Mississippi UPSTART and Head Start Evaluation

Kindergarten Impact Study

Submitted September 2, 2020

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

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

(This page intentionally left blank)

Table of Contents

Executive Summary

With a notable portion of Mississippi's preschool-aged children living in poverty or with other related risk factors, the need to mitigate potential negative outcomes through preschool educational intervention programs is clear. A recently developed partnership between The Mississippi Head Start Association (MHSA) and Waterford UPSTART aims to provide educational access and opportunity to every young child in the state. With a focus on providing services in early childhood development and education, MHSA is a critical player in the collaborative effort, while The Mississippi UPSTART Program (MSUP) is a home-based computer preschool program developed and provided by the Waterford Institute to prepare young children for school entry and future academic success.

This report is intended to document the result of this collaboration and, more specifically, MSUP's longer-term impact on kindergarten students' literacy and math skills. The current report focuses on the program's impact on this cohort of children nearly 2 years after the start of their preschool year.

Program Impact

Prior to entering the UPSTART program in the Fall of 2018, the preschool children were randomly assigned to either a reading program (treatment group) or a math program (comparison group). We hypothesized that if MSUP had a lasting effect on improving early literacy skills, then the kindergarten children who participated in the reading program (treatment) as preschoolers, would outperform the students participating in the math program (comparison) on measures of literacy. Similarly, we would expect kindergarteners who participated in the math program as preschoolers to outperform children in the reading program on measures of math.

We used the Kaufman Test of Educational Achievement, Third Edition (KTEA-3) to study reading and math outcomes. Results revealed that kindergarteners from the treatment group (reading program), scored higher on measures of literacy than did the comparison kindergarteners who participated in the math group. Similarly, the comparison children (math group) outperformed the treatment children at the end of kindergarten on measures of math skills. These significant differences suggest that the impact of the UPSTART preschool program was maintained through the spring of the kindergarten year.

In addition to our educational outcomes, we studied the continued effect of the program on social-emotional development (SED). SED outcomes were measured using the Social Skills Improvement System Rating Scales (SSIS RS), and we found that both treatment and control students had similar SED levels in kindergarten, and the majority of students in both groups were at or above national average SED levels.

These findings must be interpreted with caution, however, because the study is limited by the lack of highly related baseline measures to control for preexisting differences in learning achievement, and our final sample size was negatively impacted by lack of

participant response, complicated by the onset of a national health pandemic (COVID-19). The kindergarten study sample was representative of the program population, and there were no major demographic differences between the treatment and comparison groups. Approximately 96% of the entire sample was African American and a little less than half (41%) were 100% below the poverty level. About 68% of parents in the entire sample were never married.

Introduction

Landscape of MS Early Education

In the state of Mississippi, more than a third of pre-school aged children live in poverty and more than half receive food assistance (MCCB, 2018). Other factors, like geographic location, disability, foster care, and homelessness further increase the potential for negative outcomes. The state has defined children with one or more of these characteristics as vulnerable, at-risk and most in need of high-quality early childhood education programs (MCCB, 2018).

Funding for early childhood education in the state has continually evolved over the last two decades. The first models of early education funding focused on supplementing existing programs by providing training and mentorship opportunities to teachers and by helping to identify gaps in community resources. Excel by Five (created in 2005) and Mississippi Building Blocks (created in 2008) adopted community-based approaches to learning, like creating training and mentorship programs, all through private funding.

The state did not offer public pre-k until 2013 when the Mississippi state legislature implemented the Early Learning Collaborative Act. Existing preschool programs, partnered with a local public-school district, can apply to compete for state funding to become an Early Learning Collaborative (ELC). The intended purpose of ELC's is to increase Kindergarten readiness and increase the number of high-quality pre-k partnerships in the state. Data collected from students in the program have shown increased kindergarten readiness scores, thus demonstrating the effectiveness of the program. In 2018, all ELCs earned the average assessment scores and 78% of pre-k students were at or above the target scores. Funding for this program has steadily increased from $3 million funding eleven ELCs to $6.7 million to fund nineteen ELCs in 2019 (Wright, 2018). Despite increased funding each year since 2013, the program is only reaching about 4% of the state's 4-year-old population (Friedman et. al, 2018).

In January 2019, Mississippi received funding from the US Department of Education, *Preschool Development Grant Birth Through Five (PDG B-5)*, to further expand the network of early childhood education services across the state and increase the education and resources offered to families.

Given the objective to expand the pre-k network and further increase opportunities to serve vulnerable, at-risk populations, the recent partnership between The Mississippi

Head Start Association (MHSA) and The Mississippi UPSTART program (MSUP) aligns well with the state's mission to increase kindergarten readiness.

Mississippi UPSTART PROGRAM

The Mississippi UPSTART Program (MSUP) is a home-based computer preschool program developed and provided by the Waterford Institute to increase the school readiness skills of preschool children.

The UPSTART software uses adaptive lessons, digital books, animated songs, and activities to deliver individualized early literacy content. The reading skills taught by the Waterford Early Learning Program at Level 1 of the curriculum[1] include:

- Phonological Awareness
- Phonics
- Comprehension and Vocabulary
- Language Concepts

Children are encouraged to use the UPSTART program for 15 minutes a day, 5 days a week and families are provided with parental resources and technical support from Waterford customer service representatives.

The first year of collaboration between MHSA and MSUP was the 2018-19 school year where Waterford enrolled 684 preschool children and provided them with an adaptive program of computer-based early literacy instruction to prepare them academically for kindergarten. This first-year cohort participated in the preschool program from November 2018 through August 2019 and subsequently attended kindergarten during the 2019-2020 school year. This kindergarten study was conducted during the spring of their kindergarten year.

Evaluation Overview

The Evaluation and Training Institute (ETI), a non-profit research center, was hired as an independent, third-party evaluator to study the impact of the Mississippi UPSTART Program (MSUP) on children's literacy and math skill development. In order to assess the longer-term impact of the program, ETI recruited a sample of children who had participated in the UPSTART program during their preschool year and then tested their literacy and math abilities during the spring of their kindergarten year. The evaluation specifically sought to study how kindergarteners who were randomly assigned as preschoolers to either the reading program (treatment group) or the math program (comparison group) differed on measures of both literacy and math skills.

[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.

Research Questions

Our primary hypothesis was if MSUP had a longer-term effect on improving early literacy skills, then children at the end of kindergarten who participated in the reading program as preschoolers, would outperform the kindergarteners who participated in the math program as preschoolers on measures of literacy. We held a similar hypothesis for the comparison group and children's mathematical skills at the end of kindergarten. Additionally, with existing concern in early education communities regarding the impact of computer time on young children's social and emotional development, we've included an exploratory research question to better understand any potential influence that a computer-based educational program may have on these developing skills.

Program Impact Research Questions

Research Question 1: Do kindergarten children who received the UPSTART reading program during preschool, demonstrate higher scores on measures of **early literacy** a year following the end of the program compared to kindergarten children who received the UPSTART math program during preschool?

Research Question 2: Do kindergarten children assigned to receive the UPSTART math program in preschool have higher scores on measures of **early math skills** one year after the completion of the program compared to kindergarten children who were assigned in preschool to the UPSTART reading program?

Research Question 3: What impact (if any) does the participation in a computer-based educational program in preschool have on the social emotional development of kindergarteners?

In the impact analysis, the outcomes of interest were measures of early math and literacy skills such as phonological awareness, comprehension/ vocabulary, alphabet and language concept, and numeracy.

Methods

Research Design. The kindergarten impact study was implemented as a posttest-only repeated measures research design with nonequivalent groups and is diagramed below, where T stands for children who received the UPSTART preschool reading program, and C stands for a matched sample of Comparison children who participated in the UPSTART math program. The "X" indicates that the children either received the UPSTART reading or UPSTART math program prior to kindergarten. O1 and O2 indicate measurements taken at the beginning and end of preschool, and O3 was the measurement taken during spring of kindergarten.

T	Xr	O1	O2	O3
C	Xm	O1	O2	O3

The initial objective was to link the baseline test scores in preschool (O1 and O2) to the repeated measure at kindergarten; however, the Head Start agencies did not follow the original measurement design and the scores could not be linked. This resulted in a post-test only design. Because this introduced the possibility that group differences may be attributable to selection effects or differential dropout rates rather than a true treatment effect, we accounted for this with statistical controls using interaction effects in our ANOVA models.

Our two main research questions for the kindergarten impact study are as follows:

> D*oes UPSTART produce improvements in reading readiness through kindergarten?*
>
> - If the answer is yes, then we would expect to see:
> $T \geq C$ @ O3
> - If the answer is no, then we would expect to see:
> $T < C$ @ O3
>
> D*oes UPSTART produce improvements in math readiness through kindergarten?*
>
> - If the answer is yes, then we would expect to see:
> $C \geq T$ @ O3
> - If the answer is no, then we would expect to see:
> $C < T$ @ O3

Data Collection Procedures
Child assessments were individually administered to children by trained test administrator staff in locations convenient for the participating families[2]. Parents completed an informed consent prior to the administration of any tests with the children. Parents also completed a parent survey including a social skills questionnaire with respect to their child. Assessment sessions took approximately 30 minutes and participating families received a gift card as a reward for their time.

Outcome Measures.
In the Mississippi UPSTART evaluation, the outcomes of interest were measures of early literacy skills relevant to emerging readers such as early phonemic awareness, letter recognition, and reading comprehension (KTEA-3), along with basic mathematical concepts (KTEA-3) and social-emotional development (SSIS).

[2] Following the onset of the public health emergency caused by COVID-19, all test administration was conducted online virtually through video conferencing to prevent direct contact between the children and the test administrators.

KTEA-3

The Kaufman Test of Educational Achievement Third Edition (KTEA-3) is a widely used measure of student achievement (Lichtenberger & Breaux, 2010) and measures academic skills in reading, math, written language and oral language. The instrument has shown good psychometric properties with the overall reliability ranging from 0.87 to 0.95. The KTEA-3 includes age-appropriate early literacy scales including preschool-kindergarten aged children in the following domains:

- Phonological processing
    - Child blends word parts and word sounds into whole words; identifies rhyming word pairs; deletes beginning, ending, and middle sounds from target words; and identifies words that have the same ending sound as the target word
- Letter and word recognition
    - Child recognizes and identifies lower and uppercase letters, identifies letters that correspond to phonetic sounds, provides letter sounds for target letters, and reads basic pre-primer words.
- Reading comprehension
    - Child is presented with symbols and their meaning and is asked to "read" the symbols, child "reads" common signs, child points to a picture of a printed word.

All KTEA-3 subscales use assigned entry points based on the child's age and have discontinue rules based on the number of incorrect responses with the purpose of limiting the number of test items given to children to reduce testing time and prevent boredom or frustration. KTEA-3 raw scores are converted to standard scores by age, and the standard scores are used in our analysis. The outcome measures are not over-aligned with the intervention. Data collection procedures were identical for treatment and comparison children and outcomes were consistently defined for both groups.

SSIS RS

The Social Skills Improvement System (Gresham & Elliot, 2008) Rating Scales (SSIS) measure social skills. The measures can be used to identify specific social behavior acquisition (and performance deficits) to provide a baseline for post-intervention progress evaluation as well as to track progress. We used the SSIS parent survey to assess the following components that combine to form the social skills scale: communication, cooperation, assertion, responsibility, empathy, engagement, self-control. Parents completed the survey while their child was administered the KTEA-3 assessment.

Table 1 summarizes the alignment between the UPSTART curriculum and the literacy and numeracy constructs measured by the KTEA-3.

Table 1. Alignment of Outcome Measures with UPSTART Curriculum

Measure	Subtest	Phonemic Awareness	Comprehension/ Vocabulary	Alphabet & Language Concepts	Numeracy
KTEA-3	Phonological Processing	X			
	Letter & Word Recognition		X	X	
	Reading Comprehension		X		
	Math				X

In addition to our outcome measures, demographic and background data were collected either directly from the parents in a parent survey or in data transferred to ETI from Waterford's database of participating families. We collected data on the following independent variables: student-level covariates such as gender, parents' level of education, ethnicity, and household income.

Data Analysis. A summary of the variables examined in the kindergarten impact analysis is shown in **Table 2.**

Table 2. Variables in the Kindergarten Analysis of Reading and Math Achievement

Predictor Variables	**Data Values**
Group	Treatment vs. Control
Gender	Male vs. Female
Ethnicity	Asian Black/African American Caucasian/White Hispanic American Indian/Alaska Native Hawaiian/Pacific Islander Multi-Ethnic Other Unknown
Low Income	Yes/No
Attended Any Preschool (other than UPSTART)	Yes/No
Special Education	Yes/No
KTEA-3 Outcome Variables (Standardized)	**Data Values**
Letter-Word Recognition	0-51
Phonological Processing	0-50

Reading Comprehension	0-35
Math Concepts	0-22

Comparison of Observed Literacy Outcomes. Descriptive statistics were computed for each of the outcome scores, subset by treatment group. This enabled a comparative analysis of reading achievement trends at the end of kindergarten. The observed reading achievement outcomes for the UPSTART treatment group and the comparison group were then plotted graphically. T-tests were also computed to determine the statistical significance of the observed achievement differences between the treatment and comparison groups at each of the four outcome measures.

Analysis of Variance Analysis. The observed achievement trends are captured in the comparative analysis described above. However, because the study utilized a quasi-experimental design in which the treatment and comparison groups were not completely equivalent on all factors that influence reading achievement outcomes, a two-way analysis of variance (ANOVA) was calculated on participants' outcome scores. A two-way ANOVA was run (alpha set to .05) to examine the effect of treatment, gender, family poverty level and parent education level on outcomes in kindergarten. Main effects were studied along with interaction effects between the additional predictors.

Study Sample. Evaluation participants were drawn from five Head Start agencies across ten Mississippi counties (**Table 3).** Children in families who agreed to participate in the preschool program were randomly assigned to the treatment (reading) or a comparison condition (math). As seen in **Table 4,** these families were predominantly African American representing 96% of the sample and lower income with 41% below 100% of the poverty level.

Table 3. Participating UPSTART Head Start Agencies

Name of Agency	Counties Served
Friends and Children	Humphreys, Madison, Copiah
ICS	Noxubee
Washington County Opportunities (WCO)	Washington
AJFC Community Action	Wilkinson, Adams, Jefferson, Amite
Bolivar	Bolivar

Table 4. UPSTART Sample Usage Data separated into math and reading groups

Category	Indicator	Math (N=62)		Reading (N=50)		Total (N=112)	
		n	%	n	%	n	%
Race/ Ethnicity	% White	2	3%	1	2%	3	3%
	% African American	59	95%	48	96%	107	96%
Gender	% Female	36	58%	24	48%	60	54%
Language	% English	62	100%	50	100%	112	100%
Poverty Level	Poverty Level Below 100%	21	34%	25	50%	46	41%
	Poverty Level Below 185%	20	32%	12	24%	32	29%
Parent Marital Status	Married	11	18%	13	26%	24	21%
	Never Married	44	71%	32	64%	76	68%
Parent Educational Attainment	High School Diploma/GED	23	37%	15	30%	38	34%
	Some College	26	42%	23	46%	49	44%
Program Graduate	Graduated UPSTART	56	90%	42	84%	98	88%

Program Impact Results

In order to study the potential longer-term effects of the UPSTART program, we reviewed the reading and math scores of kindergarten students who participated in the UPSTART program in the 2018-2019 preschool year. All kindergarten students used either the reading program (treatment) or the math program (comparison). We hypothesized that if the UPSTART program had a lasting effect on improving early literacy skills, then students who participated in the reading program (treatment) would continue to outperform the students who participated in the math program (comparison) on measures of literacy in kindergarten. Similarly, we expected students who participated in the math program to outscore students who received the reading program on measures of math in kindergarten.

A two-way ANOVA was run on a sample of 112 participants to examine the effect of the treatment on outcome scores in the spring of kindergarten. There was a main effect for

treatment showing reading students scored significantly higher than comparison students on letter-word recognition, reading comprehension and phonological processing. In addition to our main research focus on literacy, an additional analysis of math learning showed that the comparison group of students (enrolled in a Math/Science program), significantly outscored the treatment children in math- a notable benefit to the comparison student group.

To provide additional context for our findings, we present effect sizes between treatment and comparison groups. An effect size (ES) takes the difference between two group means on an outcome variable and represents it in standard deviation units. Effect sizes describe the magnitude of the difference between two groups, and essentially create a standardized scale to facilitate results interpretation. Effect sizes that represent meaningful differences in a given context are referred to as having a practical or substantive significance. Following recommendations from the What Works Clearinghouse (WWC) (What Works Clearinghouse, 2017) and a meta-analysis of similar educational interventions and studies (Lipsey et al., 2012), we set an effect size threshold of .26 to denote effects that have practical significance and are substantively important. Appendix A provides greater detail on how the benchmark was determined.

The outcomes are based on KTEA-3 standard scores, which account for distance of the examinee's raw score from the mean raw score, taking into account the standard deviation (degree of variability) of the distribution of raw scores. Like a percentile rank, a standard score relates one person's performance to the performance of other students of the same age. Unlike percentile ranks, however, standard scores have the property that a particular size of difference between two scores represents the same amount of difference in the skill being measured regardless of where on the scale the scores fall. For example, the difference between standard scores of 75 and 90 has the same meaning as the difference between scores of 110 and 125.

Letter-Word Recognition

Treatment students scored significantly higher than comparison students on letter-word recognition (LWR), $F(1,110) = 4.57$, $p = .035$. Letter-word recognition is measured by the child's ability to identify a specified letter name, letter sound or correctly identify sight words[3]. Stronger letter and word recognition is a reliable predictor of later reading and spelling abilities (Piasta & Wagner, 2010). We found no effects from UPSTART treatment interacting with preschool graduation status, gender, poverty or parent education levels. For example, female and male treatment students both score higher than female and male comparison students on measures of LRW.

[3] High-frequency words that are generally learned first – e.g. "on, was, that, etc."

Table 5. Letter-Word Recognition Treatment and Comparison Standardized Mean Scores

	N	Mean	SD	ES
Tx	50	98.48*	10.27	**.41**
Cr	62	94.39	9.91	

*p<.05, **p≤.01,***p<.001

In addition to the significant main effect of treatment condition, the treatment effect size was .41, well above our predetermined threshold of .26, further indicating a substantive impact on letter-word recognition. On average, treatment students scored 4.09 points higher than comparison students demonstrating a critical skill of early literacy.

Reading Comprehension

Treatment students scored higher than control students on kindergarten tests of reading comprehension $F(1,110) = 4.16$, $p = .04$. In order to assess their reading comprehension skills, students were asked to derive meaning from printed symbols or signs, match words with pictures, and respond to command items or passages that test both their ability to read and understand what is written. Reading comprehension is critical to students' early literacy development as it involves both the capacity to read what an author has written and the child's own ability to use their current knowledge to understand what is written. Reading comprehension skills at an early age allow students to access a wider variety of media (e.g. print, tv, internet, etc.) and knowledge and think critically about both (Shanahan et al., 2010).

We found no effects from UPSTART treatment interacting with preschool graduation status, gender, poverty or parent education levels.

Table 6. Reading Comprehension Treatment and Comparison Standardized Mean Scores

	N	Mean	SD	ES
Tx	50	97.54*	15.99	**.39**
Cr	62	91.18	16.74	

*p<.05, **p≤.01,***p<.001

As shown in Table 6, students who used the reading program in preschool scored 6.36 points higher than comparison students on reading comprehension at the end of kindergarten. This difference was statistically significant and produced an effect size of .39.

Phonological Processing

Treatment students scored higher than comparison students on kindergarten measures of phonological processing, F (1,110) = 3.905, p = .05. When measuring their phonological processing skills, students are asked to listen to and manipulate whole words, word parts, and sounds as well as identify rhyming and non-rhyming words. Because phonological processing involves analyzing the relationship between sounds and letters, it is a strong indicator of early literacy performance and reading success. (Gillon, 2002). We found no effects from UPSTART treatment interacting with preschool graduation status, gender, poverty or parent education levels.

Table 7. Phonological Processing Treatment and Control Standardized Mean Scores

	N	Mean	SD	ES
Tx	50	92.74*	11.40	**.38**
Cr	62	88.46	11.39	

*$p<.05$, **$p\leq.01$,***$p<.001$

At the end of kindergarten, treatment students scored 4.28 points higher than comparison students on phonological processing. This difference was statistically significant and produced an effect size of .38 (see Table 7). See Appendix B for raw mean scores for the individual phonological processing scales.

Math Concepts and Applications

Comparison students scored higher than treatment students on kindergarten tests of math concepts and applications F(1,110) = 6.82, p = .01. To assess mathematical skills, students were asked to respond to basic concepts such as number identification, shape identification, and varying quantities. In Duncan et al's (2007) analysis of six longitudinal data sets the ability to grasp early math concepts proved an important predictor of later math achievement. At the end of kindergarten, students who used the math program in preschool scored 3.94 points higher than treatment students on math concepts and applications (**Table 8**). The math program had a statistically significant impact on mathematical skills and produced an effect size of .5.

Table 8. Math Concepts and Applications Treatment and Control Standardized Mean Scores

	N	Mean	SD	ES
Tx	50	72.16	10.67	**.5**
Cr	62	76.10*	4.69	

*p<.05, **p≤.01,***p<.001

Social Skill Results

For the purposes of this study, ETI utilized the SSIS Rating Scale (parent form) to measure children's social skills. With a combination of exposure to social and emotional skill development both through the traditional Head Start preschool initiatives and also the UPSTART intervention (treatment children), the vast majority of both treatment and comparison children scored average or above on measures of social and emotional development.

To help stakeholders review descriptive findings about social skill development, the following descriptions capture the seven different social skill subscales:

- *Communication*: child's ability to take turns and make eye contact during a conversation, using appropriate tone of voice and gestures, and being polite by saying "thank you" and "please."
- *Cooperation*: child's propensity to help others, share materials, and comply with rules and directions.
- *Assertion*: child's attempt to initiate behaviors, such as asking others for information, introducing oneself, and responding to the actions of others.
- *Responsibility*: child's ability to show regard for property or work and demonstrate the ability to communicate with adults.
- *Empathy*: child's ability to show concern and respect for others' feelings and viewpoints.
- *Engagement*: child's ability to join activities that are in progress and inviting others to join, initiating conversations, making friends, and interacting well with others.
- *Self-Control*: child's ability to respond appropriately in conflict (e.g. disagreeing, teasing) and non-conflict situations (taking turns and compromising).

As indicated in Table 9, more than 9 in 10 regardless of treatment condition demonstrated appropriate skill level indicating similar patterns in development across both groups. When comparing the overall scores across all subscales, nearly 60% of both treatment and comparison children performed at average social emotional skill level and more than a third demonstrated above average scores. While there was a nominal difference in overall scores in favor of the comparison children (98% vs 90%, respectively) we observed overwhelmingly similar results regardless of treatment

condition likely a result of simultaneous exposure for all the children as part of the Head Start program. Benchmark scores can be referenced in Appendix C.

Table 9. Level of Social Skill among Treatment (N=50) and Control (N=62) Students at Post-Test

	Below Average		Average		Above Average	
	TX	CR	TX	CR	TX	CR
Communication	2	1	32	37	16	24
	4%	2%	64%	60%	32%	39%
Cooperation	4	4	26	33	20	25
	8%	6%	52%	53%	40%	40%
Assertion	1	--	40	44	9	18
	2%	--	80%	71%	18%	29%
Responsibility	5	4	29	36	16	22
	10%	6%	58%	58%	32%	35%
Empathy	4	1	23	33	23	28
	8%	2%	46%	53%	46%	45%
Engagement	2	--	28	40	20	22
	4%	--	56%	65%	40%	35%
Self-Control	5	1	26	39	19	22
	10%	2%	52%	63%	38%	35%

Discussion

In the state of Mississippi, where a significant portion of the population of young learners are susceptible to educational hardships, the ability to offer access to high quality and effective programs is critical. Research continues to support the importance of early education, particularly critical in at-risk populations as a means to shore up academic skills and ensure vulnerable children's school readiness. The developing collaboration between MHSA and MSUP, allows at-risk and vulnerable families to gain access to both traditional preschool settings and the opportunity to supplement early learning at home with the UPSTART program. The partnership itself has the ability to unite those who need additional educational access with an online intervention that can support school readiness and academic success.

An extensive body of work underscores the long-term effects of early childhood education on positive educational outcomes. Research has documented the lasting benefits on 4-year-olds who have been part of high-quality, early learning settings (Campbell FA, Pungello EP, Miller-Johnson S, Burchinal M, Ramey CT, 2001; Reynolds et al., 2007). Studies have long documented that this early learning experience

becomes an increasingly important protective factor for children living in poverty (Barnett, 1998). There is also evidence to suggest that when the family and neighborhood environments are not able to adequately prepare children for school, early childhood becomes the relevant time for investment in educational programs (Duncan and Magnuson, 2003). With this early learning exposure contributing to a child's academic trajectory, creating an optimal experience becomes a critical factor for maximizing educational opportunities in communities that need it most.

After following kindergarten students from the beginning of their preschool year (2018) into the spring of their kindergarten year (2020), we found that the UPSTART program had a lasting effect on improving the early literacy skills of students who participated in the reading program and concurrently improved the early math skills of students who participated in the math program. Using a delayed posttest design, we observed these gains nearly 2 years after the start of the pre-K intervention. While encouraging, these results must be weighed against the study limitations.

Two significant limitations to the kindergarten study exist and are important factors when considering the results: preexisting differences between treatment and comparison students and sample attrition. First, because of an inability to connect baseline student learning measures at entrance into preschool with our kindergarten outcome measures, preexisting differences between treatment and comparison students cannot be ruled out. Unforeseen methodological limitations in the preschool baseline assessment by the Head Start agencies prevented measures taken during preschool from being linked to measures taken during kindergarten.

The second limitation is sample attrition due to parents' lack of response to kindergarten data collection requests, which was compounded with the onset of a health pandemic that prevented parents from meeting with researchers in-person[4]. A random selection of preschool program participants was not possible, which resulted in a convenience sample of parents who elected to participate in the kindergarten study. These differential dropout rates within our sample resulted in unequal treatment and comparison groups that may have differed depending on their motivation and reasons for participating in the study. To see if the group composition affected outcomes, we examined whether systematic group differences had an effect on student literacy and math outcomes (interaction effects) and found no significant differences by key demographic characteristics such as poverty level, parent education, preschool graduation status, or gender. The above limitations resulted in an observational research study without the control of a learning baseline or random assignment due to sample attrition. Given these factors, the study can be considered an observational, post-test only design.

Further Research

The findings from this study offer observational evidence that support the collaboration of the MHSA and MSUP. As Mississippi continues to fund early education initiatives it is imperative to conduct tightly controlled research that studies both short-term school

[4] The kindergarten data were to be collected starting in March, 2020, when the COVID-19/corona virus pandemic began its spread throughout the US.

readiness impacts of preschool programs, but also examines the potential sustained, longer-term benefit using more rigorous research methods.

There are several key recommendations coming from the current UPSTART kindergarten evaluation.

- Continue to examine the MHSA and MSUP collaborative effort and subsequent outcomes as more children are served in the state of Mississippi.
- As this program progresses, continue to look at both short-term and longitudinal data to understand the extent to which the UPSTART program is impacting the literacy and math skills of the children who participate.
- To improve the overall integrity of the research design, future evaluations should ensure appropriate measurement at all intended observations so that links between pretest and posttest can be made.

In order to adequately support children and families living in our most vulnerable communities in Mississippi and across the country, we look to programs that provide high quality, early learning opportunities and simultaneously depend on research to evaluate and build evidence for the programs that offer critical early educational support for our youngest learners.

References

Barnett, S. W. (1998). Long-Term Cognitive and Academic Effects of Early Childhood Education on Children in Poverty. *Preventive Medicine, 27*(2), 204-207.

Burts, D.C., Kim, D.H. (2014) Teaching Strategies GOLD Assessment System: Measurement Properties and Use. Dialog from the Field

Campbell FA, Pungello EP, Miller-Johnson S, Burchinal M, Ramey CT. The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. Developmental Psychology. 2001;37:231–242.

Claessens, A., Duncan, G., & Engel, M. (2009). Kindergarten skills and fifth-grade achievement: Evidence from the ECLS-K. *Economics of Education Review, 28*(4), 415-427. doi:10.1016/j.econedurev.2008.09.003

Cohen, J. (1988) Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Duncan GJ, Magnuson KA. Off with Hollingshead: Socioeconomic resources, parenting, and child development. In: Bornstein MH, Bradley RH, editors. Socioeconomic status, parenting, and child development. Mahwah, NJ: Erlbaum; 2003. pp. 83–106.

Friedman-Krauss, A. H, Barnett, W.S, Weusenfeld, G.G., Kasmin, R. DiCrecchio, N. Horowitz, M. (2018). The State of Preschool 2017. *National Institute for Early Education Research.*

Gillon, G. (2002). Phonological Awareness Intervention for Children: From the Research Laboratory to the Clinic. *The ASHA Leader, 7*(22), 4-17. doi:10.1044/leader.ftr2.07222002.4

Hardin, B. & Peisner- Feinburg,E. (2004). The Learning Accomplishment Profile-Third Edition (LAP-3). Lewisville, Nc: Kaplan Early Learning Company.

Hardin, B., Peisner- Feinburg,E., Weeks, S. (2005). The Learning Accomplishment Profile- Diagnostic Third Edition (LAP-3). Lewisville, Nc: Kaplan Early Learning Company.

Heroman, C., Burts, D. C., Berke, K-L., & Bickart, T. S. (2010). Teaching Strategies GOLD® objectives for development & learning. Washington, DC: Teaching Strategies LLC.

Ho, A. D., & Yu, C. C. (2015). Descriptive Statistics for Modern Test Score Distributions: Skewness, Kurtosis, Discreteness, and Ceiling Effects. *Educational and psychological measurement*, *75*(3), 365–388. doi:10.1177/0013164414548576

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman test of educational achievement, third edition.* Bloomington, MN: NCS Pearson.

Kreijns, K., Kirschner, P.A., Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior*. Volume 19, Issue 3, pages 335-353.

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.

Lipsey, M., Weiland, C., Yoshikawa, H., Wilson, S., & Hofer, K. (2015). Prekindergarten age cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis*, *37*, 296-313.

Lichtenberger, E. O., & Breaux, K. C. (2010). *Essentials of WIAT–III and KTEA–II assessment.* Hoboken, NJ: John Wiley & Sons.

Mississippi Community College Board (MCBB) (2018). Connected for Success: A Family Based Unified and Integrated Early Childhood System. Mississippi Community College Board.

Piasta, S., & Wagner, R. (2010, January). Developing Early Literacy Skills: A Meta-Analysis of Alphabet Learning and Instruction. Retrieved July 02, 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910925/

Reynolds AJ, Temple JA, Ou S-R, Robertson DL, Mersky JP, Topitzes JW, Niles MD. Effects of a school-based, early childhood intervention on adult health and well-being: A 19-year follow-up of low-income families. Archives of Pediatrics and Adolescent Medicine. 2007;161(8):730–739.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving Reading Comprehension in Kindergarten Through 3rd Grade* (pp. 1-7, Publication). What Works Clearninghouse.

What Works Clearinghouse. (2017). Procedures handbook (version 4.0). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Wright, C. M., (2018). Mississippi State Board of Education Strategic Plan 2016-2020. Mississippi: Mississippi Department of Education.

Appendix

Appendix A: Determining UPSTART Effect Size Benchmark

One way to assess the practical significance of an intervention is to compare its impact with effect sizes from similar evaluation studies – those that use analogous outcome measures, are evaluating a comparable intervention, or are evaluating interventions that target similar groups. Researchers at the Institute of Education Sciences (IES) reviewed 829 effect sizes from 124 education research studies conducted on K-12 students and reported an array of different effect size distributions that can provide insight into what constitutes a large or small effect relative to similar education evaluation studies (Lipsey et. al, 2012). They provide the following benchmarks to be used as normative comparisons:

- *Benchmark by outcome measure*. IES researchers looked at the type outcome measures (i.e., did researchers use a self-developed outcome measure, a general standardized outcome measure like an IQ test, or a subject-specific standardized outcome measure like a reading or math test) by grade level and found that the average effect size for education research studies evaluating elementary students with a standardized subject test (like the Brigance and Bader literacy tests) was .25. Average effect sizes were slightly higher for middle school students, but lower for high school students (.32 and .03, respectively)
- *Benchmark by intervention type*. Another metric for evaluating effect size was based on the type of intervention under investigation. Researchers sorted the interventions of reviewed studies into several broad categories (e.g., a whole school program, a teaching technique, a new instructional format, skill training, or an instructional program). The UPSTART program was closest to an instructional program, or "a relatively complete and comprehensive package for instruction in a content area like a curriculum or a more or less free standing program (e.g., science or math curriculum; reading programs for younger students; broad name brand programs like Reading Recovery; organized multisession tutoring program in a general subject area." (p. 35) The average effect size for research studies that evaluated a comprehensive instructional program such as UPSTART was .13. Larger effect sizes were found for interventions in the instructional component/skill training and teaching techniques and categories (.36 and .35, respectively).
- *Benchmark by intervention target*. A final yardstick to contextualize effect sizes focused on the targeted group of the intervention (e.g., individual students, small group, classroom, whole school, mixed.) that targeted individual students had average effect sizes of .40. Interventions that targeted individual students had the highest observed effect sizes, on average.

To determine a single benchmark, we took an average of the three different benchmarks (i.e., benchmark by outcome measure = .35; benchmark by intervention type = .13; and benchmark by intervention target = .40) and the resulting benchmark value was .26. This benchmark will be used to contextualize the effect sizes presented in this report and to aid the reader in determining the practical significance of the effect of UPSTART.

Appendix B: Phonological Processing Scales

Phonological Processing Raw Scores
In addition to the age-standardized phonological processing (PP) scores, we also analyzed scales that are combined to create the PP test. These scales do not have age standardized scores, so raw scores are presented (below). Raw scores are not comparable from subtest to subtest because subtests contain different numbers of items, are not equally difficult, have unique growth curves, and produce widely varying distributions of scores. Raw scores do not take into account age of the participant. These scores are shown to highlight trends in the raw PP data (letter word recognition, reading comprehension and math concept tests do not have multiple scales).

On average, treatment students tended to have higher raw scores than comparison students across all scales within phonological processing (see Figure 1). The greatest mean differences are evident in Blending, Deleting Sounds and Segmenting, notably skills targeted by the UPSTART intervention (see Table 10).

Figure 1. Phonological Processing Treatment and Control Scale Mean Scores by Subtest

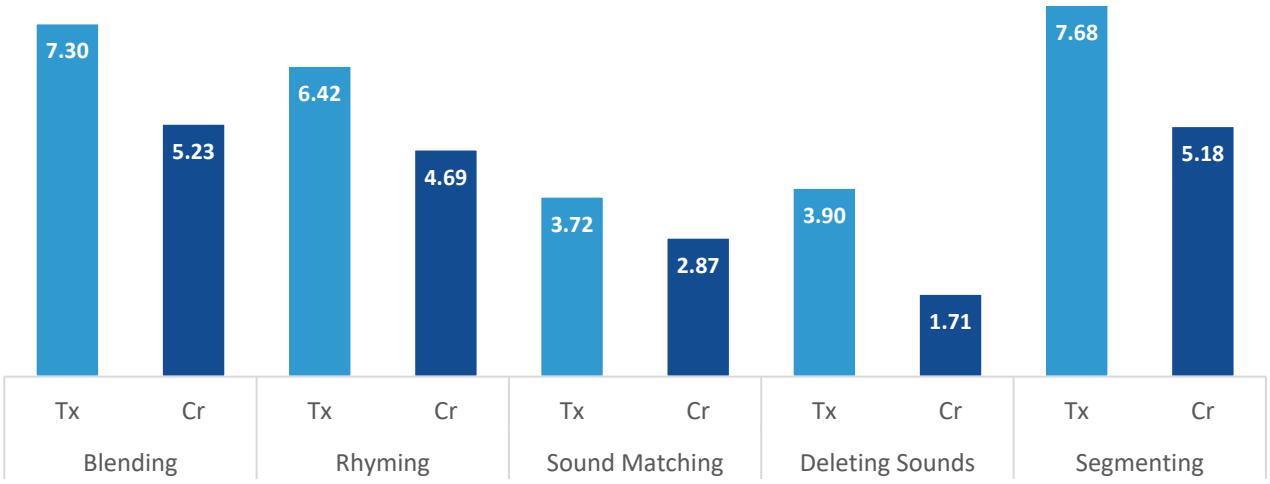

Table 10. Phonological Processing Treatment and Control Scale Mean Scores by Sub-domain Test

		N	Mean	SD
Blending	**Tx**	50	7.30	1.61
	Cr	62	5.23	2.32
Rhyming	**Tx**	50	6.42	2.57
	Cr	62	4.69	2.67
Sound Matching	**Tx**	50	3.72	1.80
	Cr	62	2.87	1.62

		N	Mean	SD
Deleting Sounds	**Tx**	50	3.90	9.01
	Cr	62	1.71	1.94
Segmenting	**Tx**	50	7.68	2.79
	Cr	62	5.18	3.36

Appendix C: SSIS Rating Scales

Behavior Levels Corresponding to Subscale Raw Scores for the Teacher and Parent Forms, Ages 3-5, by Norm Group *(pg.199 of SSIS Manual)*

Social Skills	**Below Average**	**Average**	**Above Average**
Communication	0 – 11	12 – 19	20 – 21
Cooperation	0 – 8	9 – 15	16 – 18
Assertion	0 – 10	11 – 19	20 – 21
Responsibility	0 – 7	8 – 15	16 – 18
Empathy	0 – 8	9 – 16	17 – 18
Engagement	0 – 10	11 – 19	20 – 21
Self-Control	0 – 6	7 – 14	15 – 21