

Assessing the Impact of Really Great Reading in Florida: A Quasi-Experimental Study

Study Type: ESSA Evidence Level II

Prepared for:
Really Great Reading

Prepared by LearnPlatform:
Alexandra Lee, Ph.D., Researcher, LearnPlatform
Meetal Shah, Ph.D., Senior Researcher, LearnPlatform

December 8, 2023

EXECUTIVE SUMMARY

Really Great Reading (RGR) contracted with LearnPlatform by Instructure, a third-party edtech research company, to examine the impact of usage of its reading program on student reading outcomes. LearnPlatform designed the study to satisfy Level II requirements (Moderate Evidence) according to the Every Student Succeeds Act (ESSA).

Study Sample, Measures, and Methods

This study occurred during the 2022-23 school year. The sample included 1,181 students enrolled in Kindergarten-Grade 2 from five schools in a large public school district. In terms of demographics, the sample included students who identified as female (46%), male (54%), Hispanic (45%), White (31%), Black (14%), multi-racial (7%), and Asian (2%). Nearly 10% percent of the students were receiving special education services and three percent were English language learners (ELLs).

Researchers identified students¹ that received RGR reading instruction by verifying that their school used RGR for reading instruction. Reading achievement was measured using DIBELS® 8th Edition scores. Taken together, these measures allowed researchers to investigate patterns in RGR implementation and potential impacts of program use on students' reading achievement.

Researchers used a variety of quantitative analytic approaches to answer the research questions. First, researchers used descriptive statistics to examine participant characteristics and implementation of the program. Researchers then used multilevel regression models to examine whether RGR use was associated with significant differences in students' reading scores in spring 2023, controlling for their baseline scores in fall 2022. The multilevel regression analyses included student-level covariates to control for potential selection bias (i.e., gender, race, special education, and ELL designation) and a random effects term to control for classroom-level differences. In addition, researchers calculated standardized effect sizes (Hedge's *g*) to make the model-predicted changes in student outcomes more interpretable.

¹ Students were also considered to be "users" of RGR if their teacher was in a school that used RGR for instruction.

Student Outcomes



Kindergarten students who received Really Great Reading (RGR) instruction (*Countdown*) had significantly higher DIBELS® *composite* reading scores at the end-of-the-year than students who did not receive RGR-aligned instruction in 2022-2023.



Kindergarten students who received Really Great Reading (RGR) instruction (*Countdown*) had significantly higher DIBELS® *letter naming fluency* scores at the end-of-the-year than students who did not receive RGR-aligned instruction in 2022-2023.



Kindergarten students who received Really Great Reading (RGR) instruction (*Countdown*) had significantly higher DIBELS® *nonsense word fluency (correct letter sounds)* scores at the end-of-the-year than students who did not receive RGR-aligned instruction in 2022-2023.



Teachers' use of RGR instruction in the 2022-2023 school year was not associated with significant differences in reading achievement among Grade 1 (*Blast*) and Grade 2 (*HD Word*) students.

Conclusions

This study provides results to satisfy ESSA evidence requirements for Level II (Moderate Evidence) given the quasi-experimental study design, positive statistically significant findings, and large sample size.

TABLE OF CONTENTS

Introduction	3
Program Implementation Research Question	3
Effectiveness Research Question	3
Methods	4
Program Implementation Findings	6
Program Effectiveness Findings	8
Conclusions and Recommendations	10
References	11
Appendix A. Really Great Reading Logic Model	12
Appendix B. Additional Information on Study Design and Methods	13
Appendix C. Additional Information on Kindergarten-Grade 2 Outcome Findings	20

Introduction

Recognizing that 65% of students cannot read proficiently by 4th grade (McFarland et al., 2019), which is due in part to the enduring research-practice gap (Schneider, 2018), Really Great Reading (RGR) provides teachers with the tools to implement research-based, science of reading instruction to help students develop word-level reading skills using phonics, phonemic awareness, orthographic mapping, and deciphering word meaning.

As part of their ongoing efforts to demonstrate the efficacy of its reading program, RGR contracted with LearnPlatform by Instructure, a third-party edtech research company, to examine the relationship between usage of its program and student outcomes. After collaborating on the development of an updated logic model (Appendix A) for RGR (Lee et al., 2023), LearnPlatform designed a study to satisfy ESSA Level II requirements (Moderate Evidence) with the following research questions.

Program Implementation Research Question

1. Were there significant differences between students who received RGR instruction and those who did not in terms of:
 - a. Demographics (i.e., race and gender)?
 - b. Special education designation?
 - c. English Language Learner designation?

Effectiveness Research Question

2. After controlling for students' prior reading levels, what is the magnitude of observed differences of students who received RGR instruction compared to students who did not receive RGR instruction on standardized reading assessments?

Methods

This section of the report briefly describes the setting, participants, measures, and analysis methods.

Setting

The study included one large public school district in Florida during the 2022-23 school year. The sample included 1,181 Kindergarten–Grade 2 students from five schools.

Participants

Sixty-two percent ($n = 737$) of the students in the sample attended a school where RGR was used for reading instruction; these students were considered to be in the intervention group for the purpose of the comparative analyses. The remaining 38% ($n = 444$) students were considered to be in the control group as they received reading instruction using a different program (not RGR).

In terms of demographics, the sample included students who identified as female (46%), male (54%), Hispanic (45%), White (31%), Black (14%), multi-racial (7%), and Asian (2%) (see Figure 1). Nearly 10% percent of the students were receiving special education services and three percent were English language learners (ELLs; see Figure 2).

Measures

This study included the following measures to provide insights into Really Great Reading (RGR) implementation and evidence about the potential impacts of the program on student outcomes.

RGR Use. Researchers independently verified which schools used RGR for reading instruction and which schools did not via email with the Curriculum Coordinator for the school district. Of the five schools included in the study, three schools used RGR for reading instruction and two schools did not. Among the schools that used RGR, each grade level used a different, grade specific RGR program: Countdown in Kindergarten, Blast in Grade 1, and HD Word in Grade 2. These programs were implemented for Tier I² instruction.

Standardized Student Assessments. Researchers used DIBELS® 8th Edition to assess reading outcomes. DIBELS is a standardized research-based assessment that reliably measures reading ability and progress from Kindergarten through Grade 8. DIBELS can be used to assess different reading skills, which are dependent on students' grade levels. For the purpose of the present study,

² According to the MTSS (Multi-Tiered Systems of Support), Tier I instruction is used as core/universal instruction for students who are making expected academic progress (typically 80-90% of students), Tier II instruction is the first level of intervention for student who do not show progress in Tier I, and Tier III instruction is used when a student a needs intensive individualized intervention because they do not show progress in Tier II.

researchers used composite scores, letter naming fluency, phonemic segmentation fluency, and nonsense word fluency (correct letter sounds) to assess outcomes for Kindergarten students; composite scores, letter naming fluency, and nonsense word fluency (correct letter sounds) to assess outcomes for Grade 1 students; and, oral reading fluency (words correct) to assess outcomes for Grade 2 students.

Data Analysis

Researchers used a variety of quantitative analytic approaches to answer the research questions. First, researchers used descriptive statistics to examine student characteristics and implementation of the program. Next, researchers evaluated whether linear regression models or multilevel regression models would be best for the outcomes analysis. Multilevel models had better model fit³ than regression models as they accounted for classroom-level random effects. Given this, researchers used multilevel regression models for each grade level that accounted for classroom-level differences and used students' membership in the intervention (i.e., RGR user) or comparison (i.e., RGR non-user) condition as a categorical predictor of reading achievement. All multilevel models included beginning-of-year DIBELS scores, gender, race, ELL, and SPED designation as covariates to control for potential selection bias. All findings were interpreted as statistically significant at the $p < .05$ level; Hedge's g effect sizes and improvement index conversions are included to assist with interpretation.

Baseline Equivalence

To ensure the validity of the study's findings and to adhere to ESSA Level II standards, the researchers assessed the equivalence of student demographic characteristics (i.e., race, gender, ELL, and SPED designation) and standardized assessment scores at baseline between intervention and comparison groups. The demographics factors that were not equivalent at baseline (i.e., race/ethnicity for Kindergarten-Grade 2 students and gender for Grade 2 students) were included as covariates in all analyses.

Fall 2022 DIBELS® scores were examined separately for each grade level to verify that there were no significant differences between students who used RGR and those who did not. Propensity score weighting was used to account for different predicted probabilities for each condition. The results of regression tests indicated that the differences at baseline in terms of DIBELS® scores were within the boundary for statistical adjustment according to the What Works Clearinghouse (WWC) criteria (WWC, 2022) for all models included in the outcomes analysis⁴. Specifically, statistically significant baseline differences with an effect size between 0.05 and 0.25 are acceptable with statistical adjustment. See Appendix B for more details regarding baseline equivalence.

³ The best fitting model for each grade was identified using AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and partial eta-squared effect sizes (Burnham & Anderson, 2004)

⁴ Baseline equivalence was not achieved for phonemic segmentation fluency for Kindergarten and Grade 1 students and oral reading fluency (words correct) for Grade 1 students, so these models were not used for outcomes analysis.

Program Implementation Findings

RGR was implemented for Kindergarten – Grade 2 students at three racially diverse schools in a public school district in south Florida.

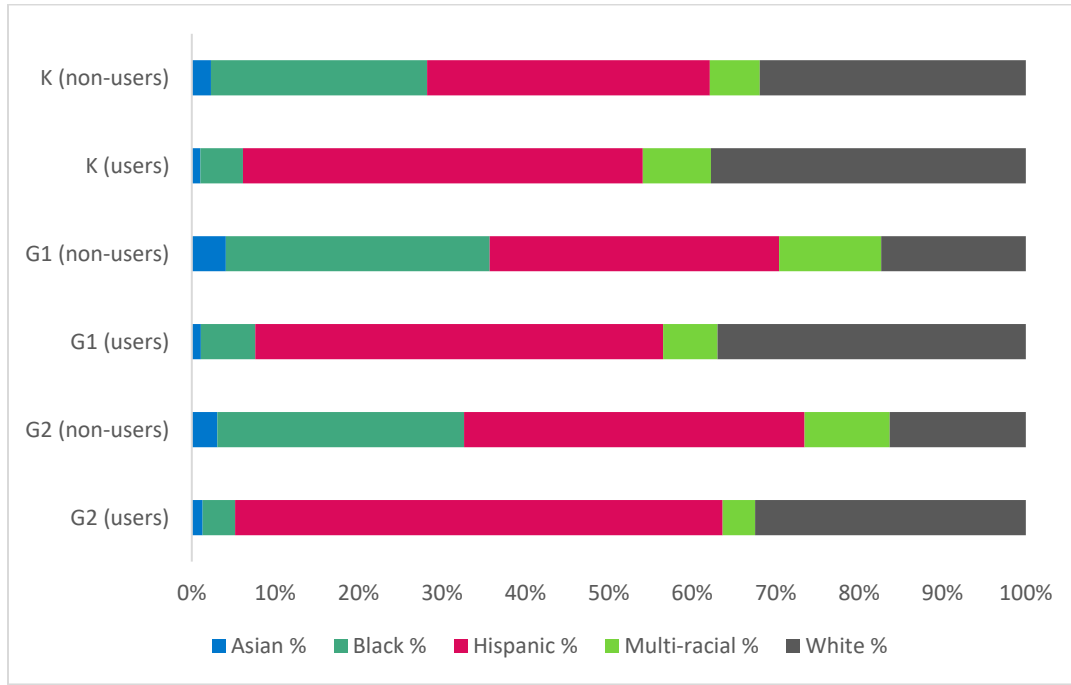


Figure 1. Overall distribution of RGR users by race and grade designation.

Students who were designated as English Language Learners (ELLs) students used RGR across Kindergarten – Grade 2.

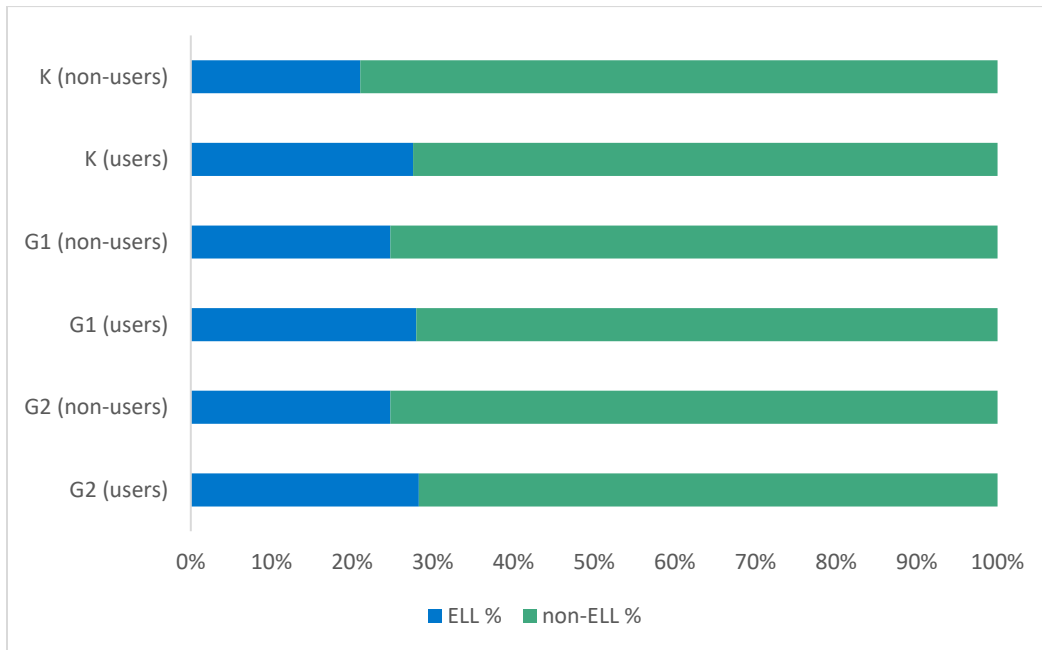


Figure 2. Overall distribution of RGR users by ELL and grade designation.

Students who were designated as Special Education (SPED) students used RGR across Kindergarten – Grade 2.

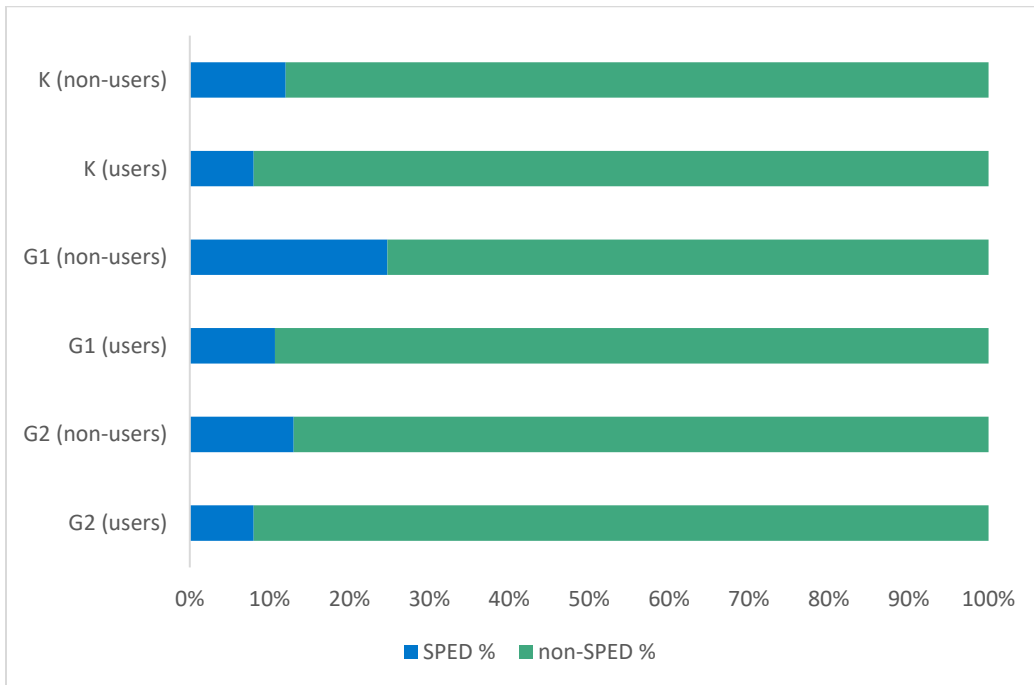


Figure 3. Overall distribution of RGR users by SPED and grade designation.

Program Effectiveness Findings

To answer the remaining study research questions, researchers used multilevel regression analysis. In addition to examining the statistical significance of the tests used, researchers calculated standardized effect sizes and used the improvement index to determine the magnitude of the relationship between RGR usage and student reading outcomes. The key study findings are included below, and the full set of results can be found in Appendix C.

Differences in Reading Outcomes Among Kindergarten Students who Received RGR Instruction and Students Who Did Not

To understand whether RGR positively impacted students' reading outcomes, researchers conducted a series of multilevel regression analyses for distinct reading outcomes (i.e., letter naming fluency, nonsense word fluency (correct letter sounds), composite scores, and oral reading fluency) controlling for baseline scores (fall 2022), gender, race, ELL and SPED designation. Overall, Kindergarten students had significantly greater reading scores compared to those who did not receive RGR-aligned instruction. There were no significant differences for Grade 1 and Grade 2 students.

Key Finding 1. Kindergarten students who used RGR (*Countdown*) had significantly higher DIBELS® composite scores compared to students who did not use RGR (Figure 3). This means that for a student at the 50th percentile who used RGR, they would be expected to move up to the 72nd percentile (i.e., 22 p.p. improvement) compared a student who did not use the program (Hedge's $g = .57, p < .001$).

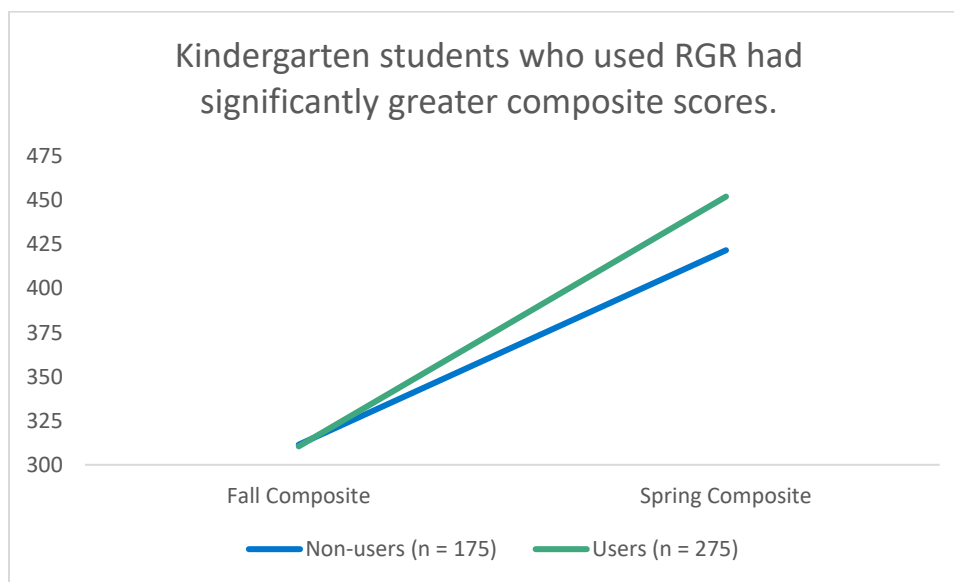


Figure 4. Adjusted marginal means by condition for baseline and outcome composite scores.

Key Finding 2. Kindergarten students who used RGR (*Countdown*) had significantly higher DIBELS® letter naming fluency scores compared to students who did not use RGR (Figure 4). This means that for a student at the 50th percentile who used RGR, they would be expected to move up to the 69th percentile (i.e., 19 p.p. improvement) compared a student who did not use the program (Hedge’s $g = .50, p < .001$).

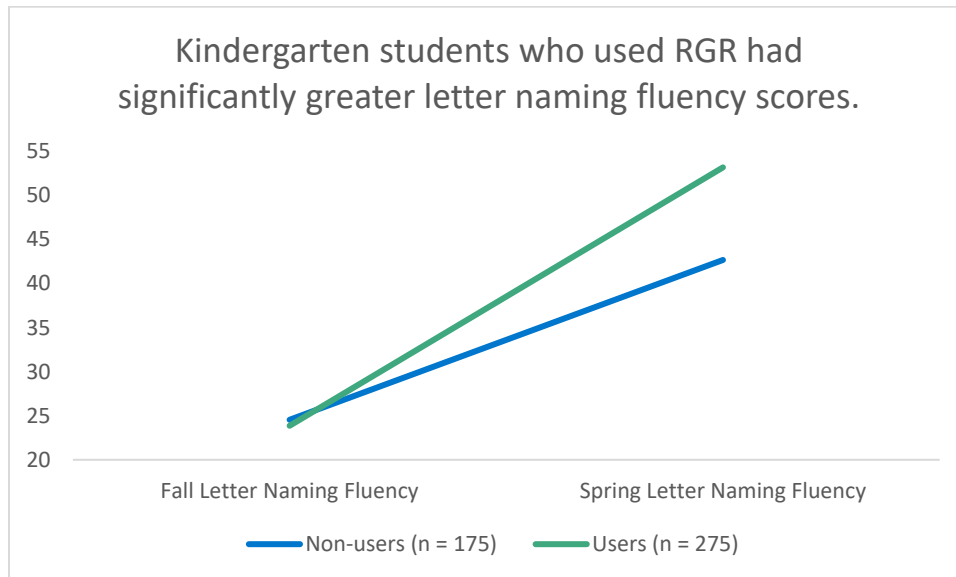


Figure 5. Adjusted marginal means by condition for baseline and outcome letter naming fluency scores.

Key Finding 3. Kindergarten students who used RGR (*Countdown*) had significantly higher DIBELS® nonsense word fluency (*correct letter sounds*) scores compared to students who did not use RGR (Figure 5). This means that for a student at the 50th percentile who used RGR, they would be expected to move up to the 67th percentile (i.e., 17 p.p. improvement) compared a student who did not use the program (Hedge’s $g = .43, p < .001$).

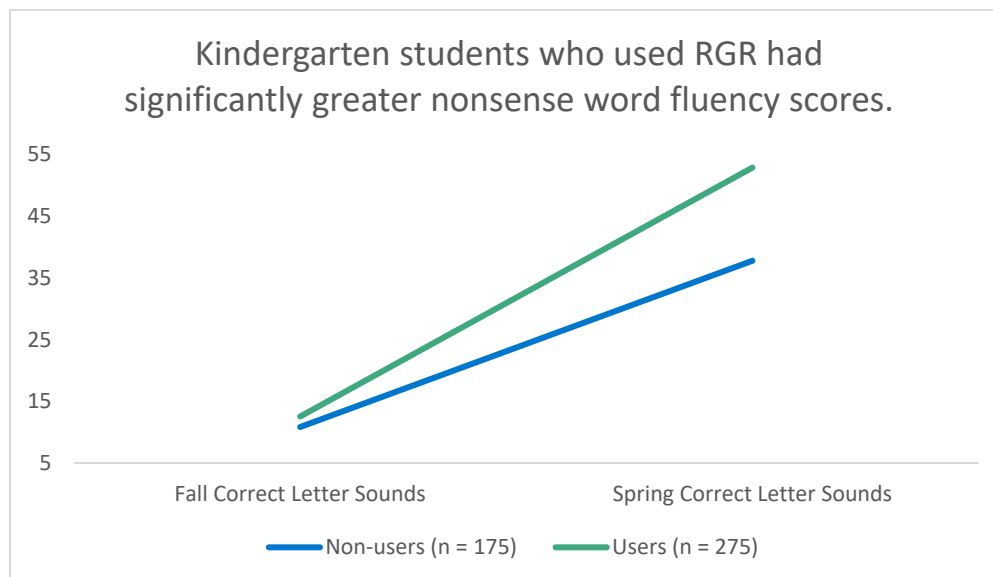


Figure 6. Adjusted marginal means by condition for baseline and outcome composite scores.

Conclusions and Recommendations

In sum, the results of this study suggest that there is a positive effect of Really Great Reading-aligned reading instruction on Kindergarten students' reading outcomes. The data indicate that students who received RGR instruction had positive effects on reading outcomes at the end of the school year compared to students who do not receive RGR instruction, even though they started the school year at equivalent levels. It is also notable that this study took place in a context that serves diverse students – this study provides evidence that RGR is beneficial for *all kindergarten students* regardless of their racial background, and special education and ELL designation.

Given the positive outcome findings of the impact analysis among the sample, this study provides results to satisfy ESSA evidence requirements for Level II (*Moderate Evidence*). Specifically, this study met the following criteria:

- ✓ Quasi-experimental design
- ✓ Proper design and implementation
- ✓ Baseline equivalence for treatment and comparison groups
- ✓ Statistical controls through covariates
- ✓ At least one statistically significant, positive finding
- ✓ No statistically significant, negative findings

Researchers recommend the following next steps for the RGR team:

- seek to gather more detailed data about the different ways educators use the program (e.g., types and frequency of RGR learning activities completed by students) to inform specific implementation recommendations.

References

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33, 261–304.

Lee, A., Wall, A., & Shah, M. (2023). *Really Great Reading logic model: ESSA level IV study*. LearnPlatform by Instructure.

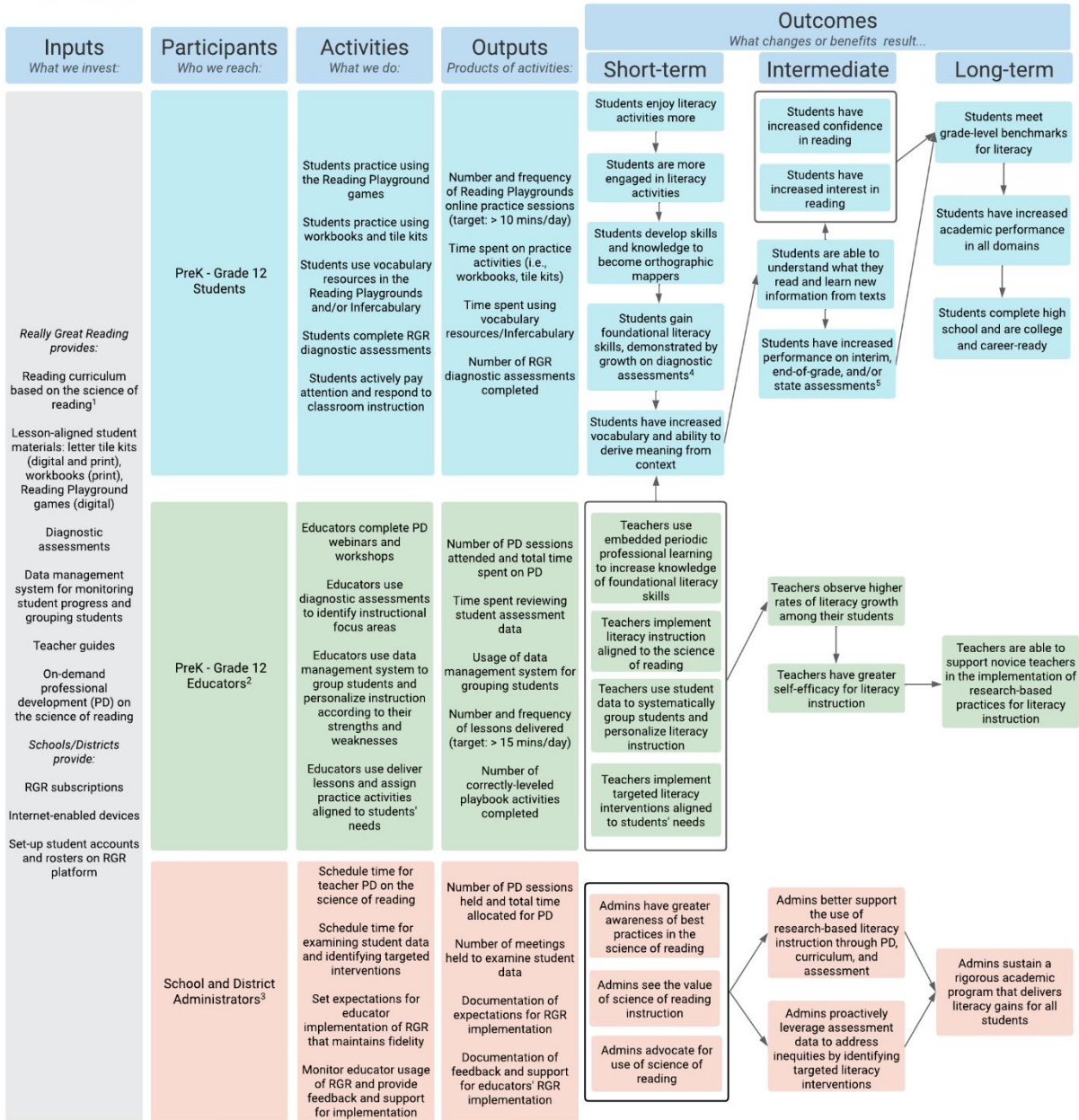
What Works Clearinghouse. (2022). What Works Clearinghouse procedures and standards handbook, version 5.0. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at <https://ies.ed.gov/ncee/wwc/Handbooks>

Appendix A. Really Great Reading Logic Model

LOGIC MODEL



Problem Statement: Sixty-five percent of students are not able to read proficiently by 4th grade, which is due in part to the enduring research-practice gap. Really Great Reading (RGR) provides teachers with the tools to implement research-based science of reading instruction, which helps students develop word-level literacy using phonics, phonemic awareness, orthographic mapping, and deciphering word meaning.



¹ The science of reading is a set of research-based practices that support the development of reading by helping students to relate written text to spoken language by focusing on malleable factors that underpin reading ability, including phonics, phonemic awareness, and decoding (Peterscher et al., 2020).

² Educators may include primary classroom teachers, interventionists, Title I coordinators, paraprofessionals, and special educators.

³ School and District Administrators may include literacy/ELA instructional coaches, curriculum specialists, special education directors, district-level PD directors, principals.

⁴ Foundational literacy skills for students in Pre-K - 2nd grade include phonemic awareness, alphabetic principals, and oral reading fluency, skills for students in 3rd grade or higher include decoding, word reading, and reading fluency.

⁵ Examples of interim, end-of-grade, state assessments include DIBELS, iReady, NWEA, and MAP.



Appendix B. Additional Information on Study Design and Methods

Propensity Score Weighting

To help make the Kindergarten-Grade 2 student groups (i.e., students who received RGR-aligned instruction and students who did not) as comparable as possible, propensity score weights were calculated for each student. To calculate propensity scores, researchers conducted binary logistic regression with student group as the dependent variable and fall 2022 DIBELS® scores gender, race, ELL and SPED designation as the covariates. The probability was saved as a new variable. Weights were calculated by dividing one by the probability (one/probability). Students without a weight were dropped from the final analytic sample. All analyses that included students who did not receive RGR-aligned instruction included these weights.

Baseline Equivalence

Researchers conducted baseline equivalence analyses to determine whether there were baseline differences in characteristics between students who used RGR and students who did not during the 2022-23 school years. Specifically, researchers used chi-square analyses on student-level demographics and linear regressions for DIBELS® reading scores for the three grade levels included in the study.

Baseline Equivalence for Kindergarten Students

As presented in Table B1, results from the chi-square analyses reveal that there was baseline equivalence for Kindergarten students for gender, SPED, and ELL designations. However, there were significant differences for race, so this variable was included as a covariate in the outcomes analysis.

Table B1. Baseline Equivalence Analysis of Kindergarten Students Demographics by User Group

Characteristic	Students who used RGR (n = 280)		Students who did not use RGR (n = 175)		Chi-squared	p-Value
	Percent	N	Percent	N		
<i>Gender</i>						
Male	48.93	137	55.43	97	4.51	.105
Female	49.29	138	44.57	78		
<i>Race</i>						
Hispanic	47.14	132	34.29	60	43.54	<.001
White	37.14	104	32.00	56		
Black	5.36	15	25.71	45		
Multi-racial	7.50	21	5.71	10		
Asian	1.07	3	2.29	4		
<i>Special Education Designation</i>						
Yes	8.21	23	12.00	21	1.77	.184
No	91.79	257	88.00	154		
<i>English Language Learner Status</i>						
Yes	26.79	75	20.57	36	5.75	.056
No	71.43	200	79.43	139		

Note. There were five students who used RGR who were missing gender, race, and ELL information.

As presented in Table B2, results from the weighted regression analysis reveal that there was baseline equivalence for Kindergarten students for DIBELS® composite, letter naming fluency, and nonsense word fluency (correct letter sounds) in fall 2022. However, baseline scores for phonemic segmentation fluency were not equivalent at baseline (i.e., Hedge’s *g* effect size > .25), so outcomes analysis was not conducted for this outcome.

Table B2. Baseline Equivalence Analysis of Kindergarten Students by User Group

Baseline Variable	Effect Size	<i>p</i> -value
DIBELS® composite scores fall 2022	-0.02	.861
DIBELS® letter naming fluency scores fall 2022	-0.04	.691
DIBELS® correct letter sounds fluency scores fall 2022	0.13	.297
DIBELS® phonemic segmentation fluency scores fall 2022	0.46	<.001

Baseline Equivalence for Grade 1 Students

As presented in Table B3, results from the chi-square analyses reveal that there was baseline equivalence for Grade 1 students for gender, SPED, and ELL designations. However, there were significant differences for race, so this variable was included as a covariate in the outcomes analysis.

Table B3. Baseline Equivalence Analysis of Grade 1 Students Demographics by User Group

Characteristic	Students who used RGR (n = 320)		Students who did not use RGR (n = 104)		Chi-squared	p-Value
	Percent	N	Percent	N		
<i>Gender</i>						
Male	48.12	154	59.62	62	5.35	.069
Female	44.69	143	37.50	39		
<i>Race</i>						
Hispanic	45.00	144	33.65	35	57.74	<.001
White	34.38	110	17.31	18		
Black	6.25	20	30.77	32		
Multi-racial	6.25	20	11.54	12		
Asian	0.94	3	3.85	4		
<i>Special Education Designation</i>						
Yes	8.44	27	12.50	13	1.52	.218
No	91.56	293	87.50	91		
<i>English Language Learner Status</i>						
Yes	26.25	84	24.04	25	3.01	.222
No	66.56	213	73.08	76		

Note. There were 23 students who used RGR and three students who did not use RGR who were missing gender, race, and ELL information.

As presented in Table B4, results from the weighted regression analysis reveal that there was baseline equivalence for Grade 1 students for DIBELS® composite, letter naming fluency, and nonsense word fluency (correct letter sounds) in fall 2022. However, baseline scores for phonemic segmentation fluency and oral reading fluency (words correct) were not equivalent at baseline (i.e., Hedge’s *g* effect size > .25), so outcomes analysis was not conducted for these outcomes.

Table B4. Baseline Equivalence Analysis of Grade 1 Students by User Group

Baseline Variable	Effect Size	<i>p</i> -value
DIBELS® composite scores fall 2022	-0.08	.634
DIBELS® letter naming fluency scores fall 2022	0.14	.238
DIBELS® correct letter sounds fluency scores fall 2022	0.03	.849
DIBELS® phonemic segmentation fluency scores fall 2022	0.80	<.001
DIBELS® oral reading (words correct) fluency scores fall 2022	-0.27	.165

Baseline Equivalence for Grade 2 Students

As presented in Table B5, results from the chi-square analyses reveal that there was baseline equivalence for Grade 2 students for SPED designation. However, there were significant differences for gender, race, and ELL designation, so these variables were included as a covariate in the outcomes analysis.

Table B5. Baseline Equivalence Analysis of Grade 2 Students Demographics by User Group

Characteristic	Students who used RGR (n = 214)		Students who did not use RGR (n = 171)		Chi-squared	p-Value
	Percent	N	Percent	N		
<i>Gender</i>						
Male	41.59	89	54.97	94	36.51	<.001
Female	35.51	76	43.27	74		
<i>Race</i>						
Hispanic	44.86	96	39.77	68	90.06	<.001
White	25.23	54	16.37	28		
Black	2.80	6	29.24	50		
Multi-racial	3.27	7	9.94	17		
Asian	0.93	2	2.92	5		
<i>Special Education Designation</i>						
Yes	7.01	15	12.87	22	3.75	.053
No	92.99	199	87.13	149		
<i>English Language Learner Status</i>						
Yes	28.04	60	33.92	58	36.50	<.001
No	49.07	105	64.33	110		

Note. There were 49 students who used RGR and three students who did not use RGR who were missing gender, race, and ELL information.

As presented in Table B6, results from the weighted regression analysis reveal that there was baseline equivalence for Grade 2 students for DIBELS® oral reading fluency (words correct) in fall 2022. However, baseline scores for nonsense word fluency (correct letter sounds) were not equivalent at baseline (i.e., Hedge's *g* effect size > .25), so outcomes analysis was not conducted for this outcome.

Table B6. Baseline Equivalence Analysis of Grade 2 Students by User Group

Baseline Variable	Effect Size	<i>p</i> -value
DIBELS® correct letter sounds fluency scores fall 2022	-0.99	<.001
DIBELS® oral reading (words correct) fluency scores fall 2022	-0.17	.418

Appendix C. Additional Information on Kindergarten-Grade 2 Outcome Findings

Differences in Reading Outcomes Among Kindergarten-Grade 2 Students who received Really Great Reading Instruction and Students Who Did Not

Table C1. DIBELS® scores by RGR usage group (covariates: fall 2022 DIBELS® scores (baseline), gender, race, SPED, and ELL designation)

Grade Level	Outcome	Coefficient	Standard Error	t-value	p> t	Effect Size
Kindergarten (Countdown)	Composite	30.40	7.11	4.27	<.001***	0.57
	Letter naming fluency	10.50	3.08	3.40	.001**	0.50
	Correct letter sounds	15.09	4.98	3.03	.004**	0.43
Grade 1 (Blast)	Composite	3.05	9.66	0.32	.754	0.06
	Letter naming fluency	1.73	2.49	0.69	.495	0.10
	Correct letter sounds	0.54	9.37	0.06	.955	0.01
Grade 2 (HD Word)	Oral reading fluency	-3.32	11.55	-0.29	.777	-0.06

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001