2024

READ USA Literacy Tutoring: Autumn 2023 RCT Outcomes Analysis

JEROME V. D'AGOSTINO & EMILY M. RODGERS
THE OHIO STATE UNIVERSITY
Spring 2024
REPORT #2024-06

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, over four thousand Grades 1-5 children who struggle in reading have been tutored by mainly DCPS high school or early college students. To our knowledge, this is the first program to use a cross-age setting in which the tutors are much older than the tutees (far-age instead of near-age). 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, they made similar or slightly better gains 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 Autumn of 2023, a randomized controlled trial (RCT) of READ USA was carried out for the first time. Students in Grades 2-5 in seven participating DCPS schools were randomly assigned to either receive the intervention in autumn or in spring, 2024. All students were administered assessments at the beginning and end of the autumn semester, 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). This report provides a description of the research design and outcomes from the autumn 2023 RCT study.

Participants

Within each of the seven participating schools, students were assigned at random to receive READ USA tutoring in the autumn or spring semesters. There were 159 students assigned to participate first (treatment) and 154 students who were assigned to wait until spring, 2024 to receive the tutoring (control). Thus, all study participants were eligible for the intervention based on their reading achievement levels.

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, and 14 in Grade 2.

Table 1. Student Participants by School and study group, Autumn 2023

DCPS School	2nd grade		3rd grade		4th grade		5th grade	
	Control	READ USA	Control	READ USA	Control	READ USA	Control	READ USA
Arlington	2	7	10	11	12	12	13	7
Beauclerc	2	2	9	13	9	13	6	6
Hogan			11	11	11	14	4	5
Lake Lucina	1		8	15	15	9		
Long Branch			3	3	4	6	0	3
Mamie			13	9	13	6		
S.A. Hull			7	6			1	1
Total	5	9	61	68	64	60	24	22

Table 2 provides an overview of the student demographic profile by group. The two groups were relatively comparable in terms of their characteristics, with the treatment group having a slightly greater proportion of boys, Hispanic students, Language other than English, Eligible for Free Meals, and having a disability than the comparison students.

Measures

Gray Oral Reading Test Version 5 (GORT-5). The GORT-5 was administered to all participating students in Grades 2-5 before and after the intervention in autumn 2023. 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 accuracy and rate, 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

Table 2. Student Demographics by Study Group, Autumn 2023

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

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 (accuracy and rate) and the sum score (fluency and comprehension).

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 co-efficient alphas ranging from .85 - .93 for six-nine-year-olds 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 was administered at the beginning of year and midyear to all students in Grades 3-5. 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 baseline measure (PM1), a midyear test (PM2), and an end of year measure (PM3). All tests yield a scale score that ranges by about 120 points per grade. The scale scores also convert to

one of five performance levels, including well below grade level, below grade level, on grade level, proficient, and exemplary.

Students' beginning of autumn 2023 GORT and FAST PM1 scores served as pretest measures, and their end of autumn 2023 GORT and FAST PM2 scores served as posttest measures. GORT grade equivalents by grade level (except for Grade 2 due to low sample size) were examined first compare the READ USA students' growth rates to the control and to national normative expectations. The GORT scale scores were used for statistical purposes to compare the pretest to posttest change between groups. Because the FAST scale score change scores (PM2 minus PM1) appeared to be rather unstable, the FAST performance levels were used instead to examine growth over the autumn period. Each student was categorized as either decreasing in performance level from PM1 to PM2, having no performance level change, or improving one or two performance levels.

Treatment, Demographic, Grade, and School Variables

A binary variable was computed to indicate control ("0") or READ USA ("1") students. 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.

Three grade-level dummy variables were created to identify second, fourth, and fifth grade students. Third-grade students were coded "0" for all three grade-level dummy variables to serve as the comparison group. School dummy variables also were created to distinguish between READ USA funding sources. Three variables were created to identify if students attended Arlington, Long Branch, and S.A. Hull schools, respectively. Students in Beauclerc, Hogan, Lake Lucina, and Mamie were coded "0" for each of the three school dummy variables, and thus served as the base comparison. To examine any possible moderator effects, interaction terms were created by multiplying the READ USA variable by each of the demographic grade, or school variables.

<div align="center">

Results

</div>

The GORT grade equivalent pretest and posttest average scores of the READ USA and control students were examined first by grade level (excluding Grade 2). The control group pre-post average change scores provide a counterfactual of the gain that likely would have occurred for the READ USA students had they not received the treatment. 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.

The statistical analyses were not conducted by grade level, school, or subgroups within any of the demographic values, because there were too few students within those blocks to provide reliable estimates of the READ USA effect. An analysis by any subgroup, therefore, could have resulted in misleading conclusions. Instead, analyses were conducted on the combined data set with the predictor variables added to examine possible moderating effect of the intervention.

Multiple linear regression analysis was performed to examine treatment effects on each of the GORT measures. The pretest-posttest change score on each subtest served as the primary outcome for those analyses. To examine treatment effects on FAST, each student was classified in one PM1 to PM2 change ordered categories, which were "decreased a performance level," "no change in performance levels," "gained one performance level," or gained two performance levels." Ordinal regression was applied to examine the effects given that the ordered categories represented a nonlinear, rank-ordered scale. For both the GORT and FAST analyses, the READ USA effect was computed with the group dummy variable, and all demographic, grade, and school identifiers entered into the equations. After the READ USA effect was computed, the interaction terms were added to explore any moderating effect of the treatment.

GORT Outcomes. Tables 3-5 present the READ USA and control average pretest, posttest, and pre-post difference score for GORT Rate, Accuracy, Fluency, and Comprehension (the GORT does not yield a sums score grade equivalent score) for 3rd-, 4th-, and 5th-grade students, respectively. As can be seen in Table 3, the 3rd-grade students from both groups were nearly two grade levels below the national average at the beginning of the school year (the average national beginning of 3rd-grade value is 3.0), with the READ USA average about one to three months behind the control average, depending on the GORT subtest. Given that the treatment duration was roughly four months long, the average national growth rate from pretest to posttest would be 0.4, or four months of learning. Although behind at the beginning of the school year, the control group kept pace with the national average, gaining about three to five months across the subtests. READ USA students, on average, gained slightly more—four months in Rate, five months in Fluency, and seven months in both Accuracy and Comprehension. They gained about one month more than the control in Rate and Fluency, and three months more in Accuracy and Comprehension.

Fourth-grade students (Table 4) were about two years behind the national average at pretest in Rate and Comprehension, about a year and one half behind in Fluency, and about a year behind in Accuracy. The control students gained about three or four months in Rate and Fluency over the intervention period, which is on pace with the national average, but they declined one month from pretest to posttest in Accuracy and gained only one month in Comprehension. READ USA 4th-grade students, however, gained five months in Rate and Fluency, seven months in Comprehension, and eight months in Accuracy. These values represent above average national gain rates, which is rather impressive given the students' beginning values.

5

The fifth-grade students (Table 5) were about two years behind the national average in Comprehension at the beginning of the schoolyear, and about one year to one- and one-half years behind on the other subtests. Other than for Accuracy, the control students, on average, made fewer than the expected four months of gain over the autumn semester, gaining two months in Comprehension, and three months in Rate and Fluency. READ USA students gained only two months in Rate, which was one less month than the control students, and four months in Fluency (one more month than the control). In Comprehension, however, they gained eight months over the semester, which is twice the number of months expected, and in Accuracy, they gained nine months, or a full school year of expected growth in half the year (students are expected to make nine months of gain during the school months and one month of growth over the summer). The fifth-grade rates of gain in Accuracy and Comprehension by READ USA students are far above national expectations, and obviously superior to the control group.

Table 3. READ USA (n=68) and Control (n=61) GORT Results in Grade Equivalents, 3rd Grade

GORT Measure	Pretest	Posttest	Difference
Rate			
READ USA	1.3 (1.1)	1.7 (1.1)	0.4
Control	1.5 (1.2)	1.8 (1.2)	0.3
Accuracy			
READ USA	1.4 (1.4)	2.1 (1.7)	0.7
Control	1.7 (1.3)	2.1 (1.5)	0.4
Fluency			
READ USA	1.3 (1.2)	1.8 (1.3)	0.5
Control	1.5 (1.2)	1.9 (1.2)	0.4
Comprehension			
READ USA	1.3 (1.2)	2.0 (1.5)	0.7
Control	1.4 (1.2)	1.9 (1.6)	0.5

Note. Cell values are mean grade equivalents (standard deviations in parentheses).

Figures 1 through 5 display the pretest and posttest average scale scores on each GORT scale by group for students in Grades 2-5. It can be seen from the tables that READ USA students, on average scored lower on each pretest, however, none of the group differences at pretest were statistically significant. It is also evident from the figures that for all GORT measures except for GORT Rate, READ USA students, on average surpassed the control students at posttest. On Rate (Figure 2), although READ USA students made greater gains, the average READ USA student remained behind the mean control student at posttest.

Table 4. READ USA (n=60) and Control (n=64) GORT Results in Grade Equivalents, 4th Grade

GORT Measure	Pretest	Posttest	Difference
Rate			
READ USA	2.2 (1.2)	2.7 (1.3)	0.5
Control	2.3 (1.5)	2.7 (1.6)	0.4
Accuracy			
READ USA	2.7 (1.9)	3.5 (2.1)	0.8
Control	3.1 (2.5)	3.0 (2.1)	-0.1
Fluency			
READ USA	2.4 (1.4)	2.9 (1.6)	0.5
Control	2.4 (1.7)	2.7 (1.7)	0.3
Comprehension			
READ USA	1.9 (1.3)	2.6 (1.5)	0.7
Control	1.9 (1.3)	2.0 (1.4)	0.1

Note. Cell values are mean grade equivalents (standard deviations in parentheses).

Table 5. READ USA (n=22) and Control (n=24) GORT Results in Grade Equivalents, 5th Grade

GORT Measure	Pretest	Posttest	Difference
Rate			
READ USA	3.6 (1.5)	3.8 (1.6)	0.2
Control	3.2 (1.2)	3.5 (1.4)	0.3
Accuracy			
READ USA	3.7 (1.7)	4.6 (2.3)	0.9
Control	3.9 (1.8)	4.4 (1.9)	0.5
Fluency			
READ USA	3.5 (1.5)	3.9 (2.0)	0.4
Control	3.5 (1.3)	3.8 (1.6)	0.3
Comprehension			
READ USA	3.0 (1.6)	3.8 (1.5)	0.8
Control	3.2 (1.3)	3.4 (1.4)	0.2

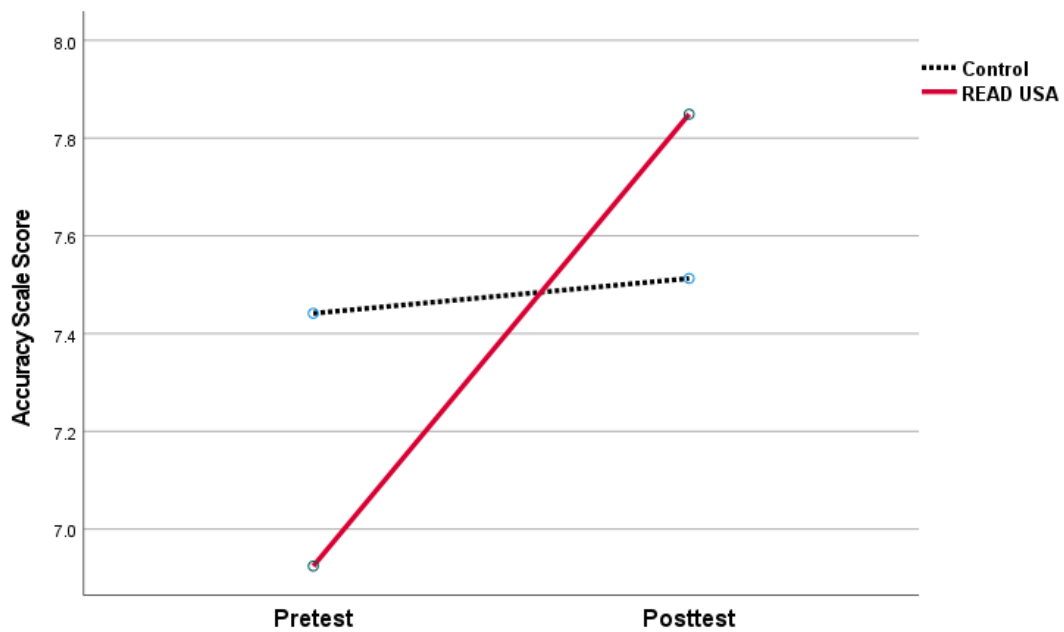*Note*. Cell values are mean grade equivalents (standard deviations in parentheses).

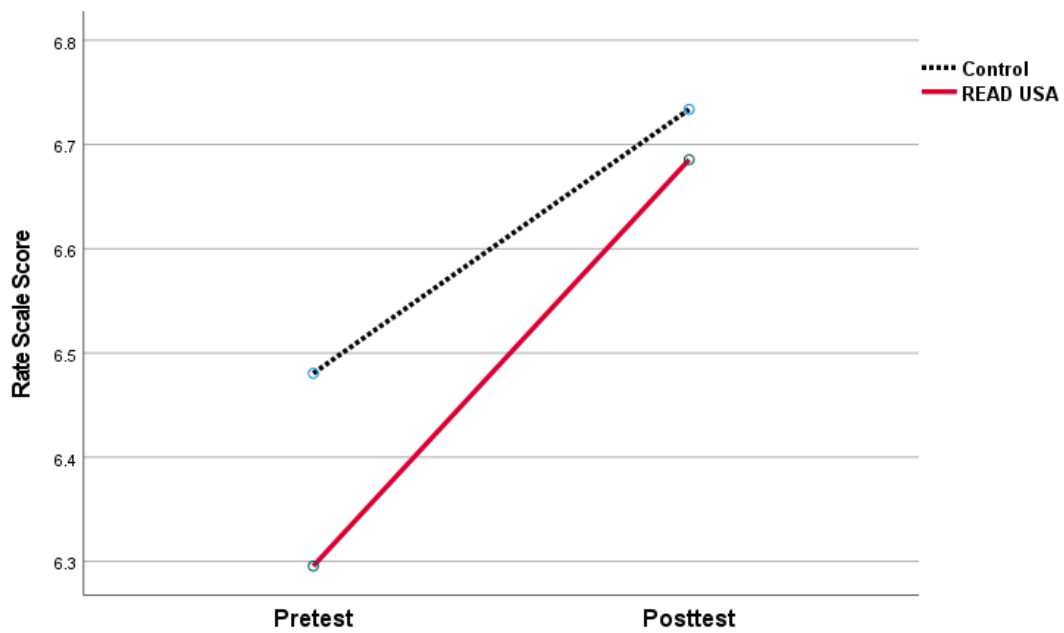*Figure 1*. GORT Accuracy Mean Scale Scores by Group and Testing Time

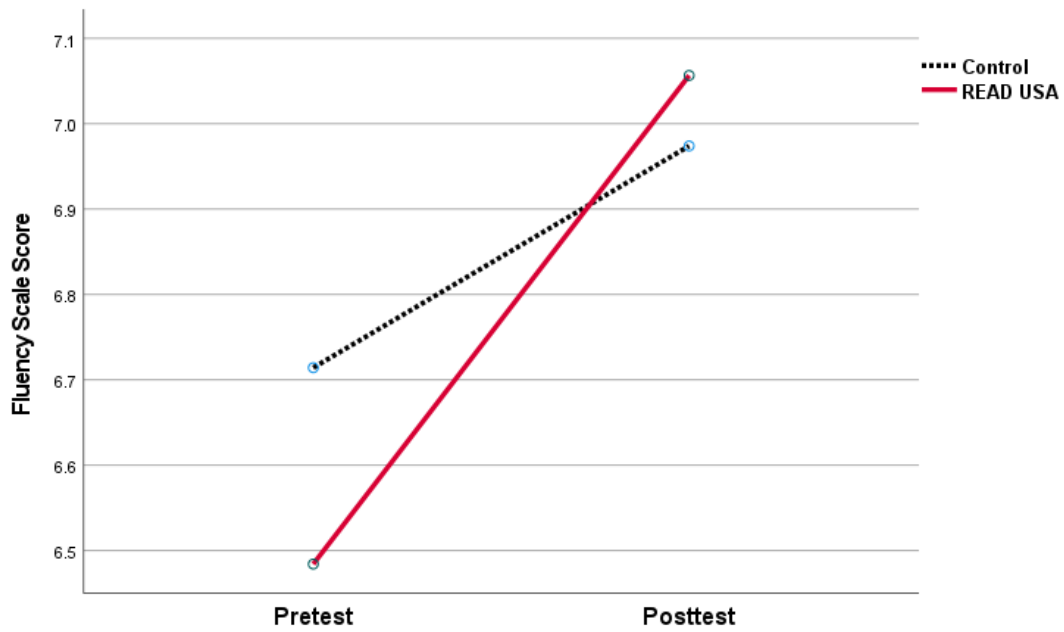*Figure 2*. GORT Rate Mean Scale Scores by Group and Testing Time

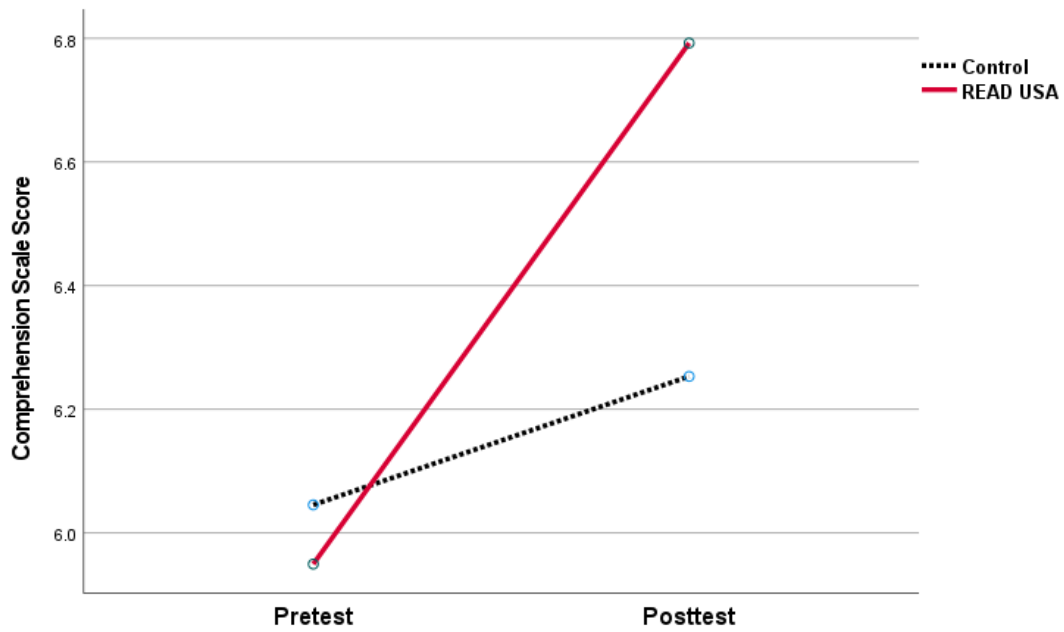*Figure 3*. GORT Fluency Mean Scale Scores by Group and Testing Time

Figure 4. GORT Comprehension Mean Scale Scores by Group and Testing Time
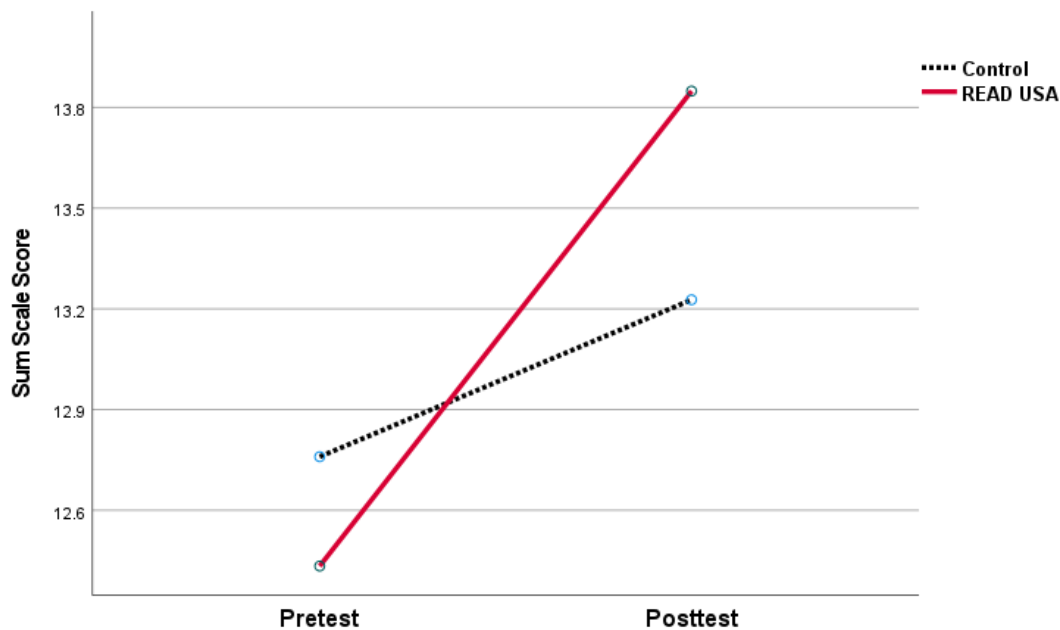
Figure 5. GORT Sum Mean Scale Scores by Group and Testing Time

The mean scores depicted in Figures 1-5 do not take grade level, school, or student background variables into account, but given that each student was assigned at random to one of the two groups, the greater gains made by READ USA students are indicative of greater reading growth for those who received the intervention. Table 6 presents the results from the regression analyses on the GORT pretest-posttest gain scores. Each column in the table represents separate regression models by GORT test. The top of the table presents the regression coefficients and whether each was statistically significant for the initial models in which the READ USA, demographic, grade, and school variables were included in the models. The bottom of the table included the regression coefficients for the interaction terms.

Confirming what was found from the graphical displays of the means (Figures 1-5), READ USA students significantly outgained their peers in the control group on all GORT measures except for Rate. Other than the READ USA variable, very few other variables were statistically significant in any of the initial models. Grade 4 students overall had lower Accuracy gains than students in the other grades, however, students across all demographic subgroups did not differ significantly in their average GORT gain scores, and the gains were rather consistent across grades and schools.

Among the interaction terms, Grade 4 treatment students outgained their Grade 4 control peers on Accuracy, and the treatment effect was found to be larger in Accuracy and the Sum at S.A. Hull, and at Long Branch, the treatment effect was larger on average on Comprehension and the Sum score. No other interaction terms were significant, revealing that the READ USA

effect was rather consistent across all demographic variables and grade levels.

The READ USA coefficients in Table 6 represent the mean gain difference between the treatment and control groups. Dividing the coefficients by the respective stand deviations of each gain scale provides and index of effect size. The effect sizes were: 0.42 for Accuracy, 0.26 for Fluency, 0.25 for Comprehension, and 0.28 for the Sum score.

FAST Outcomes. Figures 6 and 7 show the PM1 and PM2 performance levels, respectively, for READ USA and comparison students in Grades 3-5. Note that at the beginning of the year (Figure 6), about 90% of READ USA and about 75% of control students performed well below grade level expectations. Greater proportions of control students scored at either the next level up (below grade level) or on grade level, although very few students in both groups scored at grade level on PM1.

Table 6. GORT Pretest-Posttest Gain Score Linear Regression Results, Autumn 2023

Variable	Accuracy	Rate	Fluency	Comprehension	Sum
Intercept	-.10	-.14	-.01	-.02	-.02
READ USA (T)	.77***	.12	.42*	.55*	.86**
Male (G)	-.23	.15	.13	-.21	-.08
Minority (M)	.47	.17	.22	.28	.50
Free Meal (FM)	.04	.23	-.06	-.05	-.11
English Leaner (EL)	.37	.13	.30	.46	.76
Disability (D)	.19	.02	.07	.46	.53
Grade 2 (2)	-.06	.38	-.18	.41	.24
Grade 4 (4)	-.55*	.22	-.16	-.40	-.56
Grade 5 (5)	-.38	-.32	-.45	.04	-.41
Arlington (A)	.26	-.06	.44	.10	.54
Long Branch (L)	.25	-.58	-.16	.48	.32
S.A Hull (H)	-.13	.14	-.09	.94	.86
Interaction Terms					
T*G	-.48	.10	-.29	-.23	-.51
T*M	-.59	-.52	-.24	.71	.47
T*FM	.32	.08	-.31	-.93	-1.24
T*EL	.30	.08	-.36	-.21	-.57
T*D	.18	-.36	-.16	.49	.33
T*2	-.46	-.88	-.78	.13	-.65
T*4	1.07*	-.04	.45	.88	1.33
T*5	.13	-.60	-.51	.24	-.27
T*A	.85	.38	.57	.92	1.49
T*L	1.19	.40	.77	2.26*	3.03*
T*H	2.61**	1.16	1.65	2.03	3.68*

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

At PM2, there remained a greater proportion of READ USA students well below grade level, but as can be seen in Figure 7, the difference in proportions between the two groups has diminished considerably. About 58% of READ USA and 56% of control students were well below grade level, which represented a difference shift of about 15 percentage points at PM1 to two percentage points at PM2. At midyear, there was an equal percentage of students from each group who were below grade level, more students on grade level in the control group, and an equal proportion of students in each group who scored at the proficient level.

Figure 8 presents the PM1 to PM2 performance level change category by group. A slightly greater proportion of control students lost a performance level. Fewer READ USA students had no change, and a greater proportion of treatment students gained one level. An equal proportion from each group gained two performance levels. Taken together, it is evident that READ USA students made a greater shift in performance level from PM1 to PM2.
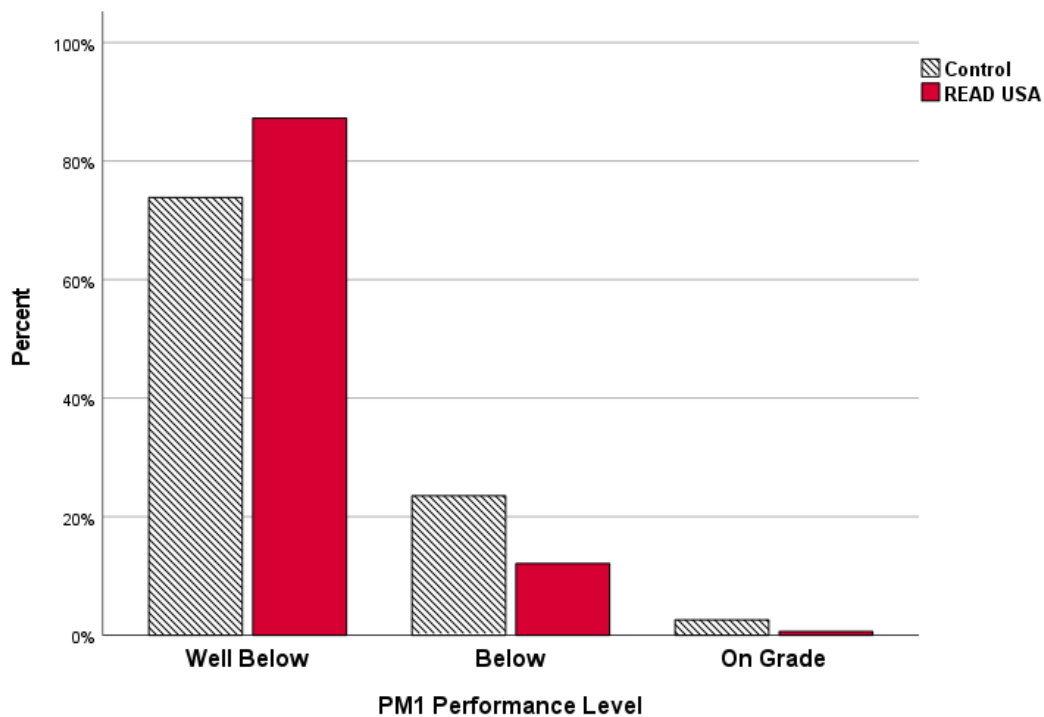

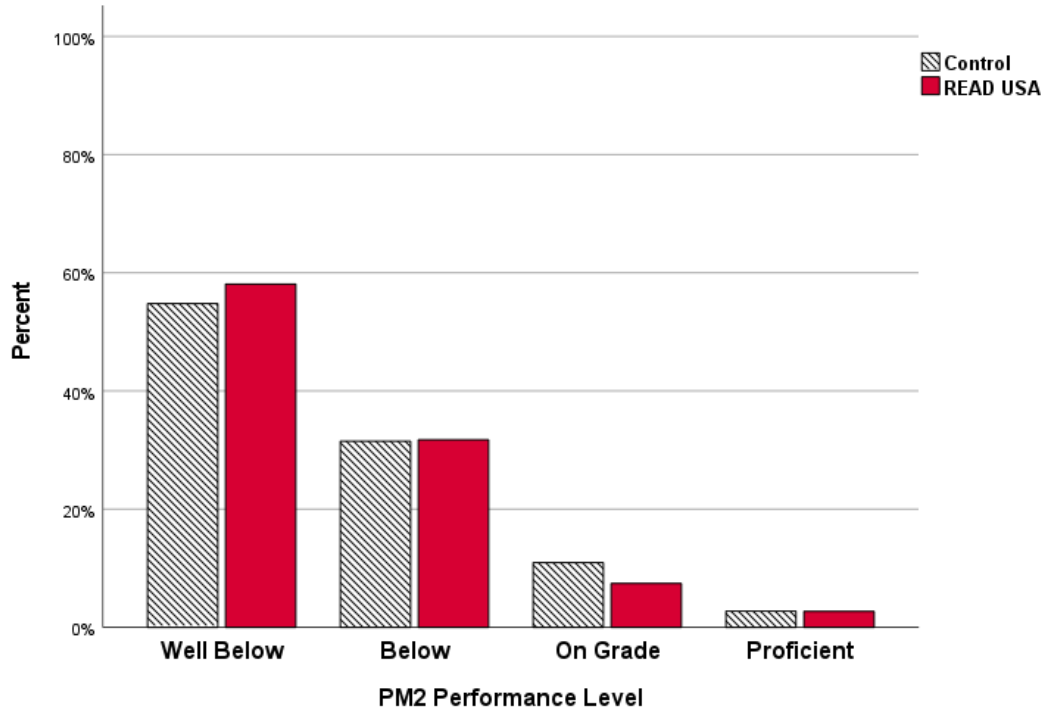*Figure 6*. FAST PM1 Performance Levels by Group at Pretest

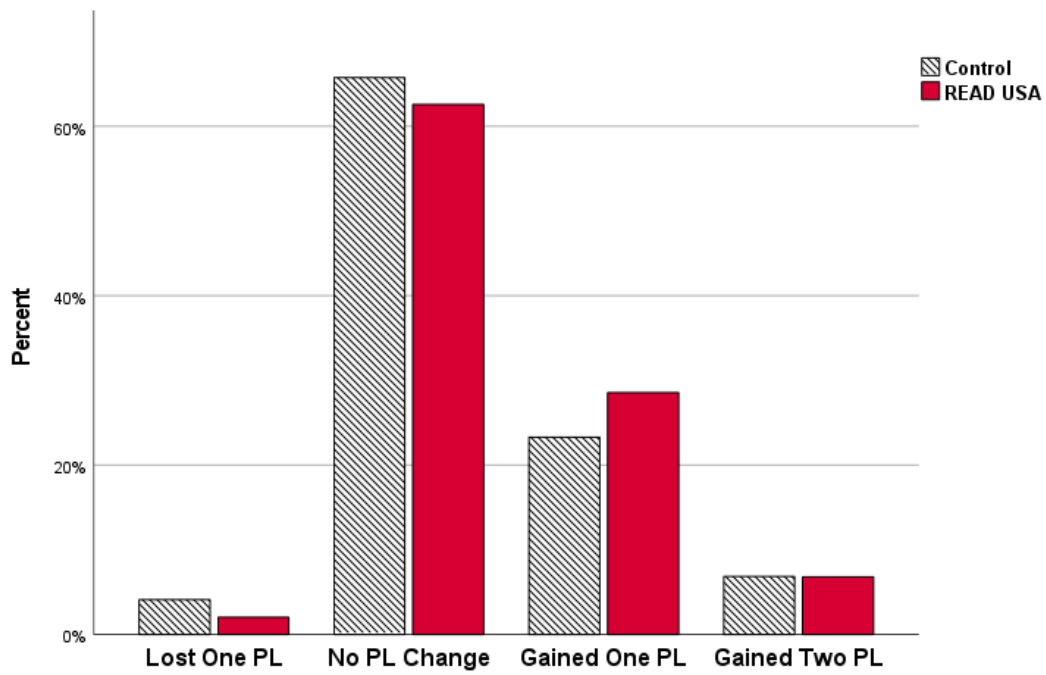*Figure 7*. FAST PM2 Performance Levels by Group at Posttest

Figure 8. FAST PM1 to PM2 Performance Level Change by Group

The ostensible greater shift for READ USA students as depicted in Figure 8 does not consider student background, grade, or school factors. Table 4 presents the results from the ordinal regression analysis, which provides an estimate of the treatment effect while considering the other variables of interest. The coefficients in the table do not represent predicted change on the outcome per unit change on the independent variable as they do in Table 3. Instead, the coefficients represent the log odds that students in the subgroup listed in the table were more likely to be at a higher category of PM1 to PM2 performance level change.

As can be seen from Table 7, READ USA students had a 0.76 greater likelihood of improving more on the FAST between PM1 and PM2 than control students. Converting the log odds into a percentage, READ USA students were 68% more likely to have a greater FAST change than their peers. Minority students also were more likely to improve more on the FAST. Minority students were 66% more likely to improve more than white and Asian American students. Students at Arlington and S.A. Hull also were more likely to improve: 78% more likely at Arlington and 85% more likely at S.A Hull. The only moderating effect found on the FAST was for 4th-grade students. READ USA was less effective for 4th-grade students than students in 3rd- and 5th-grade.

Table 7. FAST Ordinal Regression Results, Autumn 2023

Variable	FAST PL Gain
READ USA (T)	.76*
Male (G)	-.24
Minority (M)	.65*
Free Meal (FM)	-.19
English Leaner (EL)	.30
Disability (D)	.00
Grade 4 (4)	-.12
Grade 5 (5)	-.37
Arlington (A)	1.25***
Long Branch (L)	-.66
S.A Hull (H)	1.76***
Interaction Terms	
T*G	-.28
T*M	.16
T*FM	.40
T*EL	.57
T*D	-.08
T*4	-1.22*
T*5	.17
T*A	.21
T*L	-.88
T*H	.04

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

An effect size can be computed for log odds values by dividing the value by 1.81. Dividing the coefficient for the treatment effect from Table 7 (.76/1.81) yields an effect size estimate of 0.42 on the FAST performance level change variable.

Summary

Prior studies have consistently demonstrated positive effects for students who participated in the literacy tutoring program, READ USA; however, this is the first study to use a randomized control trial (RCT) design to study its impacts. This is important because an RCT employs an equivalent control group for comparison, thus allowing researchers to gauge how the students would have performed on the tests from pretest to posttest had they not been provided the intervention. As such, researchers can say with greater confidence that changes in achievement were due to participating in READ USA and less likely due to other factors unrelated to the intervention.

In the current study, students in Grades 2-5 in seven participating DCPS schools were randomly assigned to either receive the intervention in autumn 2023 or in spring 2024; students randomly selected to participate in spring 2024 served as the comparison group. All students were administered assessments at the beginning and end of the autumn semester, 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). This report provides a description of the research design and outcomes from the autumn 2023 RCT study.

Results from this experiment show that students in all grades who participated in READ USA outperformed the comparison group in terms of pretest and posttest gains on two of the three independent GORT measures (Accuracy and Comprehension), and on the two combined measures (Fluency and the Sum score). READ USA and control students did not significantly differ on the third independent measure (Rate). 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 notable.

Moreover, the READ USA effect was rather consistent across all demographic variables; meaning, the positive effect of the intervention was the same no matter student background characteristics. It is also noted that the impact of the intervention on the GORT varied across some schools, however, in all schools the effect was positive.
In terms of the Florida state progress monitoring assessment, the FAST, students participating

in READ USA again outperformed the comparison group, this time on performance level change from PM1 to PM2. There was a grade level effect observed however, in that READ USA students appeared less effective for 4th grade students and more effective for students in 3rd and 5th grade. As found with the GORT assessment, the positive effects were found no matter the student's race, gender, language status, economic status, or disability. Furthermore, unlike with the GORT, there were no differences in the FAST effect across schools.

In comparison to other literacy interventions, the magnitude of the effects on the GORT (0.42 for Accuracy,0.26 for Fluency, 0.25 for Comprehension, and 0.28 for the Sum score) and FAST (0.42) for READ USA are from average to above average. 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).

It's also important to consider that the great majority of the interventions included in the meta-analyses 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. Future research should focus on the specific interactions between tutors and tutees that produce the most powerful effects.

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.

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.