phonological awareness and alphabet knowledge that are also powerful predictors of later achievement (NELP, 2008). Similarly, an analysis of six longitudinal datasets indicated that early language skills such as vocabulary, letter knowledge, words, and word sounds were a consistent predictor of later reading achievement, along with math and attention skills. (Duncan et al, 2007).

School readiness is influenced by familial and environmental factors, and can be enhanced through effective preschool education (Ackerman & Barnett, 2005). Preschool interventions can positively influence children's academic skills (Mashburn et al., 2008, NELP, 2008) and evidence from randomized control trials suggests that the long-term benefits of high-quality early intervention include a reduction in both special education placement and grade retention (Ramey & Ramey, 2004). Educational technology interventions have shown merit in promoting school readiness with preschool children, and particularly those from socioeconomically disadvantaged families (Li, Atkins, & Stanton, 2006).

Children living in rural communities are more likely to live in low-income families compared to children living in urban areas (Addy, Englehardt, & Skinner, 2013) and over 9.7 million students are enrolled in rural school districts, which comprise 20% of all American public school students (Johnson et al., 2014). Children who live in rural environments may experience difficulties in accessing high quality preschool programs as rural schools are often limited in their course offerings, facilities, and resources due to their small student population, financial constraints, and large geographic size (Teigen et al., 2012). Underscoring this point, researchers at the Center on Enhancing Early Learning Outcomes report that about 15% of children in rural communities attend a high-quality pre-kindergarten program, compared to 30% of children in urban and suburban areas (Nores & Barnett, 2014).

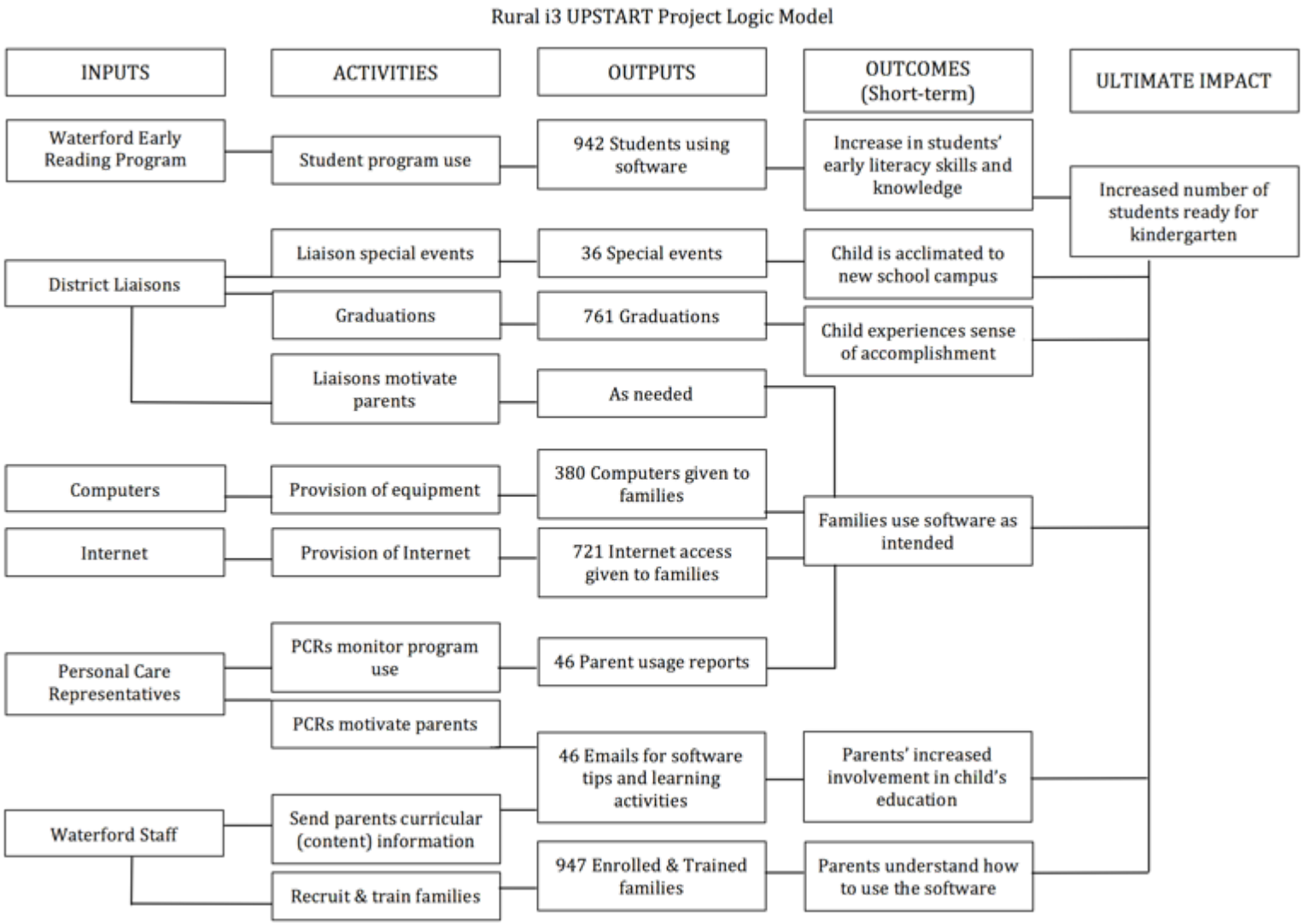
Rural UPSTART Project Description

UPSTART is a school readiness program that uses a home-based educational technology model well-suited to preschoolers and families residing in rural communities. The logic model shown in **Figure 1** illustrates the key components of the Rural UPSTART project, along with corresponding activities, resulting outputs, mediators or intermediate outcomes through which the intervention is designed to work, and the program's ultimate impact.

Below is a list of the key components of the logic model, each of which is described in further detail below:

- Waterford Early Learning software (WEL) provides adaptive early literacy curriculum via an online computer software program
- Waterford staff recruit and train families, and provide parents with weekly emails
- Waterford Personal Care Representatives monitor children's program usage and provide motivation to families
- School District Liaisons serve as UPSTART representatives at the local level
- Educational Technology, such as computers and internet access, is provided to families with financial need free of charge

Figure 1


Rural i3 UPSTART Project Logic Model
INPUTS
ACTIVITIES
OUTPUTS
OUTCOMES (Short-term)
ULTIMATE IMPACT
Waterford Early Reading Program
Student program use
942 Students using software
Increase in students' early literacy skills and knowledge
Increased number of students ready for kindergarten
District Liaisons
Liaison special events
36 Special events
Child is acclimated to new school campus
Graduations
761 Graduations
Child experiences sense of accomplishment
Liaisons motivate parents
As needed
Computers
Provision of equipment
380 Computers given to families
Internet
Provision of Internet
721 Internet access given to families
Families use software as intended
Personal Care Representatives
PCRs monitor program use
46 Parent usage reports
PCRs motivate parents
46 Emails for software tips and learning activities
Parents' increased involvement in child's education
Waterford Staff
Send parents curricular (content) information
Recruit & train families
947 Enrolled & Trained families
Parents understand how to use the software

UPSTART Waterford Early Learning software program

The primary component of UPSTART is the Waterford Early Learning (WEL) software program, an in-home preschool computer-based program that uses software to provide preschool children with reading, math, and science curriculum, with a focus on reading instruction. The program is designed to promote the development of literacy and math skills that will prepare young children for entry into school by providing an individualized learning experience that adapts to each child's skill level. Content is delivered online through adaptive lessons, digital books, songs, and activities.

While the complete Waterford Early Learning Program typically consists of curriculum in both reading and math/science, in order to craft a rigorous evaluation as required by i3 standards, the WEL program was split into two components: *Reading only* for children randomly assigned to the treatment group, and *Math/Science only* for children assigned to the control group. The format, delivery, and requirements of the software program were identical between the two groups, and the only difference consisted of curriculum content. **Table 1** showcases the reading domains and skills taught by the Waterford Early Learning Program at Level 1 of the curriculum[1].

Table 1
UPSTART Program Reading Domains and Skills

UPSTART Reading Domains	Level 1 Reading Skill
Phonics Systematically builds from not reading to confident reading at 90 words a minute	• Recognize *A* through Z, and *a* through *z* • Learn 10 letter sounds and 20 sight words to read 10 leveled readers • Spell child's name
Comprehension/Vocabulary Develops vocabulary and critical thinking skills through rich reading experiences	• Read along and understand nursery rhymes • Read along and understand alliterative books • Learn 255 target vocabulary words
Language Concepts Introduces concepts of written language (from letters and pictures to basic grammar)	• Understand print (left-to-right, letters, pictures, words, text) • Develop oral language skills (colors, shapes, numbers, sizes, etc.)
Phonological Awareness Develops awareness of individual sounds in words	• Break words into individual sounds (cat to (/*k*/ /*a*/ /*t*/) • Blend individual sounds into words (/*k*/ /*a*/ /*t*/ to cat) • Change a sound in a word to make a new word (cat to bat)

[1] Level One is the beginning point of the curriculum where the preschool child begins as a nonreader and is introduced to skills designed to teach the child to read. Levels range from one to three and the child is tested at the beginning of the program and placed in a level based on his or her performance. All information for this program was obtained from the program's web site (http://www.waterford.org).

Math curriculum covered by the mathematics portion of the UPSTART Waterford Early Learning program includes the content areas and skills presented in **Table 2**.

Table 2
UPSTART Program Math Domains and Skills

UPSTART Math Domains	Level 1[2] Math Skill
Numbers and Operations Teaches number recognition, place value, counting, and arithmetic computation	• Recognize, order, and write numbers 0 through 20 • Order, count, and sequence numbers to 100 by ones and tens • Use strategies to compare group size (more than, less than, or equal to)
Operations and Algebraic Thinking Teaches arithmetic computation	• Use objects, drawing, etc., to represent addition and subtraction • Add and subtract within 10, including solving word problems • Fluently add and subtract within 5
Measurement and Data Develops a foundational understanding of measurement, time, and money. Prepares students to analyze data.	• Compare, classify, and describe measurable attributes of objects • Use digital and analog clocks to tell time to the hour • Identify coins and their value
Geometry Teaches properties of shapes, positioning, and the identification of parts of regions or groups.	• Identify basic shapes regardless of their orientation and environment • Create composite shapes • Learn about shape positioning • Understand similarities and differences in 2- and 3-dimensional shapes

Program Usage Requirements
While the recommended use for either the Reading or Math/Science UPSTART program is 20 minutes a day for 5 days a week, children are required to use the program at home for **15 minutes a day, 5 days a week**.

The UPSTART program has several resources available to parents to assist them in meeting the usage requirement that are outlined in the next section.

[2] Level One is the beginning point of the curriculum where the preschool child is introduced to skills designed to teach the child to do basic math. Curriculum levels range from one to three and each child is tested at the beginning of the program and placed in a level based on his or her performance. All information for this program was obtained from the program's web site (http://www.waterford.org).

Waterford Support

In order to facilitate academic supervision and encourage usage of program software, families are provided with several forms of technical, motivational and curriculum support. Before beginning the program, all participating parents attend a comprehensive orientation where they review state kindergarten preparedness guidelines alongside UPSTART curricular content, discuss strategies for motivating children to use the program consistently, learn how to navigate the software, and review available resources.

For example, parents have access to a manager portal in the UPSTART software program where they can monitor children's usage on a daily basis, as well as review children's assessment scores and progress. In the Parent Manager portal, parents can also access enrichment materials and educational activities that can supplement offline learning outside of the UPSTART software.

Additionally, parents receive weekly newsletter emails that contain graphs of children's weekly usage and information about program usage, features of program curricular content, and suggestions for supplementary educational activities. Families receive either a Reading or Math/Science email, depending on their experimental condition, and emails are available in both English and Spanish.

Lastly, families receive direct telephone support from their assigned Waterford Personal Care Representatives (PCRs). These specialized staff members serve as partners with families to provide technical and motivational support. PCRs monitor children's program usage and contact parents if usage drops below the minimum requirement to provide individualized strategies to encourage consistent use. They also serve as primary contact if parents encounter technical difficulties or challenges when using the computer program.

District Liaisons

The UPSTART preschool program primarily consists of the Waterford Early Learning software, but it also includes district liaisons from each of the 18 participating school districts that represent the UPSTART program locally. Liaisons serve as a bridge between participating families, Waterford, and school districts in order to provide school readiness opportunities throughout the preschool year and to encourage program use. In addition to providing motivational support to families on an as-needed basis, district liaisons sponsor events at local elementary schools (e.g., tours of kindergarten classrooms, graduations, holiday parties) to create a seamless transition between UPSTART and the first year of kindergarten.

Provision of Technology

The UPSTART software is primarily delivered over the internet in a web-based format that allows access from any computer with high-speed internet that meets minimum operating requirements. However, if families do not possess the required technology,

grant funds allow for providing parents with computers, modems, wireless cards, and/or internet service to access UPSTART software.

Evaluation Study Design and Methods

The following section presents information about the research methods used to conduct the evaluation, including: the research questions and design, creation of treatment (UPSTART students) and control (non-UPSTART students) samples, instruments used to assess program outcomes, and data collection and analyses procedures.

Research Question

Our primary research question is as follows:

> Do pre-kindergarten children randomly assigned to receive the UPSTART Reading software program for one year have higher scores than their counterparts assigned to the UPSTART Math/Science program on measures of emerging literacy, such as phonological awareness, vocabulary and oral language, reading comprehension, and alphabet knowledge?

We hypothesized that if the UPSTART Reading program has no effect on improving early literacy skills, then we would expect children who participated in UPSTART Reading (treatment group) to perform at the same level as children who received the math/science program (control group) on measures of early literacy. Conversely, if UPSTART reading has an effect on improving early literacy, then children in UPSTART reading should perform significantly better than the comparison group on literacy outcomes.

Research Design

ETI adopted a randomized control trial research design to study the impact of the UPSTART literacy program on participants. The evaluation design is diagrammed below in **Figure 2**.

Figure 2
Evaluation Design – Preschool Study

Summer 2014			Summer 2015	
Pre-Test	Random Assignment	**UPSTART Reading Treatment**	Post-Test	Kindergarten
Pre-Test		**UPSTART Math Control**	Post-Test	

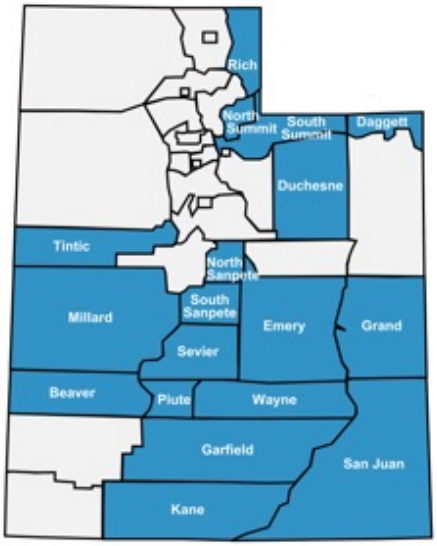
Children were recruited to participate in the evaluation in the early summer of 2014. Evaluation participants were administered two literacy instruments (Brigance IED III and PELI) to obtain baseline data and after this initial pre-testing was complete, UPSTART children were randomly assigned to either UPSTART Reading (treatment) or UPSTART Math/Science (control). Although the Math/Science group was the designated control group, children in this condition received computer-based curriculum designed to help

them learn early math and science skills and knowledge. Children in both groups completed their respective programs over a nine-month period and post-program literacy assessments were conducted after the UPSTART program was completed in Summer 2015, just before the children entered kindergarten.

Participant Recruitment and Random Assignment

Rural school districts in Utah were identified based on methodology developed by the Utah Foundation (Teigen et al., 2012) that examined the number of students in each district, the geographical size of districts, their distances from urban centers, and their proportions of Necessarily Existent Small Schools[3]. The eighteen districts classified as rural are presented in **Figure 3**.

Figure 3
Map of Utah's Rural Districts



In 2014 there were over 560,718 students in Utah's district schools. A majority of these students (53%) attended schools in five districts: Alpine, Davis, Granite, Jordan, and Canyons. The eighteen rural districts in the state had 31,979 enrolled students, or 5.7% of the state's students. To underscore this imbalance, each of the five largest districts had more students than the eighteen rural districts combined. Enrollment in rural districts has remained relatively stable over the past ten years, staying between 30,000 and 32,000 students.

[3] Schools meet necessarily existent small school standards if "one-way bus travel over Board approved bus routes from the school to the nearest school within the district of the same type requires students in kindergarten through grade six to travel more than 45 minutes, students in grades seven through twelve to travel more than one hour and 15 minutes." (Utah State Board of Education, R277-445-3). In addition to the distance requirement, schools must not exceed maximum enrollment thresholds (160 for elementary schools and 300-600 for secondary schools, depending on the number of grades served).

All eighteen rural school districts participated in the UPSTART program. Within each school district the program was offered to all age-eligible English-speaking children. Five of the smaller districts (Daggett, Piute, Rich, Tintic, and Wayne) were excluded from the preschool study due to insufficient sample size and geographical constraints.

All families who registered for the UPSTART program were contacted by ETI. During this initial interaction, we described the evaluation in detail, established whether or not the family was interested in participating, conducted a screening to determine evaluation eligibility, and if the family was eligible, assigned the parent and child to a testing session. Program families were excluded from the evaluation if their child did not speak English, had a diagnosed learning disability, or if they could not make the scheduled testing times.

During the summer months of 2014 we collected pre-test data from 549 students in 13 rural school districts across Utah, sampled from a total program group size of approximately 655 eligible families (roughly 84% of the total program group size at the start of testing).

All program participants were randomly assigned to the treatment or control condition at the end of the pre-test data collection period. Families learned of their group assignment during the UPSTART software training, an event that took place in each district at the start of the program and after all evaluation pre-testing was completed.

Participants' group membership status was kept separate from all post-test data collection activities and was only incorporated into the data file after all data had been recorded

At the beginning of the summer of 2015, we contacted the 549 families who participated in the evaluation and completed baseline pre-testing. We reminded families about their previous participation in the UPSTART evaluation, and schedulers reviewed evaluation goals and outlined post-test data collection procedures. If families were interested and available, they were assigned an appointment for a post-test session. Post-test data collection lasted from June-August 2015 and test administrators were blinded from children's group membership. A total of 497 families completed both pre- and post-tests. We provide a detailed discussion of the attrition rate in the next section of the report.

Test administrators were unaware of children's membership in either the treatment or control group during pre-test and post-test data collection, and data was collected from both groups during the same time periods and with identical procedures.

Data Collection Procedures

Assessments were individually administered to children by trained test administrators and were held at central locations within each district (e.g., elementary schools, libraries, community centers). In order to provide greater convenience for families, multiple testing locations were used in districts with populations spread across a large geographical area. All assessments utilized active parent informed consent protocols

and families were given a $40 gift card as an incentive for their child's participation. The entire assessment procedure was completed in 30-40 minutes on average.

Test Administrator Monitoring

Test Administrators participated in a comprehensive full-day training on adhering to evaluation protocols and administering standardized instruments in 2014 and a webinar review in 2015 conducted by senior project staff. All trainings used materials developed by assessment creators (e.g., instrument training videos, PowerPoint presentations, sample record forms, scoring rubrics, etc.) and incorporated mock practice assessments into the curriculum. At the conclusion of the training, test administrators conducted practice exercises under authentic testing conditions to establish reliability.

ETI senior staff members observed each test administrator during his or her first day of post-testing and used a structured observational protocol to provide specific feedback about their overall performance and administration of each subscale. At the end of the observation day, all test administrators were implementing the testing protocol with fidelity. Additionally, we regularly checked in with assessment staff during their testing period to answer any questions. Where possible, we reviewed the test administrator's data as it was entered into a database to check for any signs of scoring or coding related errors.

We engaged in a thorough data cleaning process after data were entered by test administrators, which included examining responses to be certain that test administrators were correctly scoring instruments and reviewing all qualitative scoring on the PELI oral language subscale to ensure scores were aligned to the rubrics developed by the test creators.

Minimizing Attrition

Participant loss, or attrition, is an issue that all multi-wave studies face. While statistical techniques can be employed to correct attrition bias, it is preferable to minimize sample attrition when conducting longitudinal research. We employed a variety of strategies to prevent attrition from impacting the external and validity of the research study. These strategies included offering separate incentives for participating in the pre-test and post-tests, providing multiple testing sites within districts to reduce participants' traveling burdens, scheduling participants throughout the day, and using a variety of platforms to contact participants and remind them of testing appointments (e.g., phone calls, text messages, emails, and social media postings). Additionally, because we have a network of assessment staff testing in locations across the state, we were often able to test families who had moved to other school districts.

Our efforts to minimize attrition were greatly aided by program staff at the Waterford Institute. Waterford staff were able to provide updated contact information for evaluation participants as families were actively engaged in using the software and personal care representatives had a history of regular contact with families via email and phone calls. District liaisons served as another resource for communicating with evaluation participants. Liaisons were embedded within the rural communities, had longstanding

relationships with UPSTART families, and were available to place phone calls on ETI's behalf or even conduct home visits to locate hard to reach families.

The attrition rates for the UPSTART preschool evaluation are presented in **Table 3**. Selective attrition can negatively affect the internal validity of a study if there are differential attrition rates between the treatment and control groups (Miller & Hollist, 2007). Nine percent of the total sample did not remain in the study, which was comprised of 10% of treatment families enrolled in Reading withdrawing from the study and 9% of control families enrolled in Math not continuing their participation. As the dropout rates between the two groups are comparable, the threats to internal validity due to attrition are minimal.

Table 3
Number of Participants in Evaluation Sample

Group	Reading	Math	Total Sample
Children Pre-Tested	285	264	549
Children Post-Test	257	240	497
% Attrition	**10%**	**9%**	**9%**

Families' reasons for not participating in post-testing included moving out of the testing area, being unable to travel to testing appointments due to pregnancy or having newborn children, or not responding to repeated requests from ETI and Waterford staff members.

Instruments

In the Rural UPSTART i3 evaluation, the outcomes of interest are measures of early literacy skills that are aligned to the UPSTART curriculum and considered to be important predictors of later reading ability, such as phonemic and phonological awareness, letter recognition, vocabulary, and oral language comprehension. We collected outcome data for the UPSTART preschool evaluation with two standardized instruments, the Brigance Inventory of Early Development (Brigance, 2013) and the Preschool Early Literacy Indicators (Dynamic Measurement Group, 2014).

Brigance IED III Literacy. The Brigance Inventory of Early Development III (Brigance) was selected as an early literacy measure of phonics, language concepts, and phonological awareness. Eight Brigance scales were administered to children from the academic skills/cognitive domain of literacy: alphabet knowledge, phonological awareness, auditory discrimination, phoneme manipulation, survival sight words, and pre-primer vocabulary. A composite Brigance score to create a comprehensive score of early literacy achievement was created by adding the scores from the eight child-administered subtests. Possible scores on the Brigance composite range from a low of 0 points to a high of 170 points.

PELI. The Preschool Early Literacy Indicators (PELI), designed by the creators of the DIBELS Next (Good et al., 2010), is a storybook-embedded assessment of essential

pre-literacy and oral language skills needed for kindergarten (Kaminski et al., 2014). The assessment is designed for preschool and pre-kindergarten students (ages 3-5) and measures Alphabet Knowledge, Vocabulary and Oral Language, Phonemic Awareness, and Comprehension. Potential scores on the PELI instrument range from from 0 to 99 points. Because the Alphabet Knowledge subscale in the PELI was similar to the Uppercase Letter Knowledge subscale in the Brigance, we administered the latter subscale only in an effort to reduce preschoolers' testing time.

Table 4 presents the subscales from the PELI and Brigance instruments used to measure our domains of interest and curricular content areas of the UPSTART Reading curriculum: phonics, comprehension/vocabulary, language concepts, and phonological awareness.

Table 4
Evaluation Measure Subscales

Brigance III	PELI
• Recites Alphabet • Uppercase Letter Knowledge • Pre-primer vocabulary • Survival Sight Words • Auditory Discrimination	• Alphabet Knowledge
	• Expressive Vocabulary • Oral Language • Oral Comprehension
• Visual Discrimination • Experiences with books and text	
• Phoneme Manipulation • Phonological Awareness	• Phonological Awareness

Sample

One practical issue that frequently happens in randomized control trials is noncompliance by the treatment group or incomplete data. For example, although the UPSTART computer program usage requirements are specified as 5 days a week for 15 minutes a day, not all participating families were able to use the program as intended, despite program staff's best intentions, and a few families withdrew from the

program altogether. This issue of noncompliance can threaten the integrity of randomization and the assurance that groups differ only with respect to the intervention. One potential solution to this problem is the use of a concept common in clinical research called *intention-to-treat* (ITT) sample, which includes all randomized participants, regardless of the treatment received, deviations from protocols, or withdrawal from the treatment group (Gupta, 2011). It is the most conservative approach and gives an unbiased estimate of treatment effects (Montori & Guyatt, 2001).

Table 5 illustrates the different sample compositions based on program usage. The Intention to Treat sample represents the most conservative assessment of the impact of UPSTART Reading and includes all participants with pre-test and post-test scores in the analysis, including program users and non-users. However, other researchers argue that if a participant is included in the treatment group, but did not actually receive treatment, it indicates little about the treatment's efficacy (Gupta, 2011). To that end, we report participant numbers from two other samples, based on program use. Program Users are a slightly smaller group and include all participants who used either the UPSTART Reading or Math program to some degree, regardless of amount. Finally, Program Graduates consist of participants who used the program according to usage requirements and are the most faithful adherents of the UPSTART program.

All results in this report will consist of the Intention to Treat sample, or all participants regardless of the amount of their UPSTART program use.

Table 5
Participants Based on Usage

Sample	Reading (Treatment)	Math (Control)
Intention to Treat	257	240
Program Users	251	234
Program Graduates	230	199

Data Analytic Approach

The general strategy for determining whether or not there was an impact of the UPSTART Reading program on preschoolers' literacy skills was to compare program Reading participants with their Math counterparts on post-test literacy scores collected at the beginning of kindergarten. Since our intent to treat, RCT research design framework remained intact throughout the study (low overall and differential attrition rates; both under the threshold set by WWC), this meant that our treatment and control groups were balanced by randomization and we did not have to compensate for imbalances during the analysis process. We chose the simplest data analytic model to test for group differences because it offered ease of interpretation to multiple audiences and more complicated models were not needed to contrast group outcomes.

Our data analytic approach assumes that treatment and control groups were equivalent at pre-test on factors that may influence emergent literacy skills measured at kindergarten, such as initial differences between the two groups (e.g., pre-test

achievement scores), as well as demographic factors that may differentiate between the treatment and control groups. If the treatment and control groups do not show statistically significant differences across variables effecting their post-test scores at the beginning of the UPSTART Reading program, then any observed differences can be reasonably attributed to participation program.

To test this assumption, the baseline equivalence of demographic characteristics and pre-program literacy scores between the experimental and control group were assessed using independent samples *t*-tests (for continuous variables) and chi-square tests (for categorical variables).

We hypothesized that if the UPSTART Reading program has no effect on improving early literacy skills, then children who participated in Reading (treatment group) would perform at the same level as children who received the Math/Science program (control group) on post-program measures of early literacy, or H_0: $\mu_T = \mu_C$ on post-test scores. Conversely, if UPSTART Reading has a positive effect on improving early literacy, then children in UPSTART Reading should perform significantly better than the comparison group on literacy outcomes, or H_a: $\mu_T > \mu_C$ on post-test scores.

Post-test differences on literacy outcome measures between the treatment and control groups were examined using independent sample *t*-tests. The statistical model used was:

$y_{i,j} = \mu_i + \varepsilon_{i,j}$ for i = Treatment (1), Control (0) and $j = 1, 2, \ldots n_i$ where $\varepsilon_{i,j} \sim N(0,\sigma^2)$

where each observation, j, from group i is is modeled as the sum of the group mean plus a random error term, typically modeled by a normal distribution with a mean equal to zero and a fixed variance σ^2.

We first conducted an exploratory analysis of our outcome variables of interest. We reviewed Brigance and PELI post-test score data to check for outliers, confirm that the normality assumption is met, and verify that there are mean differences between the two groups to justify further analyses. Our review determined that several outcome variables had non-normal distributions based on visual inspection of histogram plots. In addition, we found skewness and kurtosis absolute values of greater than 3.29 ($p <$ .001) and significant Shapiro-Wilk tests (Ghasemi & Zahedisl, 2012), suggesting that some of the data were not normally distributed. In order to check against potential errors estimating parameters when using parametric statistic with data that are not normally distributed (Leech & Onwueguzie, 2002), we analyzed the data with a non-parametric group comparison test, the Mann-Whitney test, and compared the results to parametric models. In cases where the models yielded different results we noted the findings.

Equivalence of Baseline Characteristics of Treatment and Control Group

Although the adoption of random assignment with participants helps to minimize pre-existing differences between groups, it does not guarantee that groups are matched and equivalent. Consequently, it is important to review sample characteristics carefully even when using a randomized experimental research design. The equivalence of the treatment (UPSTART Reading) and control (UPSTART Math) groups were examined on the basis of the Brigance and PELI pre-test scores and on the basis of demographic characteristics.

Equivalence on Demographic Variables

Table 6 presents child-level and family-level demographics for evaluation participants based on the final sample of children who had both pre-test and post-test scores and who participated in the UPSTART program. For comparison purposes, information is displayed separately for preschool children in the reading (treatment) group and for those in the math (control) group.

Group equivalence of demographic characteristics was examined with a chi-square analysis. As evidenced in **Table 6**, the proportion of girls was slightly higher in the treatment group (52% for Reading group, 48% for Math group), but this difference did not reach statistical significance. Additionally, the majority of the students in both the treatment and control samples were White (92% and 93%, respectively), with the next most frequently mentioned group being Hispanic/Latino students (4% and 5%, respectively).

In addition to data about preschool children's demographics, information was also collected about family characteristics. A slightly higher proportion of Reading families than Math families were at or under 200% of poverty level (60% and 53%, respectively). However, additional statistical analysis revealed no significant differences in poverty designation between the two groups.

Table 6
Demographics for Reading and Math Group Participants

Demographic Characteristic	Reading (N=257)	Math (N=240)
Gender (%)		
Male	48	52
Female	52	48
Race/Ethnicity* (%)		
White	94	95
Hispanic	5	5
Native American	2	3
African American	1	<1
Pacific Islander	1	<1
Asian	1	0
Other	<1	0
Household Poverty Level (%)		
At or under 100%	16	14
At or under 185%	54	48
At of under 200%	60	53
Parental Education (%)		
Did not complete high school	4	2
High school graduate/GED	13	15
Some college	46	45
Bachelor's degree	30	32
Master's degree	7	5
Doctorate/MD/JD	1	1
No response	0	1
Parental Marital Status (%)		
Married	92	93
Divorced/Separated	4	3
Unmarried	4	2
No response	0	3

* Note: Percentage of responses is greater than 100% because respondents could indicate membership in multiple racial/ethnic groups.

Equivalence on Outcome Variables

In order to ensure that the treatment and control groups were statistically equivalent on our outcome measures, we conducted a preliminary series of *t*-tests between the two groups to identify any significant pre-existing differences on early literacy instruments.

Brigance. As shown in **Table 7**, there were no significant pre-program differences between preschoolers who were assigned to UPSTART reading and those who were assigned to UPSTART Math on the Brigance Literacy Composite Score. Moreover, preschoolers were equivalent on all eight subtests of the Brigance Literacy measure.

Table 7
Brigance Literature Pre-Test Analysis of Treatment-Control Differences

Brigance Literature Pre-Test Score	Group	N	Mean	Sig
Recites Alphabet	Treatment (Reading)	257	7.19	.761
	Control (Math)	239	6.95	
Visual Discrimination	Treatment (Reading)	257	11.23	.383
	Control (Math)	239	10.78	
Identifies Uppercase Letters	Treatment (Reading)	257	8.89	.537
	Control (Math)	239	8.39	
Phonological Awareness	Treatment (Reading)	257	4.34	.280
	Control (Math)	239	4.08	
Auditory Discrimination	Treatment (Reading)	257	6.25	.361
	Control (Math)	239	6.03	
Phoneme Manipulation	Treatment (Reading)	257	2.36	.849
	Control (Math)	240	2.33	
Common Signs	Treatment (Reading)	257	1.03	.671
	Control (Math)	239	.99	
Word Recognition	Treatment (Reading)	257	.25	.615
	Control (Math)	238	.20	
Brigance Literacy Composite	**Treatment (Reading)**	**257**	**54.25**	**.408**
	Control (Math)	**238**	**52.03**	

PELI. Preliminary results from pre-test data displayed in **Table 8** indicate that the treatment and control group children were essentially equivalent on the PELI Composite, as well as on each of the four subtests (e.g., Vocabulary and Oral Language, Comprehension, and Phonological Awareness).

Table 8
PELI Pre-Test Analysis of Treatment-Control Differences

PELI Pre-Test Score	Group	N	Mean	Sig
Identifies Uppercase Letters	Treatment (Reading)	257	8.89	.537
	Control (Math)	239	8.39	
Vocabulary and Oral Language	Treatment (Reading)	257	16.37	.670
	Control (Math)	240	16.65	
Comprehension Section	Treatment (Reading)	256	14.32	.553
	Control (Math)	240	14.04	

Phonological Awareness	Treatment (Reading)	257	4.42	.622
	Control (Math)	240	4.63	
PELI Composite	**Treatment (Reading)**	**256**	**44.11**	**.890**
	Control (Math)	**239**	**43.86**	

Based on the randomized control design and our review of the demographic and baseline outcome variables for the treatment and control groups, we believe that the requirement for equivalent samples has been met and that any observed differences on the post-test between the treatment and control group can be credited to participation in UPSTART Reading.

Treatment and Control Group Differences on Outcomes

Having demonstrated that the Reading and Math Groups were statistically equivalent on demographic and baseline indicators, we proceeded to examine post-test scores to determine if UPSTART Reading has an impact on children's emerging literacy. If UPSTART has no effect on improving early literacy skills, then the children who participated in UPSTART Reading during their preschool year would be expected to perform at the same level as the comparison group on measures of early literacy prior to entering kindergarten.

In order to check the effect of non-normal distributions on our significance tests, we ran non-parametric models and compared the parameters to our parametric models. Parametric and non-parametric significance testing results were the same for all but one outcome, Phonological Awareness as measured by the Brigance (subscale assesses blending, segmenting, and rhyming skills). UPSTART Reading participants' phonological awareness scores were not significantly different from controls using the parametric test (p = .074), but were significant using our non-parametric test (p = .030).

We present t-test results for each subscale on the Brigance and PELI, as well as results from composite scores. Effect size[4] estimates are also included for all post-test differences between the two groups as recommended by the American Psychological Association publication manual and to evaluate our results across other designs, samples, and analyses (Wilkinson & Task Force, 1999).

Brigance

As seen in **Table 9**, there were significant post-program differences between preschoolers who were assigned to UPSTART Reading and those who were assigned to UPSTART Math on the Brigance literacy composite score, with UPSTART Reading children outperforming their Math counterparts. Moreover, children assigned to UPSTART Reading scored significantly higher than students in UPSTART Math on six of the eight Brigance subtests: Recites Alphabet, Identifies Uppercase Letters, Phonological Awareness, Phoneme Manipulation, Reads Common Signs, and Word Recognition.

[4] Effect size (Cohen's d) was calculated as the difference between treatment and control means divided by their pooled standard deviations.

Table 9
Brigance Literature Post-Test Analysis of Treatment-Control Differences

Brigance Literature Post-Test Score	Group	N	Mean	SD	t	df	Sig
Recites Alphabet	Treatment (Reading)	257	19.18	9.17	2.68	495	.008
	Control (Math)	240	16.86	10.14			
Visual Discrimination	Treatment (Reading)	257	16.41	3.53	0.50	495	.615
	Control (Math)	240	16.26	3.07			
Identifies Uppercase Letters	Treatment (Reading)	257	20.18	7.48	5.59	495	.000
	Control (Math)	240	16.13	8.68			
Phonological Awareness	Treatment (Reading)	257	6.46	2.56	1.79	495	.030[a]
	Control (Math)	240	6.04	2.58			
Auditory Discrimination	Treatment (Reading)	257	7.70	2.80	0.14	495	.888
	Control (Math)	240	7.67	2.50			
Phoneme Manipulation	Treatment (Reading)	257	4.02	1.56	3.30	495	.001
	Control (Math)	240	3.55	1.64			
Reads Common Signs	Treatment (Reading)	257	2.23	1.95	2.75	495	.006
	Control (Math)	240	1.74	1.97			
Word Recognition	Treatment (Reading)	257	3.31	3.98	5.53	495	.000
	Control (Math)	240	1.49	3.28			
Brigance Literacy Composite	**Treatment (Reading)**	257	**79.49**	**22.50**	**4.68**	**495**	**.000**
	Control (Math)	240	**69.74**	**23.91**			

[a] two-sample Wilcoxson rank-sum (Mann-Whitney) test

Effect sizes were calculated to show the magnitude of UPSTART Reading's impact as measured by the Brigance subtests and the Brigance composite score. The effect size estimates are presented in **Table 10** and show the magnitude of the average performance difference in standard deviation units between the treatment group and the control group on each of the post-program Brigance subtests. Overall, UPSTART Reading produced small to medium-size impacts on enhancing preschool children's emergent literacy, with the largest effects in recognizing uppercase letters and recognizing/decoding pre-primer vocabulary words.

Table 10
Brigance Post-Test Effect Size Estimates

Brigance Post-Test	Effect Size	Significance	Magnitude
Recites Alphabet	0.24	**	Small
Visual Discrimination	0.05	NS	
Identifies Uppercase Letters	0.50	***	Medium
Phonological Awareness[a]	0.16	*	
Auditory Discrimination	0.01	NS	
Phoneme Manipulation	0.29	***	Small
Reads Common Signs	0.25	***	Small
Word Recognition	0.50	***	Medium
Brigance Literacy Composite	**0.42**	***	**Small**

*p<.05, **p<.01, ***p<.001
[a] two-sample Wilcoxson rank-sum (Mann-Whitney) test

PELI

Initial results from post-test data displayed in **Table 11** indicate that the treatment children scored significantly higher than the control children on the PELI composite. In particular, children enrolled in UPSTART Reading outperformed children assigned to UPSTART Math on Phonological Awareness and Alphabet Knowledge (as measured with the Brigance). There were no differences between the two groups on subtests measuring Vocabulary and Oral Language and Listening Comprehension.

Table 11
PELI Post-Test Analysis of Treatment-Control Differences

PELI Post-Test Score	Group	N	Mean	Std. Dev	t	df	Sig
Vocabulary and Oral Language	Treatment (Reading)	257	22.04	6.87	-.93	495	.355
	Control (Math)	240	22.59	6.21			
Comprehension Section	Treatment (Reading)	257	18.02	4.34	-.11	495	.913
	Control (Math)	240	18.06	3.62			
Phonological Awareness	Treatment (Reading)	257	10.33	4.84	2.77	495	.006
	Control (Math)	240	9.15	4.62			
PELI Composite	**Treatment (Reading)**	**257**	**70.57**	**18.54**	**2.90**	**495**	**.004**
	Control (Math)	**240**	**65.93**	**17.02**			

Effect sizes were calculated to show the magnitude of UPSTART Reading's impact as measured by the PELI subtests and the PELI composite score. The effect size estimates are presented in **Table 12** and show the magnitude of the average performance difference between the treatment group and the control group on the PELI post-test. In general, the effect sizes on the phonological awareness subscale and overall PELI composite would be considered small.

Table 12
PELI Post-Test Effect Size Estimates

PELI Post-Test	Effect Size	Significance	Magnitude
Identifies Uppercase Letters	0.50	**	Medium
Vocabulary and Oral Language	-0.08	NS	
Oral Comprehension	-0.01	NS	
Phonological Awareness	0.25	***	Small
PELI Composite	**0.26**	***	**Small**

*p<.05, **p<.01, ***p<.001

Summary and Discussion

Results from initial inferential statistics indicate that preschool children randomly assigned to UPSTART Reading outperformed their counterparts who were assigned to UPSTART Math/Science on the Brigance and PELI, two standardized measures of early literacy. Specifically, UPSTART Reading children had higher scores on subscales measuring letter recognition, phoneme manipulation, phonological awareness (measured by the PELI), ability to read common words and pre-primer vocabulary compared to children enrolled in UPSTART Math. There were no differences between the two groups on subtests that measured visual discrimination, auditory discrimination, oral comprehension, vocabulary and oral language, and phonological awareness (measured by the Brigance).

While these preliminary results point to the potential of UPSTART Reading in promoting the attainment of early literacy skills, additional research needs to delve deeper into the data. It is possible that the degree to which children utilized UPSTART Reading influences their resulting outcomes. Children were required to utilize the program for 15 minutes a day, 5 days a week. However, not all children met that standard, and some managed to exceed it. How does the level of UPSTART Reading usage relate to reading readiness outcomes?

Other important questions come to mind. For example, are there other predictors aside from participation in UPSTART reading that may impact literacy post-test scores, such as demographic characteristics or baseline literacy skills? Multilevel models may improve our predictive capabilities due to the hierarchical nature of the collected data, as students are nested within districts and may be more homogeneous and share similar experiences than if individuals were randomly sampled from a larger population. Multilevel modeling can help to untangle individual and group effects on the outcome of interest (Gelman, 2006). Finally, there could be other factors, such as a child's motivation to use UPSTART reading, parental support, or home literacy environment that may mediate or moderate the relationship between UPSTART Reading participation and early literacy outcomes.

Further research will address these questions in order to present a more detailed and nuanced assessment of the UPSTART Reading program.

References

Ackerman, D. J., and Barnett, W. S. (2005). *Prepared for kindergarten: What does "readiness" mean?* New Brunswick, NJ: National Institute for Early Education Research. Retrieved from http://nieer.org/resources/policyreports/report5.pdf.

Addy, S., Englehardt, W., & Skinner, C. (2013) *Basic facts about low-income children*. New York, NY: National Center for Children in Poverty. Retrieved from http://www.nccp.org/publications/pdf/text_1074.pdf

Brigance, A. H. (2013). *Brigance Inventory of Early Development III* (IED-III) (3rd ed.). N. Billerica, MA: Curriculum Associates.

Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives*, *15*, 213- 238.

Duncan, G., Dowsett, C., Classens, A., Magnuson, K., Huston, A., Klebanov, P., Pagani, L., Feinstein, L., Engel, Brooks-Gunn, J., Sexton, H., Duckworth, K and Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428-1446.

Dynamic Measurement Group (2014). *Preschool Early Literacy Indicators*. Eugene, OR: Author.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48* (3), 432-435.

Good, R. H., Kaminski, R. A., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., Stollar, S., & Wallin, J. (2010). *DIBELS Next*. Eugene, OR: Dynamic Measurement Group. Available at http://dibels.org/.

Ghasemi, A., Zahediasl, S. (2012), Normality test for statistical analysis: A guide for non- statistician. *International Journal of Endocrinology & Metabolism*, *10*(2), 486-489. Retrived from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3693611/#A3505R11

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

Heckman, J. J. & Masterov, D. V. (2004). *The productivity argument for investing in young children*. Committee for Economic Development. Retrieved from http://jenni.uchicago.edu/human-inequality/papers/Heckman_final_all_wp_2007-03-22c_jsb.pdf

Johnson, J., Showalter, D., Klein, R., & Lester, C. (2014). *Why rural matters 2013-2014:*

The condition of rural education in the 50 states. Washington, D.C.: Rural School and Community Trust. Retrieved from http://www.ruraledu.org/user_uploads/file /2013-14-Why-Rural-Matters.pdf

Kaminski, R. A., Abbott, M., Bravo Aguayo, K., Latimer, R., Good, R. H. (2014). The Preschool Early Literacy Indicators: Validity and benchmark goals. *Topics in Early Childhood Special Education*, *34*(3), 71-82.

Leech, N. L., & Onwuegbuzie, A. J. (2002, November). *A call for greater use of nonparametric statistics.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Chattanooga, TN. Retrieved from http://eric.ed.gov/?id=ED471346

Li, X., Atkins, M. S., & Stanton, B. (2006). Effects of home and school computer use on school readiness and cognitive development among Head Start children: a randomised controlled pilot trial. *Merrill- Palmer Quarterly*, *52*(2):239- 263.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749.

Miller, R. B., & Hollist, C. S. (2007). *Attrition Bias*. Faculty Publications, Department of Child, Youth, and Family Studies. Paper 45. Retrieved from http://digitalcommons.unl.edu/famconfacpub/45

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

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Available at http://www.nifl.gov/earlychildhood/NELP/NELPreport.html

Nores, M., & Barnett, W. S. (2014). *Access to high quality early care and education: Readiness and opportunity gaps in America*. CEELO Policy Report. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from http://ceelo.org/wp-content/uploads/2014/05/ceelo_policy_report_access_ quality_ece.pdf

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly*, *50*(4), 471-491.

Stanton, B. (2006). Effects of home and school computer use on school readiness and cognitive development among Head Start children: A randomized controlled pilot trial. *Merrill- Palmer Quarterly*, *52*(2), 239- 263.

Teigen, S., Hershey Kroes, S, Lyon Cotti, M., Wald, S., & Merrill (2012). *Reaching for Educational Equity: An evaluation of Utah's rural schools*. Salt Lake City, UT: Utah Foundation. Retrieved from http://www.utahfoundation.org/img/pdfs/rr712_complete_evaluation.pdf

Wilkinson & Task Force on Statistical Inference, APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594604.

The **Evaluation and Training Institute (ETI)** is a non-profit 501(c)3 consulting firm in Los Angeles. In its 40th year, ETI has directed studies ranging in scope from $1,500 to $1,000,000, covering educational, organizational, and social policy issues, and targeting the local, state, national and international levels. ETI has successfully completed third-party evaluations of educational programs, including those that target early literacy, curricular development and improvement, school readiness, school reform, K-12 learning, after/before school programs, community-based organizations, and professional organizations.

Using information to improve communities