2024

READ USA Literacy Tutoring: 2023-2024 RCT Outcomes Analysis

JEROME V. D'AGOSTINO & EMILY M. RODGERS
THE OHIO STATE UNIVERSITY

September 2024
REPORT #2024-07

READ USA Literacy Tutoring: Outcomes Analysis

Background and Study Design

READ USA Literacy Tutoring began in Duval County Public Schools (DCPSs) in the summer of 2019. Since then, thousands of elementary-grade children who struggle in reading have been tutored by DCPS high school and early college students. The tutors, who are paid and trained by READ USA, work with each participating student for about 40 minutes per lesson, three days a week. Content specialists and the teachers at the participating schools provide ongoing support to the tutors.

The first study of READ USA was conducted on pretest and posttest data collected on participating students during the summer of 2021 (Dinsmore, 2021). Subsequent studies based on pretest and posttest change scores or comparisons with a non-equivalent control group were conducted on data from spring, 2022 (D'Agostino & Rodgers, 2023a), summer and fall, 2022 (D'Agostino & Rodgers, 2023b), and spring, 2023 (D'Agostino & Rodgers, 2023c). An analysis of participating students' attitudes toward reading also was conducted on data from the spring of 2023 (D'Agostino & Rodgers, 2023d). Taken together, these studies revealed that students who received READ USA made significant gains on literacy tests such as the Gray Oral Reading Test (GORT)—5th edition, and compared to peers who did not receive tutoring, performed similarly or slightly better on the Florida State Assessments.

The evidence produced from prior studies was promising, but without an equivalent control group, it was not possible to ascertain with a higher degree of confidence if READ USA students truly profited from the intervention in terms of academic gains. In the 2023-2024 schoolyear, a randomized controlled trial (RCT) of READ USA was carried out for the first time. Students in Grades 3-5 in seven participating DCPS schools were randomly assigned to either receive the intervention during the autumn or spring terms. All students were administered assessments at the beginning, middle, and end of the school year, which allowed for a comparison of treatment gains (the students who participated in autumn) to control gains (the students who were on the waiting list to receive the intervention in spring) in the first half of the year. Because the two groups switched conditions for the spring term, it was also possible to compare spring gains among the two groups. An analysis of outcomes during the fall semester were reported in D'Agostino and Rodgers (2024). This report provides an outcomes analysis based on the data from both the fall and spring periods.

Evaluation Method

Participants

The were 360 3rd- through 5th-grade students identified across the seven participating schools who were considered eligible for the treatment based on their prior year FAST scores. Given the average number of students in seven schools (*n*=51), an alpha level of .05, two tail test, a 0.80 power level, half of the students assigned to the Autumn condition, and five covariates that were expected to explain 50 percent of the outcome variance, a power analysis yielded a minimal detectable effect size (MDS) = 0.21. Thus, the sample size was deemed adequate to detect a modest intervention effect.

Within each of the seven participating schools, but not blocking by grade level, eligible students were assigned at random to receive READ USA tutoring in the autumn (Fall Only) or spring (Spring Only) semesters. There were 159 students assigned to participate first (treatment) and 154 students who were assigned to receive the intervention in spring. Table 1 presents the number of students by grade and condition by school. As can be seen, most study participants were in Grades Three (n=129) and Four (n=124), with 46 of the 313 total students enrolled in Grade 5.

Table 1. Initial and (Final) Number of Sampled Students by School and Group

DCPS School	3rd grade		4th grade		5th grade	
	Fall Only	Spring Only	Fall Only	Spring Only	Fall Only	Spring Only
Arlington	13 (11)	13 (10)	13 (12)	12 (12)	9 (7)	14 (13)
Beauclerc	14 (13)	13 (9)	13 (13)	11 (9)	6 (6)	8 (6)
Hogan	15 (11)	12 (11)	16 (14)	15 (11)	6 (5)	6 (4)
Lake Lucina	16 (15)	10 (8)	11 (9)	17 (15)	0	0
Long Branch	5 (3)	3 (8)	7 (6)	5 (4)	4 (3)	0
Mamie	16 (9)	17 (13)	7 (6)	16 (13)	0	0
S.A. Hull	6 (6)	8 (7)	0	0	2 (1)	1 (1)
Total	85 (68)	76 (61)	67 (60)	76 (64)	27 (22)	29 (24)

During the year, some of the students no longer remained in the study, mainly due to mobility. The final analytic sample numbers are presented in Table 1 in parentheses. In terms of attrition, 150 of 179 Fall Only students remained in the study, which represented a 16.2% attrition rate for that group. In the Spring Only condition, 149 of 181 students remained, for an attrition rate of 17.7%. The overall attrition, therefore, was 299 of 360, or 17%, and the group differential attrition was 1.5%.

Table 2 provides an overview of the student demographic profile by group. The two groups were relatively comparable in terms of their characteristics. The Fall Only group had slightly greater proportions of students who were male, Hispanic, and who spoke a Language other than English than the Spring Only group. Students with a disability and who were Eligible for Free Meals also were more prevalent in the Fall Only condition.

Measures

Gray Oral Reading Test Version 5 (GORT-5). The GORT-5 was administered to all participating students in autumn, midyear and spring. The GORT is a norm-referenced informal reading inventory. Students read grade level passages aloud and respond to comprehension questions. Scores on multiple scales can be derived, including age and grade equivalents, and scale scores. The GORT provides measures of student reading proficiency in Rate and Accuracy, which can be combined to yield a Fluency score. Students' answers to the questions about what they read are used to compute a Comprehension score, and a Sum score can be derived by totaling the fluency and comprehension scores. Thus, there are three independent subscales, Rate, Accuracy, and Comprehension, and two combined scales, Fluency (Rate and Accuracy) and the Sum score (Fluency and Comprehension).

Table 2. Student Demographics by Group

Demographic Variable	Fall Only	Spring Only
Gender		
Male	49%	52%
Female	51%	48%
Race/Ethnicity		
White	19%	23%
American Indian	0%	1%
American Island	1%	1%
Asian American	2%	3%
African American	49%	49%
Hispanic	23%	18%
Multiracial	6%	5%
Language		
English	69%	75%
Spanish	23%	20%
Other	8%	6%
Free Meal Cost	67%	53%
Student with Disability	23%	20%

The GORT-5 was normed on a sample of 2,556 students ranging in age from 6-23 years old from 33 states. The reliability of GORT-5 scores is high ranging from .99 for interscorer reliability, between .82 and .90 for retest-retest, greater than .85 for alternate forms, and with coefficient alphas ranging from .85 - .93 for children ages 6-9 and exceeding .90 for all other ages (Hall & Tannenbaum, 2013; Wiederholt & Bryant, 2012).

There is extensive validity evidence for the GORT-5, including content-, construct- and criterion-related validity (average coefficients ranging from large or very large with five other tests). These multiple sources of validity support the view that the GORT-5 is a valid measure of reading ability (Hall & Tannebaum; 2013; Wiederholt & Bryant, 2012).

Florida Assessment of Student Thinking (FAST). The FAST is a computer-administered assessment created for Florida and aligned with BEST Standards. The FAST measures students' strengths and weaknesses relative to grade-level literacy content to assess students' literacy skills. The test is administered three times during the school year, including a beginning year performance measure (PM1), a midyear performance measure (PM2), and an end of year measure (PM3). All tests yield a scale score that ranges by about 120 points per grade. Unlike many tests designed to measure student achievement change, each PM tests the full grade-level content, so many students, especially those who have fallen behind their peers, are below grade level at the first two PM testing events.

Treatment, Demographic, Grade, and School Variables

To analyze if the student background, grade level, and school variables moderated the effect of the READ USA intervention, and to control for any of the demographic differences between the treatment and control groups, a set of demographic variables were used. A Gender variable was coded "0" for females and "1" for males. A Minority variable was coded "0" for white and Asian American and "1" for American Indian, American Island, African American, Hispanic, and multiracial students. An English Language variable was coded "0" for English primary language speaker, and "1" for Spanish or other language primary speaker. Free Meal was coded "0" for not eligible and "1" for eligible, and Disability was coded "0" for no disability, and "1" for students with a disability. For the analysis of the beginning to midyear outcomes, a treatment condition variable was coded "0" for Spring only students and "1" for Fall Only students. To examine the midyear to spring outcomes, the coding of the treatment condition variable was switched so that Spring Only students were coded "1" and Fall Only students were coded "0."

Two grade-level dummy variables were created to identify fourth, and fifth grade students. Third-grade students were coded "0" on both grade-level dummy variables to serve as the reference group. To examine any possible moderator effects, interaction terms were created by multiplying the treatment condition variable by each of the demographic grade, or school variables.

Statistical Analysis

Hierarchical linear modeling (HLM) was used to analyze the data given that students (Level 1) were nested within one of the seven schools (Level 2). Two-level HLM analyses were done separately to ascertain if there were fall or spring treatment effects on each of the GORT and FAST outcomes. To examine the effects of the intervention in the autumn, the midyear scores served as the outcomes. An initial analysis was conducted by including only the intervention variable at the student level as a predictor in each model. A second model was then conducted by adding the demographic and grade-level covariates. The beginning year scores on each respective outcomes also were added as pretest covariates. A third model was then conducted by adding the interaction terms to examine if treatment effects were moderated by any of the demographic or grade variables. The same analyses were conducted to ascertain spring treatment effects, except the spring test scores served as the outcome variables and the midyear scores on each respective outcome served as the pretest covariates.

Hypotheses

The following hypotheses guided the analyses:

1. Fall Only students will score significantly higher on average than Spring Only students on the midyear GORT and FAST tests after adjusting for the covariates.
2. Spring Only students will have higher average spring GORT and FASTS scores than Fall Only students after adjusting for the covariates, but the effects will not be as strong as on the midyear tests because Fall Only students would also have had the intervention.

There were no hypotheses for the moderator analysis.

Results

Before conducting the HLM analyses, descriptive statistics were conducted on the GORT and FAST outcomes. The GORT grade equivalent means and standard deviations (in parentheses) at each of the three time points and by treatment condition are presented in Tables 3-5. The fall (autumn-midyear) and spring (midyear-spring) average gains and standard deviations also are provided in each table. Because the grade equivalents were established based on the scores from a national norming group, the values also provide for normative national comparisons, or the typical gain expected from the "average proficient" student in the nation. Given that the treatment duration was roughly four months in both autumn and spring, the average national growth rate from pretest to posttest was 0.4, or four months of learning.

As can be seen from Table 3, 3rd-grade students in both groups were roughly a grade level behind national norms in the autumn. The average grade equivalent scores across the GORT outcomes and two groups were about 2 in the autumn when the expected averages were 3. Other than for Accuracy, the fall gains were relatively comparable for the two groups, but in spring, the Spring Only group outpaced the Fall Only group on all GORT outcomes.

Overall, the 4th-grade students were farther behind normative expectations in the autumn compared to the 3rd-grade participants (Table 4) . On Rate and Fluency, both groups were over a year behind the national norm at the beginning of the year. Fall Only students were less than a year behind on Accuracy, and Spring Only students were behind two months. In comprehension, both groups were over a year and one half behind the 4th-grade norm in the autumn. The Fall Only group made greater gains than the Spring Only students, on average, on Accuracy and Comprehension, but on Rate and Fluency, the groups gained about the same during the fall. Like the 3rd-grade findings, Spring Only students outgained the Fall Only group on all GORT outcomes.

On Comprehension, 5th-grade students in both groups were about two years behind national norms in autumn (Table 5). The two 5th-grade groups were about a year or more behind on the other GORT outcomes at the beginning of the year. Like the 4th-grade Fall Only students, the 5th-grade Fall Only group outperformed the Spring Only students on Accuracy and Comprehension in the fall, but the two groups had comparable fall gain averages on Rate and Fluency. As was the case in 4th-grade, the Spring Only 5th-grade students outgained the Fall Only students in spring on all GORT outcomes.

Table 3. READ USA GORT Grade Equivalent Means (SD) by Fall Only (*n*=68) and Spring Only (*n*=61), 3rd Grade

GORT Measure	Autumn	Midyear	Spring	Fall Gain	Spring Gain
Rate					
Fall Only	1.8 (0.8)	2.3 (0.7)	2.6 (0.7)	0.5 (0.5)	0.3 (0.5)
Spring Only	2.0 (1.0)	2.4 (0.9)	3.0 (1.2)	0.4 (0.6)	0.6 (0.7)
Accuracy					
Fall Only	2.0 (1.3)	2.9 (1.2)	3.5 (1.3)	0.9 (0.8)	0.5 (1.3)
Spring Only	2.3 (1.0)	2.9 (1.2)	4.1 (1.6)	0.6 (1.2(	1.2 (1.2)
Fluency					
Fall Only	1.9 (1.0)	2.5 (0.9)	2.9 (0.9)	0.6 (0.9)	0.5 (1.0)
Spring Only	2.1 (0.9)	2.6 (0.9)	3.4 (1.2)	0.5 (0.6)	0.9 (0.7)
Comprehension					
Fall Only	2.0 (0.9)	2.7 (1.3)	2.7 (0.9)	0.7 (1.3)	0.0 (1.3)
Spring Only	2.1 (0.9)	2.7 (1.4)	3.5 (1.6)	0.6 (1.4)	0.8 (1.7)

Table 4. READ USA GORT Grade Equivalents Means (SD) by Fall Only (*n*=60) and Spring Only (*n*=64), 4th Grade

GORT Measure	Autumn	Midyear	Spring	Fall Gain	Spring Gain
Rate					
Fall Only	2.5 (1.1)	3.0 (1.1)	3.3 (1.0)	0.5 (0.6)	0.3 (0.5)
Spring Only	2.8 (1.1)	3.2 (1.2)	3.8 (1.3)	0.4 (0.7)	0.6 (0.8)
Accuracy					
Fall Only	3.1 (1.6)	3.9 (1.7)	4.4 (1.7)	0.8 (1.1)	0.5 (1.2)
Spring Only	3.8 (2.4)	3.5 (1.7)	4.7 (1.5)	-0.3 (2.0)	1.2 (1.3)
Fluency					
Fall Only	2.8 (1.2)	3.2 (1.4)	3.7 (1.2)	0.4 (1.0)	0.5 (0.9)
Spring Only	3.0 (1.5)	3.3 (1.3)	4.1 (1.3)	0.3 (0.8)	0.8 (0.7)
Comprehension					
Fall Only	2.3 (1.0)	2.9 (1.3)	3.1 (1.0)	0.6 (1.4)	0.2 (1.1)
Spring Only	2.3 (1.3)	2.5 (1.1)	3.7 (1.3)	0.2 (1.2)	1.2 (1.4)

Table 5. READ USA GORT Grade Equivalents Means (SD) by Fall Only (*n*=22) and Spring Only (*n*=24), 5th Grade

GORT Measure	Autumn	Midyear	Spring	Fall Gain	Spring Gain
Rate					
Fall Only	3.7 (1.3)	4.0 (1.4)	4.4 (1.2)	0.3 (0.8)	0.3 (0.6)
Spring Only	3.1 (1.1)	3.5 (1.2)	4.2 (1.3)	0.4 (0.6)	0.7 (0.7)
Accuracy					
Fall Only	4.0 (1.4)	5.2 (2.1)	6.1 (2.8)	1.2 (1.4)	0.9 (2.6)
Spring Only	4.0 (1.7)	4.5 (1.7)	6.1 (2.7)	0.5 (1.2)	1.6 (2.0)
Fluency					
Fall Only	3.8 (1.2)	4.2 (1.9)	5.1 (1.8)	0.4 (1.5)	0.8 (1.3)
Spring Only	3.5 (1.2)	3.9 (1.3)	5.0 (1.8)	0.4 (0.7)	1.2 (1.1)
Comprehension					
Fall Only	3.0 (1.0)	4.0 (1.1)	3.8 (1.7)	1.0 (1.2)	-0.2 (1.4)
Spring Only	3.1 (1.1)	3.5 (1.4)	4.7 (1.8)	0.2 (1.2)	1.2 (1.7)

For the HLM analyses, students' GORT and FAST scale scores were used as the dependent measures. The GORT an FAST scale score means and standard deviations (in parentheses) of the total sample are presented in Table 6 at each of the three time points and by group. Tables 7-9 present the scale score means and standard deviations disaggregated by grade level. From

Table 6, it is evident that the pretest differences between the two groups were negligible. Across the GORT measures, the difference was between .02 to .06 scale scores between the group means, which translate to effects from .04 to .18. The same was the case on the FAST—there was a 1-point difference between the groups, translating to a .05 effect size. It also can be seen that on the Fluency measures (Rate, Accuracy, & Fluency), both groups made comparable yearly gains, which was expected given that each group received the intervention for the same amount of time, but at different times during the year. On the Comprehension measure, and consequently on the Sum score, however, the Spring Only group gained more scale score points across the year than did the Fall Only group. The reverse occurred on the FAST—Fall Only students on average gained more over the year than Spring Only students.

Figures 1 and 2 graphically display the means by time and group on the GORT Sum score and FAST, respectively. On the GORT (Figure 1), the Fall Only group grew more on average from autumn to midyear but was surpassed by the Spring Only group from midyear to spring. The trend on the FAST was different. The Spring Only group gained more in the autumn than the Fall Only group, but the trend switched in spring with the Fall Only group continuing to grow at the same rate as in the autumn and the Spring Only group growing less in spring than in the first part of the year.

Table 6. READ USA GORT and FAST Scale Score Means (SD) by Fall Only (*n*=151) and Spring Only (*n*=149), Total Sample

Measure	Autumn	Midyear	Spring
Rate			
Fall Only	6.3 (2.8)	6.7 (2.7)	7.2 (2.4)
Spring Only	6.5 (3.0)	6.8 (2.9)	7.4 (3.0)
Accuracy			
Fall Only	6.9 (3.1)	7.9 (3.1)	8.4 (3.1)
Spring Only	7.5 (3.5)	7.6 (3.2)	8.8 (3.4)
Fluency			
Fall Only	6.5 (2.9)	7.1 (2.9)	7.7 (2.7)
Spring Only	6.8 (3.1)	7.0 (3.0)	8.0 (3.2)
Comprehension			
Fall Only	5.9 (2.7)	6.9 (2.6)	6.6 (2.5)
Spring Only	6.0 (2.8)	6.3 (3.0)	7.3 (3.2)
Sum			
Fall Only	12.4 (5.2)	13.9 (4.9)	14.3 (4.7)
Spring Only	12.8 (5.5)	13.3 (5.5)	15.4 (5.9)
FAST			
Fall Only	178.1 (18.4)	187.1 (20.1)	195.6 (17.3)
Spring Only	179.2 (21.3)	189.9 (20.6)	192.1 (23.3)

Table 7. READ USA GORT and FAST Scale Score Means (SD) by Fall Only (*n*=68) and Spring Only (*n*=61), 3rd Grade

Measure	Autumn	Midyear	Spring
Rate			
Fall Only	5.9 (2.9)	6.4 (2.8)	6.8 (2.6)
Spring Only	6.4 (3.1)	6.5 (3.1)	7.2 (3.1)
Accuracy			
Fall Only	6.3 (2.9)	7.3 (3.0)	7.9 (3.1)
Spring Only	6.8 (3.1)	7.3 (3.2)	8.6 (3.7)
Fluency			
Fall Only	6.0 (2.9)	6.7 (2.9)	7.3 (2.8)
Spring Only	6.4 (3.1)	6.8 (3.1)	7.9 (3.5)
Comprehension			
Fall Only	5.9 (2.8)	6.8 (2.9)	6.4 (2.6)
Spring Only	5.9 (3.0)	6.5 (3.4)	7.4 (3.5)
Sum			
Fall Only	12.0 (5.4)	13.5 (5.3)	13.7 (5.1)
Spring Only	12.3 (5.7)	13.3 (6.1)	15.2 (6.6)
FAST			
Fall Only	168.4 (16.4)	179.3 (19.6)	189.6 (18.3)
Spring Only	169.8 (20.1)	183.1 (17.2)	185.7 (21.7)

Table 8. READ USA GORT and FAST Scale Score Means (SD) by Fall Only (*n*=60) and Spring Only (*n*=64), 4th Grade

GORT Measure	Autumn	Midyear	Spring
Rate			
Fall Only	6.4 (2.6)	7.0 (2.7)	7.4 (2.4)
Spring Only	6.5 (3.2)	7.0 (3.1)	7.3 (3.2)
Accuracy			
Fall Only	7.3 (3.1)	8.3 (3.1)	8.8 (2.8)
Spring Only	7.9 (4.1)	7.4 (3.4)	8.6 (3.2)
Fluency			
Fall Only	6.7 (3.0)	7.4 (2.9)	8.0 (2.5)
Spring Only	6.9 (3.4)	7.0 (3.2)	7.9 (3.1)
Comprehension			
Fall Only	5.7 (2.5)	6.7 (2.4)	6.8 (2.2)
Spring Only	5.7 (2.8)	5.7 (2.7)	7.0 (2.9)
Sum			
Fall Only	12.4 (5.1)	14.1 (4.4)	14.8 (4.2)
Spring Only	12.5 (5.7)	12.7 (5.4)	14.9 (5.6)
FAST			
Fall Only	183.6 (14.5)	190.5 (16.6)	198.3 (14.4)
Spring Only	180.1 (17.4)	189.9 (20.7)	191.6 (21.6)

Table 9. READ USA GORT and FAST Scale Score Means (SD) by Fall Only (*n*=22) and Spring Only (*n*=24), 5th Grade

GORT Measure	Autumn	Midyear	Spring
Rate			
Fall Only	7.6 (2.6)	7.4 (2.5)	7.8 (1.9)
Spring Only	7.0 (1.8)	7.2 (2.0)	7.8 (1.8)
Accuracy			
Fall Only	7.9 (2.7)	8.4 (3.1)	9.3 (3.4)
Spring Only	8.4 (2.2)	8.7 (2.1)	10.0 (2.7)
Fluency			
Fall Only	7.6 (2.7)	7.5 (3.1)	8.5 (2.8)
Spring Only	7.5 (2.1)	7.8 (2.1)	9.0 (2.0)
Comprehension			
Fall Only	6.5 (2.6)	7.4 (2.3)	6.6 (3.1)
Spring Only	7.1 (2.3)	7.5 (2.1)	8.2 (2.9)
Sum			
Fall Only	14.1 (4.9)	14.9 (5.0)	15.1 (5.1)
Spring Only	14.6 (3.8)	15.3 (3.7)	17.1 (45)
FAST			
Fall Only	194.2 (17.3)	202.8 (19.6)	207.4 (13.7)
Spring Only	202.1 (16.1)	208.3 (18.4)	211.1 (22.2)

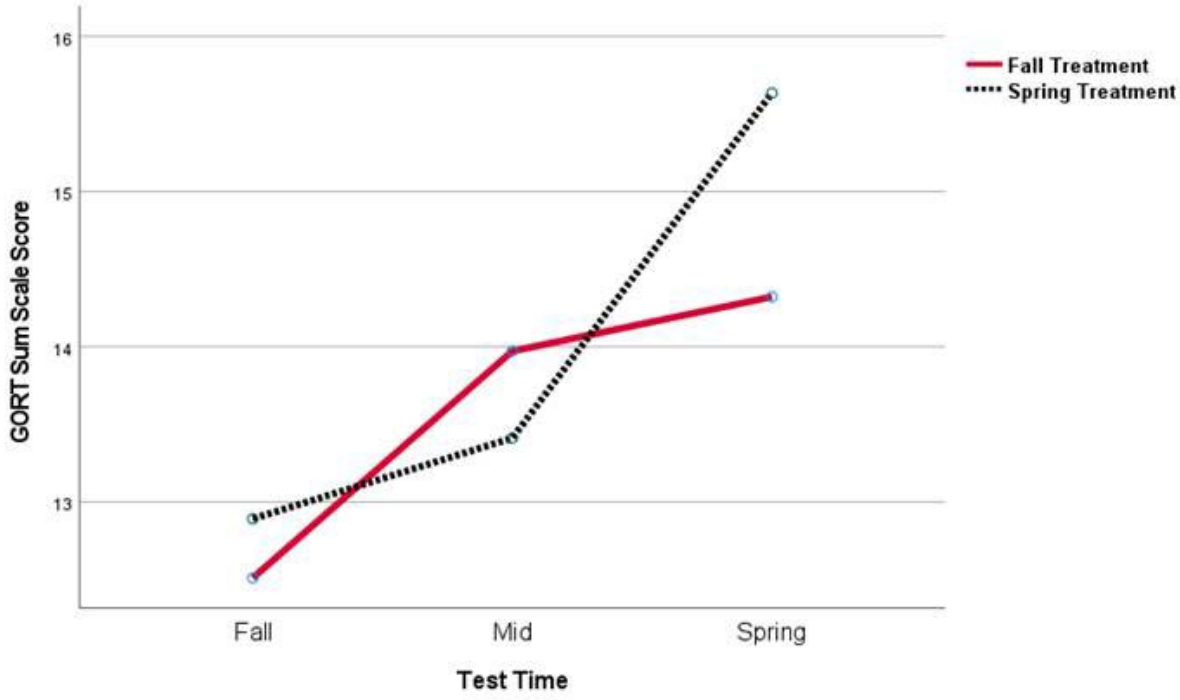

Figure 1. GORT Sum Scale Score Means by Time and Group

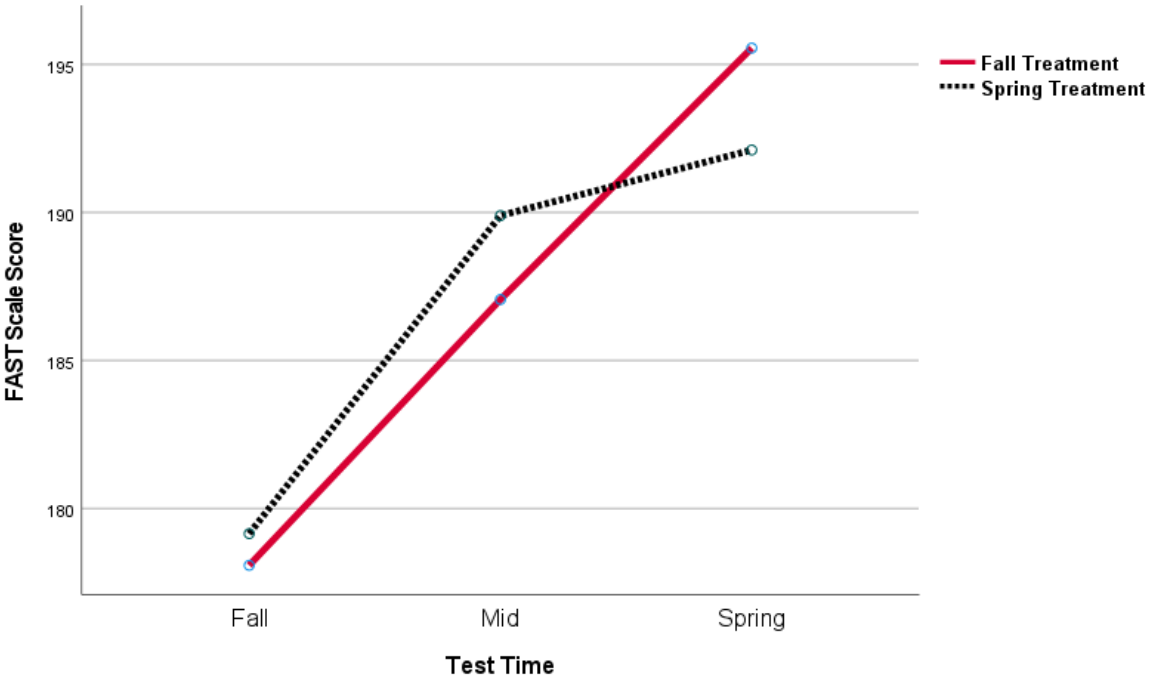

Figure 2. FAST Scale Score Means by Time and Group

Table 10. GORT Autumn Term HLM Results

Variable	Rate	Accuracy	Fluency	Comprehension	Sum
READ USA	.12	.50	.19	.67*	.87
Intercept	6.98***	8.11***	7.24***	6.92***	14.17***
READ USA (T)	.23	.71***	.31	.60*	.89*
Pretest	2.44***	.76***	.80***	.67***	.80***
Male (G)	.23	-.29	.09	-.19	-.07
Minority (M)	.01	.30	.06	.30	.46
Free Meal (FM)	.27	.13	.06	-.08	-.04
English Leaner (EL)	.09	.04	.13	-.12	.28
Disability (D)	-.12	-.25	-.22	-.08	-.10
Grade 4 (4)	.27	-.41	-.07	-.30	-.35
Grade 5 (5)	-.26	-.13	-.27	.14	-.10
Interaction Terms					
T*G	-.02	-.31	-.20	-.07	-.35
T*M	-.18	-.30	.02	1.05*	1.15
T*FM	-.35	-.42	-.68	-.68	-1.41
T*EL	.09	.07	-.36	-.17	-.59
T*D	.06	.23	.03	.71	.76
T*4	.11	1.09*	.58	.98*	1.51*
T*5	-.22	.38	-.33	.22	.15

Note. *p<.05; **p<.01, ***p<.001. The pretest was the fall GORT scores for the same subtest as the outcome for each model.

Tables 10 and 11 provide the HLM results for the GORT analyses during the fall and spring periods, respectively. The top row of both tables provides the coefficient for the treatment variable (READ USA) when included in the prediction models of each outcome (depicted in the columns) alone. The values roughly match the differences in midyear and spring means between the two groups as provided in Tabe 6. With only the treatment variable entered as a predictor, there was a midyear and spring difference in favor of the treatment group on Comprehension and the Sum score, but not on the other GORT outcomes.

The next cluster of rows provide the coefficients for the covariates and the READ USA variables with the covariates included. The variables were centered around the grand mean, which results in an intercept value that reflects the mean score on each outcome for those students who were at the mean value of each variable in the model. All the intercepts were statistically significant, which simply indicates that the values were greater than zero. Note that in all the models for the autumn to midyear analysis (Table 10), the only significant covariate, which was significant in all models, was the respective pretest scores for each outcome. Thus, students' pretest scores explained to a significant degree their posttest scores, and once

Table 11. GORT Spring Term HLM Results

Variable	Rate	Accuracy	Fluency	Comprehension	Sum
Read USA	.22	.41	.33	.80*	1.13*
Intercept	7.19	8.32	7.75***	6.36***	14.03***
READ USA (T)	.27*	.68**	.43*	1.14***	1.68***
Pretest	.85***	.78***	.81***	.59***	0.77***
Male (G)	-.02	.28	.07	.26	.39
Minority (M)	-.09	-.03	-.17	-.15	-.35
Free Meal (FM)	-.18	-.20	-.11	.26	.17
English Leaner (EL)	-.14	-.44	-.25	-1.04***	-1.08*
Disability (D)	-.27	-.96**	-.69**	-.81	-1.34**
Grade 4 (4)	-.03	.28	.16	.46	.71
Grade 5 (5)	.06	.51	.44	.02	.40
Interaction Terms					
T*G	.10	-.04	.28	-.07	.25
T*M	.10	-.26	.17	1.12*	1.35
T*FM	.68*	.43	.66	-.12	.48
T*EL	-.02	-.50	-.38	-.76	-.62
T*D	.01	-.52	-.42	-.63	-.92
T*4	-.09	.02	-.07	-.13	-.05
T*5	2.24	-.24	-.46	.11	-.38

Note. *p<.05; **p<.01, ***p<.001. The pretest was the fall GORT scores for the same subtest as the outcome for each model.

included in the models, no other background or grade level variable explained additional variance. Note that with the inclusion of the covariates, mainly the pretests, the READ USA variable remained significant for Comprehension and the Sum score, and the treatment variable also significantly predicted Accuracy scores.

In the spring, the pretest, which were the midyear scores, also predicted the outcomes on all measures, but unlike in the autumn, English Learner and Disability statuses also served as significant covariates, with EL students scoring significantly lower on spring Comprehension and the Sums, and students with identified disabilities scoring significantly lower on Accuracy, Fluency, and the Sum. The Spring Only group significantly outperformed the Fall Only group in the spring on all GORT measures with the covariates included, as evinced by the significant READ USA variable.

The last cluster of rows in Tables 10 and 11 present the results of the interaction analyses to examine if READ USA was more (or less) effective for certain subgroups. These models were produced by adding the interaction terms to the covariate models. As can be seen in Table 10, the Treatment by Minority interaction variable was significant and positive, indicating that

Table 12. FAST HLM Results

Variable	PM2	PM3
READ USA	-2.40	-3.53
Intercept	186.92***	196.27***
READ USA (T)	-1.37	-4.97**
Pretest	.53***	.60***
Male (G)	-1.60	-3.21
Minority (M)	.92	-.90
Free Meal (FM)	-1.84	1.34
English Leaner (EL)	-.45	-1.84
Disability (D)	-1.09	-6.50**
Grade 4 (4)	2.88	1.18
Grade 5 (5)	7.11*	6.57*
Interaction Terms		
T*G	-.88	-3.90
T*M	4.63	5.77
T*FM	-2.81	4.66
T*EL	-.90	-3.75
T*D	-1.39	-2.72
T*4	2.77	4.52
T*5	4.66	3.68*

Note. *p<.05; **p<.01, ***p<.001. The pretest was PM1 for PM2 and PM2 for PM3.

READ USA was even more effective for students coded as Minority status than students coded as non-Minority status. Thus, READ USA had a positive effect on all students for Comprehension, but for minoritized students, the intervention had an additional boost. The intervention also was found to be more effective for 4th-grade students in the autumn on Accuracy, Comprehension, and the Sum, which was suggested by the grade equivalent mean differences (Table 4). No other interaction terms were significant, revealing that there was no other differential effect on any of the other covariates.

In spring (Table 11), the Minority interaction on Comprehension also was positively significant, indicating that Spring Only Minority students received an additional treatment boost on Comprehension. On Rate, students eligible for free lunch also experienced an additional treatment effect relative to students who did not qualify for free lunch. No other interaction was significant.

Table 12 provides the FAST HLM results. Without covariates (the first row), group differences were not significant at midyear (PM2) or spring (PM3), but given that the coefficients were negative, the comparison group had higher outcomes scores in both cases. With the inclusion of the covariates, the pretest scores were the best predictor—PM1 significantly predicted PM2,

and PM2 significantly predicted PM3. Grade 5 students had significantly greater PM2 and PM3 scores, even after considering students' pretest scores. Students with disabilities had lower PM3 scores than students without disabilities while considering the other covariates. With the covariates included, there was no difference between the two groups on PM2, but on PM3, Fall Only students had greater scores than Spring Only students. The only significant interaction term was for Grade 5 on PM3. The interaction term was positive, indicating that the difference between the two groups was less pronounced in 5th-grade as it was in 3rd-grade. This effect can be gleaned by comparing the PM2 to PM3 gains between the two groups in Tables 7 (3rd-grade) and 9 (5th-grade). In 3rd-grade, Fall Only students gained over 10 points on the FAST from PM2 to PM3, whereas Spring Only students gained about 2.5 points. In 5th-grade, Fall Only students gained less than 5 points, while Spring Only students gained almost 3 points. Thus, the Fall Only latent effect in spring was mostly driven by 3rd-grade students.

The coefficients provided in Tables 10-12 for the READ USA variable were group mean differences adjusted for the covariates in the model. To convert the significant coefficients for the treatment variable to standardized mean differences (i.e., effect sizes), the coefficients were divided by the pooled standard deviations on the outcomes (which can be derived from the standard deviation values presented in Table 6). Table 13 provides the effect sizes for the significant READ USA variable coefficients.

Table 13. Effect Sizes (*d*) for Statistically Significant Findings

Variable	Autumn	Spring
GORT		
Rate		.10
Accuracy	.23	.21
Fluency		.14
Comprehension	.21	.41
Sum	.17	.32
FAST		-.25

Summary

Though there have been several studies conducted to examine the effectiveness of READ USA tutoring, this study employed the most rigorous design to date, and therefore, had the potential to draw the most valid inferences regarding the intervention's impact. Prior studies relied on no comparison group, or a comparison group comprised primarily of ineligible students.

In this study, eligible students were assigned at random to participate in READ USA in the autumn or spring semesters, and thus, the two groups served as control conditions for one

another. By testing the students in the beginning, middle, and end of year on a standardized reading tests and the state assessment, it was possible to ascertain the immediate and sustained effects after students received the intervention earlier in the schoolyear, and the effects of participation later in the year.

In terms of the first research hypothesis, Fall Ony students outperformed Spring Only students on GORT Accuracy, Comprehension, and the Sum, with effect sizes ranging from *d*=.17 to *d*=.23. There were no autumn effects on the GORT Rate and Fluency scales, or on the FAST. Thus, the first hypothesis was partially supported.

The second hypothesis also was partially supported. On the end-of-year assessments, Spring Only students outperformed Fall Only students on all GORT measures, with effects sizes ranging from *d*=.10 for Rate and *d*=.41 for Comprehension. Not only were there significant effects on more GORT outcomes in spring, but the effect sizes were also larger (except on Accuracy), which was contrary to Hypothesis 2. Thus, contrary to the second hypotheses, the GORT effects were stronger in spring than they were in autumn, especially on Comprehension. Although Fall Only students benefited from the autumn intervention, there gains on the GORT were not sustained to the end of the year. On the FAST, nonetheless, Fall Only students outperformed Spring Only students on PM3, seemingly indicating that autumn participation did have a lasting impact.

How do these effect sizes compare to other interventions with similar purposes? Four large-scale meta-analyses have documented that the average effect size of early literacy programs range from 0.23 (Neitzel, et al., 2022) to 0.34 (D'Agostino & Johnson, 2021) to 0.39 (Gersten, et al., 2020; Wanzek et al., 2018). When publication bias was considered, the average effects ranged from .21 to .32 across the meta-analyses. The average effect on Comprehension among all early literacy interventions reviewed by the What Works Clearinghouse was .22 (D'Agostino & Johnson). Thus, the autumn and most of the spring GORT effects were comparable to average effects detected for other primary-grade literacy interventions, but the spring Comprehension effect was larger than most effects of other interventions.

It is important to note that theoretically, comprehension is considered the outcome of good fluency, thus outperforming comparison students on the comprehension measure carries more practical weight than the two independent fluency measures, rate, and accuracy. Comprehension is the ultimate measure, while rate and accuracy might be considered penultimate measures. This finding is significant because studies of other literacy interventions typically find smaller effects on comprehension compared to word reading and fluency (Hall et al.'s, 2022). It is demonstrably more difficult to positively impact comprehension, thus Read USA's positive impact on comprehension is even more commendable.

It is also important to consider that most interventions included in the meta-analyses of elementary-grade reading were delivered by trained teachers rather than by high school and college students. The cost per pupil difference, therefore, is quite dramatic between READ USA and typical literacy interventions and given that READ USA produces average to above average effects, the cost-benefit of the intervention is quite staggering.

The reported effects are based on statistical adjustments for several background and grade covariates, but other than pretest scores, there were no other significant covariates in autumn, and only EL and disability statuses in the spring for the GORT. Disability status also served as a significant covariate in the spring on the FAST. Fifth-grade students scored higher on the FAST at midyear and spring with the other covariates considered.

The READ USA effects were consistent across most demographic and background variables. Minoritized students, however, benefitted more from the intervention in both autumn and spring than other students, which reveals that READ USA appears to produce even greater effects for students who may be considered more marginalized. The differential effects in favor of minority students were considerable in magnitude—and additional *d*=.38 effect in autumn and *d*=.40 in spring, which essentially triple and double the overall effect for those students. There also were a few grade-level modifying effects. In autumn, the intervention effects were most pronounced for 4th-grade students on all three significant GORT outcomes, and the gap between Spring Only and Fall Only students on the FAST at PM3 was less in favor of the latter group in 5th-grade. The results of the spring FAST interaction analyses by grade revealed that most of the positive gain made by Fall Only students were among 3rd-grade students.

The significant negative effect for the READ USA variable for FAST PM3 does not indicate that the intervention had a deleterious effect on students. By spring, all students in the study received the treatment, but at different times. The negative effect seemed to reveal that Fall Only students, who gained more comprehension and accuracy skills in the autumn, were able to carry those skills over to the spring where they performed better on the FAST than Spring Only students.

If Fall Only students' gains from the intervention paid off on the PM3 FAST test, however, why did they not fare better than Spring Only students on GORT Comprehension in the spring? In fact, they not only did not outperform Spring Only students at the end of the year, but their average FAST Comprehension score was lower in spring than at midyear (see Table 6). One possible explanation for this seemingly discordant finding relates to the differences in instructional sensitivity of the GORT and FAST. The GORT seems to detect a more immediate effect of READ USA, as it results in strong, positive effect sizes at a test that follows the treatment period. Without the intervention, students appear to not maintain their growth trajectory, and in some cases, may decline over time. Intervention effects on the FAST seem to

follow an opposite trend, where there is little to no immediate effect registered, followed by a latency effect at the subsequent testing period. The reasons for this phenomenon are not known, but future evaluations of READ USA should be cautious in drawing conclusions about the intervention's effectiveness at the posttest immediately following the treatment period.

In consideration of the entirety of evidence produced by this study, it can be concluded that READ USA has strong immediate effects on students' reading proficiency levels, and most importantly, on their comprehension skills. The intervention is particularly beneficial to minoritized students in developing their comprehension skills. READ USA may also have longer-term effects on students' achievement of ELA state standards, but a replication of the spring finding would be in order before a more definitive conclusion can be made.

References

D'Agostino, J.V., & Johnson, T. (2021). *A critical review of recent meta-analyses on effective reading programs and interventions*. Paper presented at the annual meeting of the Literacy Research Association, Atlanta, GA.

D'Agostino, J.V., & Rodgers, E.M. (2023a). *READ USA peer to peer tutoring: Spring 2022 implementation and outcomes evaluation, Report #2023-01*. Columbus, OH.

D'Agostino, J.V., & Rodgers, E.M. (2023b). *READ USA peer to peer tutoring: Summer-fall 2022 Outcome analysis, Report #2023-02.* Columbus, OH.

D'Agostino, J.V., & Rodgers, E.M. (2023c). *READ USA peer to peer tutoring: Spring 2023 outcome analysis, Report #2023-03*. Columbus, OH.

D'Agostino, J.V., & Rodgers, E.M. (2023d). *READ USA peer to peer tutoring: Results from the spring 2023 Elementary Reading Attitude Survey, Report #2023-04*. Columbus, OH.

D'Agostino, J.V., & Rodgers, E.M. (2024). *READ USA literacy tutoring: Autumn 2023 RCT Outcomes Analysis*. Report #2024-06, Columbus, OH.

Dinsmore, D.L. (2021). *Read USA, Inc. summer tutoring program validation report*. Jacksonville, FL: University of North Florida.

Gersten et al. (2020). Meta-Analysis of the impact of reading interventions for students in the primary grades. *Journal of Research on Educational Effectiveness*, *13*, 401-427.

Hall, A. H., & Tannebaum, R. P. (2013). Test review: J. L. Wiederholt & B. R. Bryant. (2012). Gray Oral Reading Tests—Fifth Edition (GORT-5). Austin, TX: Pro-Ed. *Journal of Psychoeducational Assessment, 31*(5), 516–520. https://doi.org/10.1177/0734282912468578

Neitzel, et al. (2022). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly, 57, 149-179*.

Wanzek, et al. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, *51*, 612-624.

Wiederholt J. L., Bryant B. R. (2012). Gray oral reading test: Examiner's manual (5th ed.). Austin, TX: Pro-Ed.