

Literacy First

Evaluation Summary Report

Sarah Caverly, Ginger Stoker

JULY 2021

Literacy First

Evaluation Summary Report

JULY 2021

Authors:

Sarah Caverly, Ginger Stoker



American Institutes for Research®

1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | [AIR.ORG](https://www.air.org)

Notice of Trademark: "American Institutes for Research" and "AIR" are registered trademarks. All other brand, product, or company names are trademarks or registered trademarks of their respective owners.

Copyright © 2021 American Institutes for Research®. All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, website display, or other electronic or mechanical methods, without the prior written permission of the American Institutes for Research. For permission requests, please use the Contact Us form on [AIR.ORG](https://www.air.org).

Contents

Introduction	1
Background	1
Study Design.....	2
Research Questions	3
Data.....	4
Sample.....	4
Attrition.....	9
Sample Characteristics.....	10
Baseline Equivalence	10
Analysis	17
Findings	18
Short-Term Literacy Skills.....	18
Long-Term Literacy Skills	20
Mathematics	21
Interpretation	21
Summary	23
References.....	24
Appendix A. Data Tables	25

Exhibits

Exhibit 1. Student CONSORT Diagram for Kindergarten Outcomes	5
Exhibit 2. Student CONSORT Diagram for Grade 1 Outcomes.....	6
Exhibit 3. Student CONSORT Diagram for Grade 2 Reading Fluency and Reading Comprehension Outcomes (Cohorts 1 and 2)	8
Exhibit 4. Student CONSORT Diagram for STAAR Grade 3 Reading and Mathematics Outcomes (Cohort 1 Only)	9
Exhibit 5. Overall and Differential Attrition by Outcome for Kindergarten, Grade 1, and Grade 2 Students	10
Exhibit 6. Sample Descriptive Statistics and Effect Size Differences for Kindergarten Students	12
Exhibit 7. Sample Descriptive Statistics and Effect Size Differences for Grade 1 Students	13
Exhibit 8. Sample Descriptive Statistics and Effect Size Differences for Grade 2 Oral Reading Fluency.....	14
Exhibit 9. Sample Descriptive Statistics and Effect Size Differences for Grade 2 Reading Comprehension	15
Exhibit 10. Sample Descriptive Statistics and Effect Size Differences for STAAR Grade 3 Reading and Mathematics	16
Exhibit 11. Treatment Effects of Literacy First for Kindergarten Students.....	19
Exhibit 12. Treatment Effects of Literacy First for Grade 1 Students.....	19
Exhibit 13. Treatment Effects of Literacy First for Grade 2 Students.....	20
Exhibit 14. Treatment Effects of Literacy First for Grade 2 Students When in Grade 3	21
Exhibit 15. Treatment Effects of Literacy First for Grade 2 Students When in Grade 3	21
Exhibit 16. Effect Sizes and Improvement Indices	22
Exhibit A1. Impact Analysis Results for Letter Sound Fluency (LSF), Kindergarten Students.....	25
Exhibit A2. Impact Analysis Results for Kindergarten Decoding Fluency (KDF), Kindergarten Students	25

Exhibit A3. Impact Analysis Results for Nonsense Word Fluency (NWL/PS), Grade 1 Students	26
Exhibit A4. Impact Analysis Results for Oral Reading Fluency (ORF/FRO), Grade 1 Students	26
Exhibit A5. Impact Analysis Results for Whole Words Read (WWR/PCL), Grade 1 Students	27
Exhibit A6. Impact Analysis Results for Oral Reading Fluency (ORF/FRO), Grade 2 Students	27
Exhibit A7. Impact Analysis Results for Reading Comprehension (ITBS/Logramos), Grade 2 Students	28
Exhibit A8. Impact Analysis Results for STAAR Grade 3 Reading, Initial Grade 2 Students (Cohort 1 Only)	29
Exhibit A9. Impact Analysis Results for STAAR Grade 3 Mathematics, Initial Grade 2 Students (Cohort 1 Only)	29

Introduction

Literacy First is an AmeriCorps program that provides trained tutors for students in kindergarten through Grade 2 in the Austin, Texas, area. Literacy First's early literacy tutoring program is designed to strengthen students' early reading and comprehension skills (e.g., phonemic awareness, letter sound identification) through daily, 30-minute tutoring sessions. The program tracks student data, including benchmark assessments, weekly progress monitoring, attendance, and demographics, as well as program fidelity of implementation data from tutors. Literacy First is currently implemented in more than 25 elementary schools in the Austin area, serving more than 1,600 Grades K–2 students annually. Fifty percent of Literacy First's tutors are bilingual, allowing the program to provide support to both Spanish- and English-speaking students. The program identifies students reading below grade level each fall and provides tutoring across the school year. Literacy First has established itself as a program that supports highly trained volunteers to implement an intensive tutoring intervention that relies on data to guide instruction.

Beginning in the 2017–18 school year, the American Institutes for Research (AIR) partnered with Literacy First to conduct an independent evaluation of the impact of the Literacy First program on student outcomes. The study has three primary aims: (1) to estimate the impact of Literacy First on kindergarten and Grade 1 students' early reading skills, (2) to estimate the impact of Literacy First on Grade 2 students' short-term reading fluency and comprehension skills, and (3) to estimate the impact of Literacy First on Grade 2 students' long-term reading comprehension.¹ In addition, the study investigated whether participation in Literacy First had an impact on students' State of Texas Assessments of Academic Readiness (STAAR) Grade 3 Mathematics test scores.

Background

Throughout the past 2 decades, national attention has emphasized the critical role of early literacy and language instruction in preventing reading difficulties and improved our understanding of how young children learn to read (Connor et al., 2014; National Reading Panel & Institute of Child Health and Human Development, 2000; Raynor et al., 2001). The research has established that students who are not strong readers by Grade 3 are less likely to build vocabulary and interact with a wide variety of texts (Good et al., 2001). In addition, recent

¹ Spanish versions of these tests were administered to Spanish-speaking students who were not ready to complete the test in English.

national data indicate that 64% of fourth graders fail to reach proficient-level reading scores as measured by the National Assessment of Educational Progress, which is designed to measure students' reading comprehension (National Center for Education Statistics, 2015). A recent study by Jacob et al. (2016) demonstrated the impact of a volunteer tutoring program on Grade 2 students' reading comprehension skills. Despite these findings, few supplemental early literacy tutoring programs have conducted efficacy trials to demonstrate their potential impact on early literacy, comprehension, and language skills.

Literacy First has established itself as a unique tutoring program in three key areas. First, Literacy First tutors are highly trained volunteers who receive more than 70 hours of training in best practices, as well as weekly follow-up visits from Literacy First experts and coaches. Second, Literacy First is intensive; each child is seen daily for approximately 30 minutes and receives tailored tutoring to address literacy/language needs. Third, Literacy First uses data to drive instruction. Tutors use a response-to-intervention model, with benchmark assessments (which take place three times per year) and weekly progress monitoring.

Literacy First has invested in rigorous research to evaluate program impacts by implementing regression discontinuity design (RDD) studies across the past 3 years, using students' beginning-of-the-year reading ability to determine program eligibility. Findings from the evaluations provide strong evidence that Literacy First significantly accelerated students' reading skills across Grades K–2 in two Austin-area school districts and charter schools (Tackett et al., 2013; Tidd, 2014, 2015). Effect sizes found across the annual evaluations range from 0.20 to 0.40 (Tidd, 2014, 2015). In addition, a quasi-experimental study investigating the impact of Literacy First on students' reading skills found that first- and second-grade students who graduated from Literacy First scored statistically significantly higher than matched comparison students on the Developmental Reading Assessment (Agile Analytics, 2018). However, there are limitations to this research base: None of the studies used an experimental design, and the RDD and propensity score matching studies have the potential for sampling and selection bias.

Study Design

The evaluation employed a multisite randomized controlled trial to estimate the impact of Literacy First on Grades K–2 students' outcomes. The study was conducted within 22 elementary schools in the Austin Independent School District (AISD) in Texas. Each of the participating schools was oversubscribed and would not have been able to provide services to all eligible students. As a result, the evaluation study design did not cause a reduction in services provided to students. Each of the schools continued to serve the same number of students as it would have in the absence of the evaluation.

Following established procedures, all Grades K–2 students from participating schools were screened for Literacy First eligibility by the Literacy First tutors during the first week of September. Eligibility for participation in Literacy First tutoring was determined by students' scores on the beginning-of-year (BOY) assessments. Kindergarten students completed the Kindergarten Decoding Fluency (KDF) and Letter Sound Fluency (LSF) assessments, which were created by the program administrators. The KDF was created to measure decoding skills aligned with program content, whereas the LSF is a modified version of the AIMSweb LSF assessment. Grade 1 students completed the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Nonsense Word Fluency (NWF), DIBELS Oral Reading Fluency (ORF), and Whole Words Read assessments (WWR). The WWR is a component of the DIBELS NWF. Grade 2 students completed the DIBELS ORF assessment.

All students who scored at the Tier 2 level on the assessment were eligible to receive Literacy First services. Assessments were administered in Spanish or English, depending on students' current language abilities as determined by the elementary school. Following the assessments, parental consent was obtained for students identified as eligible to receive the program prior to random assignment.²

Once the sample of students in each school was determined, Grades K–2 students were randomly assigned, within grade at each school, to one of two groups: (1) a treatment group that received Literacy First tutoring and (2) a control group that conducted “business as usual.” Students in the control condition did not receive Literacy First tutoring but were eligible for other types of assistance typically available to students in the participating schools (e.g., literacy coach support).

Research Questions

The evaluation was guided by the following research questions:

1. Does participation in Literacy First have a statistically significant impact on kindergarten students' early reading skills?
2. Does participation in Literacy First have a statistically significant impact on Grade 1 students' early reading skills?
3. Does participation in Literacy First have a statistically significant impact on Grade 2 students' oral reading fluency after 1 year?
4. Does participation in Literacy First have a statistically significant and positive impact on students' reading comprehension skills?

² Parental consent was passive, as approved by our institutional review board (IRB) for this study. Parent opt-out letters were sent home to all families approximately 3 weeks before Literacy First's administration of the BOY assessments.

5. Does participation in Literacy First have a statistically significant and positive longitudinal (2-year) impact on students' reading skills?
6. Does participation in Literacy First have a statistically significant and positive longitudinal (2-year) impact on students' mathematics skills?

Data

The study used extant administrative data from AISD, as well as primary data collected from students participating in the study. The administrative data from AISD consisted of student background characteristics (i.e., race/ethnicity, economic disadvantage, special education status, English language learner status, and gender) and STAAR Grade 3 Reading and Mathematics data. Grades K–1 early literacy skills data and Grade 2 oral reading fluency data were collected by Literacy First staff. Grade 2 reading comprehension data were collected by AIR staff.

Students completed either the English or Spanish versions of the assessments depending on their English proficiency level. At the beginning and end of the year, kindergarten students completed English and Spanish versions of the KDF and LSF created by the program administrators. Similarly, at the beginning and end of the year, Grade 1 students completed the DIBELS NWF, ORF, and the WWR or the Indicadores Dinámicos del Éxito en la Lectura (IDEL) Fluidez en las Palabras sin Sentido (FPS), Fluidez en el Relato Oral (FRO), and the Palabras Completas Leídas (PCL). Grade 2 students completed the DIBELS ORF or the IDEL FRO at the beginning of the year. At the end of the year, Grade 2 students completed the DIBELS ORF or FRO as well as the Iowa Test of Basic Skills (ITBS) or Logramos reading comprehension subtest.

In Grade 3, Cohort 1 students completed either the English or Spanish version of the STAAR Grade 3 Reading and Mathematics assessments, which were administered by school staff using the standard processes for state assessments. Although English and Spanish versions of all assessments covered the same content and skills, the assessments are not direct translations of each other, are not on the same scales, and are not considered equivalent.

Sample

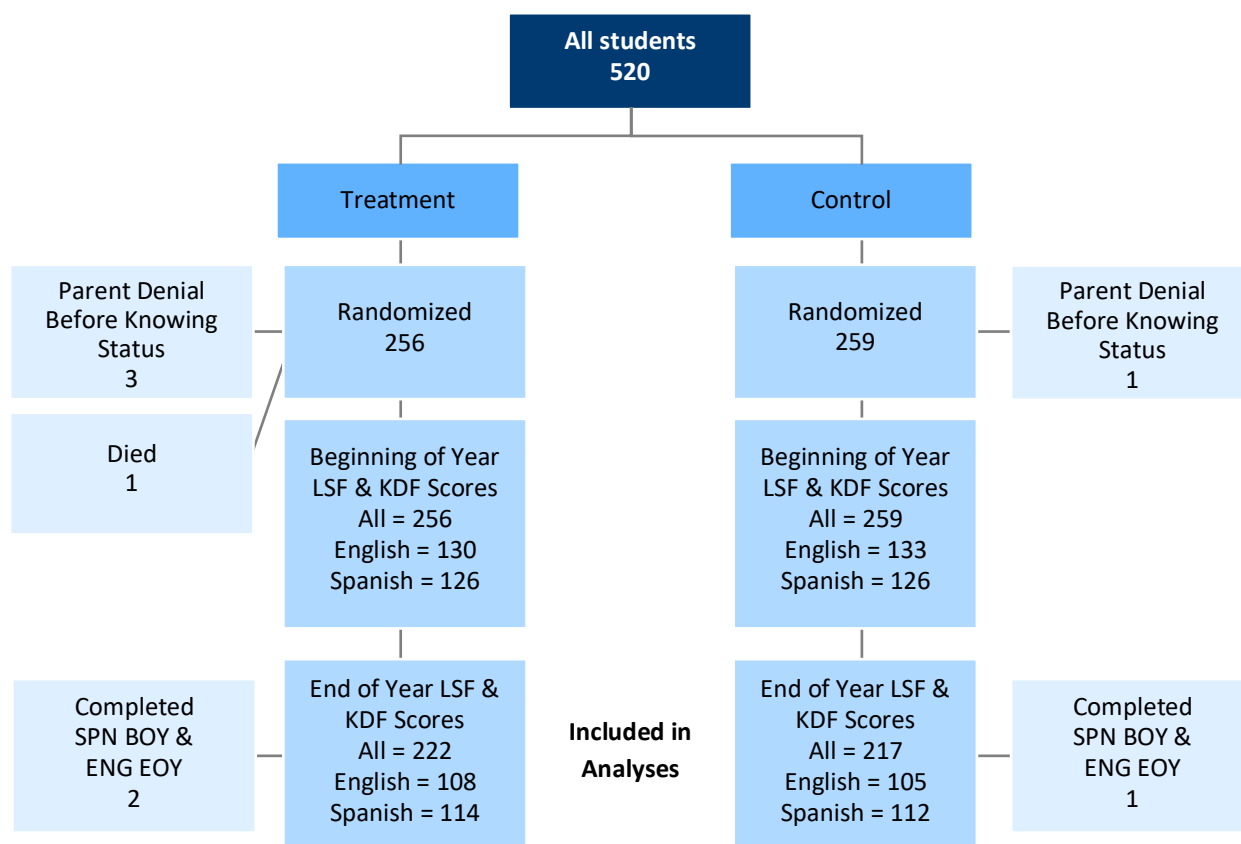
The study was conducted with one cohort of kindergarten and Grade 1 students and two cohorts of Grade 2 students. Grade 2 students in Cohort 1 were followed through Grade 3. The original study design followed Cohort 2 through Grade 3 also. However, due to coronavirus disease 2019 (COVID-19), students did not complete the STAAR in the 2019–20 school year.

Kindergarten

Across all 22 AISD schools, 520 kindergarten students were determined to be eligible on the basis of their BOY assessment scores and were randomly assigned to either the treatment group or the

control group (Exhibit 1). Using simple random assignment, 260 students were assigned to each group.³ Following random assignment, four parents (three in the treatment group and one in the control group) opted out of the study prior to knowing their student’s group assignment. In addition, one student died. These students were excluded from any data analysis. Pretest data were therefore available for 256 treatment group and 259 control group kindergarten students. At the end of the year, 224 treatment group and 218 control group students completed the end-of-year (EOY) assessments. Three of these students (two in the treatment group and one in the control group) completed the BOY assessments in Spanish and the EOY assessments in English. These students were not included in the outcomes analyses due to issues related to the use of standardized scores at BOY and EOY.⁴ The final sample of kindergarten students consisted of 222 treatment group students and 217 control group students.

Exhibit 1. Student CONSORT Diagram for Kindergarten Outcomes



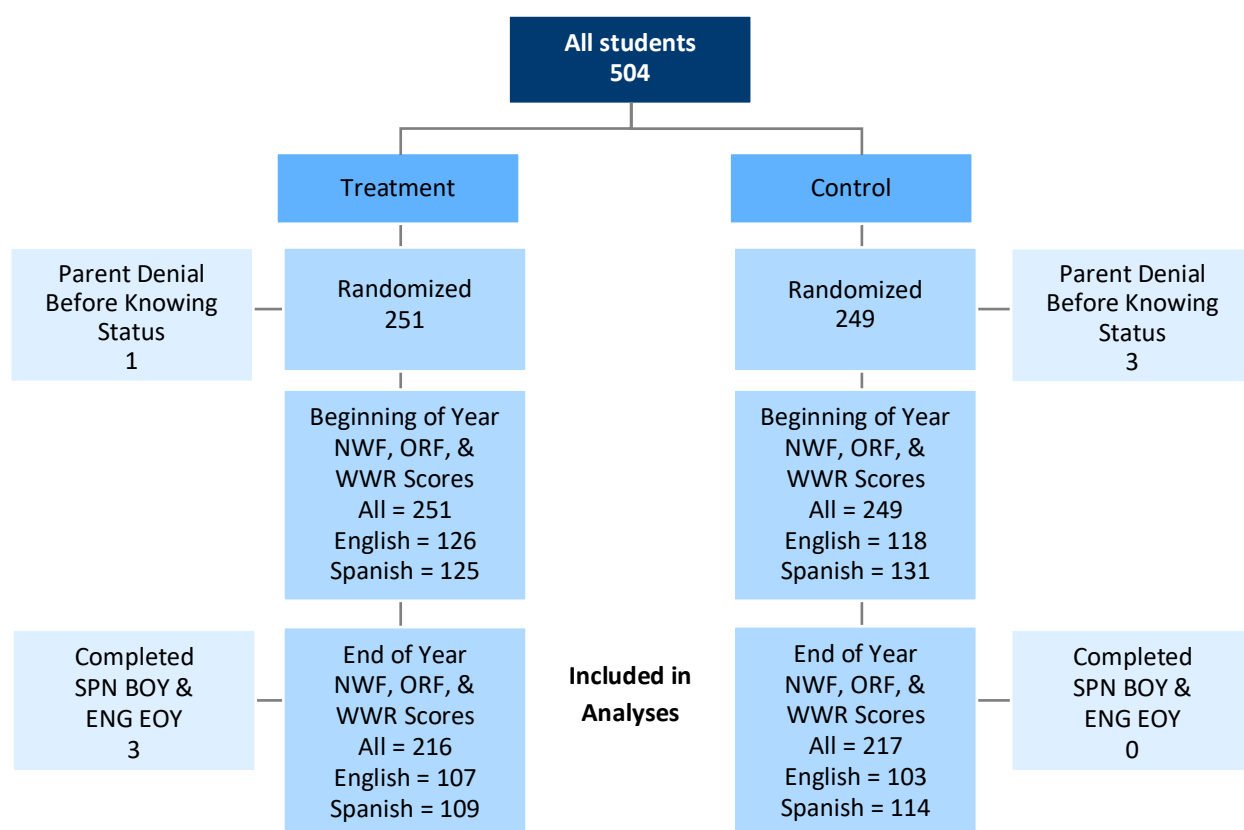
³ Due to small sample sizes within each school, as well as data availability, it was not possible to use a stratified random assignment procedure.

⁴ We constructed z scores based on the study sample separately for the students completing the English and Spanish versions of each of the assessments. For students who completed the BOY assessments in Spanish and EOY assessments in English, the z scores for the BOY and EOY assessments were not calculated relative to the same groups of students or based on the same scales.

Grade 1

For Grade 1, 504 students were determined to be eligible based on their BOY assessment scores (Exhibit 2). Using the same procedure, 252 students were assigned to the treatment group, and 252 students were assigned to the control group. Following random assignment, four parents (one in the treatment group and three in the control group) opted their students out of the evaluation prior to knowing their student's group assignment. These students were excluded from any data analysis. Pretest data were therefore available for 251 treatment group and 249 control group Grade 1 students. At the end of the year, 219 treatment group and 217 control group students completed the EOY assessments. Three of the treatment group students completed the BOY assessments in Spanish and the EOY assessments in English. These students were not included in the outcomes analyses due to issues related to the use of standardized scores at the beginning and end of the year.⁵ The final sample of Grade 1 students consisted of 216 treatment group students and 217 control group students.

Exhibit 2. Student CONSORT Diagram for Grade 1 Outcomes



⁵ We constructed z scores based on the study sample separately for the students completing the English and Spanish versions of each of the assessments. For students who completed the BOY assessments in Spanish and EOY assessments in English, the z scores for the BOY and EOY assessments were not calculated relative to the same groups of students or based on the same scales.

Grade 2

Across all 22 AISD schools, 1,046 Grade 2 students (723 in Cohort 1 and 323 in Cohort 2) were determined to be eligible to participate in the study based on their beginning-of-the-year test scores. During the consent process, parents were able to opt their students out of the evaluation, which led to 22 Grade 2 students in Cohort 1 and three students in Cohort 2 (out of the 1,046) who were not eligible for random assignment. The number of Grade 2 students eligible for participation varied considerably across schools and cohorts, ranging from two to 62 students. In 19 elementary schools in Cohort 1 and six elementary schools in Cohort 2, the number of eligible students exceeded the tutoring capacity for the school, which is typically eight students per tutor (approximately 3.5 tutors per school). When the number of eligible students exceeded the tutoring capacity of the school, students were randomly selected from the school's eligible pool of Grade 2 students to participate in the study based on each school's tutoring capacity. Otherwise, all eligible Grade 2 students were included in the study. The initial Cohort 1 sample comprised 525 Grade 2 students from 22 elementary schools, and the initial Cohort 2 sample comprised 273 students from 20 elementary schools.⁶

At the end of both Grade 2 school years, May 2018 (Cohort 1) and 2019 (Cohort 2), a total of 331 treatment students and 339 control students completed the ORF and FRO as a posttest. Of these students, 26 treatment students and 15 control students completed the FRO at the beginning of the year and the ORF at the end of the year. These students were not included in the impact analysis for this outcome. There was concern that the ORF was capturing differences in these students' English fluency in addition to students' oral reading fluency and that changing test versions reduced students' scores from what the scores would have been had students completed the pretest and posttest using the same versions of the test. The final analytic sample included 305 treatment students and 324 control students.

Similarly, a total of 319 treatment and 341 control students completed the Grade 2 ITBS and Logramos reading comprehension assessment at the end of Grade 2. Of these students, 17 treatment students and 13 control students completed the FRO at the beginning of the year and the ITBS reading comprehension section at the end of the year. These students were not included in the outcomes analyses because it was determined that, similar to the measure of oral reading fluency, the English version of the reading comprehension assessment may not accurately measure these students' progress in reading comprehension.

Finally, a total of 194 Cohort 1 students completed the STAAR Grade 3 Reading and Mathematics assessments. Of these, 34 treatment students and 43 control students completed

⁶ Two schools opted to stop implementing Literacy First after the school district increased the amount schools had to contribute toward the cost of the program.

the FRO at the beginning of Grade 2 and the English version of the STAAR Grade 3 Reading and Mathematics assessments. As previously noted, these students were not included in the outcomes analyses because there was concern that the English versions of the STAAR may not accurately measure these students' academic progress.

CONSORT diagrams for the study summarizing Grade 2 student progression through the study are shown in Exhibits 3 and 4.

Exhibit 3. Student CONSORT Diagram for Grade 2 Reading Fluency and Reading Comprehension Outcomes (Cohorts 1 and 2)

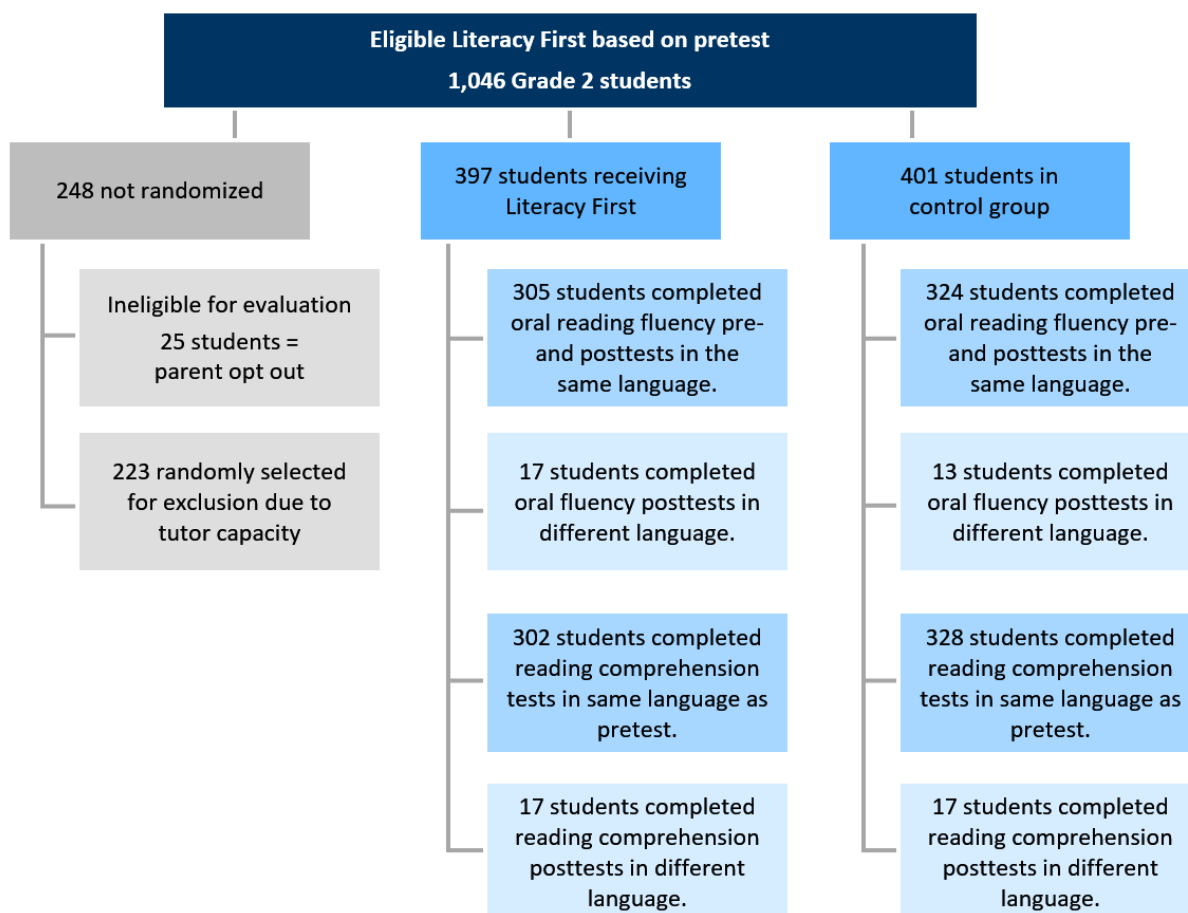
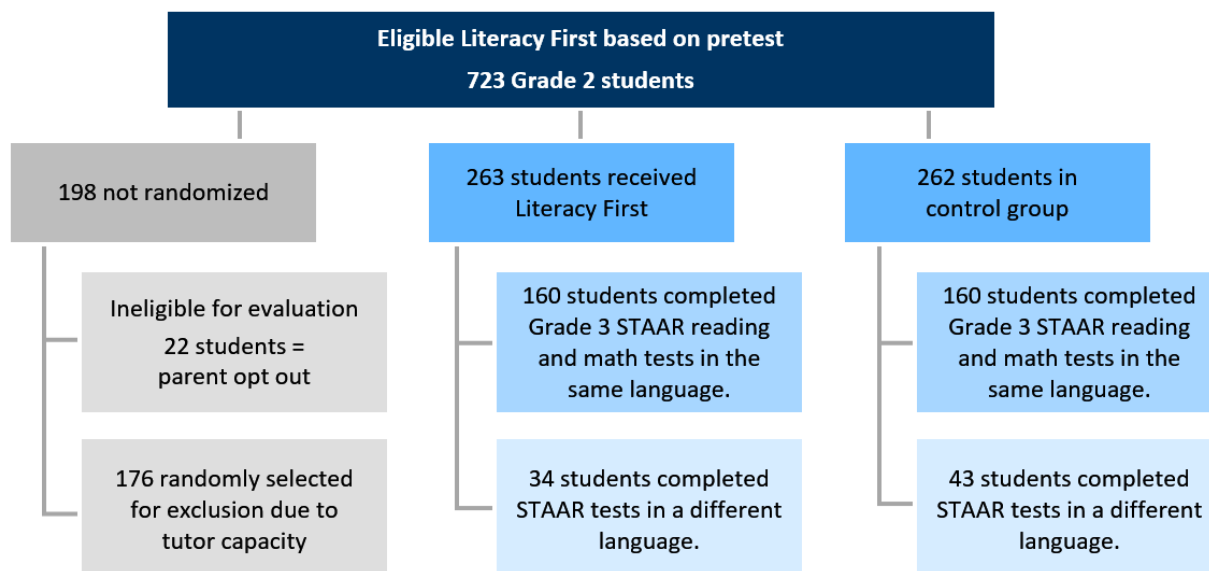


Exhibit 4. Student CONSORT Diagram for STAAR Grade 3 Reading and Mathematics Outcomes (Cohort 1 Only)



Attrition

The overall attrition rate for kindergarten students was 14.8%, and the differential attrition rate between treatment and control groups was 2.9%. Similarly, the overall attrition rate for Grade 1 students was 13.4%, and the differential attrition rate between treatment and control groups was 1.1%. These estimates are indicative of low attrition (What Works Clearinghouse, 2020a).

For Grade 2 students, the overall attrition rate for Cohorts 1 and 2 students at the end of Grade 2 was 21.2% for the oral fluency assessment and 21.1% for the reading comprehension assessment, and the differential attrition rate between treatment and control group students was 4.0% for the oral fluency assessment and 5.7% for the reading comprehension assessment (Exhibit 5). The overall attrition rate for Cohort 1 students at the end of Grade 3 was 39.0%, and the differential attrition rate between treatment and control group students was 0.2 (see Exhibit 5). All these rates are indicative of low attrition using the What Works Clearinghouse (WWC) optimistic boundary (What Works Clearinghouse, 2020a).

Exhibit 5. Overall and Differential Attrition by Outcome for Kindergarten, Grade 1, and Grade 2 Students

	Randomized sample		Analytic sample		Attrition rate	
	Treatment	Control	Treatment	Control	Overall	Differential
Kindergarten						
Letter Sound Fluency (LSF); Kindergarten Decoding Fluency (KDF)	256	259	222	217	14.8%	2.9%
Grade 1						
Nonsense Word Fluency (NWF/PS); Oral Reading Fluency (ORF/FRO); Whole Words Read (WWR/PCL)	251	249	216	217	13.4%	1.1%
Grade 2						
Oral Reading Fluency (ORF/FRO)	397	401	305	324	21.2%	4.0%
Reading comprehension (ITBS/Logramos)	397	401	302	328	21.1%	5.7%
Grade 3 (Cohort 1 of initial Grade 2 sample)						
STAAR Reading and Mathematics	263	262	160	160	39.0%	0.2%

Sample Characteristics

Overall, the kindergarten students were primarily Hispanic (85%) and economically disadvantaged (94.5%), had limited English proficiency (57.5%), and did not participate in special education (89.5%). Similar to the kindergarten students, Grade 1 students were primarily Hispanic (84%) and economically disadvantaged (92%), had limited English proficiency (52%), and did not participate in special education (88.5%). The majority of students (58.8%) had not previously received Literacy First tutoring in kindergarten. Most of the Grade 2 students who participated in the study were also Hispanic (88%), and slightly more than half the students were female (54%). About 69% of the students in the sample were identified as limited English proficient, and about 10% of the students were receiving special education services. Almost all students in the study were identified as economically disadvantaged (93%).

Baseline Equivalence

Following WWC guidelines (What Works Clearinghouse, 2020a), baseline equivalence on key demographic and achievement variables was assessed using Hedges's *g* for continuous variables

(i.e., ORF/FRO scores) and Cox's index for dichotomous variables (i.e., economic disadvantage, gender, and limited English proficiency). Exhibits 6 through 10 present the effect size differences between students in the Literacy First and control groups for each assessment. Effect size differences were calculated using group percentages, means, standard deviations, and sample sizes.

For students in kindergarten and Grade 1, most of the effect size differences were in the acceptable range but were large enough to warrant inclusion in the final analytic impact models to control for baseline differences, based on WWC threshold standards (What Works Clearinghouse, 2020a).⁷ The only variable exceeding WWC thresholds was race/ethnicity. Exhibits 6 and 7 present *p* values from analyses comparing the percentages of students in each category for the treatment and control groups, as well as differences in students' BOY assessment scores. There were no statistically significant differences between students in the treatment or control groups on any of the student characteristics or outcomes.

For the sample of Grade 2 students used to examine the effect of Literacy First on oral reading fluency, the differences in the percentage of students who were male, were limited English proficient, and received Literacy First tutoring in kindergarten and Grade 1 were large enough to warrant inclusion in the final analytic impact model to control for these baseline differences. For the sample of Grade 2 students used to investigate the effect of Literacy First on reading comprehension, the differences in the percentage of students who were male, received special education services, were economically disadvantaged, and received prior Literacy First tutoring in kindergarten and Grade 1 were large enough to warrant inclusion in the analytic model. Finally, for the sample of Grade 2 students used to examine the effect of Literacy First on STAAR Grade 3 Reading and Mathematics assessments, the difference in the percentage of students who received prior Literacy First tutoring in kindergarten was large enough to warrant inclusion in the analytic model—that is, the absolute values of the effect size differences were larger than 0.05 but smaller than 0.25.

For all samples of Grade 2 students, the differences in percentages of Black and Hispanic students exceeded the WWC threshold of 0.25 for both analytic samples. Likewise, the percentage of students from other racial/ethnic backgrounds exceeded the WWC threshold of 0.25 for the samples used to estimate the impact of Literacy First on students' reading comprehension and STAAR Grade 3 Reading and Mathematics assessments. In addition, the percentage of students who were limited English proficient and economically disadvantaged exceeded the WWC threshold of 0.25 for the samples used to estimate the impact of Literacy First on students' STAAR Grade 3 Reading and Mathematics assessments. Given the low level of

⁷ The WWC does not require variables with effect size differences less than 0.05 to be included in analytic models. The WWC requires variables with values between 0.05 and 0.25 to be included in analytic models. The WWC considers variables with effect size differences larger than 0.25 to not be equivalent. Absolute values are used for effect size differences.

attrition in the study, the differences in student characteristics at baseline should not have an impact on study results, especially because students' scores on the prior measures of achievement—ORF and FRO—do not show large effect sizes. Regardless, all student baseline variables were included in analytic models to control for these differences.

We conducted *t* tests for proportions and continuous variables, and *p* values for these analyses are included in Exhibits 8 through 10 for reference. None of the *p* values were statistically significant.

Exhibit 6. Sample Descriptive Statistics and Effect Size Differences for Kindergarten Students

	Treatment		Control		<i>p</i> value	Effect size
	<i>N</i>	%	<i>N</i>	%		
Race/ethnicity						
Black	19	8.6	18	8.3	0.71	0.08+
Hispanic	182	82.0	190	88.0	0.08	-0.29++
White	14	6.3	7	3.2	0.13	0.44++
Other race/ethnicity	7	3.2	2	0.9	0.14	0.74++
Limited English proficiency						
Yes	134	60.4	119	54.8	0.29	0.12+
No	88	39.6	98	45.2	0.29	-0.12+
Special education						
Yes	22	9.9	25	11.5	0.50	-0.12+
No	200	90.1	192	88.5	0.50	0.12+
Gender						
Male	114	51.4	107	49.3	0.68	0.05+
Female	108	48.6	110	50.7	0.68	-0.05+
Economic disadvantage						
Yes	206	93.6	204	94.9	0.65	-0.12+
No	16	6.4	13	5.1	0.65	0.12+
Beginning-of-year assessment scores						
	Mean	Std dev	Mean	Std dev	<i>p</i> value	Effect size
LSF score (standardized)	0.01	0.99	-0.02	1.01	0.75	0.03
KDF score (standardized)	0.02	1.12	-0.01	0.86	0.76	0.03

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

+0.05 < absolute value of effect size ≤ 0.25, ++Absolute value of effect size > 0.25

Exhibit 7. Sample Descriptive Statistics and Effect Size Differences for Grade 1 Students

	Treatment		Control		p value	Effect size
	N	%	N	%		
Race/ethnicity						
Black	21	9.7	20	9.2	0.72	0.07+
Hispanic	178	82.4	186	85.7	0.26	-0.18+
White	12	5.6	5	2.3	0.03*	0.69++
Other race/ethnicity	4	1.9	5	2.3	1.00	0.00
Limited English proficiency						
Yes	113	52.3	113	52.1	1.00	0.00
No	103	47.7	104	47.9	1.00	0.00
Special education						
Yes	28	13.0	21	9.7	0.33	0.18+
No	188	87.0	196	90.3	0.33	-0.18+
Gender						
Male	111	51.4	114	52.5	0.68	-0.05+
Female	105	48.6	103	47.5	0.68	0.05+
Economic disadvantage						
Yes	198	92.1	199	91.7	1.00	0.00
No	18	7.9	18	8.3	1.00	0.00
Prior Literacy First tutoring						
Yes	93	43.0	86	39.6	0.53	0.07+
No	123	57.0	131	60.4	0.53	-0.07+
Beginning-of-year assessment scores						
	Mean	Std dev	Mean	Std dev	p value	Effect size
ORF score (standardized)	-0.02	0.92	0.02	1.07	0.64	-0.04
NWF score (standardized)	-0.03	1.03	0.03	0.97	0.55	-0.06+
WWR score (standardized)	0.03	0.93	-0.03	1.06	0.58	0.06+

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

+0.05 < absolute value of effect size ≤ 0.25, ++Absolute value of effect size > 0.25

Exhibit 8. Sample Descriptive Statistics and Effect Size Differences for Grade 2 Oral Reading Fluency

	Treatment		Control		<i>p</i> value	Effect size
	<i>N</i>	%	<i>N</i>	%		
Race/ethnicity						
Black	26	8.5	17	5.3	0.10	0.38++
Hispanic	260	85.3	290	89.5	0.11	-0.28++
Other race/ethnicity	19	6.2	17	5.3	0.60	0.12+
Limited English proficiency						
Yes	209	68.5	229	70.7	0.55	-0.06+
No	96	31.5	95	29.3		
Special education						
Yes	26	8.5	28	8.6	0.96	0.00
No	279	91.5	296	91.4		
Gender						
Male	136	55.4	164	50.6	0.13	0.10+
Female	169	44.6	160	49.4		
Economic disadvantage						
Yes	287	94.1	303	93.5	0.54	0.00
No	18	5.9	21	6.5		
Prior Literacy First tutoring						
Kindergarten	90	29.5	105	32.4	0.43	-0.06+
Grade 1	109	35.7	106	32.7	0.42	0.08+
Oral fluency beginning of year						
	Mean	Std dev	Mean	Std dev	<i>p</i> value	Effect size
ORF/FRO Score (standardized)	-0.02	0.98	0.03	1.0	0.66	-0.05

+0.05 < absolute value of effect size ≤ 0.25, ++Absolute value of effect size > 0.25

Exhibit 9. Sample Descriptive Statistics and Effect Size Differences for Grade 2 Reading Comprehension

	Treatment		Control		<i>p</i> value	Effect size
	<i>N</i>	%	<i>N</i>	%		
Race/ethnicity						
Black	25	8.3	17	5.2	0.12	0.30++
Hispanic	256	84.8	295	90.0	0.05	-0.28++
Other race/ethnicity	21	6.9	16	4.9	0.37	0.22+
Limited English proficiency						
Yes	205	67.9	227	69.2	0.72	-0.03
No	97	32.1	101	30.8		
Special education						
Yes	30	90.1	30	90.8	0.73	-0.07+
No	272	9.9	298	9.2		
Gender						
Male	133	44.0	159	51.5	0.27	-0.19+
Female	169	56.0	169	48.5		
Economic disadvantage						
Yes	280	92.7	307	93.6	0.66	-0.10+
No	22	7.3	21	6.4		
Prior Literacy First tutoring						
Kindergarten	88	29.1	102	31.1	0.59	-0.06+
Grade 1	108	35.8	105	32.0	0.33	0.11+
Oral fluency beginning of year						
	Mean	Std dev	Mean	Std dev	<i>p</i> value	Effect size
ITBS/Logramos Score (standardized)	-0.01	0.97	0.03	0.99	0.60	-0.04

+0.05 < absolute value of effect size ≤ 0.25, ++Absolute value of effect size > 0.25

Exhibit 10. Sample Descriptive Statistics and Effect Size Differences for STAAR Grade 3 Reading and Mathematics

	Treatment		Control		<i>p</i> value	Effect size
	<i>N</i>	%	<i>N</i>	%		
Race/ethnicity ^a						
Hispanic	126	78.8	141	88.1	0.93	-0.40++
Other race/ethnicity	34	21.3	19	11.9	0.93	0.40++
Limited English proficiency						
Yes	91	56.9	107	66.9	0.93	-0.26++
No	69	43.1	53	33.1		
Special education						
Yes	18	11.3	18	11.3	1.0	0.00
No	142	88.7	142	88.7		
Gender						
Male	75	46.9	74	46.3	0.99	0.02
Female	85	53.1	86	53.7		
Economic disadvantage						
Yes	143	89.4	150	93.8	0.97	-0.40++
No	17	10.6	10	6.3		
Prior tutoring						
Kindergarten	48	30.0	44	27.5	0.98	0.06+
Grade 1	59	36.9	59	36.9	1.0	0.00
Oral fluency beginning of year						
	Mean	Std dev	Mean	Std dev	<i>p</i> value	Effect size
ORF/FRO Score (standardized)	0.01	0.93	-0.01	1.06	0.86	0.02

^aTo avoid cell sizes with fewer than 10 students identified as Black, White, Native American, or two or more races/ethnicities were combined into an “other race/ethnicity” group.

+0.05 < absolute value of effect size ≤ 0.25, ++Absolute value of effect size > 0.25

p* < .05, *p* < .05

Analysis

The study used a multisite design in which students were randomly assigned within schools to treatment and control groups. Because students were nested within classrooms, a two-level hierarchical linear model was used to analyze the data, with students at Level 1 and schools at Level 2. Equations 1 and 2 were used to estimate the impact of Literacy First tutoring on student outcomes. Prior to conducting the analyses, students' test scores were converted to standardized z scores using sample-based means and standard deviations on each of the assessments, separately by language. This process was necessary to place assessment scores on the same scales. Although the Spanish and English versions of the assessments were substantively similar, they are not considered equivalent.

Next, to reduce variation, student-level background characteristics and BOY assessment scores were included in the analytic model as covariates. These student-level background characteristics included race/ethnicity, gender, special education status, and economic disadvantage. Students' beginning-of-the-year assessment scores, indicators for prior Literacy First tutoring (Grades 1 and 2 only), and an indicator for students who completed the assessments in Spanish were also included in the final analytic model. For students in Grade 2, an indicator for Cohort 2 was included in the model as well. The final analytic model with covariates for Grade 2 students is shown in Equation 2.

Equation 1. Final Analytic Model With Covariates

Level 1 Students

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Treatment}_{ij}) + \beta_{2j}(\text{Cohort 2}_{ij}) + \beta_{3j}(\text{Male}_{ij}) + \beta_{4j}(\text{Black}_{ij}) + \beta_{5j}(\text{Other Race/Ethnicity}_{ij}) + \beta_{6j}(\text{Economic Disadvantage}_{ij}) + \beta_{7j}(\text{Limited English Proficient}_{ij}) + \beta_{8j}(\text{Special Education}_{ij}) + \beta_{9j}(\text{Tutoring Kindergarten}_{ij}) + \beta_{10j}(\text{Tutoring Grade 1}_{ij}) + \beta_{11j}(\text{BOY assessment(s)}_{ij}) + \beta_{12j}(\text{Spanish}_{ij}) + r_{ij}$$

Level 2 Schools

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} \dots \beta_{11j} = \gamma_{100}\end{aligned}$$

where Y_{ij} is the standardized student outcome (e.g., ORF/FRO, ITBS/Logramos, STAAR Grade 3 Reading or Mathematics), β_{0j} is the adjusted mean outcome score across school sites, and β_{1j} is the adjusted treatment effect across school sites. All variables, with exception of the treatment indicator, were group-mean centered. As shown, the treatment effect was allowed to vary across sites.

Variations of the same two analytic models were used for all outcomes with one exception. The outcomes analyses for STAAR Grade 3 Reading and Mathematics did not include a separate

indicator for Black students. In these analyses, Black students were included in the “Other race/ethnicity” category due to the low number of Black students included in the sample for analyses.

Findings

This study examined the effect of Literacy First tutoring on students’ reading and mathematics skills. The primary focus of the study was to determine whether participation in Literacy First tutoring had an effect on students’ short-term early reading, oral reading fluency, and reading comprehension outcomes. The study also sought to assess the impact of Literacy First on longer term reading skills as measured by the STAAR Grade 3 Reading Assessment. In addition, the study examined whether students’ improvement in reading would translate into gains in other subjects tested in Grade 3. The focus in this study is on Grade 3 mathematics skills as measured by the STAAR Grade 3 Mathematics Assessment. For all outcomes, additional analyses were conducted to assess whether there was a significant interaction between the treatment effect and taking the Spanish versions of the assessments. These analyses did not reveal any statistically significant differences in treatment effect for students who completed Spanish versions of the assessments.

Short-Term Literacy Skills

The first set of analyses was designed to answer research questions 1 through 4, which focus on the impact of Literacy First on students’ short-term early reading skills, oral reading fluency, and reading comprehension.

Kindergarten

Results of the analyses showed statistically significant, positive effects of Literacy First participation on kindergarten students’ letter sound fluency and decoding fluency relative to students who received business-as-usual instruction. As shown in Exhibit 11, treatment group students scored 0.83 standard deviations higher than control group students on the EOY LSF assessment, on average across sites, controlling for baseline student characteristics.⁸ Similarly, treatment group students scored 0.53 standard deviations higher than control group students on the EOY KDF assessment, on average across sites (see Appendix A for data tables). The differences between treatment and control group students were statistically significant.

⁸ Detailed data tables are provided in Appendix A.

Exhibit 11. Treatment Effects of Literacy First for Kindergarten Students

Kindergarten assessments	Treatment effect (in standard deviation units)
Letter Sound Fluency (LSF)	0.83**
Kindergarten Decoding Fluency (KDF)	0.53**

* $p < .05$, ** $p < .01$

Grade 1

Results of the analyses also showed statistically significant, positive effects of Literacy First participation on Grade 1 students' nonsense word fluency, oral reading fluency, and whole word reading relative to students who received supplemental reading supports from their teachers. Exhibit 12 shows the results of the analyses examining the effect of Literacy First on Grade 1 students' nonsense word fluency, oral reading fluency, and whole word reading. Grade 1 students who received Literacy First tutoring scored 0.33 standard deviations higher on the EOY DIBELS NWF assessment than students who did not receive Literacy First tutoring, on average across sites (see Appendix A). Similarly, Grade 1 students in the treatment group scored 0.38 standard deviations higher on the EOY DIBELS ORF assessment than students in the control group, on average across sites. Grade 1 students in the treatment group also scored 0.24 standard deviations higher on the EOY WWR assessment than students in the control group, on average across sites. The differences between treatment and control group students were statistically significant.

Exhibit 12. Treatment Effects of Literacy First for Grade 1 Students

Grade 1 assessments	Treatment effect (in standard deviation units)
Nonsense Word Fluency (NWF)	0.33**
Oral Reading Fluency (ORF)	0.38**
Whole Words Read (WWR)	0.24**

* $p < .05$, ** $p < .01$

Grade 2

Exhibit 13 shows the results of the analysis examining the effect of Literacy First on students' oral reading fluency and reading comprehension at the end of Grade 2, as measured by the ORF/FRO and ITBS/Logramos tests. On average across sites, students receiving Literacy First tutoring scored about 0.33 standard deviations higher than students in these sites who did not receive Literacy First tutoring, controlling for student background characteristics and other

covariates in the analytic model. Similarly, the results of the analysis examining the effect of Literacy First on students' reading comprehension, as measured by the ITBS and Logramos, showed that students receiving Literacy First tutoring scored about 0.19 standard deviations higher than students in those sites who did not receive Literacy First tutoring, controlling for student background characteristics and other covariates. The analyses demonstrated positive and statistically significant effects of Literacy First on students' oral reading fluency and reading comprehension.

Exhibit 13. Treatment Effects of Literacy First for Grade 2 Students

Grade assessments	Treatment effect (in standard deviation units)
Oral Reading Fluency (ORF/FRO)	0.33**
Reading Comprehension (ITBS/Logramos)	0.19*

* $p < .05$, ** $p < .01$

Long-Term Literacy Skills

An additional goal of the study was to examine whether the effects of Literacy First tutoring extend past the school year in which students receive tutoring. To do so, we examined the impact of Literacy First tutoring on students who were in the Grade 2 sample a year later when they were enrolled in Grade 3. The analysis was designed to answer research question 5, which focuses on the impact of Literacy First on students' longer-term reading skills. The outcome for this analysis is the Texas state standardized reading assessment for students in Grade 3—the STAAR Grade 3 Reading Assessment. These analyses only included students who were part of Cohort 1 because students in Cohort 2 did not complete the STAAR Grade 3 Reading Assessment due to the COVID-19 pandemic.

Exhibit 14 shows the results of the analysis examining the effect of Literacy First on students' STAAR Grade 3 Reading assessments controlling for student background characteristics or other covariates. On average across sites, students receiving Literacy First tutoring scored about 0.21 standard deviations higher on the STAAR Grade 3 Reading Assessment than students in those sites who did not receive Literacy First tutoring when controlling for student background characteristics and other covariates in the analytic model. Although this standard deviation is quite large, the results of the analysis were not statistically significant. As mentioned previously, due to COVID-19, Cohort 2 students did not complete the STAAR Grade 3 Reading assessments. Therefore, these analyses are somewhat underpowered because it was not possible to combine test scores for Cohorts 1 and 2.

Exhibit 14. Treatment Effects of Literacy First for Grade 2 Students When in Grade 3

Grade 3 assessments	Treatment effect (in standard deviation units)
STAAR Grade 3 Reading	0.21

Mathematics

In addition to examining the effect of Literacy First on students' reading skills, analyses were conducted to assess whether students' improvement in their reading skills would translate into success in other subject areas. For this study, we investigated whether improvements in students' reading skills would translate to their mathematics skills, as measured by the STAAR Grade 3 Mathematics Assessment. The analysis was designed to answer research question 6.

On average across sites, students receiving Literacy First tutoring scored about 0.14 standard deviations higher on the STAAR Grade 3 Mathematics than students in these sites who did not receive Literacy First tutoring when controlling for student background characteristics and other covariates in the analytic model (Exhibit 15). The results of these analyses, though somewhat large, were not statistically significant. However, as previously, these analyses are somewhat underpowered because Cohort 2 students did not complete the STAAR Grade 3 Mathematics and are not included in the analyses.

Exhibit 15. Treatment Effects of Literacy First for Grade 2 Students When in Grade 3

Grade 3 assessment	Treatment effect (in standard deviation units)
STAAR Grade 3 Mathematics	0.14

Interpretation

To help judge the practical importance of the effect of Literacy First on students' reading skills, findings were translated into effect sizes and an improvement index. Effect sizes were calculated using Hedges's g . Hedges's g is calculated by dividing the difference between the two group means by their pooled standard deviation.⁹ The improvement indices were calculated using the process described by the WWC (What Works Clearinghouse, 2020b). That is, the Hedges's g effect sizes were converted to Cohen's U_3 index, and the improvement index was calculated by subtracting 50% from the U_3 index value.¹⁰ The improvement index is interpreted

$$^9 g = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}}$$

$$^{10} U_3 - 50\% = \text{improvement index}$$

as the difference in percentile rank between an average treatment group member and an average control group member in the control group distribution. Exhibit 16 shows the effect sizes and improvement index for each outcome.

Exhibit 16. Effect Sizes and Improvement Indices

Assessment	Improvement index
Kindergarten	
Letter Sound Fluency (LSF)	+29
Kindergarten Decoding Fluency (KDF)	+19
Grade 1	
Nonsense Word Fluency (NWF/PS)	+12
Oral Reading Fluency (ORF/FRO)	+16
Whole Words Read (WWR/PCL)	+8
Grade 2	
Oral Reading Fluency (ORF/FRO)	+12
Reading comprehension (ITBS/Logramos)	+8
Grade 3 (Cohort 1 of initial Grade 2 sample)	
STAAR Grade 3 Reading	+8
STAAR Grade 3 Mathematics	+4

The improvement index helps put these numbers into somewhat more understandable terms. For example, the improvement index for oral reading fluency for Grade 2 students is +12. This means that an average treatment student would rank at the 62nd percentile in the control group with regard to oral reading fluency, or equivalently, that an average treatment student would rank 12 percentage points higher on oral reading fluency than an average control student, who by definition ranks at the 50th percentile. Similarly, an average treatment student would rank at the 58th percentile in the control group with regard to reading comprehension and STAAR Grade 3 Reading, or equivalently, an average treatment student would rank 8 percentage points higher on reading comprehension and STAAR Grade 3 Reading than an average control student.

Summary

The results of the current evaluation show statistically significant positive effects of Literacy First tutoring on Grades K–2 students' early reading skills, oral reading fluency, and reading comprehension within the same school year. This finding reinforces evidence that one-on-one tutoring, occurring frequently, is an effective strategy to improve students' literacy skills. In addition, it demonstrates the positive impact of the Literacy First program, including its curricula and approach to training volunteer tutors within AISD elementary schools.

An additional focus of the current study was to determine the potential long-term impact of tutoring on students' literacy and mathematics skills. Findings indicate that students who had received Literacy First tutoring in Grade 2 scored higher on the STAAR Grade 3 for both reading and mathematics. Although the effect size differences between treatment and control group students were quite large, these differences were not statistically significant. An important note is that COVID-19 impacted the final sample size used for the longitudinal analysis because the students in Cohort 2 did not complete the STAAR.

Given the results of the current study, Literacy First is an effective program for boosting reading comprehension and other early literacy skills for students in Grade 2. However, additional rigorous evaluation is needed to determine the long-term impact of Literacy First on students' literacy skills.

References

- Agile Analytics. (2018). *Literacy First evaluation: Austin ISD impact analysis*.
- Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). *Improving reading outcomes for students with or at risk for reading disabilities: A synthesis of the contributions from the Institute of Education Sciences Research Centers* (NCSER 2014-3000). U.S. Department of Education, Institute of Education Sciences, National Center for Special Education Research.
- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5(3), 257–288.
- Jacob, R., Armstrong, C., Bowden, A. B., & Pan, Y. (2016). Leveraging volunteers: An experimental evaluation of a tutoring program for struggling readers. *Journal of Research on Educational Effectiveness*, 9(1), 67–92.
- National Center for Education Statistics. (2015). *A first look: 2015 mathematics and reading (National Assessment of Educational Progress at grades 4 and 8)*. U.S. Department of Education, Institute of Education Sciences.
- National Reading Panel & Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implication for reading instruction* (NH Publication No. 00-4769). U.S. Government Printing Office.
- Raynor, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, 2(2), 31–37.
- Tackett, K. K., Leroux, A. J., & McFarland, L. (2013). *ACE: A Community for Education—evaluation report*.
- Tidd, S. (2014). *ACE: A Community for Education evaluation report, 2013–2014 school year*.
- Tidd, S. (2015). *ACE: A Community for Education evaluation report, 2014–2015 school year*.
- What Works Clearinghouse. (2020a). *Standards handbook* (Version 4.1). U.S. Department of Education, Institute of Education Sciences.
- What Works Clearinghouse. (2020b). *Procedures handbook* (Version 4.1). U.S. Department of Education, Institute of Education Sciences.

Appendix A. Data Tables

Exhibit A1. Impact Analysis Results for Letter Sound Fluency (LSF), Kindergarten Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.45	0.12	-3.76	21	0.00**
Treatment	0.83	0.76	10.91	21	0.00**
BOY LSF	0.25	0.04	6.86	386	0.00**
Black	-0.11	0.14	-0.78	386	0.44
White	-0.08	0.19	-0.41	386	0.68
Other race/ethnicity	-0.27	0.27	-0.99	386	0.32
Male	-0.04	0.07	-0.54	386	0.59
Limited English proficiency	-0.13	0.17	-0.77	386	0.44
Special education	-0.37	0.12	-3.06	386	0.00**
Economic disadvantage	-0.28	0.17	-1.63	386	0.11
Spanish	0.14	0.17	0.88	386	0.38

* $p < .05$, ** $p < .01$

Exhibit A2. Impact Analysis Results for Kindergarten Decoding Fluency (KDF), Kindergarten Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.29	0.12	-2.32	21	0.03*
Treatment	0.53	0.08	6.39	21	0.00**
BOY KDF	0.18	0.04	4.38	385	0.00**
Black	-0.00	0.16	-0.03	385	0.98
White	-0.06	0.22	-0.29	385	0.77
Other race/ethnicity	0.18	0.30	0.61	385	0.54
Male	-0.06	0.08	-0.70	385	0.49
Limited English proficiency	-0.28	0.19	-1.51	385	0.13
Special education	-0.35	0.14	-2.59	385	0.10

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Economic disadvantage	-0.53	0.19	-2.81	385	0.01**
Spanish	0.37	0.19	1.98	385	0.05

* $p < .05$, ** $p < .01$

Exhibit A3. Impact Analysis Results for Nonsense Word Fluency (NWL/PS), Grade 1 Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.14	0.07	-2.05	21	0.05*
Treatment	0.33	0.09	3.70	21	0.00**
BOY NWF	0.42	0.04	9.63	378	0.00**
Tutoring 2016–17	-0.10	0.09	-1.05	378	0.30
Black	0.10	0.07	-2.05	378	0.54
White	-0.03	0.24	-0.13	378	0.90
Other race/ethnicity	0.34	0.32	1.10	378	0.27
Male	0.20	0.09	2.34	378	0.02*
Limited English proficiency	-0.07	0.09	-0.77	378	0.44
Special education	0.01	0.14	0.05	378	0.96
Economic disadvantage	-0.05	0.17	-0.30	378	0.77
Spanish	-0.01	0.11	-0.10	378	0.92

* $p < .05$, ** $p < .01$

Exhibit A4. Impact Analysis Results for Oral Reading Fluency (ORF/FRO), Grade 1 Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.17	0.08	-2.26	21	0.04*
Treatment	0.38	0.08	4.75	21	0.00**
BOY ORF	0.51	0.04	12.39	378	0.00**
Tutoring 2016–17	0.08	0.08	0.91	378	0.36
Black	0.01	0.15	0.09	378	0.93
White	0.20	0.22	0.90	378	0.37
Other race/ethnicity	0.28	0.29	0.96	378	0.34
Male	0.07	0.08	0.90	378	0.37

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Limited English proficiency	-0.05	0.09	-0.57	378	0.57
Special education	0.03	0.13	0.21	378	0.84
Economic disadvantage	-0.49	0.16	-3.08	378	0.00**
Spanish	-0.06	0.10	-0.56	378	0.57

* $p < .05$, ** $p < .01$

Exhibit A5. Impact Analysis Results for Whole Words Read (WWR/PCL), Grade 1 Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.11	0.07	-1.46	21	0.16
Treatment	0.24	0.09	2.67	21	0.01**
BOY WWR	0.38	0.05	8.39	378	0.00**
Tutoring 2016–17	-0.10	0.09	-0.99	378	0.33
Black	0.15	0.17	0.87	378	0.38
White	0.07	0.25	0.28	378	0.78
Other race/ethnicity	0.50	0.32	1.56	378	0.12
Male	0.22	0.09	2.53	378	0.01*
Limited English proficiency	0.03	0.09	0.32	378	0.75
Special education	-0.04	0.14	2.67	378	0.77
Economic disadvantage	0.05	0.17	0.28	378	0.78
Spanish	-0.00	0.11	-0.04	378	0.97

* $p < .05$, ** $p < .01$

Exhibit A6. Impact Analysis Results for Oral Reading Fluency (ORF/FRO), Grade 2 Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.15	0.06	-2.40	21	0.03*
Treatment	0.33	0.06	5.61	21	<0.01**
Cohort 2	-0.03	0.06	-0.47	574	0.64

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Male	0.00	0.06	0.02	574	0.98
Black	0.04	0.12	0.35	574	0.73
Other race/ethnicity	0.10	0.13	0.75	574	0.45
Economic disadvantage	-0.17	0.12	-1.44	574	0.15
Limited English proficiency	0.33	0.09	3.68	574	<0.01**
Special education	-0.07	0.10	-0.68	574	0.50
Tutoring kindergarten	0.06	0.06	1.03	574	0.31
Tutoring Grade 1	-0.14	0.06	-2.27	574	0.02*
BOY ORF/FRO	0.70	0.03	24.67	574	<0.01**
Spanish	-0.22	0.09	2.34	574	0.02*

* $p < .05$, ** $p < .01$

Exhibit A7. Impact Analysis Results for Reading Comprehension (ITBS/Logramos), Grade 2 Students

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	-0.12	0.06	-2.14	21	0.44
Treatment	0.19	0.07	2.83	21	0.01*
Cohort 2	0.02	0.07	0.29	575	0.78
Male	-0.14	0.07	-2.06	575	0.04*
Black	0.17	0.15	1.13	575	0.26
Other race/ethnicity	-0.03	0.16	-0.22	575	0.83
Economic disadvantage	-0.30	0.15	-2.04	575	0.04*
Limited English proficiency	0.20	0.11	1.78	575	0.08
Special education	-0.15	0.12	-1.26	575	0.21
Tutoring kindergarten	-0.02	0.08	-0.24	575	0.81
Tutoring Grade 1	0.01	0.08	0.14	575	0.89
BOY ORF/FRO	0.49	0.04	13.93	575	<0.01**
Spanish	-0.05	0.12	-0.46	575	0.64

* $p < .05$, ** $p < .01$

Exhibit A8. Impact Analysis Results for STAAR Grade 3 Reading, Initial Grade 2 Students (Cohort 1 Only)

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	0.06	0.10	0.59	21	0.56
Treatment	0.21	0.12	1.78	21	0.09
Male	-0.36	0.10	-3.41	267	<0.01**
Other race/ethnicity	0.15	0.15	0.99	267	0.32
Economic disadvantage	-0.14	0.19	-0.72	267	0.48
Limited English proficiency	0.02	0.14	0.14	267	0.89
Special education	-0.15	0.17	-0.85	267	0.39
Tutoring kindergarten	0.03	0.12	0.25	267	0.80
Tutoring Grade 1	0.05	0.11	0.42	267	0.68
BOY ORF/FRO	0.25	0.05	4.79	267	<0.01**
Spanish	0.26	0.15	1.67	267	0.10

* $p < .05$, ** $p < .01$

Exhibit A9. Impact Analysis Results for STAAR Grade 3 Mathematics, Initial Grade 2 Students (Cohort 1 Only)

Fixed effect	Coefficient	Standard error	t ratio	Approx. df	p value
Intercept	0.15	0.11	1.33	21	0.20
Treatment	0.14	0.11	1.23	21	0.20
Male	-0.03	0.11	-0.24	267	0.81
Other race/ethnicity	-0.03	0.16	-0.22	267	0.83
Economic disadvantage	-0.03	0.20	-1.67	267	0.87
Limited English proficiency	0.24	0.15	1.62	267	0.11
Special education	-0.40	0.18	-2.25	267	0.03*
Tutoring kindergarten	0.08	0.12	0.68	267	0.50
Tutoring Grade 1	-0.02	0.12	-0.19	267	0.85
BOY ORF/FRO	0.18	0.05	3.23	267	<0.01**
Spanish	0.23	0.16	1.48	267	0.14

* $p < .05$, ** $p < .01$

About the American Institutes for Research

Established in 1946, with headquarters in Arlington, Virginia, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance to solve some of the most urgent challenges in the U.S. and around the world. We advance evidence in the areas of education, health, the workforce, human services, and international development to create a better, more equitable world. The AIR family of organizations now includes IMPAQ, Maher & Maher, and Kimetrica. For more information, visit [AIR.ORG](https://www.air.org).



American Institutes for Research®
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | [AIR.ORG](https://www.air.org)