

FINAL REPORT

Impact Evaluation of the Minnesota Reading Corps K-3 Program (2017-18)

OCTOBER 2018

PRESENTED TO:
ServeMinnesota
120 South Sixth Street,
Suite 2260
Minneapolis, MN 55402

PRESENTED BY:
NORC at the University of Chicago
1155 E 60th Street
Chicago, IL 60637

AUTHORS:
Carrie E. Markovitz
Marc W. Hernandez
Eric C. Hedberg
Heidi H. Whitmore
Jennifer L. Satorius

Acknowledgement

Many individuals and organizations contributed to the design and implementation of the *Impact Evaluation of the Minnesota Reading Corps K-3 Program (2017-18)*. While it is not possible to name everyone, we would like to acknowledge some of the individuals and organizations who played a significant role in the successful execution of this study. We would like to begin by extending a special thank you to the select schools, their staff, and the Reading Corps tutors serving in the Minnesota Reading Corps program who agreed to have their students randomly assigned for program participation and for allowing our field assessors to collect assessment data from their students. We would like to acknowledge the contributions of the study site liaisons, Kelly Holt, Julie Kubelka, and Jennifer Scolese, who coordinated and scheduled the random assignment process and data collection at individual schools, and the contributions of the field assessors who collected assessment data from the K-3rd grade students. We would like to thank the Corporation for National and Community Service (CNCS) for providing the study data from the 2014 *Impact Evaluation of the Minnesota Reading Corps K-3 Program*, so we could conduct a comparison study on the findings from the 2014 and 2018 programs. Finally, the impact evaluation would not have been possible without the assistance of the Master Coaches, Internal Coaches, and staff of the Minnesota Reading Corps program, and without the strong support of Karin Charles, David Parker, Patrick Kaiser, and Audrey Suker from ServeMinnesota and Sadie O'Connor from Reading & Math, Inc.

Table of Contents

Acknowledgement i

Executive Summary 1

 About Reading Corps 1

 Study Methodology 3

 Findings and Conclusions 5

 Concluding Thoughts 10

Introduction 12

The Reading Corps Program Model 13

 Response to Intervention Framework 13

 K-3 Program Literacy Focus 14

 Use of Data in Program Implementation 15

 Coaching and Support 16

 Training 16

About the Study 17

 Evaluation Logic Model 17

 Evaluation Design Updates 18

 Research Questions 19

 Selection of Reading Corps Sites 19

 Random Assignment of Students within Schools 20

 Data Sources 24

 Analysis Methods 25

 Study Limitations 26

Study Findings 28

 What is the impact of Minnesota Reading Corps on Kindergarten students? 28

 What is the impact of Minnesota Reading Corps on first grade students? 29

 What is the impact of Minnesota Reading Corps on second and third grade students? 30

 Findings by Major Demographic Group 32

 Findings by Dosage Characteristics 34

 Findings from 2014 and 2018 Study Comparisons 35

Conclusions and Discussion 37

 Discussion 39

 Recommendations for Future Research 40

Appendix A: Reading Corps Logic Model (for Impact Evaluation).....41

Appendix B: Randomization Process.....42

Appendix C: Baseline Analysis.....43

Appendix D: Sample Attrition46

Appendix E: Detailed Analysis Methods49

Appendix F: Results Tables53

Appendix G: Descriptive Statistics for the 2014/2018 Analysis.....57

Appendix H: Balance Statistics for the 2014/2018 Analysis59

List of Figures

Figure i.1.	Mean scores for Kindergarten program and control students.....	5
Figure i.2.	Mean scores for first grade program and control students	6
Figure i.3.	Mean scores for first grade program and control students (Oral Reading Fluency).....	7
Figure i.4.	Mean scores for second and third grade program and control students by grade	8
Figure 1.	Response to Intervention – Three Tier Model.....	14
Figure 2.	Mean scores for Kindergarten program and control students.....	28
Figure 3.	Mean scores for first grade program and control students	29
Figure 4.	Mean scores for first grade program and control students (Oral Reading Fluency).....	30
Figure 5.	Mean scores for second and third grade program and control students by grade	31

List of Tables

Table i.1.	Expected Assessment Benchmarks by Grade and Season	4
Table i.2.	Oral Reading Fluency Results, Pooled Sample	7
Table i.3.	Reading Fluency by Subgroups, Pooled Second and Third Grades Sample.....	10
Table 1.	Expected Assessment Benchmarks by Grade and Season	15
Table 2.	Characteristics of schools participating in the Minnesota Reading Corps Impact Evaluation (Fall 2017).....	22
Table 3.	Student participants for the Minnesota Reading Corps Impact Evaluation (Fall 2017) ..	23
Table 4.	Reading Fluency Results, Pooled Sample	31
Table 5.	Reading Fluency by Subgroups, Pooled Second and Third Grades Sample.....	33
Table 6.	Treatment Effects in Propensity Weighted Kindergarten Samples.....	35
Table 7.	Treatment Effects in Propensity Weighted First Grade Samples.....	36
Table C.1.	Baseline Analysis of Minnesota Evaluation Participants, Kindergarten.....	43
Table C.2.	Baseline Analysis of Minnesota Evaluation Participants, 1 st Grade.....	44
Table C.3.	Baseline Analysis of Minnesota Evaluation Participants, 2 nd Grade.....	44
Table C.4.	Baseline Analysis of Minnesota Evaluation Participants, 3 rd Grade	45
Table D.1.	Attrition in Kindergarten Sample.....	47

Table D.2. Attrition in 1st Grade Sample..... 47

Table D.3. Attrition in 2nd Grade Sample..... 48

Table D.4. Attrition in 3rd Grade Sample 48

Table F.1. Results from Analysis 53

Table F.2. Reading Fluency by Subgroups, Pooled Second and Third Grades Sample 54

Table F.3. Partial correlations between outcomes and dosage indicators (K-1st Grade)..... 55

Table F.4. Partial correlations between outcomes and dosage indicators (2nd – 3rd Grades) 56

Table G.1. Descriptive Statistics of the Kindergarten Students from each Study Group used in the Analysis..... 57

Table G.2. Descriptive Statistics of the First Grade Students from each Study Group used in the Analysis..... 58

Table H.1. Standardized Differences for 2014 and 2018 Minnesota Reading Corps Evaluations' Kindergarten Treatment Groups (Reference to 2014 Control)..... 60

Table H.2. Standardized Differences for 2014 and 2018 Minnesota Reading Corps Evaluations' First Grade Treatment Groups (Reference to 2014 Control) 61

Executive Summary

Reading Corps is a multi-state AmeriCorps literacy program that helps students become successful readers and meet reading proficiency targets by the end of the third grade. Begun in 2003 in Minnesota and currently operating in 12 states and Washington D.C., trained literacy tutors implement research-based literacy interventions in school-based settings for more than 36,000 at-risk students in Prekindergarten (PreK) through grade 3 each year. The Minnesota Reading Corps program served over 600 schools throughout the state, reaching more than 20,000 students, in the 2017-2018 school year.

As part of an Innovative Approaches to Literacy (IAL) Grant through the U.S. Department of Education (ED), NORC at the University of Chicago, in partnership with Reading & Math, Inc. (RMI) and ServeMinnesota, conducted a multi-site impact evaluation of the Reading Corps program in three states during the 2017-2018 school year. This evaluation builds on prior independent evaluations of the Minnesota Reading Corps Program, including a 2014 randomized controlled trial (RCT) *Impact Evaluation of the Minnesota Reading Corps Kindergarten to third grade (K-3) program*.¹ Given recent developments in the Reading Corps program, including enhancements made to Reading Corps programming,² this evaluation of 622 Kindergarten, first, second, and third grade students during the 2017-2018 school year assessed the impact of the enhanced Reading Corps K-3 program model on program participants compared to students who did not receive the Reading Corps K-3 program in Minnesota.

About Reading Corps

Since 2003, Reading Corps recruits, trains, places, and monitors AmeriCorps members that serve as tutors in school-based settings to implement research-based literacy activities and interventions for struggling readers in PreK through grade 3. The goal of Reading Corps is to ensure that students become successful readers and achieve grade-level reading proficiency by the end of third grade. Tutors are supported by a multi-level coaching model that includes site-based (“Internal”) and external (“Master”) Coaches.

Literacy Focused Response to Intervention Framework

Reading Corps uses a Response to Intervention (RtI) framework that employs a multi-tier approach to the early identification and provision of support to struggling readers. Key features of the Reading Corps RtI framework include:

- Clear literacy targets at each age level from PreK through grade 3
- Benchmark assessment data to identify students eligible for one-on-one interventions

¹ Markovitz, C.; Hernandez, M.; Hedberg, E.; Silbergitt, B. (2014). *Outcome Evaluation of the Minnesota Reading Corps K3 Program*. NORC at the University of Chicago: Chicago, IL.

² Enhancements to the Reading Corps programming include prescribed guidance on the selection of specific reading interventions for older students, revised intervention exit criteria, and a book distribution and family engagement component to support reading at home.

- Evidence-based interventions
- Frequent progress monitoring during intervention delivery
- High-quality training, coaching, and observations to support fidelity of implementation

The Reading Corps K-3 program consists of one-on-one tutoring where tutors provide supplemental individualized literacy interventions focused on the National Reading Panel’s “Big Five Ideas in Literacy” to struggling emergent readers. Reading Corps tutors work with approximately 15-18 K-3 students for 20 minutes each day, delivering supplemental literacy interventions consisting of a set of prescribed, research-validated activities.

Use of Data

Assessment data play a key role in the implementation of Reading Corps. The program uses two types of curriculum-based measures (CBM) to determine students’ program eligibility and monitor progress once services are received.

- **Benchmark Assessments.** Tutors administer 1-minute benchmark assessments to students in the fall, winter, and spring. Benchmark standards (i.e., target scores) are built into the assessments to assist in determining which students are at-risk for academic failure versus those who are on track to be successful. Kindergarten through 3rd grade students scoring “below target” on benchmark assessment probes are eligible for Reading Corps services.
- **Progress Monitoring Assessments.** Once selected to receive Reading Corps services, tutors administer weekly 1-minute reading probes to students. Progress monitoring data allows tutors to chart student progress, assess effectiveness of current interventions, gauge if students require a change in interventions, or determine if they are ready to exit the program.

The Master Coach, Internal Coach, and Reading Corps tutor use each student’s progress monitoring assessment results over time to inform instructional changes and whether the student can exit service. Students in grades 1-3 may be exited from the program once their progress monitoring data shows that they have achieved 3 to 5 consecutive data points above their projected growth trajectory and two scores at or above the upcoming season benchmark target. Similar criteria are used for Kindergarten students, although students must have two scores at or above the Spring target to exit from the program.

Coaching, Support, and Training

Reading Corps provides multiple layers of supervision, support, and training to ensure integrity of program implementation. Site-specific Internal Coaches, who are typically staff literacy specialists, teachers, or curriculum directors, serve as immediate on-site supervisors, mentors, and advocates for tutors. The Internal Coach’s role is to monitor tutors and provide guidance in the implementation of Reading Corps’ assessments and interventions. The external, or Master Coach, is a literacy expert who provides on-site staff (i.e., Internal Coaches and Reading Corps tutors) with expert consultation on literacy instruction. The Master Coach also ensures implementation integrity of Reading Corps program

elements. In addition to these two coaching layers, a third layer consisting of program support staff provide administrative oversight for program implementation to sites participating in Reading Corps.

Prior to the start of each school year, Reading Corps hosts a three-day Institute to train returning and new Master Coaches, Internal Coaches, and Reading Corps tutors in the assessments and research-based literacy interventions employed by Reading Corps. During several sessions at the Institute, tutors learn the skills, knowledge, and tools needed to serve as literacy interventionists. Tutors are provided with detailed literacy manuals, as well as online resources that mirror and supplement the contents of the manual (e.g., videos of model interventions and best practices). Additional training and coaching sessions are provided throughout the tutors' year of service.

Study Methodology

The goal of the Minnesota Reading Corps impact evaluation was to independently and experimentally assess the impact of the Minnesota Reading Corps program on Kindergarten, first grade, second grade, and third grade students' literacy proficiency scores. The current (2017-2018) evaluation replicated many design elements of the original 2014 evaluation, but included four key differences: 1) an extension of the duration of the evaluation for second and third grade students from one semester to a full school year (i.e., examining program impact through Spring benchmark rather than Winter); 2) inclusion of second and third grade students who scored farthest from benchmark in Fall 2017; 3) inclusion of only IAL grant recipient schools (i.e., rural and high-poverty urban), and 4) independent outcome assessment data collection.

The primary research questions for the current evaluation are:

1. For Kindergarten and first grade students, what is the impact of a single semester of the Minnesota Reading Corps on program participants compared to similar students who did not receive Reading Corps?
2. For second and third grade Tier 2 students who are farthest from the Fall reading proficiency benchmark, what is the impact of a full school year of Minnesota Reading Corps on program participants compared to similar students who did not receive Reading Corps?³
3. Does program impact vary by participant characteristics, such as demographics?

Random Assignment of Students within Schools

A randomized controlled trial (RCT) design was used to answer these research questions, which was largely informed by the 2014 Impact Evaluation of the Minnesota Reading Corps K-3 Program.⁴ At the beginning of the school year (prior to program implementation), a select number of eligible students enrolled in Kindergarten through third grade at the 24 participating Minnesota Reading Corps schools

³ To maximize statistical power, the study was able to make use of the common outcome of *oral reading fluency* for second and third grade students and create a pooled sample with grade level controls in the model.

⁴ Markovitz, C.; Hernandez, M.; Hedberg, E.; Silbergliitt, B. (2014). *Impact Evaluation of the Minnesota Reading Corps K-3 Program*. NORC at the University of Chicago: Chicago, IL.

were included in the study sample and randomly assigned to two groups: a program group and a control group. In Fall 2017, students at each participating school were assessed to determine program eligibility. Eligible students in each grade within a school were matched in pairs based on their Fall benchmark scores (see **Table i.1 below**). Next, pairs of students within each grade were selected to participate in the evaluation. The Kindergarten and first grade study participants were randomly drawn from the pool of eligible students. The second and third grade study participants were purposively selected from the pool of eligible students so as to include those students with scores furthest from the Fall benchmark. After all pairs were selected, one student in each pair was randomized into the Reading Corps program group while the other was assigned to the control group. In total, 60 Kindergarten students, 160 first grade students, 190 second grade students, and 212 third grade students participated in the evaluation.

Table i.1. Expected Assessment Benchmarks by Grade and Season

Grade	Measure	Fall	Winter	Spring
Kindergarten	Test of Letter Sounds	8	27	48
Grade 1	Test of Nonsense Words (English)	36	63	N/A
Grade 1	CBMReading	N/A	52	82
Grade 2	CBMReading	63	97	116
Grade 3	CBMReading	100	122	135

Data Sources

Since the primary goal of the Minnesota Reading Corps impact evaluation was to assess the impact of the program on K-3 students’ literacy proficiency scores, the main source of data for the evaluation was comprised of the grade-appropriate and semester-specific literacy assessment scores described above. The evaluation team requested access to the 2017 Fall benchmark data that Reading Corps tutors collected to identify program-eligible Kindergarten, first, second, and third grade students enrolled at the 24 participating schools.⁵ Independent assessors from the NORC evaluation team collected the Winter (grades K-1) and Spring (grades 2-3) benchmark assessment data from students in both the program and control groups at each school to ensure independent measures of outcomes and mitigate any potential bias at posttest. Additionally, Minnesota Reading Corps program staff at the 24 participating schools provided the evaluation team with other administrative data, including students’ AmeriCorps tutoring session attendance, average length and number of sessions, and student demographic characteristics.

Analysis

To estimate program effects, the average Winter (grades K-1) and Spring (combined grades 2-3) benchmark scores of students who participated in the Minnesota Reading Corps program were statistically compared with the average Winter (grades K-1) and Spring (combined grades 2-3) benchmark scores of the control group of students who did not. The calculated difference in group averages is the causal effect of program participation. The impact evaluation analyses were conducted using mixed-regression models, also known as hierarchical linear models. This type of model uses maximum likelihood to estimate differences in outcomes associated with predictor variables while also

⁵ Fall benchmark data was collected prior to randomization, ensuring an unbiased assessment of student baseline proficiency.

modeling the multi-level structure of the data, which in this case is students nested within schools. The analytic approach accounted for the effect of program status (i.e., assignment to program or control group) in isolation as well as when relevant control variables (i.e., demographic characteristics) were included.

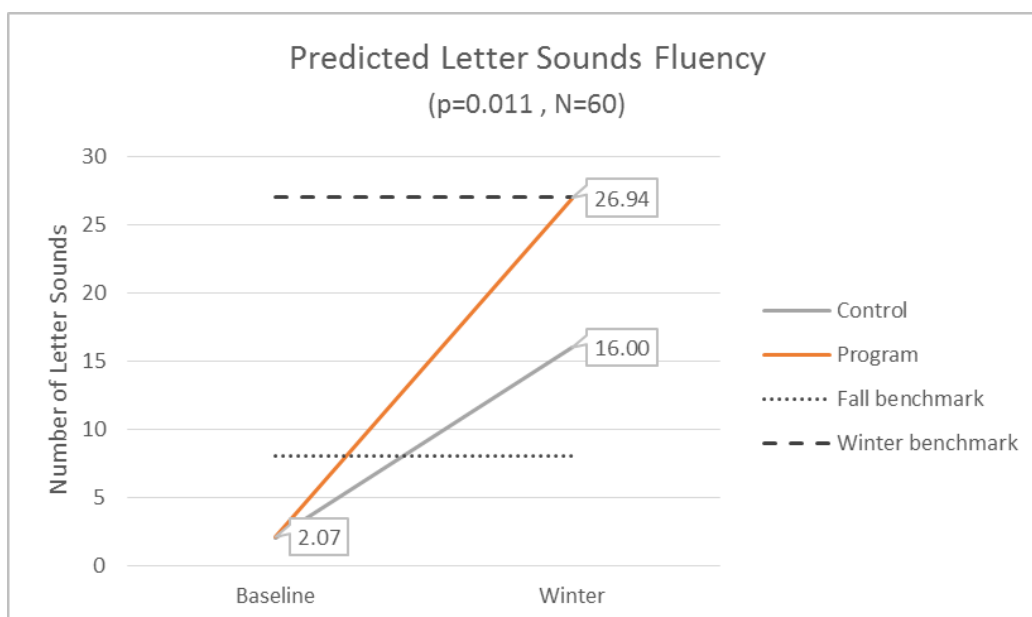
Findings and Conclusions

Below, the evaluation team presents the study findings, followed by final thoughts on the implications of these findings for the future of the Minnesota Reading Corps program.

What is the impact of Minnesota Reading Corps on Kindergarten students?

Kindergarten students who received Minnesota Reading Corps tutoring achieved significantly higher *letter sound fluency* scores by the end of the first semester (Winter 2018) than Kindergarten control students who did not receive the program (see **Figure i.1**). On average, program students correctly identified an average of 10.9 more letter sounds in a one minute period than control group students ($p=0.01$, $N=60$). This difference had a positive, meaningful effect size of 0.85, indicating that the average program student scored over three-fourths of a standard deviation unit higher than the average control group student in identifying letter sounds. Also by the end of the first semester, the program group achieved Winter benchmark expectations, whereas the control group did not reach the expected benchmark.

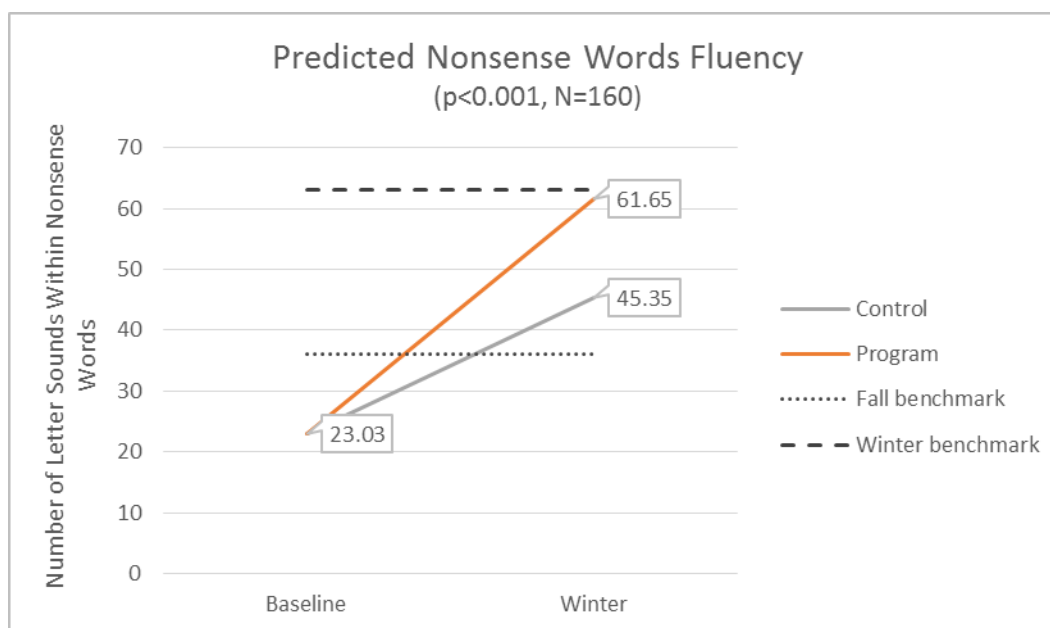
Figure i.1. Mean scores for Kindergarten program and control students



What is the impact of Minnesota Reading Corps on first grade students?

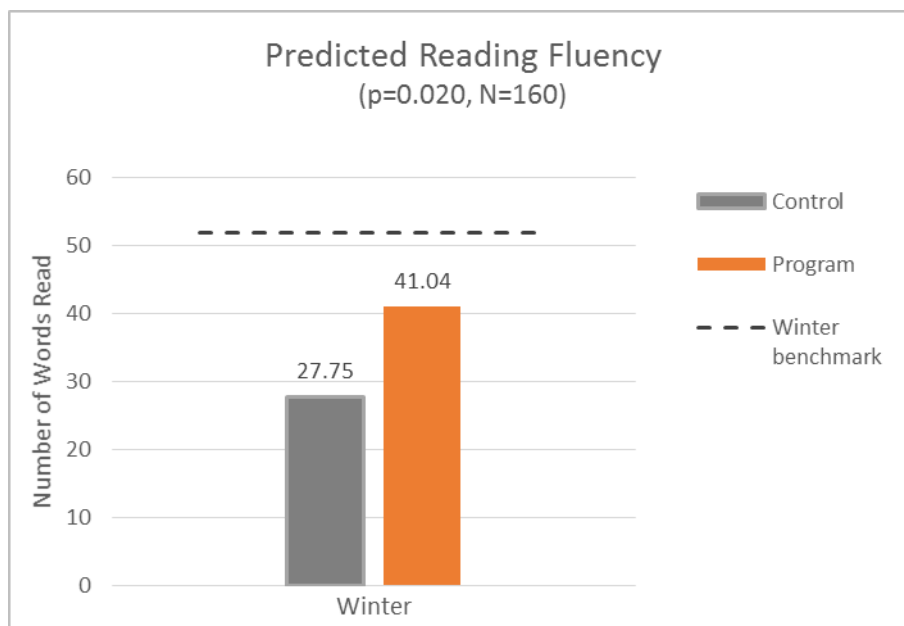
First grade Minnesota Reading Corps students attained significantly higher *nonsense word fluency* scores by Winter benchmark than first grade students in the control group. On average, program students correctly identified in one minute 16.3 more letter sounds embedded within non-real words than first grade control group students ($p < 0.001$, $N=160$). This difference represents a large, positive effect size of 0.81, indicating that the average program student scored more than three-fourths of a standard deviation unit higher than the average control group student in identifying letter sounds within nonsense words (see **Figure i.2**). Additionally, the first grade program students' average score nearly met the Winter benchmark.

Figure i.2. Mean scores for first grade program and control students



First grade students enrolled in the Minnesota Reading Corps program achieved substantially and significantly higher *oral reading fluency* scores at Winter benchmark than first grade control group students (see **Figure i.3**). On average, program students correctly read aloud 13.3 more words than the control group ($p=0.02$, $N=160$). This represents a large, positive effect size of 0.61.

Figure i.3. Mean scores for first grade program and control students (Oral Reading Fluency)



What is the impact of Minnesota Reading Corps on second and third grade students?

The pooled group of Minnesota Reading Corps second and third grade students attained significantly higher *oral reading fluency* scores by Spring benchmark than control group students. By the end of the school year, second and third grade program students read aloud an average of 6.4 more words in a one minute period than second and third grade control group students ($p < 0.005$, $N=402$). This difference is meaningful, and represents a positive effect size of 0.28, indicating that the average program student scored more than one-quarter standard deviation unit higher than the average control group student in the number of words read aloud (see **Table i.2**). In both grades, program students’ average scores were substantially higher than control students’ average scores, however, no group reached their respective Spring benchmark (see **Figure i.4**).

Table i.2. Oral Reading Fluency Results, Pooled Sample

Grade and Outcome	Treatment effect ^{a,b} (β_1)	p-value ^a	Effect size ^c (α_1/σ)	Prog. Mean ^d	Control mean ^d	Total N
Second and Third Grades^e						
Reading fluency	6.44	0.005	0.28	102.6	96.2	402
	(2.291)					

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary

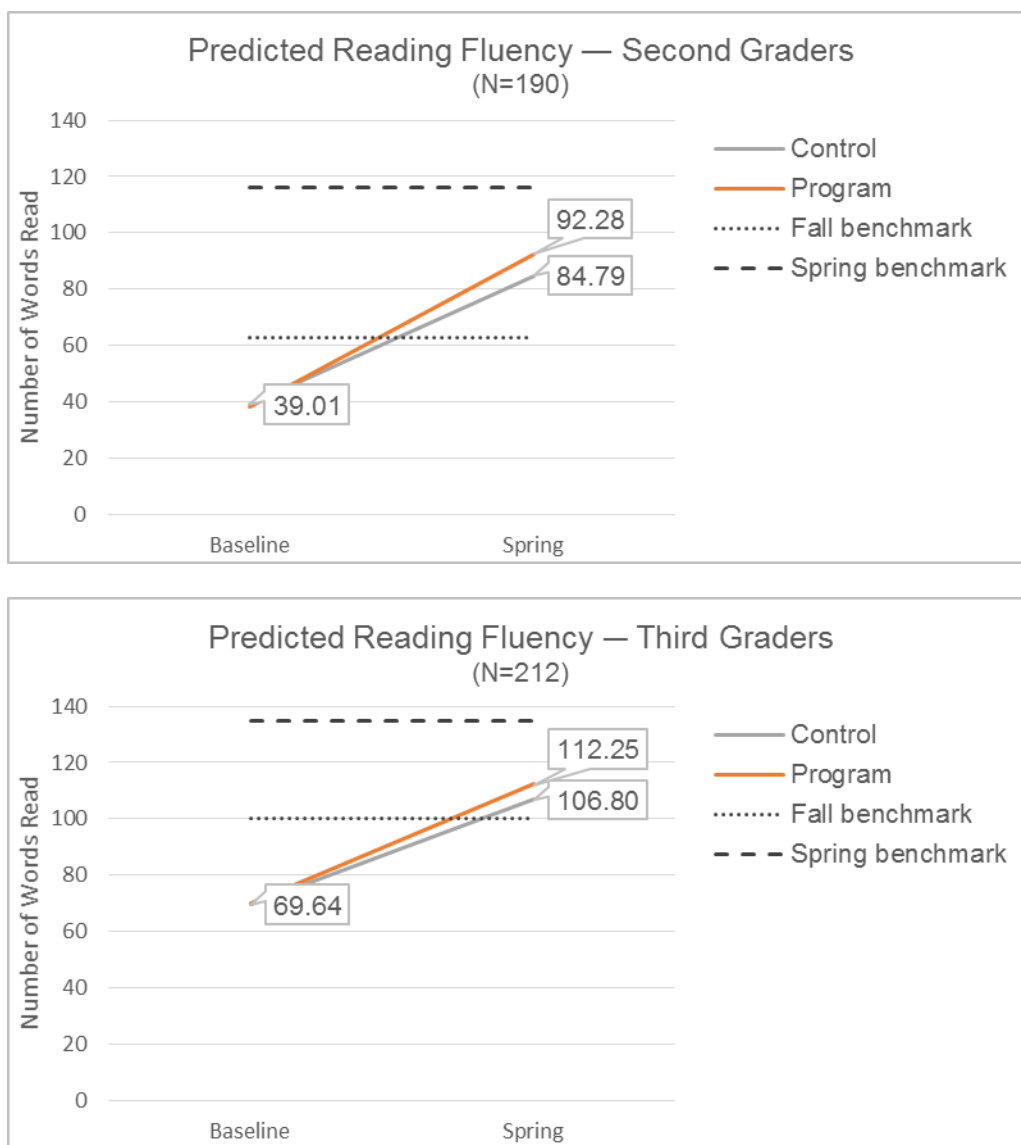
b: Standard errors in parentheses

c: Effect size reported from unconditional model without control variables

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: The Minnesota model includes sampling weights for generalizability

Figure i.4. Mean scores for second and third grade program and control students by grade



Does program impact vary by participant characteristics, such as demographics?

Below are the results by race, gender, and English Language Learner (ELL) status for the combined second and third grade sample⁶ to address the study’s research question on whether program impacts vary by participant demographics. **Table i.3** below provides the detailed results.

⁶ Sample sizes for demographic groups in Kindergarten and first grade lacked the statistical power to detect differences both within and between subgroups; therefore, we were unable to conduct analyses on subgroups for these grades.

Race/Ethnicity

For the analysis of second and third grade *oral reading fluency* by race, we found that Black/African-American students in the program group read aloud an average of 10.7 more words in a one minute testing period by the Spring benchmarking period compared to similar students in the control group ($p=.03$). This difference had a positive, meaningful effect size, indicating that the average African American program student scored one half standard deviation unit higher than the average African American control group student in reading fluency (effect size = 0.50).

Gender

The effect size in *oral reading fluency* between the program and control groups was descriptively larger for males than females. While all second and third grade students in the program group read aloud an average of 6.4 more words correctly in a one minute period ($p=.005$, $N=402$) by the Spring benchmarking period, among second and third grade males, the average difference between the program and control groups was 7.9 words ($p=.013$). This difference had a positive, medium effect size of 0.47, indicating that the average male program student scored nearly one half standard deviation unit higher than the average male control group student in reading fluency. For female students in second and third grade, the findings showed the average difference between treatment and control groups to be slightly lower (5.7 more words read correctly and effect size = 0.22).

English Language Learners

Some important differences also were detected in the patterns of findings for second and third grade ELL students. By the end of the Spring benchmarking period, second and third grade ELL students in the Minnesota Reading Corps program read aloud correctly an average of 16.3 more words in a one minute period than students in the control group ($p=.001$). This difference resulted in a positive, medium effect size, indicating that the average ELL program student scored over one half standard deviation unit higher in *oral reading fluency* than the average ELL control group student (effect size = 0.53).

Table i.3. Reading Fluency by Subgroups, Pooled Second and Third Grades Sample

Subgroup Characteristic ^f	Total N ^e	Reading Fluency				
		Treatment effect ^{a,b} (β_1)	p-value ^a	Effect size ^c (α/σ)	Prog. Mean ^d	Control mean ^d
Racial or Ethnic subgroups						
American Indian or AK Native	18	5.914 (5.414)	0.275	g	97.259	91.345
Asian	25	9.065 (8.023)	0.259	g	97.835	88.770
Black or African American	66	10.650 (5.031)	0.034	0.497	104.697	94.047
Hispanic/Latino	46	3.771 (5.181)	0.467	g	100.667	96.896
Gender subgroups						
Female	204	5.735 (2.486)	0.021	0.220	101.014	95.279
Male	198	7.878 (3.169)	0.013	0.467	102.311	94.433
Language subgroups						
ELL	94	16.279 (4.672)	0.000	0.528	113.992	97.714

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary.

b: Standard errors in parentheses.

c: Effect size reported from unconditional model without control variables.

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: Subgroups containing fewer than 10 students were excluded from analysis.

f: The Minnesota model includes sampling weights for generalizability.

g: Non-significant effect sizes not reported.

Concluding Thoughts

Kindergarten and first grade students who received the Minnesota Reading Corps program achieved significantly higher literacy assessment scores by the Winter benchmark than control students who did not participate in the program. Specifically, Kindergarten students produced more correct letter sounds than did students in the control group and reached the Winter benchmark score. First grade program students produced more letter sound segments within nonsense words than students in the control group and nearly reached the Winter benchmark score. While *oral reading fluency* is not typically a focus of Reading Corps tutoring until the second semester in first grade, Reading Corps students scored significantly higher at the Winter benchmark on average for *oral reading fluency* than control group students. This finding may indicate that first grade Reading Corps students are transferring early phonics skills gained from Reading Corps tutoring to the more complex skill of reading fluency.

Positive results were likewise found among second and third grade students. After a full year of tutoring, second and third grade program students achieved significantly higher literacy assessment scores by the Spring benchmark than did control students. Participating second and third grade students were selected

because they had *oral reading fluency* scores furthest from the Fall benchmark, and while neither grade level achieved the Spring benchmark, program students showed significant improvements in *oral reading fluency*. In addition, a statistically significant impact of Reading Corps was found among second and third grade students despite minority status (i.e., Black), gender, or ELL status. For each of these characteristics, second and third grade students who received the program significantly outperformed control students on *oral reading fluency* scores and produced larger effect sizes than those found for the average student in the program. Together, the findings show that the Minnesota Reading Corps program can have a substantial and positive impact on at-risk students who may be underserved.

Overall, the results of the evaluation showed that the Minnesota Reading Corps program positively impacted Kindergarten through third grade students' literacy outcomes. Similar to the previous 2014 study's findings, the magnitude of the effect of the program varied by grade, such that students in Kindergarten and first grade on average achieved larger gains within a single semester, and second and third grade students furthest from benchmark showed moderate gains in a more complex literary skill over a longer period of time. Therefore, based on findings from two highly rigorous studies of the program, it is possible to conclude that the Minnesota Reading Corps K-3 program is effective at improving K-3 students' early reading skills. Such results provide support for the continued implementation of the program and replication of the Reading Corps model for eligible K-3 students.

Introduction

Reading Corps is a multi-state AmeriCorps literacy program that helps students become successful readers and meet reading proficiency targets by the end of the third grade. Each year, the program, which began in 2003 with just 250 students in Minnesota and is currently operating in 12 states and Washington D.C., provides trained literacy tutors to implement research-based literacy interventions in school-based settings for more than 36,000 at-risk students in Prekindergarten (PreK) through grade 3.

As part of an Innovative Approaches to Literacy (IAL) Grant through the U.S. Department of Education (ED), NORC at the University of Chicago, in partnership with Reading & Math, Inc. (RMI) and ServeMinnesota, conducted a multi-site impact evaluation of the Reading Corps program in Minnesota, Wisconsin, and Florida during the 2017-2018 school year. This multi-state impact evaluation builds on prior independent evaluations of the Minnesota Reading Corps Program. This includes a 2014 randomized controlled trial (RCT) *Impact Evaluation of the Minnesota Reading Corps Kindergarten to third grade (K-3) program*, which found that Kindergarten, first, and third grade students who received Minnesota Reading Corps K-3 services during the first semester of the 2012-2013 school year accelerated their literacy skills development and performed better than comparable students who did not.⁷ Given recent developments in the Reading Corps program, including enhancements made to Reading Corps programming,⁸ this evaluation assessed the impact of the enhanced Reading Corps K-3 program model on program participants compared to students who did not receive the Reading Corps K-3 program in Minnesota.

This report describes the findings from the impact evaluation of the Minnesota Reading Corps K-3 Program on 622 Kindergarten, first, second, and third grade students during the 2017-2018 school year. The purpose of the evaluation was: 1) to evaluate the impact of the Minnesota Reading Corps K-3 program on program participants compared to similar students who did not receive the Reading Corps program; and 2) to assess the degree to which the impact varies by participant characteristics (e.g., demographics).

In the sections that follow, we describe the Reading Corps program model and specifically the Minnesota Reading Corps program. We then present the methodology used to conduct the evaluation, the study findings, and the our conclusions, as well as a discussion on the implications of the findings for the Reading Corps program. The appendices to this report are located at the end and include: Appendix A presents the Reading Corps logic model; Appendix B illustrates the randomization process; Appendices C through E provide additional methodological detail (i.e., baseline analysis, attrition, and outcome analyses); Appendix F presents detailed results tables; and Appendices G and H provide additional detail concerning the 2014/2018 comparative analysis.

⁷ Markovitz, C.; Hernandez, M.; Hedberg, E.; Silbergliitt, B. (2014). *Outcome Evaluation of the Minnesota Reading Corps K3 Program*. NORC at the University of Chicago: Chicago, IL.

⁸ Enhancements to the Reading Corps programming include prescribed guidance on the selection of specific reading interventions for older students, revised intervention exit criteria, and a book distribution and family engagement component to support reading at home.

The Reading Corps Program Model

Reading Corps is the largest AmeriCorps tutoring program in the country, in which trained AmeriCorps members serve as literacy tutors to implement evidence-based literacy instruction and assessment protocols for students in PreK through grade 3 in school-based settings. Beginning in 2003 with just 250 students in Minnesota, Reading Corps has since expanded to twelve states and Washington D.C. Reading Corps aims to broadly impact literacy outcomes for children. A primary goal for the program is that all children, age 3 to grade 3, who qualify for Reading Corps, will meet reading proficiency standards by third grade.

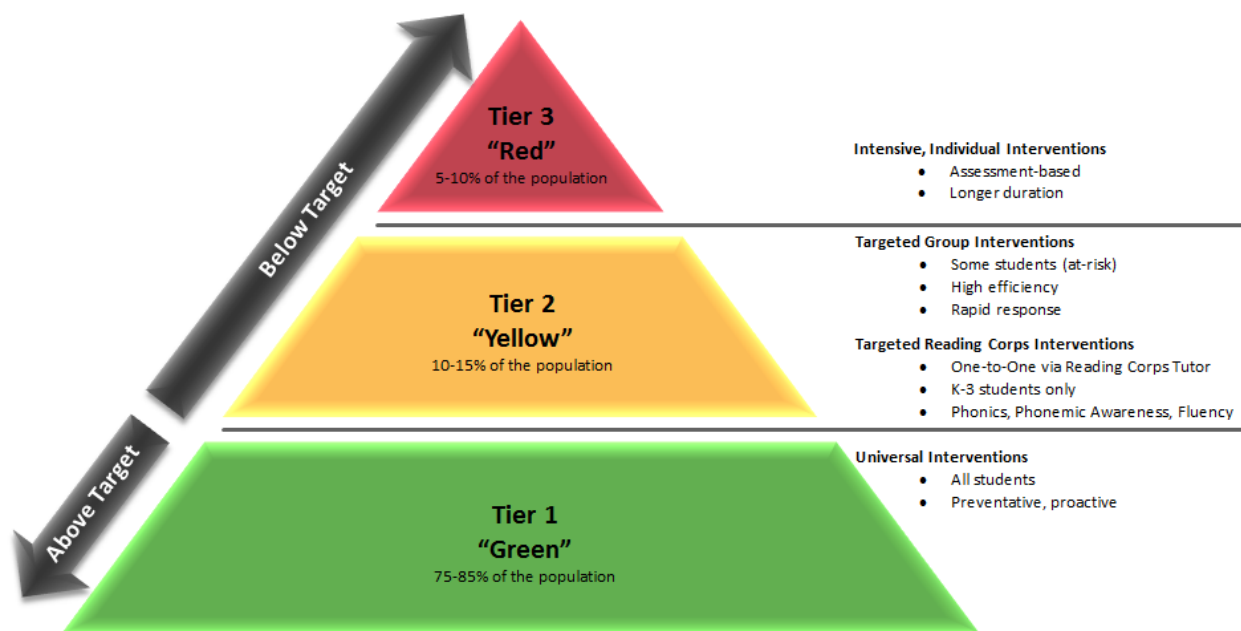
Response to Intervention Framework

Reading Corps is based on a Response to Intervention (RtI) framework which uses a multi-tier approach to the early identification and support of struggling learners. The key aspects of the Reading Corps RtI framework are:

- Clear literacy targets at each age level from PreK through grade 3
- Benchmark assessment to identify students eligible for one-on-one interventions
- Evidence-based interventions
- Frequent progress monitoring during intervention delivery
- High-quality training, coaching, and observations to support fidelity of implementation

In the RtI framework, data play the key roles of determining students' eligibility for additional services and then monitoring students' progress towards achieving academic expectations (i.e., benchmarks). Reading Corps screens students for program eligibility three times a year (i.e., fall, winter, spring) using grade-specific, literacy-focused measures that possess criterion-referenced grade- and content-specific performance benchmarks. These assessments are collectively called curriculum-based measures (CBM) because they correspond closely with curricular expectations for literacy skills at each developmental level. Program staff use scores from these general outcome measures to categorize students into one of three possible tiers (i.e., proficiency levels; see **Figure 1**): Tier 1 students score at or above benchmark and benefit from typical classroom instruction (75-80% of students score in this category); Tier 2 students score below benchmark and require specific supplemental interventions until they meet benchmarks (15-20% of students fall into this category); and Tier 3 students require intensive intervention provided by a special education teacher or literacy specialist and often have individualized educational plans (5-10% of students qualify for this category).

Figure 1. Response to Intervention – Three Tier Model



The Reading Corps K-3 program provides one-on-one tutoring where Reading Corps tutors provide supplemental individualized literacy interventions to primarily Tier 2 students in Kindergarten through third grade. The Reading Corps PreK program includes both an immersive “push-in” component, where tutors provide whole-class literacy enrichment for all students (i.e., Tier 1), and a targeted one-on-one component, where tutors provide individualized interventions to students struggling with emergent literacy skills (i.e., Tiers 2 and 3). Although Reading Corps provides both PreK and K-3 interventions to students, the focus of this report is on the Minnesota Reading Corps K-3 program. Therefore, the remainder of this report will focus on describing the K-3 program and evaluation.

K-3 Program Literacy Focus

At the K-3 level, the program is focused on the “Big Five Ideas in Literacy” as identified by the National Reading Panel, including phonological awareness, phonics, fluency, vocabulary, and comprehension.^{9, 10} Reading Corps tutors serve as one-on-one literacy interventionists, working with approximately 15-18 K-3 students below grade-specific literacy benchmarks for 20 minutes each day. Tutors are trained to deliver ten research-based literacy interventions that align with the Big Five literacy targets. The interventions share a common theme in that they focus on building fluency for foundational reading skills such as phonological awareness, letter sound knowledge, decoding skills, and *oral reading fluency*. The literacy interventions are supplemental to the core reading instruction provided at each school and consist of a set

⁹ Snow, C., Burns, M., & Griffin, P. (Eds.), (1998). *Preventing reading difficulties in young children*. Washington DC: National Academy Press.

¹⁰ Shanahan, T.; Cunningham, A.; Escamilla, KC.; Fischel, J.; Landry, S.; Lonigan, CJ., et al. (2008). *Developing early literacy: Report of the national early literacy panel*. Washington DC: National Institute for Literacy.

of prescribed, research-validated activities such as “Repeated Reading with Comprehension Strategy Practice” or “Duet Reading.”

The decision to change a student’s interventions is based upon weekly progress monitoring data. The goal of the tutoring is to raise individual students’ literacy levels so that they are on track to meet or exceed the next program-specified literacy benchmark. Meeting benchmark will allow the student to benefit fully from general (i.e., Tier 1) literacy instruction already provided in the classroom.

Use of Data in Program Implementation

Assessment data are an integral component of Reading Corps’ program implementation. The assessment tools used by Reading Corps are curriculum-based measures (CBM) from FastBridge Learning,¹¹ selected because of their well-established statistical reliability and validity in evaluating critical literacy skills appropriate for specific grade levels and seasons. Reading Corps collects the following two types of CBM assessment data to determine students’ program eligibility and monitor progress once services are received.

- **Benchmark Assessments.** Tutors administer 1-minute benchmark assessments to students in the fall, winter, and spring. Benchmark standards (i.e., target scores) are built into the assessments to assist in determining which students are at-risk for academic failure versus those who are on track to be successful. Kindergarten through 3rd grade students scoring “below target” on benchmark assessment probes are eligible for Reading Corps services. **Table 1** lists the specific Fastbridge CBM assessments and corresponding benchmark scores used to identify program eligible students in Kindergarten through third grades.
- **Progress Monitoring Assessments.** Once selected to receive Reading Corps services, tutors administer weekly 1-minute reading probes to students. This progress monitoring data allows tutors to chart student progress, assess effectiveness of current interventions, gauge if students require a change in intervention, or determine if they are ready to exit the program. For each participating student, these progress monitoring scores are graphed and then reviewed monthly by a collaborative team consisting of the tutors, Internal Coach, and Master Coach.

Table 1. Expected Assessment Benchmarks by Grade and Season

Grade	Measure	Fall	Winter	Spring
Kindergarten	Test of Letter Sounds	8	27	48
Grade 1	Test of Nonsense Words (English)	36	63	N/A
Grade 1	CBMReading	N/A	52	82
Grade 2	CBMReading	63	97	116
Grade 3	CBMReading	100	122	135

In the K-3 program, eligible students receive intervention services until their progress monitoring data shows that they have achieved 3 to 5 consecutive data points above their projected growth trajectory and

¹¹ <http://www.fastbridge.org/>

two scores at or above the upcoming season benchmark target. Similar criteria are used for the discontinuation of services with Kindergarten students, although students must have two scores at or above the Spring target to be eligible to exit from the program. Once these criteria are met, a student is deemed “on-track” to achieve appropriate grade-level benchmark at the next assessment window, and is “exited” from the Reading Corps program (i.e., the tutor no longer provides intervention services). The Master Coach, Internal Coach, and Reading Corps tutor discuss each student’s assessment results over time before deciding to exit the student from service.

The data intensive nature of the Reading Corps program provides tutors, coaches, teachers and principals/directors with a consistent, objective means of identifying students to receive program services, tracking their progress toward achieving academic goals related to critical literacy skills, and informing instruction. The data also provide tutors and coaches with objective information about the efficacy of the interventions with individual students, which can in turn be used to tailor the most effective instruction to the student’s current skill level.

Coaching and Support

Reading Corps provides multiple layers of supervision to ensure integrity of program implementation. Site-specific Internal Coaches, who are typically staff literacy specialists, teachers, or curriculum directors, serve as immediate on-site supervisors, mentors, and advocates for tutors. The Internal Coach’s role is to monitor tutors and provide guidance in the implementation of Reading Corps’ assessments and interventions. As the front-line supervisor, the Internal Coach is a critical component of the program’s supervisory structure.

The external, or Master Coach, is a literacy expert who provides site staff (e.g., Internal Coaches and Reading Corps tutors) with expert consultation on literacy instruction and ensures implementation integrity of Reading Corps program elements. In addition to these two coaching layers, a third layer consisting of Reading Corps program support staff helps ensure a successful year of AmeriCorps service. Program support staff are Reading Corps employees who provide administrative oversight for program implementation to sites participating in Reading Corps.

Training

Prior to the start of each school year, Minnesota Reading Corps hosts a three-day Institute to train returning and new Master Coaches, Internal Coaches, and Reading Corps tutors. This intensive, information-filled training provides foundational training in the assessments and research-based literacy interventions employed by Reading Corps. During several sessions at the Institute, tutors learn the skills, knowledge, and tools needed to serve as literacy interventionists. Tutors are provided with detailed literacy manuals as well as online resources that mirror and supplement the contents of the manual (e.g., videos of model interventions and best practices). Both the manuals and online resources are intended to provide tutors with just-in-time support and opportunities for continued professional development and skill refinement. Additional training and coaching sessions are provided throughout the tutors’ year of service.

About the Study

This section provides a detailed description of the methodology used to implement the 2018 Minnesota Reading Corps K-3 program impact evaluation. The methodology was largely informed by the 2014 *Impact Evaluation of the Minnesota Reading Corps K-3 Program*.¹²

Evaluation Logic Model

A logic model for the Reading Corps program illustrating key program and school inputs and activities, as well as the program’s desired short-term outcomes and long-term goals for PreK and K-3 students is provided in **Appendix A**. The Reading Corps logic model serves as the conceptual framework for the design of the Minnesota Reading Corps impact evaluation. The logic model presents a comprehensive illustration of the complete Reading Corps program and includes inputs, activities, short-term outcomes, and long-term goals for PreK and K-3 students. The focus of the Minnesota Reading Corps K-3 program impact evaluation was to assess the impact of program participation on Kindergarten, first, second, and third grade students’ literacy scores. As such, the evaluation focused on only those components of the logic model relevant to K-3 students.

Four key program- and school-based inputs and resources are essential to successful K-3 program implementation: 1) the selection of schools based on degree of student need and school capacity to partner effectively with the program; 2) school identification of at-risk K-3 students within the school based on benchmark assessments of students’ literacy skills; 3) web-based data management systems to track and monitor student progress with literacy interventions; and 4) school use of research-based core literacy curriculum.

Three important Reading Corps K-3 program inputs related to Reading Corps tutors include: 1) joint program and school recruitment, screening and placement of tutors in schools; 2) comprehensive Reading Corps training of tutors and Internal Coaches in literacy interventions, assessment, data-driven decision-making and program rules; and 3) school identification and assignment of dedicated Internal Coaches to support and monitor the tutors. The logic model also illustrates the multiple layers of supervision and coaching the Reading Corps program provides to its school-based Internal Coaches and Reading Corps tutors.

As shown in the logic model, the Reading Corps program’s primary activities include: 1) conducting benchmark assessment three times per year (fall, winter and spring) to identify students in need of literacy tutoring; 2) delivering one-on-one tutoring to eligible students 20 minutes a day, 5 days a week; 3) assessing and charting weekly student progress on grade-specific literacy skills; 4) “exiting” students from the program once they achieve assessment scores putting them on track to meet or exceed the next benchmark; and 5) identifying and tutoring new students eligible for the program. The intended short-term outcomes of these activities are demonstrated improvement on FastBridge measures at the

¹² Markovitz, C.; Hernandez, M.; Hedberg, E.; Silbergliitt, B. (2014). *Impact Evaluation of the Minnesota Reading Corps K-3 Program*. NORC at the University of Chicago: Chicago, IL.

subsequent benchmarking period (i.e., Winter, Spring) and a successful (permanent) exit from Reading Corps tutoring services. The desired long-term outcome of the Reading Corps program is for all third grade students to meet or exceed grade-level proficiency on the Minnesota state third grade reading test (MCA-III).

Evaluation Design Updates

The current (2017-2018) evaluation of the Minnesota Reading Corps K-3 program replicated many design elements of the original 2014 evaluation, but included four key differences: 1) an extension of the duration of the evaluation for second and third grade students from one semester to a full school year (i.e., examining program impact through Spring benchmark rather than Winter); 2) inclusion of second and third grade students who scored farthest from benchmark in Fall 2017; 3) inclusion of only IAL grant recipient schools (i.e., rural and high-poverty urban), and 4) independent outcome assessment data collection.

In the previous 2014 Minnesota Reading Corps study, *all* students (K-3) were followed for a single semester. The results showed positive and large effects for Kindergarten and first grade students. For this reason, the current evaluation, again assessed the impact of Reading Corps tutoring over a single semester (i.e., Fall 2017 to Winter 2018) on Kindergarten and first grade students' literacy outcomes. In contrast, for the second and third grade students, the current evaluation assessed the impact of tutoring over a full school year (i.e., through the Spring 2018 benchmark) on second and third grade students literacy outcomes. The duration of the second and third grade evaluation was extended for two critical reasons. First, the majority of students in these grades receive Reading Corps services for more than one semester. As such, a single semester did not provide an accurate representation of an average student's experience with or impact of the program. Second, the previous (2014) evaluation's exploratory growth model suggested that second and third grade students would continue to improve their literacy proficiency beyond the first semester. For these reasons, we chose to evaluate Minnesota Reading Corps' impact on second and third grade students over the course of a full school year (i.e., through Spring 2018 benchmark). In addition to extending the duration of the second and third grade evaluation through Spring benchmark, the study explored the effectiveness of the Reading Corps program on second and third grade Tier 2 students¹³ furthest from benchmark.

Another difference between the past and current evaluation was the sample selection criteria applied. Due to funding requirements for the current multi-state study, sampled schools were required to be IAL recipients. As a result of IAL program eligibility, the 2017-2018 sample was comprised of only urban and rural schools.

Finally, this evaluation employed independent assessors to collect outcome assessment data in Winter 2018 (K-1) and Spring (2-3), rather than rely on program administrative data collected by Reading Corps

¹³ To maximize statistical power, the study was able to make use of the common outcome of *oral reading fluency* for second and third grade students and create a pooled sample with grade level controls in the model.

tutors, as was done in the previous 2014 study. The use of independent assessors was intended to allay any possible concerns about assessor bias.

In addition to these design differences, the current study also includes a comparison study between the previous 2014 and current 2017-2018 findings for Minnesota Kindergarten and first grade students.

Research Questions

As the logic model illustrates, the Reading Corps program's short-term objective is to improve program eligible students' literacy skills so they are on track to achieve grade-level proficiency. The primary goal of the Minnesota Reading Corps K-3 program impact evaluation was to independently and experimentally assess the impact of the Minnesota Reading Corps K-3 program on students' literacy proficiency scores.

Thus, the primary research questions for the evaluation are as follows:

1. For Kindergarten and first grade students, what is the impact of a single semester of the Minnesota Reading Corps on program participants compared to similar students who did not receive Reading Corps?
2. For second and third grade Tier 2 students who are farthest from the Fall reading proficiency benchmark, what is the impact of a full school year of Minnesota Reading Corps on program participants compared to similar students who did not receive Reading Corps?
3. Does program impact vary by participant characteristics, such as demographics, program dosage, and program attendance?

To answer these research questions, the evaluation team analyzed grade-appropriate and semester-specific literacy assessment scores collected from 60 Kindergarten students, 160 first grade students, 190 second grade students, and 212 third grade students enrolled at 24 schools participating in the Minnesota Reading Corps during the 2017-2018 school year.

Within each school, a specific number of program eligible students in the same grade were matched into pairs based on their Fall benchmark scores, with one student in each pair randomized into the Minnesota Reading Corps K-3 program, while the other student was assigned to the control group. This process ensured baseline equivalence across treatment and control groups on the Fall, pre-intervention, grade-appropriate outcome of interest. Data from the Winter benchmark assessment were used to answer the first research question for Kindergarten and first grade students, while data from the Spring benchmark assessment were used to answer the second research question for second and third grade students. The third question on impacts by participant and other characteristics, was examined for the second and third grade students only because the larger sample sizes allowed us to meaningfully examine representative subgroups of the overall sample.

Selection of Reading Corps Sites

There are several hundred Minnesota Reading Corps schools operating in the state. Power analyses determined that the study required 24 schools to participate in the impact evaluation of the Minnesota

Reading Corps K-3 Program.¹⁴ All of the schools considered for participation in the study met IAL eligibility criteria. IAL directs resources to communities that need them most, such as districts with high poverty rates and/or small rural¹⁵ schools. Reading Corps received IAL funding to conduct evaluations in two other states. Therefore, they used the same IAL selection criteria for the Minnesota evaluation. Roughly 124 Minnesota Reading Corps schools met these eligibility criteria and, therefore, were considered for participation in the study. Of the 124 identified schools, several had to be eliminated from the sampling frame due to either their ability to participate or appropriateness for the study.¹⁶ After these criteria were applied, 55 schools remained eligible for the evaluation. Equal numbers of schools were selected by urbanicity (10 urban and 14 rural schools) using probability-proportional-to-size (PPS)¹⁷ and agreed to participate in the study. Of the schools selected, all 24 provided second and third grade students for the study. Since smaller samples of students were required for the Kindergarten and first grade analyses, a subset of these schools (16 for first grade and 6 for Kindergarten) were asked to provide students from the lower grades for the study. The 24 schools selected and recruited to participate in the study are presented in **Table 2** below.¹⁸

Random Assignment of Students within Schools

At the beginning of the 2017-2018 school year (prior to program implementation), a select number of eligible students enrolled in Kindergarten through third grade at the 24 participating schools were randomly assigned to two groups: a program group that participated in the Reading Corps K-3 program and a control group that did not participate in the program.^{19,20} This process began by first matching program eligible students within each school within the same grade into pairs based on the similarity of their Fall benchmark scores.²¹ Next, pairs of students were selected for participation within each grade. To address the study's two main research questions, the selection of pairs for Kindergarten and first grade

¹⁴ The study is designed to have the power to detect an effect size that is equivalent to the effect size found among Kindergarten, first, and second and third grade students in the previous Minnesota Reading Corps study.

¹⁵ A recurring inquiry about our analysis concerns the difference in the impact of the Reading Corps program between the urban and rural schools. While we agree that this is an important topic, the sample does not support testing differences in the impact between urban and rural schools because urbanicity is confounded with a number of other factors related to achievement (e.g., race/ethnicity, FRPL eligibility, teacher to student ratios). These confounding variables make any observed differences difficult to interpret and attribute to urbanicity alone.

¹⁶ Schools were eliminated from the sampling frame because they had low numbers of eligible students (i.e., inability to provide both a program and control group), were implementing a pilot or alternative program design (not the typical Reading Corps program), already were participating in another overlapping evaluation study, and/or were in a remote location (i.e., more than 200 miles from the Twin Cities).

¹⁷ PPS sampling allows for schools with larger numbers of eligible students to have a higher likelihood of being selected.

¹⁸ Six St. Paul Public Schools (SPPS) were originally selected as part of the school sample; however, SPPS declined to have their schools participate in the study. In some cases, replacement schools were identified to participate or schools had to be dropped and not replaced in the study.

¹⁹ Students who participated in Reading Corps in previous years were eligible for participation in the evaluation. The important eligibility criterion for students was not whether they had received Reading Corps services in the past, but instead, whether they were eligible to receive services at the beginning of the 2017-2018 school year. Furthermore, since students were randomly assigned to condition, it was equally likely that a student who previously received services would be assigned to the program and control groups. As such, each group should have a roughly equal number of students who had and had not participated previously in the Reading Corps program.

²⁰ Processes for identifying which students should be administered the Fall benchmark assessment can somewhat vary across schools. While it was recommended that all students in grades K-3 be assessed, some schools may have used previous years' test scores or other more subjective means for identifying students to be assessed.

²¹ Tier 1 scores and low Tier 3 scores requiring support from a school staff member were not eligible for the program, and thus, not eligible for the study.

students was randomly conducted to ensure a diversity of scores (closest and furthest from the expected Fall benchmark), while pairs of second and third grade students were purposely selected to include students whose scores were furthest from the Fall benchmark. Then, one student in each selected pair was randomized into the Reading Corps program while the other was assigned to the control group.²² This matched pair design ensured that students in the program and control groups had similar distributions of Fall benchmark scores at the start of the school year.²³ Key characteristics of the 24 participating schools are provided in **Table 2** below. The final column lists the number of students who were included in the final analytic sample (N=622), while the next to last column lists the number of eligible students randomly assigned to receive Reading Corps services at each of the 24 schools (N=724). In total, 724 eligible K-3 students were selected to participate in the evaluation.

²² A figure illustrating the randomization process is provided in **Appendix B**.

²³ A baseline equivalence analysis was conducted to confirm that the two groups formed through random assignment were similar at baseline (see **Appendix C**).

Table 2. Characteristics of schools participating in the Minnesota Reading Corps Impact Evaluation (Fall 2017)

School, Urbanicity	Number of Reading Corps Tutors ^a			% Free and Reduced-Price Lunch ^b	%Non-White ^b	School Enrollment ^{b,c}	Study Participants (N) ^d	
	Full Time	3/4 Time	Part Time				Randomized Sample	Analytic Sample
Urban							83.8% (K), 44.2% (F), 53.1% (S/T)	80.0% (K), 42.5% (F), 51.2% (S/T)
Achieve Language Academy	2	0	0	82.80%	88.45%	407	32 (K), 8 (F), 20 (S/T)	26 (K), 8 (F), 20 (S/T)
Burroughs Community School	1	0	1	11.78%	21.20%	764	8 (F), 34 (S/T)	8 (F), 28 (S/T)
College Preparatory Elementary	1	1	0	91.09%	98.28%	348	12 (F), 18 (S/T)	10 (F), 14 (S/T)
Dowling School, Urban Environmental Magnet	2	0	0	48.51%	47.91%	503	34 (S/T)	28 (S/T)
Earle Brown Elementary	2	0	1	81.61%	87.86%	1104	12 (K), 14 (F), 52 (S/T)	8 (K), 14 (F), 42 (S/T)
Hiawatha Leadership Academy - Morris Park	1	0	2	91.47%	98.97%	387	8 (K), 6 (F), 18 (S/T)	4 (K), 6 (F), 10 (S/T)
Lake Nokomis Community School - Keewaydin Campus	0	0	2	51.63%	50.93%	430	20 (T)	14 (T)
Northrop Urban Environmental Learning Center	1	1	0	33.07%	33.66%	511	12 (F), 24 (S/T)	12 (F), 22 (S/T)
Stonebridge World School	2	0	0	91.13%	97.61%	293	10 (K), 12 (F), 22 (S/T)	10 (K), 8 (F), 16 (S/T)
Valentine Hills	0	0	1	45.43%	44.16%	634	4 (F), 12 (S/T)	2 (F), 12 (S/T)
Rural							16.2% (K), 55.8% (F), 46.9% (S/T)	20.0% (K), 57.5% (F), 48.8% (S/T)
Adrian Elementary	1	0	0	36.92%	11.92%	260	10 (F), 14 (S/T)	10 (F), 14 (S/T)
Brandon Elementary	1	0	0	31.58%	5.85%	171	14 (S/T)	10 (S/T)
Breckenridge Elementary	1	1	0	42.05%	10.93%	302	12 (F), 18 (S/T)	12 (F), 12 (S/T)
Janesville-Waldorf-Pemberton Elementary	1	0	0	32.77%	4.52%	354	14 (S/T)	12 (S/T)
Lyle Elementary	0	0	1	59.68%	20.16%	124	12 (S/T)	8 (S/T)
McGregor Elementary	1	0	0	66.43%	32.13%	277	12 (F), 18 (S/T)	12 (F), 16 (S/T)
Parkers Prairie Elementary	1	0	0	37.72%	5.54%	289	12 (F), 12 (S/T)	12 (F), 12 (S/T)
Pine City Elementary	2	0	1	43.07%	7.48%	815	24 (F), 36 (S/T)	22 (F), 32 (S/T)
Rippleside Elementary	1	0	0	49.92%	5.69%	633	8 (K), 12 (F), 12 (S/T)	8 (K), 12 (F), 12 (S/T)
Rothsay Elementary	1	0	0	47.85%	4.84%	186	14 (S/T)	14 (S/T)
Southland Elementary	0	0	1	40.72%	10.82%	194	4 (K), 4 (F), 8 (S/T)	4 (K), 4 (F), 8 (S/T)
Swanville Elementary	1	0	0	40.98%	3.28%	183	10 (F), 22 (S/T)	8 (F), 22 (S/T)
Underwood Elementary	0	1	0	44.41%	2.80%	322	18 (S/T)	14 (S/T)
Walker-Hackensack-Akeley Elementary	0	0	1	e	e	e	12 (S/T)	10 (S/T)

a: 2017-2018 data directly from RC, including updates and changes documented by email as noted throughout the year.

b: 2015-2016 data from CCD (NCES)

c: All students in all grades, excluding Alternative Education

d: (K) indicates Kindergarten; (F) indicates first grade; (S/T) indicates second/third grades.

e: CCD (NCES) data not available for any year 2001-2016.

Table 3 presents descriptive statistics for the K-3 students included in the evaluation. Demographics include gender, race/ethnicity, English Language Learner (ELL) status, and Fall benchmark scores. The analytic sample size for the evaluation (N=622) is smaller than the 724 eligible students initially randomized into the study. During the school year, some students left the school (i.e., moved) or were chronically absent and did not receive regular Minnesota Reading Corps tutoring or assessments. Additionally, some parents did not consent to their child’s participation in the evaluation. These students and their matched pair were removed from the analytic sample (i.e., pairwise deletion). Thus, the final sample of students included in the evaluation totaled 622 students.²⁴

Table 3. Student participants for the Minnesota Reading Corps Impact Evaluation (Fall 2017)

	Kindergarten (N=60)		1st Grade (N=160)		2nd Grade (N=190) ^a		3rd Grade (N=212) ^a	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female	52%	—	51%	—	54%	—	48%	—
Race/Ethnicity		—		—		—		—
White	27%	—	66%	—	62%	—	59%	—
Black	23%	—	11%	—	15%	—	18%	—
Asian	12%	—	7%	—	7%	—	6%	—
Hispanic/Latino	37%	—	10%	—	11%	—	12%	—
American Indian or AK Native	2%	—	3%	—	5%	—	4%	—
Multi-racial	—	—	3%	—	1%	—	1%	—
Native HI or Other Pacific Islander	—	—	—	—	—	—	<1%	—
Other	—	—	—	—	—	—	<1%	—
English Language Learner (ELL)	37%	—	31%	—	23%	—	24%	—
Fall Benchmark Score ^b	2.02	2.26	23.04	6.89	39.01	12.74	69.64	18.42

a: 2nd and 3rd grades were pooled for the analysis; Total N=402

b: Kindergarten = Letter sound fluency; 1st grade = Nonsense word fluency; 2nd and 3rd grades = Oral reading fluency

The evaluation team conducted a power analysis prior to developing the evaluation’s sampling plan, in which the number of students required to detect a difference between the treatment and control groups was calculated.²⁵ As shown in **Table 3** above, 60 Kindergarten students, 160 first grade students, 190 second grade students, and 212 third grade students were included in the evaluation, thus the minimum sample sizes for the evaluation were achieved or nearly achieved, and the study was adequately powered to detect plausible effect sizes.

The evaluation team used a NORC proprietary, centralized password-protected website to conduct the within-pair randomization of students to the program or control groups. Minnesota Reading Corps Internal Coaches entered Fall benchmark and demographic data on all program-eligible students into the website. The website included grade-specific Fall benchmark assessment score range checks to ensure that only those students eligible for Minnesota Reading Corps services were included in the sample. Once

²⁴ See **Appendix D** for details on sample attrition.

²⁵ For Kindergarten, a minimum of 40 students (i.e., 20 pairs of students) was needed to detect an effect size of 0.67, for first grade, a minimum of 140 students (i.e., 70 pairs of students) was needed to detect an effect size of 0.39, and for the combined second and third grade analysis, a minimum of 420 students (i.e., 210 pairs of students) was needed to detect an effect size of 0.2.

all eligible students' data were entered into the website and confirmed, the website then automatically randomized students within grades within pairs to either the program or control group and displayed each student's assignment. The Internal Coach then shared the list of students with the Reading Corps tutors.

Because the evaluation was designed to measure the impact of Minnesota Reading Corps K-3 program participation relative to nonparticipation, students in the control group were embargoed from receiving Reading Corps tutoring services during the first semester of the school year (grades K-1), or until after the Spring benchmarking period (grades 2-3). Thus, Internal Coaches and Reading Corps tutors were asked to tutor the students assigned to the program group as usual and to not tutor the control group students until after the Winter (grades K and 1) or Spring (grades 2 and 3) benchmark. The evaluation team remained in regular contact with the schools to verify that program students received Minnesota Reading Corps tutoring services and control students did not.

Data Sources

The primary data for the evaluation are comprised of the Fastbridge reading assessments described above. The evaluation team partnered with the Minnesota Reading Corps program to utilize the Fall benchmark assessment data that Reading Corps tutors collected on all K through 3rd grade students enrolled at the 24 eligible Reading Corps schools in Minnesota. Minnesota Reading Corps has been completing Fastbridge reading assessments with students for the past five years, so the Fall benchmark data collection was standard practice for the program. Also, bias in the collection of pretest Fall scores by program staff was highly unlikely because students had yet to be assigned to treatment or control groups. As mentioned previously, Reading Corps tutors collect Fall benchmark data on all K-3 students who the school identifies as potentially eligible for Reading Corps services.²⁶ This procedure identifies objectively those students who meet Reading Corps eligibility criteria. Given that tutors already collect this data, and so as not to duplicate assessment efforts, the evaluation team requested access to the 2017 Fall benchmark data in order to identify the pool of students eligible to participate in the evaluation.

In order to address potential bias at posttest, independent assessors from the NORC evaluation team collected the Winter (grades K-1) and Spring (grades 2-3) benchmark assessment data from students in both the treatment and control groups at each of the 24 schools. The independent assessors completed the exact same assessments using the same tools as the Reading Corps program staff and participated in an assessment training that is similar to the training provided to Reading Corps tutors. Informed consent from students' parents was obtained using an opt-out consent procedure. Additionally, program staff at the 24 participating schools provided the evaluation team with other administrative data including students' tutoring attendance, average length and number of sessions, and demographic characteristics.

²⁶ In many cases, schools assess all of their K-3 students for eligibility.

Analysis Methods

The main study analysis was designed to determine if there is an impact of the Reading Corps program on student literacy outcomes, and if so, to estimate the average causal treatment effect of the program for Kindergarten students, first grade students, and second and third grade students combined. We estimated program effects by statistically comparing the average score of students who received Reading Corps tutoring with the average score of a similar control group of students who did not. The calculated difference in group averages is the causal effect of program participation. This effect is considered “causal” because the primary predictor variable (the program) was randomly assigned to students in the two groups, and therefore, not related to any other student characteristics that may affect the outcome. For this report, to estimate a plausible effect, we fit a statistical model that includes information about the student’s Fall test score and membership in major demographic groups (e.g., race, gender, and English Language Learner status), which improves the precision of our estimates. From this model, we predict the average Winter (for Kindergarten and first grade students) or Spring (for second and third grade students) scores for the program, and control for program participation and the associated sampling error. We then compute the likelihood²⁷ of observing our results when the program has no effect (noted as a p-value). If this likelihood is low (less than 0.05), we conclude that it is likely that the program had an effect on the winter or spring scores.

The impact evaluation analyses were conducted using mixed-regression models, also known as hierarchical linear models.²⁸ This model uses maximum likelihood to estimate differences in outcomes associated with predictor variables, while also modeling the multi-level structure of the data, which in this case, are students nested within schools. Two models are estimated for each outcome. The first model uses program status as the only predictor. From this first model, we calculate an effect size based on typical standard deviation-based effect sizes such as the Cohen’s d or Hedges’ g parameter for multisite randomized designs.²⁹ The second model includes control variables to remove any spurious effects of demographic characteristics on outcome values. While the program and control groups are well balanced, they are not perfectly equal, and thus, the treatment effect is slightly different. From this second model, we then predict the scores for program and control groups based on the program variable and holding all other controls as their sample means. This is the information that was used to produce the plots and tables provided in the report. More detailed information about the specific statistical models employed is available in **Appendix C**.

Comparison Study Methods. For comparing the treatment effects of the 2014 and 2018 evaluations, we are limited in our ability to meet causal assumptions because students were not randomly assigned to a particular study cohort and unknown factors may correlate with the “history” associated with each

²⁷ The p-value summarizes the likelihood of the observed data’s test statistic or a statistic of higher magnitude under the assumption of the null hypothesis.

²⁸ Also note that our models include weights to adjust for student non-response and school generalizability.

²⁹ Hedges, L. V., & Rhoads, C. (2010). *Statistical Power Analysis in Education Research*. NCSER 2010-3006. *National Center for Special Education Research*; Hedberg, E. C. (2017). *Introduction to Power Analysis: Two-group Studies* (Vol. 176). SAGE Publications.

cohort.³⁰ Therefore, we relied upon quasi-experimental techniques to construct plausible comparisons of study impacts. However, we were only able to conduct a statistical comparison between the findings for Kindergarten and first grade students due to differences in the evaluation periods for second and third grade students (i.e., one semester in 2014 vs. a full-year for 2018). Similar Fall and Winter benchmark assessments were collected for the 2014 and 2018 evaluations, so these measures were used to make comparisons between the Kindergarten and first grade student effects.³¹ Thus, 151 Kindergarten and 356 first grade students from both studies were included in the analysis resulting in four study groups:³²

- The control group in 2014,
- The treatment group in 2014,
- The control group in 2018, and
- The treatment group in 2018.

We used quasi-experimental methods to assess whether program impact changed as the program changed (2014 to 2018) by examining the interaction between the randomly assigned treatment/control differences between non-randomly assigned cohorts of the program. To maximize the data utilized, we used propensity score weighting to weight the data based on key covariates to produce comparable groups.³³ We fit a multinomial probability model to model membership in each of the four groups using a set of key covariates, weighted the data accordingly, and then calculated regression adjusted means to estimate treatment effects. Our key covariates included demographic group membership (i.e., a combination of race, gender, and ELL status), Fall baseline measurement, and the quantile group of fall baseline.

Appendix G includes descriptive statistics of the sample without weights.

Study Limitations

The design of this study is strong because the primary predictor of interest, receiving the program, was randomly assigned, and thus, is exogenous to all other variables in the model. That is, a strong argument can be made that improvements in Winter and Spring benchmark scores are the result of the program and not from unobserved factors, which leads to high levels of internal validity. However, there are always limitations to any study.

As noted above, the evaluation team conducted a power analysis prior to developing the sampling plan, to ensure an adequate number of students in order to detect a difference between the treatment and control

³⁰ Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

³¹ In 2014, the second and third grade students were assessed in Winter when assignment to randomized groups was maintained, but the spring assessment was after membership in the randomized groups was relaxed. In 2018, randomization was maintained through spring, and no winter assessment was collected. As a result, we are not able to compare effects for 2nd and 3rd grade students.

³² To maintain comparability with the 2018 study, we removed 2014 students who attended St. Paul Public Schools (See footnote #26). In addition, 8% of the 2014 students who had very high Fall benchmarks that exceeded the maximum value of the 2018 Fall benchmark scores were excluded. Students with abnormally small predicted probabilities of membership in any of the four groups were also removed.

³³ Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155, 138–154.

groups. However, while the sample is adequately powered to detect effects, it is still limited by the availability of program eligible students. This situation raises the possibility that the effects represented in this report only reflect the effects for a specific population of eligible students and schools, and that results will be difficult to replicate in other locations.³⁴ The small sample sizes in Kindergarten and first grade also limited our ability to conduct subgroup analyses on demographic groups of students in these grades.

Another common limitation in education impact evaluations is student attrition. As mentioned above, the evaluation team implemented a matched-pair experimental design for the study. Students within a school within grade were matched on Fall benchmark scores and then randomly assigned within pair to either the program or control group. Statistical tests showed that the program and control groups did not differ on important demographic variables and proficiency measures at the beginning of the school year. As is inevitable, students moved away or were chronically absent from school (i.e., attrition). When this occurred, the affected student and the matched student within the pair were both dropped from the study's analytic sample. This pairwise deletion procedure reduced the overall sample size, but ensured the integrity of the RCT design. A detailed description of attrition by grade level may be found in **Appendix D**.

³⁴ Exacerbating this general concern over sample representativeness is that the St. Paul Public Schools (SPPS) declined to participate in the evaluation, thereby excluding one of the largest and most diverse urban school districts in the state. To overcome this limitation, the study included additional urban charter schools located in St. Paul and urban schools from other districts to diversify the urban school sample.

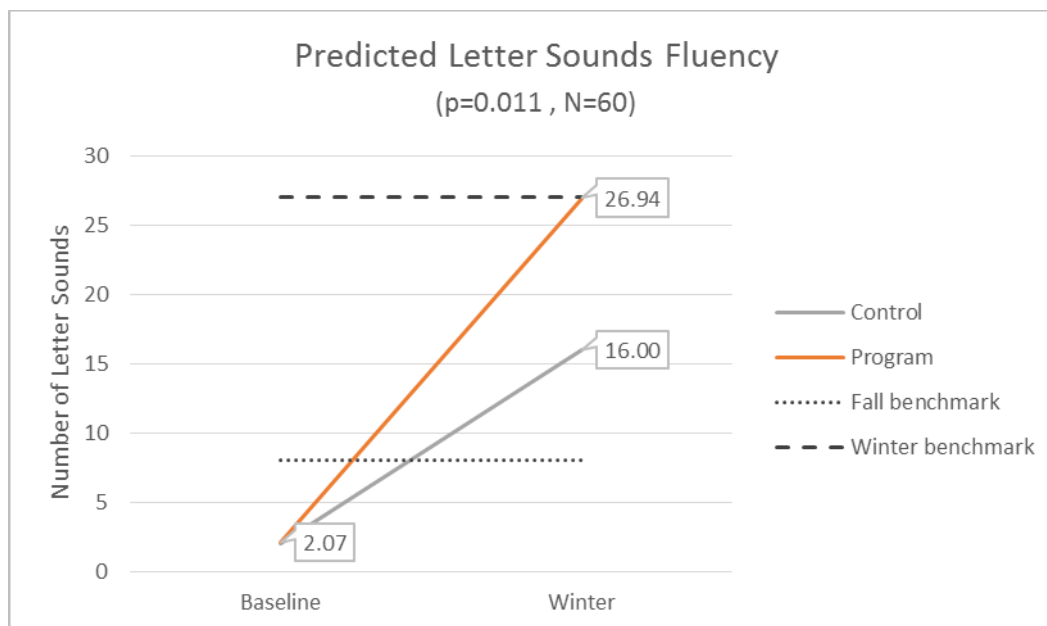
Study Findings

In this section, we present the major findings from the analysis of Kindergarten, first grade, and second and third grade study participants. The “p-value” represents the probability of observing the data if we assume the null hypothesis that the program and control groups have the same average outcomes. The “effect size” indicates the results from a model without controls that is converted into standard deviation units. Below we summarize the study’s key findings. Detailed tables presenting the results of our analysis are provided in **Appendix F**.

What is the impact of Minnesota Reading Corps on Kindergarten students?

The Minnesota Reading Corps program had a significant and large impact on Kindergarten students’ *letter sound fluency* scores between the Fall benchmark (September 2017) and the Winter benchmark (January 2018). By the end of the first semester, Kindergarten students in the program correctly identified an average of 10.9 more letter sounds correctly in a one minute period than Kindergarten students in the control group ($p=0.01$, $N=60$). This represents a large effect size of 0.85 standard deviations, indicating that the average program student scored over three-fourths of a standard deviation unit higher than the average control group student in identifying letter sounds. **Figure 2** presents the findings visually and shows that the program Kindergarten students’ average score at Winter benchmark was 26.9 letter sounds compared to 16.0 letter sounds for the average control student. Also by the end of the first semester, the program group achieved Winter benchmark expectations of 27 letter sounds, whereas the control students’ average score did not reach the expected benchmark.

Figure 2. Mean scores for Kindergarten program and control students



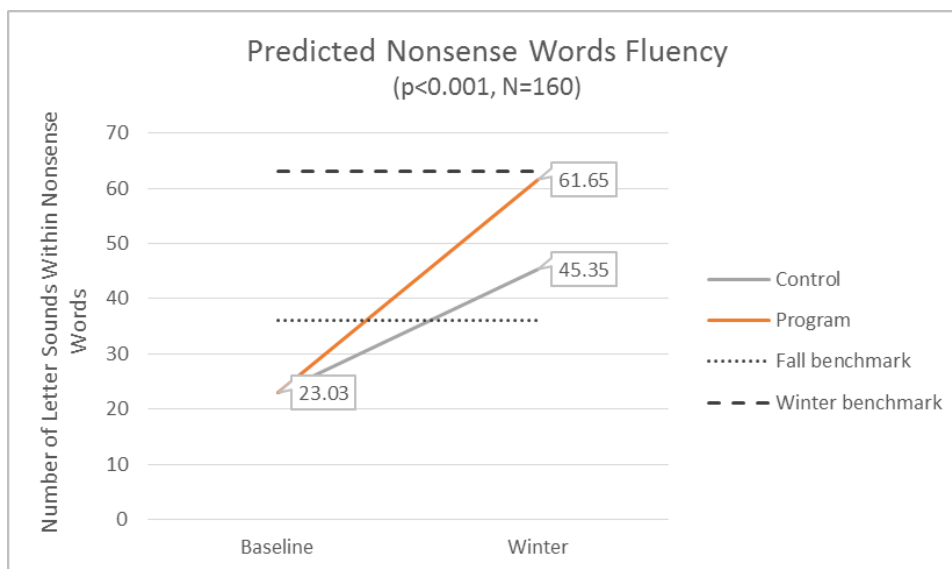
What is the impact of Minnesota Reading Corps on first grade students?

For first grade students, the study examined two measures, *nonsense word fluency* and *oral reading fluency* (Winter benchmark only). While reading fluency is not typically a focus of Reading Corps tutoring until the second semester in first grade, the Winter benchmark for *oral reading fluency* was collected to serve as a baseline measure of this skill in first grade. However, *oral reading fluency* scores would theoretically be impacted by the intervention, as students begin to apply early phonics skills to the more complex skill of reading fluency, demonstrating potential generalization of tutoring to a more complex task.

Nonsense Word Fluency

Consistent with the findings for Kindergarten participants, first grade students enrolled in the Minnesota Reading Corps program attained significantly higher *nonsense word fluency* scores by Winter benchmark than first grade students in the control group. By the end of the first semester, first grade students who received Reading Corps tutoring correctly identified in one minute an average of 16.3 more letter sounds embedded within non-real words than first grade control group students identified ($p < 0.001$, $N=160$). This represents a large, positive effect size of 0.81 standard deviations, indicating that the average program student scored more than three-fourths of a standard deviation unit higher than the average control group student in identifying letter sounds within nonsense words. **Figure 3** presents the findings visually, which shows that the program students' average score of 61.7 was substantially higher than the control group's average score of 45.4. Furthermore, the first grade program students' average score nearly met the Winter benchmark of 63 letter sounds within nonsense words in a one-minute period. The study observed a meaningful difference between the two study groups in terms of their progress between Fall and Winter benchmarks, so first grade program students are likely to continue to grow at an accelerated pace compared to the control group students.

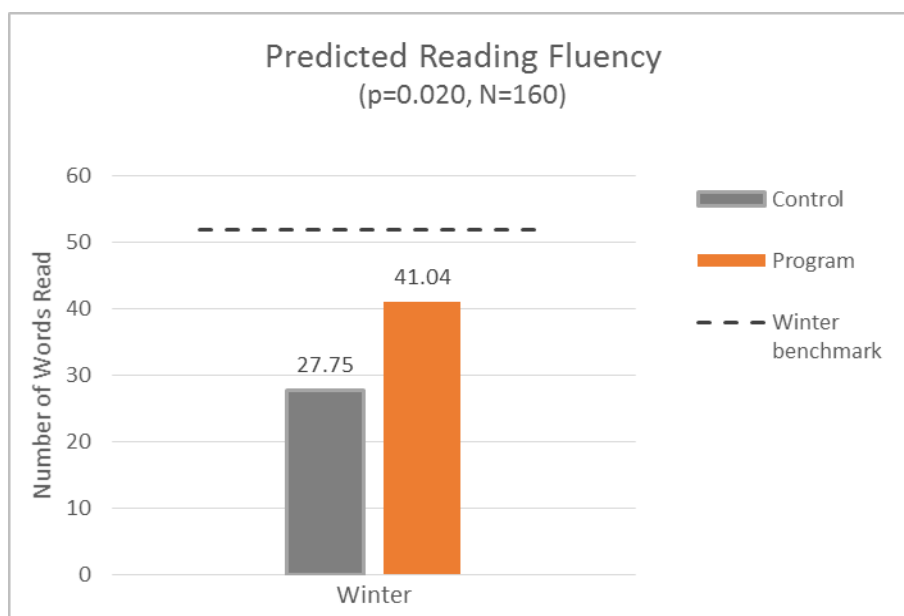
Figure 3. Mean scores for first grade program and control students



Oral Reading Fluency

First grade students enrolled in the Minnesota Reading Corps program achieved substantially and significantly higher reading fluency scores at Winter benchmark than first grade control group students (see Figure 4). On average, program students correctly read aloud 13.3 more words than the control group ($p=0.02$, $N=160$). This represents a large, positive effect size of 0.61.

Figure 4. Mean scores for first grade program and control students (Oral Reading Fluency)



What is the impact of Minnesota Reading Corps on second and third grade students?

Among the combined group of second and third grade study participants, students enrolled in the Minnesota Reading Corps program attained significantly higher *oral reading fluency* scores by Spring benchmark than similar students in the control group. By the end of the school year, second and third grade students who received Reading Corps tutoring read aloud an average of 6.4 more words in a one minute period than second and third grade students in the control group ($p < 0.005$, $N=402$). This represents a meaningful, positive effect size of 0.28 standard deviations, indicating that the average program student scored more than one-quarter standard deviation unit higher than the average control group student in the number of words read aloud. Although the program students' average scores (92.3 for second grade and 112.3 for third grade) were substantially higher than the control group's average scores (84.8 for second grade and 106.8 for third grade) in both grades, as shown visually in Figure 5, neither the program nor control groups for second or third grade students' average scores reached the Spring benchmarks of 116 and 135 words read aloud, respectively. The study did observe a meaningful difference between the pooled study groups in terms of their progress between Fall and Spring benchmarks, so second and third grade program students, if they continue to receive Reading Corps through the end of the school year, are likely to continue to grow at an accelerated pace compared to their similar control group peers.

Table 4. Reading Fluency Results, Pooled Sample

Grade and Outcome	Treatment effect ^{a,b} (β_1)	p-value ^a	Effect size ^c (α_1/σ)	Prog. Mean ^d	Control mean ^d	Total N
Second and Third Grades^e						
Reading fluency	6.44	0.005	0.28	102.6	96.2	402
	(2.291)					

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary

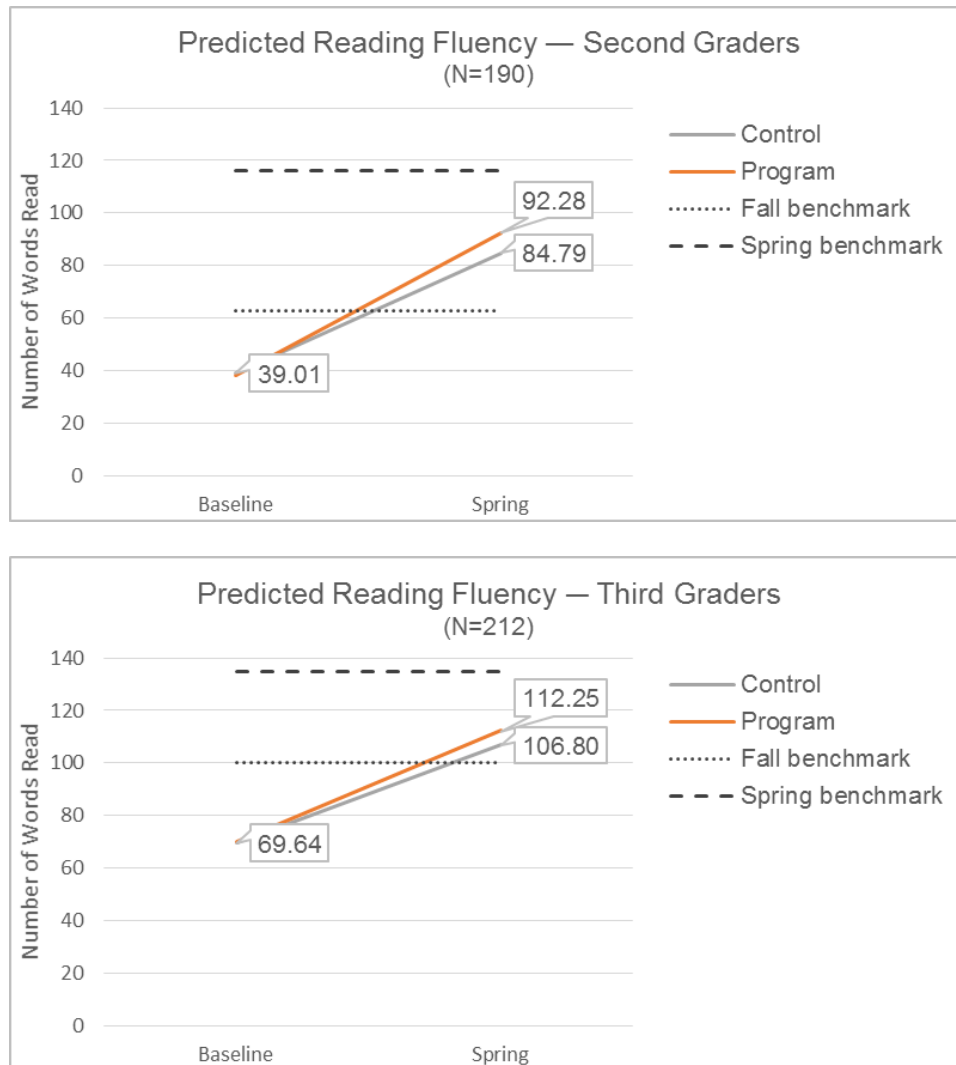
b: Standard errors in parentheses

c: Effect size reported from unconditional model without control variables

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: The Minnesota model includes sampling weights for generalizability

Figure 5. Mean scores for second and third grade program and control students by grade



Findings by Major Demographic Group

To address the study's third research question focused on whether program impacts vary by participant demographics, we originally planned to include a set of subgroup analyses of K-3 student outcomes. However, the sample sizes for demographic groups in Kindergarten and first grade were relatively small and lacked the statistical power to detect differences both within and between subgroups. Therefore, we were unable to conduct analyses on subgroups for Kindergarten and first grade students.

Due to the larger sample sizes for the combined second and third grade sample, our findings include a set of subgroup analyses of study participants in these combined grades. The evaluation team identified subgroups for further examination based on the major demographic groups generally of interest in education research. We provide results by race, gender, and English Language Learner (ELL) status. For race/ethnicity, we were only able to examine differences between the program and control groups in *oral reading fluency* for second and third grade Black/African American students due to the small sample sizes for additional subgroups of Hispanic/Latino, Asian, and Native American students.

For the subgroup analysis, the evaluation team fit models to the second and third grade data that were moderated by major subgroup (e.g., females, males, etc.), to estimate how the differences between the program and control groups varied by demographic group, and conducted significance tests to examine treatment effects for each subgroup. Results from the subgroup statistical tests are provided in **Table 5** and **Table F.2 in Appendix F**.

Race/Ethnicity

For the analysis of second and third grade *oral reading fluency* by race, we found that African American second and third grade students in the program group read aloud an average of 10.7 more words in a one minute testing period by the Spring benchmarking period compared to similar students in the control group ($p=.03$). This difference had a positive, meaningful effect size, indicating that the average African American program student scored one half standard deviation unit higher than the average African American control group student in reading fluency (effect size = 0.50).

Gender

While the subgroup analysis found statistically significant effects of the program for both female and male students in second and third grades, **Table 5** shows that the effect size in *oral reading fluency* between the treatment and control groups was descriptively larger for males than females. Most notably, by the Spring benchmarking period, the average difference between program and control groups among second and third grade male students was larger than the impact found for the general second and third grade population. While all second and third grade students receiving Minnesota Reading Corps tutoring read aloud an average of 6.4 more words correctly in a one minute period ($p=.005$, $N=402$) by the Spring benchmarking period, among second and third grade males, the average difference between the program and control groups was 7.9 words ($p=.013$). Furthermore, this difference had a positive, medium effect size, indicating that the average male program student scored nearly one half standard deviation units higher than the average male control group student in reading fluency (effect size = 0.47). In contrast, for

female students in second and third grade, the findings showed the average difference between treatment and control groups to be slightly lower (5.7 more words read correctly and effect size = 0.22).

English Language Learners

Compared to the analysis of all second and third grade students, some important differences also were detected in the patterns of findings for second and third grade ELL students. By the end of the Spring benchmarking period, second and third grade ELL students in the Minnesota Reading Corps program read aloud correctly an average of 16.3 more words in a one minute period than students in the control group (p=.001). This difference resulted in a positive, substantial effect size, indicating that the average ELL program student scored over one half standard deviation unit higher in *oral reading fluency* than the average ELL control group student (effect size = 0.53). In contrast, *all* second and third grade students receiving Minnesota Reading Corps tutoring read aloud an average of 6.4 more words correctly in a one minute period (p=.005, N=402) by the Spring benchmarking period compared to students in the control group.

Table 5. Reading Fluency by Subgroups, Pooled Second and Third Grades Sample

Subgroup Characteristic ^f	Total N ^e	Reading Fluency				
		Treatment effect ^{a,b} (β ₁)	p-value ^a	Effect size ^c (α ₁ /σ)	Prog. Mean ^d	Control mean ^d
Racial or Ethnic subgroups						
American Indian or AK Native	18	5.914 (5.414)	0.275	G	97.259	91.345
Asian	25	9.065 (8.023)	0.259	G	97.835	88.770
Black or African American	66	10.650 (5.031)	0.034	0.497	104.697	94.047
Hispanic/Latino	46	3.771 (5.181)	0.467	g	100.667	96.896
Gender subgroups						
Female	204	5.735 (2.486)	0.021	0.220	101.014	95.279
Male	198	7.878 (3.169)	0.013	0.467	102.311	94.433
Language subgroups						
ELL	94	16.279 (4.672)	0.000	0.528	113.992	97.714

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary.

b: Standard errors in parentheses.

c: Effect size reported from unconditional model without control variables.

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: Subgroups containing fewer than ten students were excluded from analysis.

f: The Minnesota model includes sampling weights for generalizability.

g: Non-significant effect sizes not reported.

Findings by Dosage Characteristics

In addition to the overall effect of the Minnesota Reading Corps K-3 program on students, the study also explored the effect of dosage (i.e., amount of intervention received) on student outcomes to identify how much intervention is needed to produce anticipated effects. As explained in the detailed analysis section in **Appendix E**, dosage analyses are often difficult to perform and interpret with reactive programs (i.e., programs that determine dosage based on changing need). In this study, we attempted to surmount these issues by incorporating three aspects of “dosage” into the model: 1) Number of sessions; 2) Average length of sessions; and 3) Attendance. Attendance is defined as the proportion of planned sessions completed. We present these effects from a multivariate mixed regression model where the effect of any one aspect of dosage holds constant the influence of the other aspects. That is, for example, we present the effect for attendance while holding constant the average length of sessions. Tables presenting the results of our analysis of dosage are provided in the **Appendix F**.

The dosage analysis for Kindergarten students did not yield significant findings.³⁵ The direction of the correlations with student benchmark scores was not unexpected, but requires explanation. For example, the association between student outcomes and total number of sessions was negative, reflecting the reactive nature of the Reading Corps intervention. It is often the case that programs provide more sessions to those students who are farthest from benchmark or most in need. This finding demonstrates this inverse relationship. A more promising finding (but still not statistically significant) was the positive association between Kindergarten student outcomes and attendance. This relationship indicates that when students attended the planned number of sessions appropriate for their level of need, they tended to score higher on the Winter benchmark assessment.

The dosage analysis for first grade students yielded similar findings to those found among the Minnesota Kindergarten students. The association between outcomes and total number of sessions was negative, significant, and strong (correlation = $-.66$; $p = 0.001$), reflecting the reactive nature of providing more sessions to students who were most in need. The association with attendance was positive, similar to Kindergarten students, suggesting that when there was good coordination between student need and planned number of sessions, students who received tutoring with high fidelity produced higher Winter benchmark scores.

Consistent with the findings in the younger grades, the association between outcomes and attendance among second and third grade students was again positive, albeit not significant, and the association between outcomes and total number of sessions was negative and significant (correlation = $-.32$; $p = 0.000$). These findings once again support the understanding that students who attended their planned number of sessions appropriate for their level of need produced higher Spring benchmark scores.

³⁵ Sample sizes were again small for the Kindergarten subgroup analysis, which most likely contributed to the lack of statistically significant findings. Another contributing factor is the fact that many of the predictors used in the dosage analysis are correlated, which produces larger statistical uncertainty in the parameters and makes significance more challenging to achieve.

Findings from 2014 and 2018 Study Comparisons

As noted previously, NORC evaluated the Minnesota Reading Corps K-3 program in 2014.³⁶ Therefore, we statistically compared the results from the present study with those from 2014. Due to differences in schools, sample sizes, and data frequency, it was important to employ quasi-experimental methods to construct comparable groups to compare the results of the two evaluations. This was possible for Kindergarten and first grade; however, we were unable to perform a statistical comparison between 2014 and 2018 for the second and third grades due to the difference in evaluation periods (one semester in 2014 vs. a full-year for 2018). Therefore, below we compare the effects from a matched subset of the 2014 evaluation Kindergarten and first grade sample to the effects found in the 2018 sample using propensity weights.

Kindergarten

Table 6 presents the treatment effects for the propensity weighted 2014 and 2018 Kindergarten samples, as well as the difference between study samples' treatment effects. In the 2014 study sample, program group students identified 11 more letter sounds compared to control group students (difference = 11.0; SE=4.405); this effect was statistically significant (p=0.037) with the adjusted p-values. For the 2018 study sample, there also was a statistically significant treatment effect in which program group students identified 10.74 more letter sounds than control group counterparts (difference = 10.7; SE=3.3; p=0.004). In comparing the treatment effects for the 2014 and 2018 study samples, the 2018 study impact differed by 0.28 fewer letter sounds than their 2014 matched counterparts (difference in difference = -0.28; SE=4.308). Given this small difference, there was no statistically significant difference between the sizes of the 2014 and 2018 study samples' impacts (p=1.000). Therefore, both studies demonstrate comparably large impacts on Kindergarten students in the program.

Table 6. Treatment Effects in Propensity Weighted Kindergarten Samples

Outcome	Treatment effect ^{a,b}	p-value ^{a,c}	Control Mean	Program Mean
2014 Study Sample				
Letter Sounds fluency	11.019	0.037	24.791	35.810
Winter benchmark = 27	(4.405)			
2018 Study Sample				
Letter Sounds fluency	10.735	0.004	18.080	28.815
Winter benchmark = 27	(3.304)			
Difference Between Study Samples' Effects (2018-2014)				
Letter Sounds fluency	-0.284	1.000		
Winter benchmark = 27	(4.308)			

a: Reported statistics from predictive model that includes the Fall assessment, gender, race, ELL status

b: Cluster robust standard errors in parentheses.

c: Bonferroni-adjusted p-values

³⁶ Markovitz, C.; Hernandez, M.; Hedberg, E.; Silbergliitt, B. (2014). *Impact Evaluation of the Minnesota Reading Corps K-3 Program*. NORC at the University of Chicago: Chicago, IL.

First Grade

Table 7 displays the treatment effects for the propensity weighted 2014 and 2018 first grade samples, as well as the difference between study samples’ treatment effects. In the 2014 study sample, there is a statistically significant treatment effect in which program group students identified 9 more letter sounds within nonsense words than their control group counterparts (difference =9.0; SE=2.7; p=0.002). For the 2018 study sample, there was a statistically significant treatment effect in which program group students identified 18.1 more letter sounds within nonsense words than control group counterparts (difference =18.1; SE=2.0; p<0.001). Finally, there was a statistically significant difference between the 2014 and 2018 study samples’ treatment effects, in which the 2018 study impact resulted in an increase of 9.1 nonsense words compared to their 2014 counterparts (difference in difference = 9.1; SE=3.5; p=0.027).

The larger effect size for first grade students in the 2018 study may be due to several reasons. An obvious explanation is that the Reading Corps program’s impact on students continues to improve as the program matures and the intervention model is further refined. However, another possible explanation is that the benchmark target scores, which help determine program eligibility and exit criteria, have changed between 2014 and 2018; thus, students may be staying in the program longer and possibly experiencing a larger impact of the program. Nevertheless, it is important to remember that the first grade study duration was for only one semester, and any differences resulting from changes to Winter benchmark exit targets occurred over a relatively short period of time.

Table 7. Treatment Effects in Propensity Weighted First Grade Samples

Outcome	Treatment effect ^{a,b}	p-value ^a	Control Mean	Program Mean
2014 Study Sample				
Nonsense Words fluency	8.986	0.002	50.939	59.925
Winter benchmark = 63	(2.667)			
2018 Study Sample				
Nonsense Words fluency	18.079	<0.001	46.990	65.069
Winter benchmark = 63	(2.006)			
Difference Between Study Samples (2018-2014)				
Nonsense Words fluency	9.093	0.027		
Winter benchmark = 63	(3.487)			

a: Reported statistics from predictive model that includes the Fall assessment, gender, race, ELL status

b: Cluster robust standard errors in parentheses

c: Bonferroni-adjusted p-values

Conclusions and Discussion

The findings from the 2018 Minnesota Reading Corps K-3 program impact evaluation provide important evidence for addressing the study’s research questions. Below, the evaluation team offers our conclusions on the impact of the Minnesota Reading Corps K-3 program on Kindergarten, first grade, and second and third grade students’ literacy proficiency scores, organized by major research question. Following our assessment of the questions is a discussion on the implications of our findings for the Minnesota Reading Corps K-3 program and its replication sites.

1. For Kindergarten and first grade students, what is the impact of a single semester of the Minnesota Reading Corps on program participants compared to similar students who did not receive Reading Corps?

The results of the study showed that Kindergarten and first grade students who received Minnesota Reading Corps tutoring achieved significantly higher literacy assessment scores by the Winter benchmark than did control students who did not participate in the program.

Kindergarten students who received tutoring produced more correct letter sounds by the end of the first semester than did students in the control group, and they reached the Winter benchmark. The effect size of .85 represents more than half (56%) a year’s worth of typical growth in reading skills.³⁷ This effect was not unexpected, given similar findings from the earlier 2014 impact evaluation of the Minnesota Reading Corps K-3 program. Kindergarten students who qualified to receive Reading Corps services were chosen because they knew few letter sounds at the beginning of the school year, as evidenced by low Fall benchmark scores. Students’ low scores at the beginning of the school year are often due to a lack of exposure to or instruction in letter sound correspondence at home or in early education settings. Previously unexposed Kindergarten students who are explicitly taught can quickly learn the correspondence between letters and their sounds.³⁸ Kindergarten students in the Reading Corps program are scheduled to receive at least 100 minutes a week of tutoring focused explicitly on mastering letter sound correspondence and related skills (e.g., phonological awareness). This intensive, targeted intervention produced large gains in letter sound knowledge, which resulted in the average student achieving the Winter benchmark. These findings, coupled with similarly positive findings from the previous 2014 evaluation of the program, confirm that the Minnesota Reading Corps program achieves its goal of accelerating at-risk Kindergarten students’ *letter sound fluency*, setting them on track to exceed grade level expectations by the end of the first semester.

In addition, first grade students participating in the Minnesota Reading Corps program produced similarly large effects on *nonsense word fluency* scores compared to students in the control group by Winter

³⁷ See Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172-177.

³⁸ National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health

benchmark. As with the Kindergartners, first grade students demonstrated both statistically significant and large effects of the program. By the end of the first semester, students who received Minnesota Reading Corps tutoring were able to produce more letter sound segments in nonsense words than students in the control group. The effect size of .81 represents a large proportion (84%) of a typical first grade students' growth in reading ability. Further, first grade students on average nearly reached the Winter benchmark proficiency score by the end of the first semester. In addition, first grade Reading Corps students' higher *oral reading fluency* scores suggested that they were successfully transferring early phonics skills gained from Reading Corps nonsense words fluency tutoring to the more complex skill of reading fluency. These recent study findings, which substantially surpassed the positive and significant findings from the previous 2014 evaluation of the program, confirm the Minnesota Reading Corps program's effectiveness in setting first grade students on average on an accelerated pace to reach benchmark by the end of the first semester compared to first grade control students.

In sum, the results of the 2018 Minnesota Reading Corps impact evaluation on Kindergarten and first grade students suggest that the program produces large, meaningful effects for its younger students within a single school semester. These early phonics skills are theoretically³⁹ and empirically⁴⁰ understood to be necessary for later reading proficiency. Reading Corps tutors' intensive one-on-one tutoring produces increases in Kindergarten students' *letter sound fluency* and first grade students' *nonsense word fluency*, helping them reach expected benchmark scores in Kindergarten (and nearly reach benchmark scores in first grade) within a single semester. The first grade *nonsense word fluency* and *oral reading fluency* results suggest that Minnesota Reading Corps' focus on foundational skills measured by *nonsense word fluency* may transfer to the more complex skill of oral reading within a single semester. Therefore, these findings contribute to the growing body of *empirical* evidence that demonstrates foundational skills directly contribute to reading comprehension later in 3rd grade and beyond.

2. For second and third grade Tier 2 students who are farthest from the Fall reading proficiency benchmark, what is the impact of a full school year of Minnesota Reading Corps on program participants compared to similar students who did not receive Reading Corps?

The effect of the Minnesota Reading Corps program was meaningful and significant for the combined second and third grade students after full-school year access to the program intervention. Second and third grade students who received the program achieved significantly higher literacy assessment scores by the Spring benchmark than did control students who did not participate in the program. In contrast to the Kindergarten and first grade students, however, neither the second nor third grade students on average achieved their expected Spring benchmarks by the end of the school year. While the average second and third grade student did not reach the Spring benchmark, it is important to highlight the magnitude of the program's impact on participating students' *oral reading fluency*, equivalent to 47% of the average

³⁹ Gough, P. B. & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, 7: 6-10.

⁴⁰ Catts, H. W.; Herrera, S.; Nielsen, D. C.; & Bridges, M. S. (2015). *Reading and Writing: An Interdisciplinary Journal*, v28 n9 p1407-1425.; Stanley, C. T.; Petscher, Y.; & Catts, H. W. (2018). *Reading and Writing: An Interdisciplinary Journal*, v31 n1 p133-153.

student's annual growth in reading.⁴¹ Also, as discussed above, recent studies support the understanding that *oral reading fluency* is both theoretically and empirically related to overall reading proficiency.⁴²

The second and third grade analysis also purposely selected pairs of students at each site who were furthest from the Fall benchmark in order to better understand the impact of the program on these students. The findings indicate that on average second and third grade students who were furthest from benchmark at the beginning of the school year show significant improvements in *oral reading fluency* as a result of receiving Reading Corps tutoring. As such, it is reasonable to conclude that the program can produce positive, meaningful results for eligible second and third grade students who were furthest from benchmark.

3. Does program impact vary by participant characteristics, such as demographics, program dosage, and program attendance?

After completing two rigorous evaluations of the Minnesota Reading Corps program, which establish an overall positive impact of the program on Kindergarten, first grade, and second and third grade students' literacy outcomes, we also conducted analyses to examine whether differential effects of the program existed for specific subgroups of students based on student characteristics, including gender (male, female), race (Black/African American), and ELL status. However, we were only able to complete these subgroup analyses for the combined sample of second and third grade students due to the smaller than acceptable sample sizes in the lower grades. A statistically significant impact of the program was found among second and third grade students despite minority status (i.e., Black), gender, or ELL status. For each of these characteristics, second and third grade students who received the program significantly outperformed control students on *oral reading fluency* scores and produced larger effect sizes than those found for the average student in the program.

Given the consistency of these findings with the previous 2014 study's subgroup analyses, which showed similar significant and positive subgroup findings, an important conclusion for the evaluation is that the program does appear to have a particularly positive effect on important subgroups of students typically considered at-risk for academic failure.

Discussion

Overall, the results of the evaluation showed that the Minnesota Reading Corps K-3 program positively impacts Kindergarten through third grade students' literacy outcomes. Similar to the previous 2014 study's findings, the magnitude of the effect of the program varies by grade, such that students in Kindergarten and first grade on average achieve larger gains within a single semester, a relatively short

⁴¹ See Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172-177.

⁴² Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, 5, 239-256; Kim, Y-S; Petscher, Y; Schatschneider, C; Foorman, B. *Journal of Educational Psychology*, v102 n3 p652-667 Aug 2010

period of time; whereas older students show moderate gains over a longer period of time, an entire school year.

The large program effects found among Kindergarten and first grade students indicate that the Reading Corps program can rapidly and effectively shift younger students' learning trajectories, so they become on track to meet or exceed grade level expectations within a single semester of school. The large and significant impacts of the program's early, intensive intervention suggests that the program helps students who are not Kindergarten-ready catch up to their peers on the critical emergent literacy skill of *letter sound fluency* and helps first grade students extend their letter sound knowledge to the more complicated task of sounding out nonsense words.

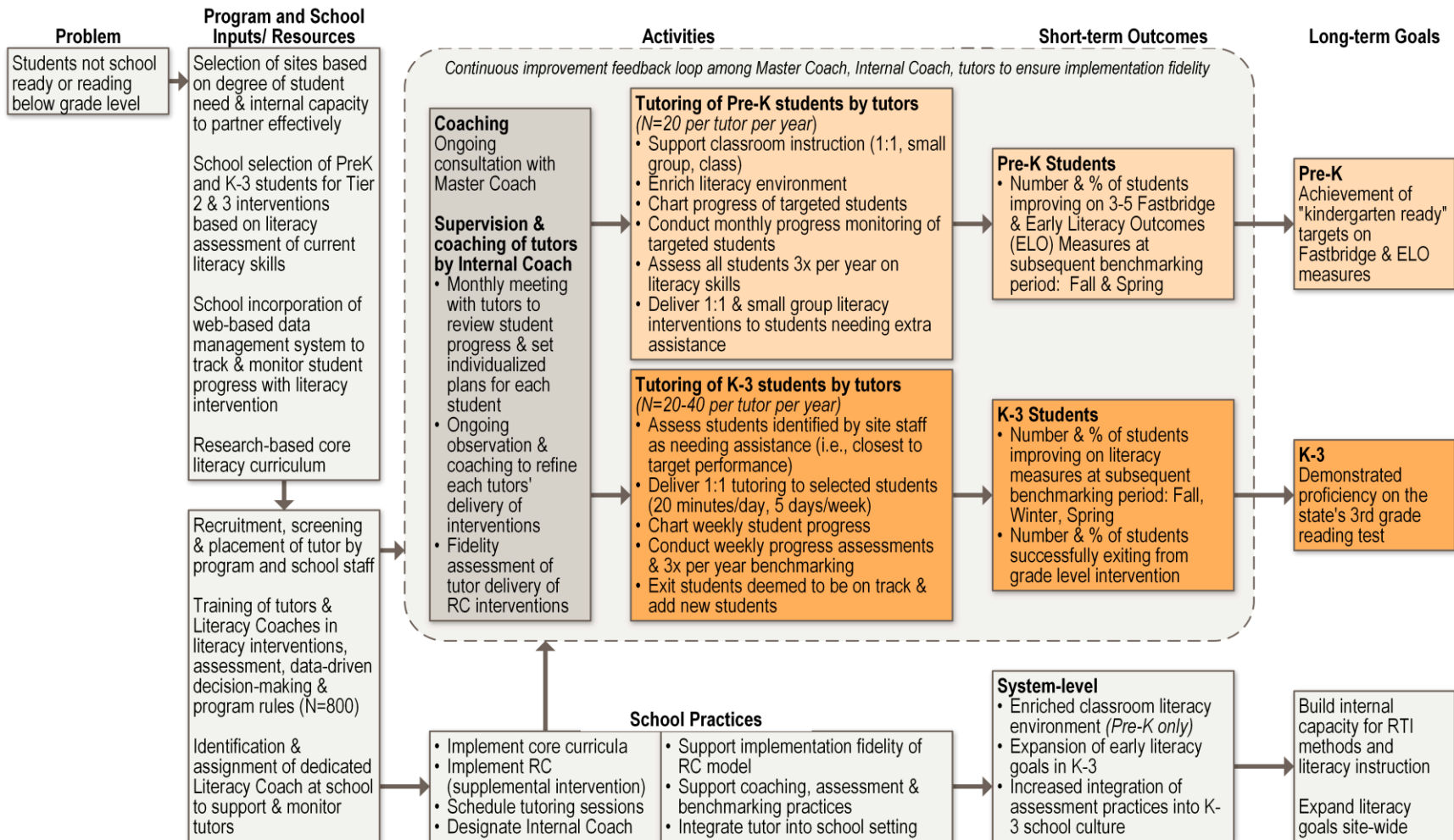
For eligible second and third grade students, the program produced meaningful effects on *oral reading fluency* over an entire school year. The 2018 evaluation results confirm the previous 2014 study's findings that suggested a lengthier period of time is required for producing impacts on second and third grade students' reading fluency, a more complex literacy skill. The second and third grade analysis also purposely selected pairs of students who were furthest from the Fall benchmark, demonstrating that the program can have a positive and meaningful impacts on a more at-risk segment of Tier 2 students' reading fluency. These findings may cause the Reading Corps program to consider shifting how they select and prioritize later-grade students for intervention.

Lastly, this most recent examination of the program confirms our previous 2014 study's conclusion that the Minnesota Reading Corps K-3 program produces positive impacts on K-3 student literacy outcomes. Thus, based on the findings from two highly rigorous randomized controlled trials of the program, it is possible to conclude that the Minnesota Reading Corps K-3 program is effective at improving K-3 students' early reading skills. The combination of Reading Corps program elements, including the utilization of Reading Corps tutors lacking any specialized background in education or literacy, that have resulted in positive impacts on student literacy outcomes in two independent studies of the program provide support for the continued implementation and replication of the Reading Corps model for eligible K-3 students.

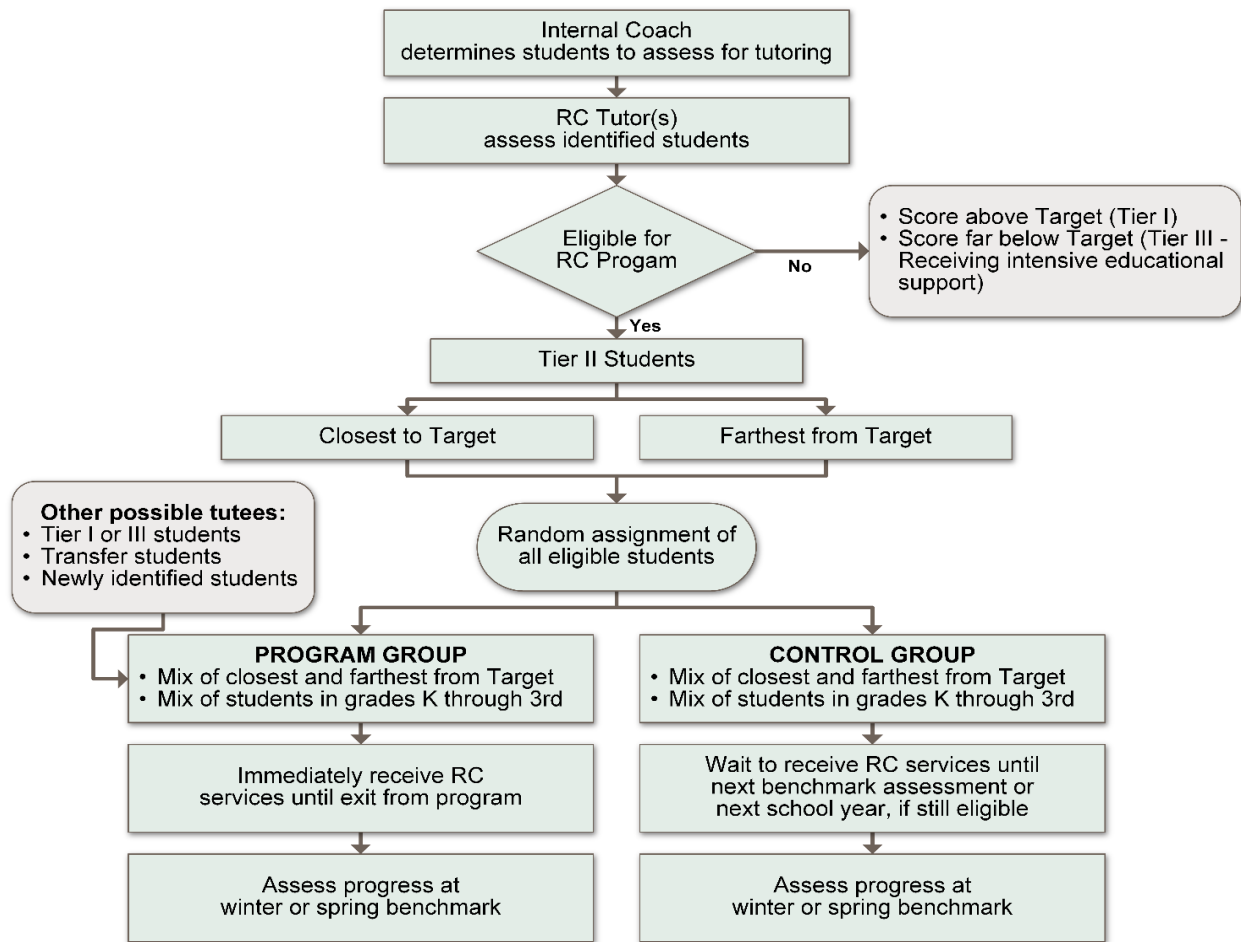
Recommendations for Future Research

Finally, in order to assess the long-term impact of the Minnesota Reading Corps program on student literacy outcomes, we recommend following the same group of randomized students into the future. A longitudinal study of students' literacy and academic outcomes on state proficiency tests, graduation rates, and other more distal educational and economic outcomes, would provide important information about the impact of the Reading Corps program on students' learning and life course trajectories. Also, it would be of interest to better understand the effect of multi-year participation in the program. For example, does early Reading Corps intervention prevent later program participation? Is there an effect of receiving the Reading Corps' PreK program on the later need for K-3 tutoring intervention? In other words, what are the possible compounding effects of program participation over time?

Appendix A: Reading Corps Logic Model (for Impact Evaluation)



Appendix B: Randomization Process



Appendix C: Baseline Analysis

The study team conducted baseline equivalence analyses, including chi-square tests of independence⁴³, t-test analyses of two populations⁴⁴, and Hedges’ g⁴⁵ to determine effect size, in order to ensure that the two groups formed through random assignment were indeed similar at baseline (i.e., prior to program intervention). Based on these analyses, it was determined that the program and control groups for Kindergarten, first, second, and third grade were comparable. While some differences among smaller subgroups of demographic variables were found, the overall balance between the program and control groups for Minnesota is sound, indicating that the two groups were in fact similar on average at baseline. Differences in these smaller subgroups were accounted for by including these less balanced demographic characteristics as control variables in the study’s final analysis. The most important measure, students’ Fall assessment scores was balanced between the treatment and control groups.

Table C.1. Baseline Analysis of Minnesota Evaluation Participants, Kindergarten

Covariate	Chi-sq	p-value	Test	p-value2	Hedges g	Total N
Kindergarten						60
<i>Gender</i>						
Male	0.07	0.80	0.26	0.80	0.06	29
Female	0.07	0.80	-0.26	0.80	-0.06	31
<i>Race/Ethnicity</i>						
Black or African American	0.37	0.54	0.61	0.65	0.15	14
White	0.34	0.56	-0.58	0.56	-0.15	16
Hispanic/Latino	0.29	0.59	-0.54	0.59	-0.13	22
Asian	0.16	0.69	0.40	0.69	0.10	13
Multi-racial	—	—	—	—	—	0
American Indian, Alaskan Native, Native Hawaiian	1.02	0.31	1.01	0.31	0.25	1
<i>Student ELL status</i>						
ELL-eligible	0.00	1.00	0.00	1.00	0.00	22
Not ELL-eligible	0.00	1.00	0.00	1.00	0.00	38
<i>Fall Assessment Scores</i>	—	—	—	0.87	-0.04	60

⁴³ Conducted as appropriate based on variable type, e.g. categorical, interval, ratio, etc. A chi-square test of independence determines whether any significant relationship exists between two nominal (categorical) variables.

⁴⁴ Conducted as appropriate based on degrees of freedom, related to sample size and number of response categories. A two-sample t- or z-test for independence determines whether there is a statistically significant difference between means in two unrelated populations, or in this analysis, between two sample groups. This helps to establish whether the samples are significantly different in terms of any specific characteristics, as measured by variables.

⁴⁵ Hedges’ g is a statistical measure of effect size which is quite similar to Cohen’s d. That is, Hedges’ g describes how much one sample group differs from another in standard deviation units. We employed Hedges g rather than Cohen’s d for the equivalence tests due to small samples.

Table C.2. Baseline Analysis of Minnesota Evaluation Participants, 1st Grade

Covariate	Chi-sq	p-value	Test	p-value2	Hedges g	Total N
First grade						160
<i>Gender</i>						
Male	0.03	0.87	0.16	0.87	0.02	79
Female	0.03	0.87	-0.16	0.87	-0.02	81
<i>Race/Ethnicity</i>						
Black or African American	0.25	0.62	-0.50	0.62	-0.08	18
White	0.69	0.41	0.83	0.41	0.13	105
Hispanic/Latino	0.00	1.00	0.00	1.00	0.00	16
Asian	0.10	0.75	0.31	0.75	0.05	11
Multi-racial	1.86	0.17	-1.36	0.17	-0.21	5
American Indian, Alaskan Native, Native Hawaiian	0.21	0.65	-0.45	0.65	-0.07	5
<i>Student ELL status</i>						
ELL-eligible	0.47	0.50	0.68	0.50	0.11	50
Not ELL-eligible	0.47	0.50	-0.68	0.50	-0.11	110
<i>Fall Assessment Scores</i>	—	—	—	0.97	0.01	160

Table C.3. Baseline Analysis of Minnesota Evaluation Participants, 2nd Grade

Covariate	Chi-sq	p-value	Test	p-value2	Hedges g	Total N
Second grade						234
<i>Gender</i>						
Male	0.43	0.51	0.66	0.51	0.09	127
Female	0.43	0.51	-0.66	0.51	-0.09	107
<i>Race/Ethnicity</i>						
White	0.07	0.79	-0.26	0.79	-0.03	134
Black or African American	2.90	0.09	1.70	0.09	0.22	42
Hispanic/Latino	0.33	0.56	-0.58	0.56	-0.08	31
Asian	0.30	0.58	-0.55	0.58	-0.07	14
American Indian or Alaskan Native	0.12	0.73	-0.34	0.73	-0.04	9
Native Hawaiian or Other Pacific Islander	1.00	0.32	-1.00	0.32	-0.13	1
Multi-racial	0.00	1.00	0.00	1.00	0.00	2
Other	1.00	0.32	-1.00	0.32	-0.13	1
<i>Student ELL status</i>						
ELL-eligible	0.09	0.76	-0.31	0.76	-0.04	56
Not ELL-eligible	0.09	0.76	0.31	0.76	0.04	178
<i>Fall Assessment Scores</i>	—	—	—	0.97	-0.01	234

Table C.4. Baseline Analysis of Minnesota Evaluation Participants, 3rd Grade

Covariate	Chi-sq	p-value	Test	p-value2	Hedges g	Total N
Third Grade						204
<i>Gender</i>						
Male	1.26	0.26	1.12	0.26	0.16	96
Female	1.26	0.26	-1.12	0.26	-0.16	108
<i>Race/Ethnicity</i>						
White	0.32	0.57	0.57	0.57	0.08	120
Black or African American	0.34	0.56	-0.59	0.56	-0.08	31
Hispanic/Latino	0.05	0.83	-0.21	0.83	-0.03	25
Asian	0.07	0.79	-0.27	0.79	-0.04	15
American Indian or Alaskan Native	0.00	1.00	0.00	1.00	0.00	10
Native Hawaiian or Other Pacific Islander	—	—	—	—	—	0
Multi-racial	0.34	0.56	0.58	0.56	0.08	3
Other	—	—	—	—	—	0
<i>Student ELL status</i>						
ELL-eligible	0.03	0.87	-0.16	0.87	-0.02	49
Not ELL-eligible	0.03	0.87	0.16	0.87	0.02	155
<i>Fall Assessment Scores</i>	—	—	—	0.72	0.05	204

Appendix D: Sample Attrition

The What Works Clearinghouse places a large emphasis on attrition when qualifying their standards for causal evidence. To that end, we present our attrition rates and detail the reasons why we experienced sample shrinkage for the final analysis. Differential attrition is also important to consider, as differences in sample loss between program and control groups can bias the results. To avoid this, we took advantage of the matched pair design and removed incomplete pairs, which results in zero differential attrition.

A total of 74 Kindergarten students were randomized into program or control conditions through the matched pair design. Of those randomized, two program students were removed from the program (e.g. behavioral issues), and five withdrew from the school (four in the program group and one in the control group). Table D.1 summarizes these findings. To avoid differential attrition, if a member of a student's pair left the sample, the remaining pair member was also removed from the analysis pool. This situation resulted in the removal of one extra student in the program condition and six extra students in the control condition due to incomplete pairs. As such, the analysis sample was reduced to 60 Kindergarten students, an attrition rate of 19%, which given the lack of differential attrition, is well within the norms outlined by the U.S. Department of Education's Institute of Education Sciences' What Works Clearinghouse (WWC).

Table D.1. Attrition in Kindergarten Sample

	Overall	Program	Control
Total Randomized	74	37	37
Attrition			
Refused	0	0	0
Removed from Program	2	2	0
Withdrawn	5	4	1
Pair member Attrition	7	1	6
Total Attrition	14	7	7
Analysis sample	60	30	30
Attrition Rate	19%	19%	19%

A total of 172 first grade students were randomized into program or control conditions through the matched pair design. Of those randomized, two students refused the study (one in the program group and one in the control group), one program student was removed from the program (e.g. no longer eligible for program), and four withdrew (one in the program group and three in the control group); Table D.2 summarizes these findings. To avoid differential attrition, if a member of a student’s pair left the sample, the remaining pair member was also removed from the analysis pool. This means that we removed three extra students in the program condition and two extra students in the control condition due to incomplete pairs. As such, the analysis sample was reduced to 160 first grade students, an attrition rate of 7%, which given the zero differential attrition, is well within the norms outlined by the WWC.

Table D.2. Attrition in 1st Grade Sample

	Overall	Program	Control
Total Randomized	172	86	86
Attrition			
Refused	2	1	1
Removed from Program	1	1	0
Withdrawn	4	1	3
Pair member Attrition	5	3	2
Total Attrition	12	6	6
Analysis Sample	160	80	80
Attrition Rate	7%	7%	7%

A total of 222 second grade students were randomized into program or control conditions through the matched pair design. Of those randomized, two program students refused the study, two program students were removed from the program (e.g. low attendance, no longer eligible for program), and 13 withdrew (seven in the program group and six in the control group); Table D.3 summarizes these findings. To avoid differential attrition, if a member of a student’s pair left the sample, the remaining pair member was also removed from the analysis pool. This means that we removed five extra students in the program condition and ten extra students in the control condition due to incomplete pairs. As such, the analysis sample was reduced to 190 second grade students, an attrition rate of 14%, which given the zero differential attrition, is well within the norms outlined by the WWC.

Table D.3. Attrition in 2nd Grade Sample

	Overall	Program	Control
Total Randomized	222	111	111
Attrition			
Refused	2	2	0
Removed from Program	2	2	0
Withdrawn	13	7	6
Pair member Attrition	15	5	10
Total Attrition	32	16	16
Analysis Sample	190	95	95
Attrition Rate	14%	14%	14%

A total of 256 third grade students were randomized into program or control conditions through the matched pair design. Of those randomized, three program students refused the study, seven program students were removed from the program (e.g. low attendance, parent refused program participation, behavioral problems, no longer eligible for program), ten withdrew (five in the program group and five in the control group), and two control students were removed for other reasons (e.g. clerical error in student records, moved to lower grade); Table D.4 summarizes these findings. To avoid differential attrition, if a member of a student’s pair left the sample, the remaining pair member was also removed from the analysis pool. This means that we removed seven extra students in the program condition and 15 extra students in the control condition due to incomplete pairs. As such, the analysis sample was reduced to 212 third grade students, an attrition rate of 17%, which given the zero differential attrition, is well within the norms outlined by the WWC.

Table D.4. Attrition in 3rd Grade Sample

	Overall	Program	Control
Total Randomized	256	128	128
Attrition			
Refused	3	3	0
Removed from Program	7	7	0
Withdrawn	10	5	5
Other	2	0	2
Pair member Attrition	22	7	15
Total Attrition	44	22	22
Analysis Sample	212	106	106
Attrition Rate	17%	17%	17%

Appendix E: Detailed Analysis Methods

Impact Analysis (Kindergarten and First Grade). The analysis for the impact evaluation of the Kindergarten and first grades employed a mixed model (defined below) to estimate the treatment effect for each outcome using data from specific Kindergarten and 1st grades individually. The statistical model uses assignment to treatment as the primary predictor in a linear equation to estimate the difference between the averages of those assigned treatment to those assigned control. This predictor is coded as

$$P = \begin{cases} -.5 & \text{if assigned control} \\ .5 & \text{if assigned treatment} \end{cases}$$

Specifically, we fit two models. The first model for an outcome y for student i in school j is

$$(1) y_{ij} = \alpha_0 + \alpha_1 P_{ij} + u_{0j} + u_{1j} P_{ij} + e_{ij},$$

where α_0 is the overall average of school means of student outcomes, α_1 is the average of the school-specific differences in the means of treatment and control students, u_{0j} is the school mean effect (i.e., the difference between the overall average and the school-specific average), u_{1j} is the school effect on the treatment effect (i.e., the difference between school-specific treatment difference and the average of school treatment differences), and e_{ij} is the student residual.

From this first model (1), we estimate an effect size, δ , defined as

$$\delta = \frac{\alpha_1}{\sigma} = \frac{\alpha_1}{\sqrt{\text{var}(u_{0j}) + \text{var}(e_{ij})}},$$

which defines the difference in standard deviation units; the standard deviation is estimated with

$$\sigma = \sqrt{\text{var}(u_{0j}) + \text{var}(e_{ij})}.$$

The second model includes controls, which were used for two reasons. First, when there is balance between treatment and control groups (i.e., they have similar averages of control variables), including controls in the model improves the statistical significance by reducing the variances of the random effects (u_{0j} , u_{1j} , and e_{ij}) which comprise (in part) the estimated sample variance of the treatment effect. The second reason controls were employed was to account for any small imbalances that exist between treatment and control groups. When small differences exist, the concern with estimated treatment effects is that some of the difference between treatment and control groups may be attributed to differences in sample demographics. By entering these variables into the model, these factors are held “constant” and thus estimate a treatment effect net of the influences of the entered variables.

The variables entered as controls include English Language Learner status, gender, race, and most importantly, the Fall benchmark score. In addition to entering the student values into the equation, the study team also calculated the school averages of these variables and entered them into the model as well. The model which includes controls can be noted as

$$(2) y_{ij} = \beta_0 + \beta_1 P_{ij} + \sum_c (\gamma_c X_{cij} + \lambda_c \bar{X}_{cj}) + q_{0j} + q_{1j} P_{ij} + r_{ij},$$

where β_0 is the overall average of school means of student outcomes net of controls, β_1 is the average of the school-specific differences in the means of treatment and control students net of controls, q_{0j} is the school mean effect (i.e., the difference between the overall average and the school-specific average) net of controls, q_{1j} is the school effect on the treatment effect (i.e., the difference between school-specific treatment difference and the average of school treatment differences) net of controls, r_{ij} is the student residual net of controls, and $\sum_c (\gamma_c X_{cij} + \lambda_c \bar{X}_{cj})$ represents all the effects of control variables at both the student and school levels. Note that, for example, $var(u_{0j}) > var(q_{0j})$, which is why we do not derive effect sizes from the second model (2).

In this report, we present estimates and statistical tests for the β_1 parameter for each outcome and grade. We also present marginal predictions of y_{ij} based on the specific values of the treatment predictor (-.5 for control and .5 for treatment) and setting all controls to sample average values. These represent the expected control and treatment averages for typical students. Finally, we present the effect size parameter estimates, δ , defined above.

Impact Analysis (Pooled 2nd and 3rd grade samples). The analysis for the pooled 2nd and 3rd grade samples followed the same procedures as the Kindergarten and 1st grade analyses, except that we pooled information from students in both 2nd and 3rd grades. Thus, our models for the 2nd and 3rd grades include a moderating factor based on grade for all effects. Our overall treatment effect for this model is based on the marginal change in the Spring scores for students in both grades based on program assignment.

Dosage Analysis. The relationship between the amount, or dosage, of an intervention received by participants and their outcomes is often difficult to estimate and interpret. This is especially true for programs that are, in part, reactive in nature in that they provide more services to those with higher need. As a result, we often find negative associations between dosage and outcomes that can be counter intuitive to interpret. However, these negative findings are often the result of the responsive program structure that leads to positive outcomes—those who need more (and score lower) receive more services.

As a result, we attempted to conceptualize dosage in different ways and used statistical adjustments to better understand the relationship between “more” dosage and outcomes. For this report, we used three methods to quantify how much of the program a student received. The first variable is the “number of sessions,” which is simply a count of the number of tutoring sessions the participant received. The next dosage variable “attendance” was created to measure the proportion of attempted intervention sessions which were successfully administered to the student. Attempted sessions, including those that were unable to be administered, were documented.⁴⁶ Thus, the attendance variable is the fraction of attempted sessions that were successfully administered to a student. Next, to measure reactive need, we computed the “Average length of sessions,” or the mean number of minutes each intervention session lasted. Each of

⁴⁶ We also documented the reasons for missing sessions, which included student's absence, school cancellation, Reading Corps tutor's absence, conflicting student assessments times, among other conflicts.

these variables were used in a regression model to predict the outcome values for program participants. Thus, the regression model is specified as

$$y_{ij} = \lambda_0 + \lambda_1 \left(\frac{\text{Total number of sessions}}{\text{Total number of sessions}} \right)_{ij} + \lambda_2 \left(\frac{\text{Total min of sessions}}{\text{Total number of sessions}} \right)_{ij} + \lambda_3 \left(\frac{\text{Total number of sessions of attempted sessions}}{\text{Total number of attempted sessions}} \right)_{ij} + \sum_c \gamma_c X_{cij} + u_{0j} + e_{ij},$$

where we report the standardized coefficients to present partial correlations.⁴⁷ Our model also includes control variables represented by $\sum_c \gamma_c X_{cij}$.

The 2014/2018 Comparison Study. The comparison study involved four groups. We note each of the four groups using two indices: the first is the study index, noted at 0 for 2014 and 1 for 2018, and the second index is assignment to treatment or control, noted as 0 for control and 1 for treatment. Thus, we grouped the data as:

- $Y(0,0)$, the control group in 2014,
- $Y(0,1)$, the treatment group in 2014,
- $Y(1,0)$, the control group in 2018, and
- $Y(1,1)$, the treatment group in 2018.

Suppose we have a set of $i = \{1, 2, \dots, N\}$ observations in which each observation is coded as a member of one of the four groups; that is, the treatment variable is $T_i \in J$ where $J = \{(0,0); (0,1); (1,0); (1,1)\}$. We then can fit a multinomial regression model where the probabilities of all except the first group are defined as

$$\begin{aligned} \Pr(T_i = (0,1)) &= \frac{\exp[X_i \beta_{(0,1)}]}{1 + \exp[X_i \beta_{(0,1)}] + \exp[X_i \beta_{(1,0)}] + \exp[X_i \beta_{(1,1)}]} \\ \Pr(T_i = (1,0)) &= \frac{\exp[X_i \beta_{(1,0)}]}{1 + \exp[X_i \beta_{(0,1)}] + \exp[X_i \beta_{(1,0)}] + \exp[X_i \beta_{(1,1)}]} \\ \Pr(T_i = (1,1)) &= \frac{\exp[X_i \beta_{(1,1)}]}{1 + \exp[X_i \beta_{(0,1)}] + \exp[X_i \beta_{(1,0)}] + \exp[X_i \beta_{(1,1)}]} \end{aligned}$$

and membership in the first group is defined as

$$\Pr(T_i = (0,0)) = 1 - (\Pr(T_i = (0,1)) + \Pr(T_i = (1,0)) + \Pr(T_i = (1,1)))$$

and each β is a set of coefficients that relate the covariates, X , to group membership.

⁴⁷ We achieved standardized coefficients by z-scoring both the outcome and predictors into standard deviation units (i.e., $z = (x - \bar{x}) / s_x$).

Next, we defined the weight for each case as the inverse of the estimated probability of membership in their observed group, that is

$$w_i = \begin{cases} \frac{1}{\Pr(T_i = (0,0))} & \text{if } T_i = (0,0) \\ \frac{1}{\Pr(T_i = (0,1))} & \text{if } T_i = (0,1) \\ \frac{1}{\Pr(T_i = (1,0))} & \text{if } T_i = (1,0) \\ \frac{1}{\Pr(T_i = (1,1))} & \text{if } T_i = (1,1) \end{cases}$$

We then estimated the means of each group using a weighted regression that also includes the covariates, $X, \bar{Y}(T)$.

For propensity weighting purposes, we created several covariates. First, we created an 8-valued group variable that was a combination of race, gender, and ELL status (we collapsed all non-White racial groups into a single category). We also used the Fall benchmark scores in two ways: using both a coarse 3-value quantile grouping variable and then, within those groups, the continuous score. We found that using both the quantile grouping in conjunction with the continuous score improved the matching statistics on the Fall benchmarks conditional on the demographic groups. Again, these covariates were also included in the regression model that produced the weighted means. Appendix Y details the standardized differences among these variable before and after the weighting procedure.

Tests

This study included three tests for each grade, including the treatment effect among the matched 2014 sample using the propensity weights

$$\delta_{2014} = \bar{Y}(0,1) - \bar{Y}(0,0),$$

and the treatment effect among the matched 2018 sample using the propensity weights

$$\delta_{2018} = \bar{Y}(1,1) - \bar{Y}(1,0).$$

However, we were primarily concerned with the following difference in difference effect, which compared the treatment effect in 2018 with that of 2014,

$$\delta_{2018} - \delta_{2014} = [\bar{Y}(1,1) - \bar{Y}(1,0)] - [\bar{Y}(0,1) - \bar{Y}(0,0)].$$

Since we performed three tests from the same estimation procedure, we also adjusted the p-values using the Bonferroni method of increasing the rigor of the tests proportional to the number of tests performed. We also clustered the standard errors by school membership.

Appendix F: Results Tables

Table F.1. Results from Analysis

Grade and Outcome	Treatment effect ^{a,b} (β_1)	p-value ^a	Effect size ^c ($\frac{\alpha_1}{\sigma}$)	Prog. Mean ^d	Control Mean ^d	Total N
Kindergarten						
Letter Sounds fluency	10.940 (4.314)	0.011	0.850	26.938	15.995	60
First grade						
Nonsense Words fluency	16.303 (2.751)	< 0.001	0.812	61.653	45.350	160
Reading fluency	13.293 (5.714)	0.020	0.608	41.043	27.750	160
Second & Third grades						
Reading fluency ^e	6.437 (2.291)	0.005	0.280	102.590	96.153	402

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary.

b: Standard errors in parentheses.

c: Effect size reported from unconditional model without control variables.

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: The Minnesota model includes sampling weights for generalizability.

Table F.2. Reading Fluency by Subgroups, Pooled Second and Third Grades Sample

Subgroup Characteristic ^f	Total N ^e	Reading Fluency				
		Treatment effect ^{a,b} (β_1)	p-value ^a	Effect size ^c (α_1/σ)	Prog. Mean ^d	Control Mean ^d
Racial or Ethnic subgroups						
American Indian or AK Native	18	5.914 (5.414)	0.275	g	97.259	91.345
Asian	25	9.065 (8.023)	0.259	g	97.835	88.770
Black or African American	66	10.650 (5.031)	0.034	0.497	104.697	94.047
Hispanic/Latino	46	3.771 (5.181)	0.467	g	100.667	96.896
Gender subgroups						
Female	204	5.735 (2.486)	0.021	0.220	101.014	95.279
Male	198	7.878 (3.169)	0.013	0.467	102.311	94.433
Language subgroups						
ELL	94	16.279 (4.672)	0.000	0.528	113.992	97.714

a: Reported statistics from model that includes the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; the treatment effect is allowed to vary.

b: Standard errors in parentheses.

c: Effect size reported from unconditional model without control variables.

d: Reported statistics represent marginal predictions based on the model that controls for the Fall assessment, gender, race, ELL status, and the school averages of the Fall assessment, gender, race, and ELL status; all controls are held at their sample averages.

e: Subgroups containing fewer than ten students were excluded from analysis.

f: The Minnesota model includes sampling weights for generalizability.

g: Non-significant effect sizes not reported.

Table F.3. Partial correlations between outcomes and dosage indicators (K-1st Grade)

Grade, Outcome, and Covariates ^{a,b}	Partial Correlation ^c	p-value
Kindergarten (Letter Sounds, N = 30)		
Number of sessions (λ_1)	-0.398	0.486
	(0.572)	
Average length of session (λ_2)	-0.065	0.796
	(0.251)	
Attendance (λ_3)	0.206	0.429
	(0.261)	
First Grade (Nonsense Words, N = 80)		
Number of sessions (λ_1)	-0.660	0.001
	(0.191)	
Average length of session (λ_2)	-0.015	0.689
	(0.036)	
Attendance (λ_3)	0.159	0.106
	(0.134)	

a: Model includes Fall assessment, gender, race, and ELL status control variables.

b. The Minnesota model includes sampling weights for generalizability

c: Variables coded into standard deviation units to produce partial correlations; Standard errors in parentheses

Table F.4. Partial correlations between outcomes and dosage indicators (2nd – 3rd Grades)

Grade, Outcome, and Covariates ^{a,b}	Partial Correlation ^c	p-value
Pooled sample^d (Reading Fluency, N = 201)		
Number of sessions (λ_1)	-0.321	0.000
	(0.084)	
Average length of session (λ_2)	0.065	0.291
	(0.061)	
Attendance (λ_3)	0.096	0.092
	(0.057)	
Second Grade (Reading Fluency, N = 95)		
Number of sessions (λ_1)	-0.500	0.000
	(0.131)	
Average length of session (λ_2)	0.024	0.747
	(0.074)	
Attendance (λ_3)	0.204	0.012
	(0.081)	
Third Grade (Reading Fluency, N = 106)		
Number of sessions (λ_1)	-0.249	0.016
	(0.103)	
Average length of session (λ_2)	0.035	0.573
	(0.062)	
Attendance (λ_3)	-0.028	0.732
	(0.081)	

a: Model includes Fall assessment, gender, race, and ELL status control variables.

b: The Minnesota model includes sampling weights for generalizability.

c: Variables coded into standard deviation units to produce partial correlations; standard errors in parentheses

d: Pooled sample model also includes grade control variable.

Appendix G: Descriptive Statistics for the 2014/2018 Analysis

Table G.1 presents the descriptive statistics of the Kindergartners used for this comparative analysis. Among the 2014 study sample (N=84), approximately two-thirds of Kindergartners in the control group are male (64%), while half of those in the program group are male (51%). Almost three-fourths of the control group are non-White (72%), though less than half of the program group are non-White (47%). The proportion of English Language Learner (ELL) students in the 2014 study sample is lower in the program group (5%) than in the control group (10%).

The Kindergartners’ reading fluency is measured by letter sounds correctly identified. At Fall baseline, the control group identified an average of 2.46 letter sounds, and the program group averaged 2.78 letter sounds. The control group correctly identified 23.08 letter sounds on average at their winter benchmark, while the program group correctly identified an average of 37.46 letter sounds. Among the 2018 study sample (N=67), almost half of Kindergartners in the control group are male (47%), and just over half of those in the program group are male (55%).

A large majority of the control group students are non-White (81%), and similarly, almost three-fourths of the program group are non-White (71%). Approximately one-third of the control group students in the 2018 study sample is identified as ELL (36%), while 39% of the program group is ELL. At Fall baseline, the 2018 study sample control group identified an average of 1.69 letter sounds, and the program group averaged 2.13 letter sounds. The control group correctly identified 16.47 letter sounds on average at their winter benchmark, while the program group correctly identified an average of 27.84 letter sounds.

Table G.1. Descriptive Statistics of the Kindergarten Students from each Study Group used in the Analysis

Characteristic	Control Mean	Control SD	Program Mean	Program SD
2014 Study Sample				
	N=39		N=45	
Male	64%	—	51%	—
non-White	72%	—	47%	—
ELL	10%	—	4%	—
<i>Letter Sounds scores</i>				
Baseline	2.46	2.00	2.78	1.92
Winter benchmark	23.08	16.71	37.36	20.09
2018 Study Sample				
	N=36		N=31	
Male	47%	—	55%	—
non-White	81%	—	71%	—
ELL	36%	—	39%	—
<i>Letter Sounds scores</i>				
Baseline	1.69	2.16	2.13	2.29
Winter benchmark	16.47	12.11	27.84	15.62

Table G.2 presents the descriptive statistics of the first grade students included in the comparative analysis. In the 2014 study sample (N=214), 59% of first grade students in the control group are male, and 55% of those in the program group are male. Almost half of the control group are non-White (47%), though one-third of the program group are non-White (38%). The proportion of English Language Learner (ELL) students in the 2014 study sample is lower in the control group (12%) than in the program group (16%).

The first grade students’ reading fluency is measured by correctly identified nonsense words. At Fall baseline, the control group identified an average of 23.30 nonsense words, and the program group averaged 23.62 nonsense words. The control group correctly identified 51.52 nonsense words on average at their winter benchmark, while the program group correctly identified an average of 59.92 nonsense words. Among the 2018 study sample (N=142), half of first grade students’ in the control group are male (51%), and almost half of those in the program group are male (49%). Approximately one-third of the control group students are non-White (37%), however, almost half of the program group are non-White (45%). Approximately one-fifth of the control group (21%) and the program group (22%) in the 2018 study sample is identified as ELL.

At Fall baseline, the 2018 study sample control group identified an average of 23.15 nonsense words, and the program group averaged 23.11 nonsense words. The control group correctly identified 47.18 nonsense words on average at their winter benchmark, while the program group correctly identified an average of 64.31 nonsense words.

Table G.2. Descriptive Statistics of the First Grade Students from each Study Group used in the Analysis

Characteristic	Control Mean	Control SD	Program Mean	Program SD
2014 Study Sample				
	N=86		N=128	
Male	59%	—	55%	—
non-White	47%	—	38%	—
ELL	12%	—	16%	—
<i>Nonsense words scores</i>				
Baseline	23.302	6.53	23.617	6.18
Winter benchmark	51.523	16.29	59.922	16.58
2018 Study Sample				
	N=68		N=74	
Male	51%	—	49%	—
non-White	37%	—	45%	—
ELL	21%	—	22%	—
<i>Nonsense words scores</i>				
Baseline	23.147	7.12	23.108	7.14
Winter benchmark	47.176	25.27	64.311	26.22

Appendix H: Balance Statistics for the 2014/2018 Analysis

For each study group, two columns are presented. The first shows the standardized difference (the difference in standard deviation units) of the raw sample data and the second shows the standardized difference using the propensity weights. Differences less than 0.1 in magnitude are considered acceptable by convention, though if greater than 0.1 but less than 0.15 use of the doubly robust method is encouraged (Rubin & Thomas 2000).

Table H.1 displays standardized differences between 2014 and 2018 Kindergarten sample groups prior to and after propensity weighting, using the 2014 control group as the reference group. For the 2014 program group, standardized differences in just three characteristics measured using the raw sample data fall within acceptable magnitude (non-White, female, non-ELL; Baseline Quantile 2; and Baseline Quantile 3). All raw standardized differences measured for the remaining five characteristics are 0.15 or greater. Using propensity weighting, all standardized differences of characteristics measured are less than 0.1 in magnitude, and therefore are considered acceptable by convention. Compared to the 2014 control group, the 2018 control group raw sample data demonstrates standardized differences for no characteristics that fall within acceptable range.

Furthermore, just one characteristic falls within the range for which doubly robust methods are encouraged (non-White, female, non-ELL), with all other characteristics showing raw standardized differences greater than 0.15. However, propensity weighted standardized differences for the 2018 control group are less than 0.1 magnitude for every characteristic measured, and therefore are acceptable.

Finally, for the 2018 program group, raw standardized differences for two characteristics measured fall within acceptable magnitude (White, male, non-ELL; and Baseline Quantile 3). Moreover, only one characteristic falls within the range for which doubly robust methods are encouraged (non-White, female, non-ELL), with all other characteristics showing raw standardized differences of 0.15 or greater. Yet again, propensity weighted standardized differences for the 2018 program group are less than 0.1 magnitude for every characteristic measured, and therefore are acceptable. In sum, for each sample group in the 2014 and 2018 study samples, using propensity weighting improves standardized differences among every characteristics measured to fall below the threshold of conventionally acceptable magnitude (0.1).

Table H.1. Standardized Differences for 2014 and 2018 Minnesota Reading Corps Evaluations' Kindergarten Treatment Groups (Reference to 2014 Control)

Characteristic	Control		Program	
	Raw Std. Diff.	Weighted Std. Diff.	Raw Std. Diff.	Weighted Std. Diff.
2014 Study Sample				
White, male, non-ELL			0.21	-0.02
non-White, female, non-ELL			-0.01	-0.02
non-White, female, ELL			-0.15	0.08
non-White, male, non-ELL			-0.41	0.03
non-White, male, ELL			-0.15	-0.07
<i>Letter Sounds scores</i>				
Baseline Quantile 2			0.01	-0.03
Baseline Quantile 3			0.08	-0.08
Baseline (continuous)			0.16	-0.08
2018 Study Sample				
White, male, non-ELL	-0.39	-0.03	0.04	0.01
non-White, female, non-ELL	-0.10	-0.01	-0.11	-0.04
non-White, female, ELL	0.51	-0.01	0.44	-0.03
non-White, male, non-ELL	-0.28	0.03	-0.57	0.08
non-White, male, ELL	0.30	0.04	0.44	0.05
<i>Letter Sounds scores</i>				
Baseline Quantile 2	-0.20	-0.06	-0.29	0.06
Baseline Quantile 3	-0.30	0.02	-0.08	-0.07
Baseline (continuous)	-0.37	0.06	-0.15	0.02

Table H.2 presents standardized differences between 2014 and 2018 first grade sample groups prior to and after propensity weighting, using the 2014 control group as the reference group. For the 2014 program group, standardized differences in five characteristics measured from the raw sample data fall within acceptable magnitude: White, male, non-ELL; non-White, male, non-ELL; non-White, male, ELL; Baseline Quantile 3; and the continuous Baseline measure. For one characteristic, the raw standardized difference falls within the range for which doubly robust methods are encouraged (Baseline Quantile 2), but raw standardized differences for the remaining two measures are much greater than the doubly robust range (non-White, female, non-ELL; and non-White, female, ELL). Using propensity weighting, all standardized differences of characteristics measured are less than 0.1 in magnitude, and therefore are considered acceptable.

In comparing the 2018 control group to the 2014 control group, there are no characteristics measured for which raw standardized differences fall within the range of acceptable magnitude. Moreover, there is just one characteristic for which the raw standardized difference falls within the range in which doubly robust methods are encouraged (non-White, female, non-ELL), with all other characteristics showing raw standardized differences greater than 0.15. However, propensity weighted standardized differences for the 2018 control group are less than 0.1 magnitude for every characteristic measured, and therefore are acceptable. Finally, for the 2018 program group, raw standardized differences for three characteristics measured fall within acceptable magnitude (White, male, non-ELL; non-White, male, ELL; and the continuous Baseline measure). All other characteristics show raw standardized differences of 0.15 or

greater. Again, propensity weighted standardized differences for the 2018 program group are less than 0.1 magnitude for almost all characteristics measured, and therefore are acceptable. The propensity weighted standardized difference for one characteristic measured fell within the range for which doubly robust methods are encouraged (Baseline Quantile 3). This is the only instance in any of the 2014 and 2018 first grade sample groups in which the propensity weighted standardized difference fell above the conventionally acceptable threshold (0.1).

Table H.2. Standardized Differences for 2014 and 2018 Minnesota Reading Corps Evaluations' First Grade Treatment Groups (Reference to 2014 Control)

Characteristic	Control		Program	
	Raw Std. Diff.	Weighted Std. Diff.	Raw Std. Diff.	Weighted Std. Diff.
2014 Study Sample				
White, male, non-ELL			0.01	-0.02
non-White, female, non-ELL			-0.27	-0.01
non-White, female, ELL			0.27	0.05
non-White, male, non-ELL			-0.08	0.00
non-White, male, ELL			-0.04	-0.01
<i>Nonsense words scores</i>				
Baseline Quantile 2			0.14	0.02
Baseline Quantile 3			-0.06	-0.09
Baseline (continuous)			0.05	-0.08
2018 Study Sample				
White, male, non-ELL	-0.07	-0.02	-0.06	-0.02
non-White, female, non-ELL	-0.39	-0.01	-0.18	-0.01
non-White, female, ELL	0.35	0.05	0.48	0.05
non-White, male, non-ELL	-0.18	0.01	-0.16	0.01
non-White, male, ELL	0.04	-0.01	-0.08	-0.01
<i>Nonsense words scores</i>				
Baseline Quantile 2	0.02	-0.02	0.15	0.01
Baseline Quantile 3	-0.10	-0.04	-0.19	-0.11
Baseline (continuous)	-0.02	-0.04	-0.03	-0.10