FINAL REPORT

Impact Evaluation of the Wisconsin Reading Corps Program

AUTHORS:
Carrie E. Markovitz
Marc W. Hernandez
Eric C. Hedberg
Kristen Neishi

Acknowledgement

Table of Contents

List of Figures

List of Tables

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. 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.

As part of an Innovative Approaches to Literacy (IAL) Grant through the U.S. Department of Education (DOE), a multi-site impact evaluation of the Reading Corps program was conducted in three states during the 2017-2018 school year. While prior independent evaluations have been conducted of the original *Minnesota* Reading Corps program, this multi-state evaluation assessed the impact of the Reading Corps program model on students in Minnesota, as well as two replication sites, Florida and Wisconsin. This report describes the findings from the impact evaluation of the Wisconsin Reading Corps Program on 176 Kindergarten and first grade (K-1) students attending ten public schools in Milwaukee during the first semester of the 2017-2018 school year.

The Wisconsin Reading Corps program is a replication site of the Minnesota Reading Corps model. In its fourth year of operation, Wisconsin Reading Corps will expand from serving 12 schools in Milwaukee to several more schools and districts in southeastern Wisconsin, reaching more than 1,000 students in the 2018-2019 school year. The program is supported locally through a partnership with Milwaukee Succeeds, a collaborative effort launched by a group of organizations to unite the community around a common goal of supporting "cradle to career" strategies that foster success for every child. Tutors are supported by a multi-level coaching model that includes site-based ("Internal") and external ("Master") coaches. Wisconsin Reading Corps is managed by an Executive Director, a Senior Program Manager, and two Master Coaches.

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.

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

- 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) 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 from 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 Wisconsin Reading Corps impact evaluation was to independently and experimentally assess the impact of the Wisconsin Reading Corps program on Kindergarten and first grade students' literacy proficiency scores. The primary research question for the evaluation was: What is the impact of Wisconsin Reading Corps on K-1 program participants compared to similar students who did not receive Reading Corps?

Random Assignment of Students within Schools

A randomized controlled trial (RCT) design was used to answer this research question, which was largely informed by the 2014 Impact Evaluation of the Minnesota Reading Corps K-3 Program.[1] At the beginning of the school year (prior to program implementation), all program eligible students enrolled in Kindergarten and first grade at ten participating public schools in Milwaukee were included in the study sample.[2] Eligible students in each grade within a school were matched in pairs based on their Fall benchmark scores. Students within pairs were then randomly assigned to either the program or control condition. This matched-pair design ensured that the program and control groups had similar distributions of Fall benchmark scores at the start of the school year. In total, 176 Kindergarten and first grade students participated in the evaluation.

Data Sources

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 and first grade students enrolled at the ten participating schools.[3] Independent assessors from the NORC evaluation

---

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

[2] Two schools in Milwaukee were not eligible for the study because they did not meet IAL eligibility criteria.

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

team collected the Winter benchmark assessment data from students in both the treatment and control groups at each of the ten schools to ensure independent measures of outcomes and mitigate any potential bias at posttest. Additionally, Wisconsin Reading Corps program staff at the ten participating schools provided the evaluation team with other administrative data, including students' Reading Corps tutoring session attendance, average length and number of sessions, and student demographic characteristics.

Analysis

To estimate program effects, the average Winter benchmark scores of students who participated in the Reading Corps program were statistically compared with the average Winter 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 modeling the multi-level structure of the data, which in this case are 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 Wisconsin Reading Corps program.

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

Kindergarten students who received Wisconsin Reading Corps tutoring achieved significantly higher *letter sound fluency* scores by the end of the first semester than Kindergarten control students who did not participate in the program (see **Figure i.1**). On average, program students correctly identified 6.5 more letter sounds in a one minute period than the control group (p=0.035, N=64). This difference had a positive, meaningful effect size of 0.55, indicating that the average program student scored over one half 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 on average achieved Winter benchmark expectations, whereas the control students' average score did not reach the expected benchmark.

Figure i.1. Mean scores for Kindergarten program and control students
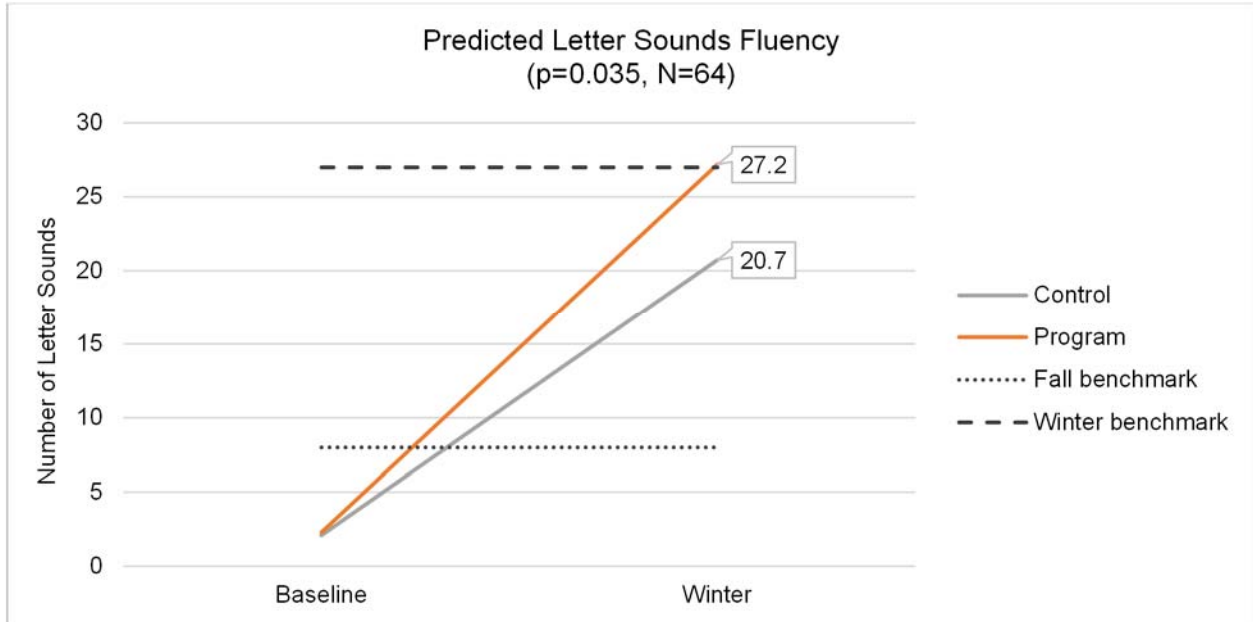
What is the impact of Wisconsin Reading Corps on first grade students?

First grade students enrolled in Wisconsin Reading Corps achieved substantially and significantly higher *nonsense word fluency* scores by Winter benchmark than first grade control group students (see **Figure i.2**). On average, program students correctly identified 8.7 more letter sounds within nonsense words than the control group (p=0.004, N=112), resulting in an effect size of 0.46. Although neither the program group nor the control group achieved Winter benchmark expectations, the difference in progress between the two study groups from Fall to Winter benchmark indicates that first grade program students are likely to continue to progress at an accelerated pace compared to control group students.

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

Predicted Nonsense Words Fluency
(p=0.004, N=112)
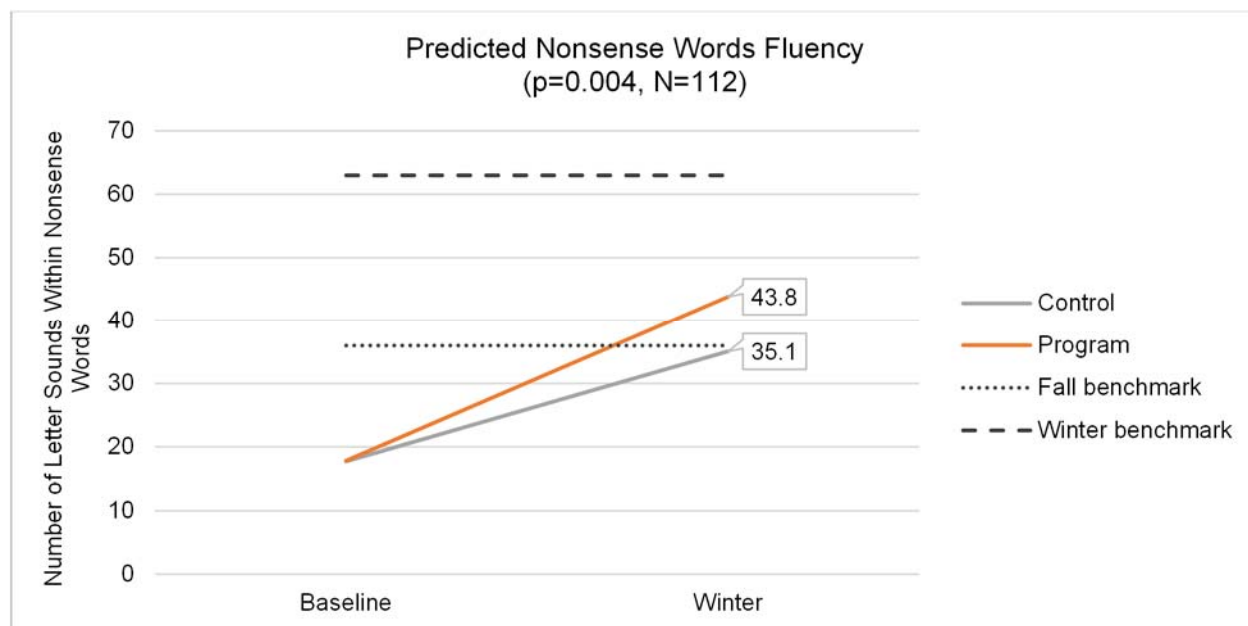
Concluding Thoughts

In sum, the results of the evaluation show that the Wisconsin Reading Corps program has a considerable, positive effect on both Kindergarten and first grade students' literacy outcomes. Compared to their control group peers, Wisconsin Reading Corps Kindergarten students scored significantly higher in *letter sound fluency* and met the Winter benchmark expectation on average. These findings indicate that the program is achieving its goal of accelerating Kindergarten students' literacy development, setting them on track to exceed grade level expectations within one semester. Wisconsin Reading Corps' first grade students also exhibited greater *nonsense word fluency* compared to their control group peers. Even though first grade students on average did not achieve the first grade Winter benchmark target by the end of the first semester, itis important to consider that the average first grade student in the program began the school year far below the Fall program eligibility benchmark, and the program students should continue to grow at an accelerated pace throughout the remainder of the school year compared to control students. Altogether, these results suggest that the Wisconsin Reading Corps program produces meaningful effects for K-1 students within a single school semester, which is a notable achievement for a replicated program in its third year of implementation.

Another equally notable implication of the study results is that it provides evidence of the Minnesota Reading Corps' successful replication of the program model in a different state – Wisconsin. The Wisconsin program includes similar program elements, which were found critical to the Minnesota Reading Corps program's success in our previous *Process Assessment of the Minnesota Reading Corps*

Program research.[4] The key Reading Corps program elements that previously resulted in positive impacts on student literacy outcomes in Minnesota have now been shown to be a robust and effective model in another location, which can help guide the development of other successful reading intervention programs for Kindergarten and first grade students. The current study findings, coupled with the findings from previous studies of the Minnesota Reading Corps program, add important and useful information to the literature on successful replications of effective reading interventions.

---

[4] Hafford, C., Markovitz, C., et al. (October 2014). *Process Assessment of the Minnesota Reading Corps PreK Program: Phase II.* (Prepared under contract to the Corporation for National and Community Service). Chicago, IL: NORC at the University of Chicago.

Introduction

Reading Corps is an AmeriCorps program that provides trained literacy tutors to support reading development for students in PreKindergarten (PreK) through grade 3, with the ultimate goal of ensuring that students become successful readers and meet reading proficiency targets by the end of the third grade. The core activities of the program are training, placing, and monitoring literacy tutors in school-based settings to implement research-based, early-literacy strategies. Tutors are supported by a multi-level coaching model that includes site-based ("Internal") and external ("Master") coaches. Since 2012, the PreK and the Kindergarten through third grade (K-3) Reading Corps model has been replicated by partners in several locations, including Colorado; Washington, D.C.; Milwaukee; and Miami.

As part of an Innovative Approaches to Literacy (IAL) Grant through the U.S. Department of Education (DOE), a multi-site impact evaluation of the Reading Corps program was conducted 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 K-3 program that found that students who received Minnesota Reading Corps 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.[5] Given recent developments in the Reading Corps program, specifically its replication in Wisconsin and Florida, and enhancements made to Reading Corps programming,[6] this evaluation assessed the impact of the enhanced Reading Corps program model on program participants compared to students who did not receive the Reading Corps program in Minnesota (K-3), as well as two replication sites, Florida (PreK) and Wisconsin (K-3).

This report describes the findings from the impact evaluation of the Wisconsin Reading Corps Program on 176 Kindergarten and first grade (K-1) students during the first semester of the 2017-2018 school year.[7] The purpose of the evaluation was to 1) evaluate the impact of the Wisconsin Reading Corps 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 (e.g., demographic) characteristics. Although the Wisconsin Reading Corps program serves students in grades K-3, the scope of this study was limited to Kindergarten and first grade students because there were not enough program eligible second and third grade students in the Milwaukee Reading Corps sites in 2017-2018 to meet the study's minimum sample sizes for detecting statistical results of program effects.

---

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

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

[7] Due to sample size limitations on eligible students in the second and third grades, the evaluation of the Wisconsin Reading Corps program only assessed the effects of the program on Kindergarten and first grade students.

In the sections that follow, we describe the Reading Corps program model and specifically the Wisconsin Reading Corps program. We then present the methodology used to conduct the evaluation, the study findings, and conclusions and discussion of the implications of the findings of the evaluation. The appendices included with this report provide additional details: Appendix A presents the Reading Corps logic model; Appendix B illustrates the randomization process; Appendices C through E describe methodological details (i.e., baseline analysis, attrition, and outcome analyses); and Appendix F presents the detailed results tables.

The Reading Corps Program Model

Reading Corps is the largest AmeriCorps tutoring program in the country, providing trained literacy tutors to implement evidence-based literacy instruction and assessment protocols for students in Prekindergarten (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 PreKindergarten 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
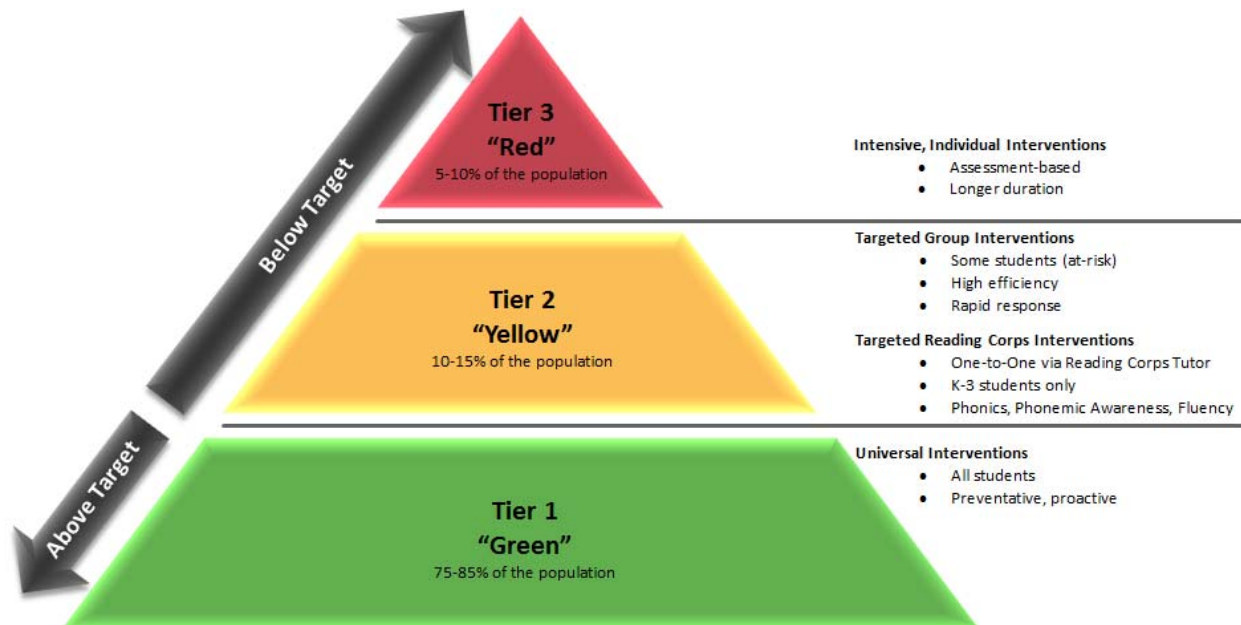
The Reading Corps K-3 program provides one-on-one tutoring where Reading Corps tutors provide supplemental individualized literacy interventions to primarily Tier 2 students in Kindergarten through third grade. Generally, those Tier 2 students who score closest to the benchmark are offered Reading Corps' intervention services first because they should require the least amount of intervention (i.e., time in program) to be set on a learning trajectory to achieve grade level proficiency. The students closest to the benchmark can be moved through the program more quickly than those students with greater need, allowing the schools to maximize support for students needing more intensive services. 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 Wisconsin 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.[8,9] Reading Corps tutors serve as one-on-one literacy interventionists, working with approximately 15-18 K-3 students who scored below the grade-specific literacy benchmark 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 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,[10] 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 and first grades.

---

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

[9] 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.

[10] http://www.fastbridge.org/

■ **Progress Monitoring Assessments.** Once selected to receive Reading Corps services, tutors administer weekly 1-minute reading probes to students. These progress monitoring data allow 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. 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 Benchmarks by Grade

Grade	Measure	Fall	Winter	Spring
Kindergarten	Test of Letter Sounds	8	27	48
Grade 1	Test of Nonsense Words (English)	36	63	n/a

For students in grades 1-3, 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 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 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 for the student's 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 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, 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.

Wisconsin Reading Corps

As the focus of this report, the Wisconsin Reading Corps program is a replication of the successful Minnesota Reading Corps model. Currently beginning its fourth year of operation, Wisconsin Reading Corps is implemented in twelve schools in Milwaukee. The program receives local support through a partnership with Milwaukee Succeeds, a collaborative effort launched by a group of organizations, including the Greater Milwaukee Foundation, as a way to unite the community around a common goal of supporting "cradle to career" strategies that foster success for every child. Wisconsin Reading Corps is managed by an Executive Director, a Senior Program Manager, and two Master Coaches. In the 2018-2019 school year, the program is expected to triple in size, expanding to additional schools in Milwaukee and Racine to serve more than 1,000 students.

About the Study

This section provides a detailed description of the methodology used to implement the Wisconsin Reading Corps program impact evaluation. The methodology was largely informed by the 2014 Impact Evaluation of the Minnesota Reading Corps K-3 Program.[11]

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 is provided in **Appendix A.** The Reading Corps logic model serves as the conceptual framework for the design of the Wisconsin 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 Wisconsin Reading Corps program impact evaluation was to assess the impact of program participation on Kindergarten and first 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 (primarily Tier 2) K-3 students within the school based on benchmark assessment of students' literacy skills; 3) web-based data management systems to track and monitor student progress with literacy interventions; and 4) school use of a research-based core literacy curriculum.

Three important Reading Corps 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 (i.e., weekly progress monitoring); 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

---

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

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

Research Questions

As the logic model illustrates, the Reading Corps program's short-term objective is to improve at-risk students' literacy skills so they are on track to achieve grade-level proficiency. The primary goal of the Wisconsin Reading Corps impact evaluation was to independently and experimentally assess the impact of the Wisconsin Reading Corps program on K-1 students' literacy proficiency scores. Thus, to achieve this goal, the research questions for the evaluation are as follows:

1. What is the impact of Wisconsin Reading Corps on program participants compared to students who did not receive Reading Corps?
2. Does the 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 176 Kindergarten and first grade students enrolled at the ten schools participating in the Wisconsin Reading Corps IAL program during the first semester of the 2017-2018 school year. Specifically, within each school, all 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 Wisconsin Reading Corps program while the other 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 was used to answer the evaluation's two research questions, which assesses the impact of the Wisconsin Reading Corps program on student literacy outcomes.

Random Assignment of Students within Schools

Of the 12 Wisconsin Reading Corps schools operating in Milwaukee, ten of those schools participated in the study.[12] The ten participating schools and their key characteristics are provided in Table 2 below.

---

[12] Two schools were not eligible for the study because they did not meet IAL eligibility criteria.

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

School	Number of Reading Corps tutors[a]		% Free and Reduced-Price Lunch[b]	%Non-White[b]	School Enroll-ment[b,c]	Study Participants (N)[f]	
	Full Time	Part Time				Randomized Sample	Analytic Sample
Auer Avenue	0	2	82.46%	99.65%	285	10 (K), 12 (F)	10 (K), 10 (F)
Forest Home Avenue	1	0	84.53%	94.03%	821	10 (K), 14 (F)	6 (K), 14 (F)
Gwen T. Jackson	1	0	88.15%	98.78%	329	8 (K), 18 (F)	8 (K), 18 (F)
Hopkins Lloyd	2	0	86.16%	97.23%	289	8 (K), 16 (F)	2 (K), 10 (F)
Milwaukee Academy of Science	0	2	93.86%	99.62%	1042	14(K), 14 (F)	12 (K), 10 (F)
Milwaukee College Prep 36th St.	0	1	68.79%	99.40%	503	4 (K), 6 (F)	4 (K), 6 (F)
Milwaukee College Prep 38th St.	0	1	81.49%[d]	98.27%	521	6 (K), 8 (F)	6 (K), 8 (F)
Milwaukee College Prep Lloyd St.	0	1	83.81%[e]	99.20%	499	10 (K), 6 (F)	10 (K), 6 (F)
Milwaukee College Prep Lola Rowe	0	1	73.09%	99.15%	472	4 (K), 8 (F)	6 (F)
Parkview	1	0	71.53%	94.91%	432	8 (K), 24 (F)	6 (K), 24 (F)

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: 2014-2015 data; 2015-2016 FRPL enrollment numbers were abnormally low. Total enrollment for 2014-2015: 443 students.
e: 2014-2015 data; 2015-2016 data do not meet NCES data quality standards. Total enrollment for 2014-2015: 494 students.
f: (K) indicates Kindergarten; (F) indicates first grade.

At the beginning of the school year (prior to program implementation), all program eligible students enrolled in Kindergarten and first grade at the ten participating schools were randomly assigned to two groups: a program group that participated in the Reading Corps program and a control group that did not participate in the program.[13] [14] Specifically, the eligibility criteria for the Reading Corps program was typically a Tier 2 score on the baseline (i.e., Fall benchmark) assessment for the program.[15] Eligible students were included in the study sample by grade in pairs, whereby each student in each grade within a school was matched with another student based upon their Fall benchmark score. Students within pairs were then randomly assigned to either the program or control condition.[16] 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.[17] The final column in Table 2 lists the number of students at each of

---

[13] Students who participated in Reading Corps in previous years were eligible for participation in the evaluation. The important eligibility criteria 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.

[14] 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-1 were assessed, some schools may have used previous years' test scores or other more subjective means for identifying students to be assessed.

[15] 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.

[16] A figure illustrating the randomization process is provided in **Appendix B**.

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

the ten schools whose Fall benchmark scores made them eligible to receive Reading Corps services during the Fall semester. In total, 202 eligible K-1 students were selected to participate in the evaluation.

Table 3 presents descriptive statistics for the K-1 students included in the evaluation. Demographics include gender, race/ethnicity, English Language Learner (ELL) status, Free- and Reduced-Price Lunch (FRPL) status, and Fall benchmark scores. The analytic sample size for the evaluation (N=176) is smaller than the 202 students eligible for program services. During the school year, some students left the school (i.e., moved), or were chronically absent and did not receive regular Wisconsin 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 176 students. See **Appendix D** for details on sample attrition.

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

	Kindergarten (N=64)		1st Grade (N=112)	
	Mean	**SD**	**Mean**	**SD**
Female	45%	—	51%	—
Race/Ethnicity				
White	2%	—	2%	—
Black	91%	—	79%	—
Asian	2%	—	9%	—
Hispanic/Latino	6%	—	9%	—
Multi-Racial	0%	—	1%	—
English Language Learner (ELL)	6%	—	10%	—
Free and Reduced Price Lunch (FRPL)[a]	82%	—	81%	—
Fall Benchmark Score[b]	2.19	2.48	17.76	10.36

a: School average
b: Kindergarten = Letter sound fluency; 1st Grade = Nonsense word 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. For Kindergarten, a minimum of 40 students (i.e., 20 pairs of students) across the ten schools was needed to detect an effect size of 0.67, and for first grade, a minimum of 140 students (i.e., 70 pairs of students) across the ten schools was needed to detect an effect size of 0.39. As shown in Table 3 above, 64 Kindergarten students and 112 first 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. Wisconsin Reading Corps Master 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 Wisconsin Reading Corps services were included in the sample. Once 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 Master Coach then shared the list of students to be tutored with the school's Internal Coach and Reading Corps tutors.

Because the evaluation was designed to measure the impact of Wisconsin Reading Corps program participation relative to nonparticipation, students in the control group were embargoed from receiving Wisconsin Reading Corps tutoring services during the first semester of the school year. 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 benchmark. The evaluation team remained in regular contact to verify that program students received Wisconsin 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 in the *Use of Data in Program Implementation* section above. The evaluation team partnered with the Wisconsin Reading Corps program to utilize the Fall benchmark assessment data that Reading Corps tutors collected on all program-eligible K-1students enrolled at the ten eligible Reading Corps schools in Wisconsin. Wisconsin Reading Corps has been completing Fastbridge reading assessments with students for the past two 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.[18] This procedure identifies objectively those students who meet Reading Corps eligibility criteria (i.e., primarily Tier 2). 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 benchmark assessment data from students in both the treatment and control groups at each of the ten 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 ten participating schools provided the evaluation team with other administrative data, including students' tutoring attendance, average length and number of sessions, and demographic characteristics.

---

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

Analysis Methods

The 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. We estimate program effects by statistically comparing the average score of students who received Reading Corps tutoring with the average score of a similar comparison 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 scores for the program and control groups and the associated sampling error. We then compute the likelihood[19] 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 scores.

The impact evaluation analyses were conducted using mixed-regression models, also known as a hierarchical linear models. 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.[20] Due to its robustness with small sample sizes, we selected Hedges' g for calculating effect sizes. The second model includes control variables to remove any spurious effects due to 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 E**.

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 students' improvements in Winter benchmark scores are the result of the program and

---

[19] 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.

[20] 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.

not from unobserved factors, which leads to high levels of internal validity. However, there are always limitations to any study.

Because all eligible IAL schools and Kindergarten and first grade students were recruited to participate in the study, we would expect high levels of external validity. 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 very specific population of eligible students and schools, and that results will be difficult to replicate in other locations. On the other hand the Wisconsin Reading Corps program itself is a successful replication of the Minnesota Reading Corps program, and so this concern may not be warranted.

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 (See **Appendix C**). As is inevitable, there was attrition due to students moving away or being chronically absent from school. 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 by two students, but ensured the integrity of the RCT design (See **Appendix D**).

Study Findings

In this section, we present the major findings from the analysis of Kindergarten and first 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 Wisconsin Reading Corps on Kindergarten students?

The Wisconsin Reading Corps program had a significant and substantial impact on Kindergarten students' *letter sound fluency* scores between the Fall benchmark (September 2017) and the Winter benchmark (January 2018). On average, program students correctly identified 6.5 more letter sounds in a one minute period than the control group after a single semester of the program (p=0.035, N=64). Figure 2 presents the findings visually and shows that the program Kindergarten students' average score was 27.2 letter sounds compared to 20.7 letter sounds for the average control student. Also by the end of the first semester, the program group achieved the Winter benchmark expectations of 27 letter sounds, whereas the control students' average score did not reach the expected benchmark. Furthermore, this

difference had a positive, meaningful effect size, indicating that the average program student scored over one half of a standard deviation unit higher than the average control group student in identifying letter sounds (effect size = 0.55).

Figure 2. Mean scores for Kindergarten program and control students
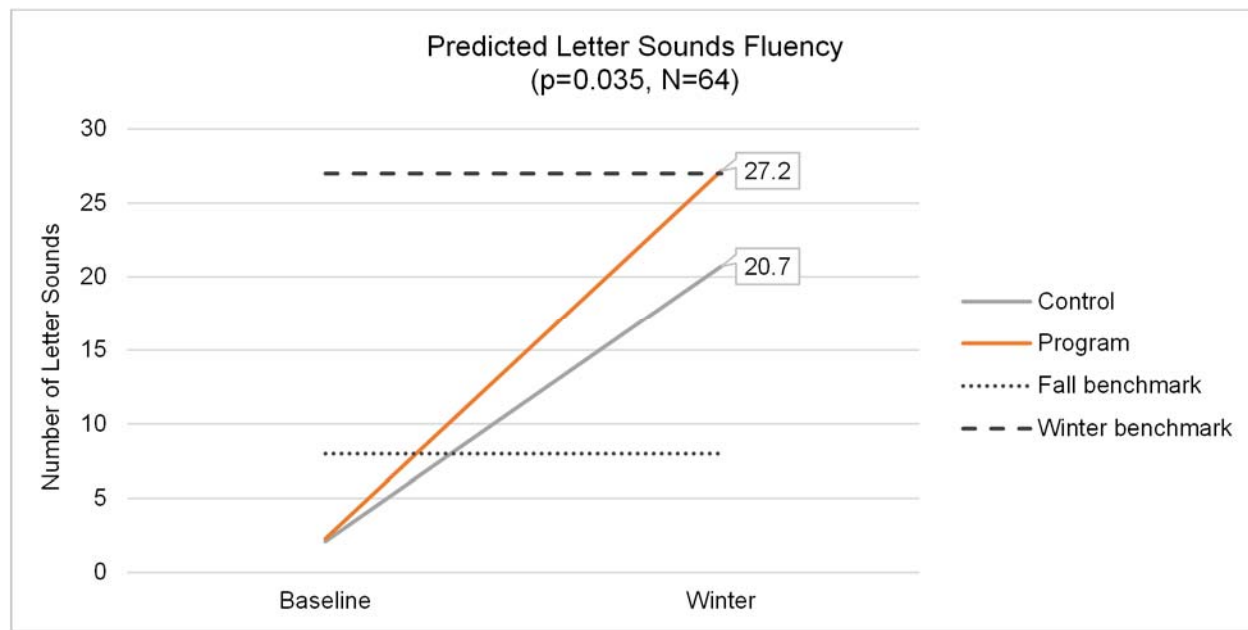
What is the impact of Wisconsin Reading Corps on first grade students?

Similar to the findings in Kindergarten, first grade students enrolled in the Wisconsin Reading Corps program attained significantly higher *nonsense word fluency* scores by Winter benchmark than first grade control group students. On average, program students correctly identified 8.7 more letter sounds embedded within non-real words than the control group (p=0.004, N=112). Figure 3 presents the findings visually, which shows that the program students' average score of 43.8 was substantially higher than the control group's average score of 35.1. While neither the program group nor the control group achieved Winter benchmark expectations, it is important to note that the average first grade students' Fall score was far below expected benchmarks prior to program intervention. 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. A similarly positive effect size was calculated, indicating that the average program student scored about one half standard deviation units higher than the average control group student in identifying letter sounds within nonsense words (effect size =0.46).

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

Predicted Nonsense Words Fluency
(p=0.004, N=112)
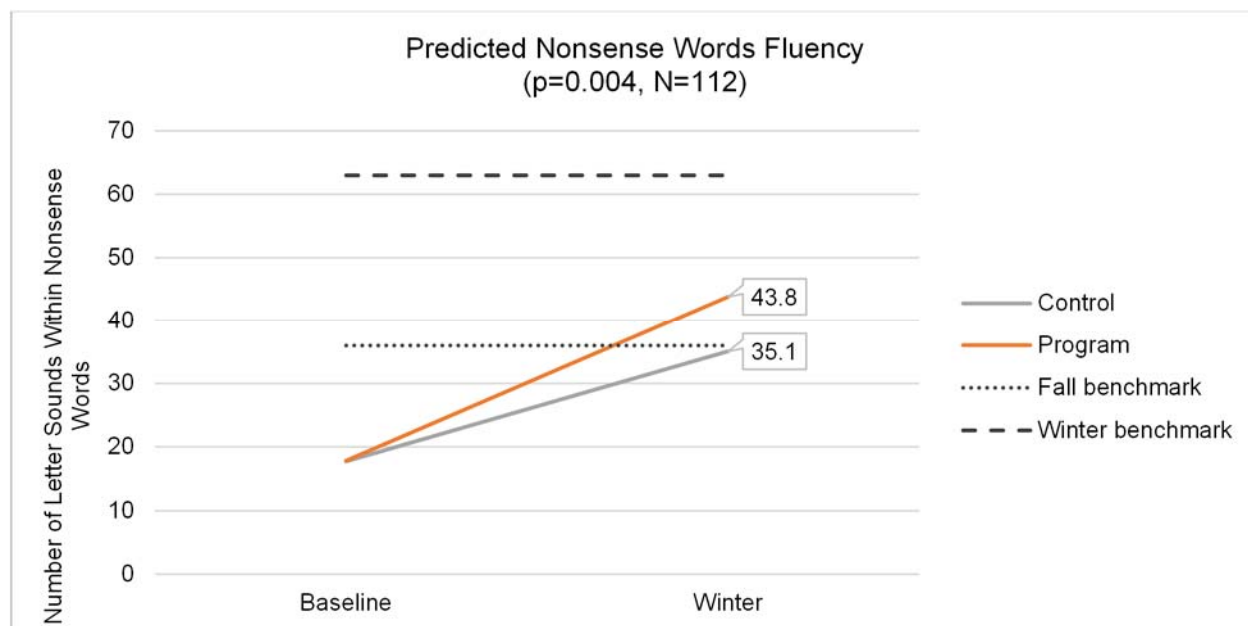
Findings by Major Demographic Group

To address the study's second research question focused on varying program impacts by participant group, we originally planned to include a set of subgroup analyses of student outcomes. As stated in the limitations section, however, the sample sizes by grade for this study were relatively small, and the subgroup sizes by demographic group were not sufficient (e.g., the vast majority of students were African American). As a result, we lack the statistical power to detect differences both within and between subgroups. Second, many of the subgroups with more than two categories are comprised of less than five students, which means reporting even aggregate results for these groups could result in participant disclosure. Therefore, we were unable to conduct analyses on subgroups for this study.

Findings by Dosage Characteristics

In addition to the overall effect of the Wisconsin Reading Corps 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 **Appendix F**.

The dosage analysis for Kindergarten students did not yield significant findings.[21] 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 education interventions. It is often the case that programs must 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 and received tutoring with fidelity, they tended to score higher on the Winter benchmark.

While the dosage analysis for first grade students yielded similar findings to those found among the Wisconsin Kindergarten students, the findings were statistically significant. The direction of the correlations were consistent with the Kindergarten findings. The association with total number of sessions was negative, significant, and strong (correlation = -.53; p = 0.011), reflecting the reactive nature of providing more sessions to students who were most in need. The association with attendance was positive, significant, and strong (correlation = .44; p = .002), indicating 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.

Conclusions and Discussion

The findings from the Wisconsin Reading Corps impact evaluation provide important evidence for addressing the study's two primary research questions. Below, the evaluation team offers our conclusions on the impact of the Wisconsin Reading Corps program on Kindergarten and first grade students' literacy proficiency scores and organizes them by major research question. Following our assessment of the questions is a discussion on the implications of our findings for the Wisconsin program and other Reading Corps replication sites.

1. What is the impact of Wisconsin Reading Corps on program participants compared to students who did not receive Reading Corps?

The results of the study showed that Kindergarten and first grade students who received the Wisconsin Reading Corps program achieve significantly higher literacy assessment scores by the Winter benchmark than did control students who did not participate in the program. The tutoring effects were similar across grades, with substantial effects found among both Kindergarten and first grade students.

---

[21] Sample sizes were again small for the Kindergarten subgroup analysis, which most likely contributed to the lack of statistically significant findings.

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 reached the Winter benchmark score. This effect was not unexpected given similar findings from the earlier impact evaluation of the Minnesota Reading Corps K-3 program. Both programs recruited Kindergarten students who qualified to receive Reading Corps services because they knew few letter sounds at the beginning of the school year, as evidenced by extremely 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.[22] Kindergarten students in the Reading Corps program are provided with 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, including achieving the Winter benchmark expectation on average. These findings indicate that the Wisconsin Reading Corps program, a replication of the successful Minnesota Reading Corps program, more than achieved 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 Wisconsin Reading Corps program produced significantly higher *nonsense word fluency* scores than students in the control group by Winter benchmark. As with the Kindergartners, first grade students demonstrated both statistically significant and sizeable effects of the program. By the end of the first semester, students who received Wisconsin Reading Corps tutoring were able to produce more letter sound segments in nonsense words than students in the control group. While Wisconsin Reading Corps first grade students exhibited greater *nonsense word fluency*, unlike the Kindergarten students, first grade students on average did not reach the first grade Winter benchmark proficiency score by the end of the first semester. However, it is important to consider that the average first grade student in the program began the school year far below the Fall expected benchmark, and the program students should continue to grow at an accelerated pace throughout the remainder of the school year compared to the control students.

In sum, the results of the Wisconsin Reading Corps impact evaluation on Kindergarten and first grade students suggest that the program produces meaningful effects for its younger students within a single school semester. The intensive one-on-one tutoring produces increases in Kindergarten students' *letter sound fluency*, helping them reach expected benchmark scores within a single semester. For first grade students, the program produces sizeable gains in *nonsense word fluency* compared to control students.

---

[22] 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

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

After establishing an overall positive impact of the Wisconsin Reading Corps program on Kindergarten and first grade students' literacy outcomes, we planned to conduct analyses to examine whether differential effects of the program existed for specific subgroups of students. However, we were unable to complete these analyses in most cases due to small sample sizes.

In addition to demographic subgroups, the study also examined the effect of dosage and attendance on student outcomes to identify what levels of program intensity are appropriate for producing anticipated effects. As is often the case, we found that higher dosages of the Wisconsin Reading Corps program did not necessarily lead to better outcomes for students because the students furthest from target required greater program dosage to reach benchmark. Thus, we found that dosage levels were often more highly correlated with student need rather than literacy outcomes. On the other hand, when looking at attendance, which is defined as students attending the planned number of sessions appropriate for their level of need, we found that the students with higher attendance levels tend to score higher on the Winter benchmark. Based on these findings, we conclude that the Wisconsin Reading Corps program is assigning the correct amount of planned sessions for their students because when those students complete the planned number of sessions, they achieve higher scores on literacy outcomes.

Discussion

Overall, the results of the evaluation show that the Wisconsin Reading Corps program positively impacts Kindergarten and first grade students' literacy outcomes, and the magnitude of the effect of the program was similarly large in both grades. These findings have important implications for both the Wisconsin program and its planned expansion in the 2018-19 school year and for the Reading Corps replication effort. In our previous study on the Minnesota Reading Corps, we concluded that the large and significant impacts of the program's early, intensive intervention on Kindergarten and first grade students helps students catch up to their peers on critical literacy skill. While the students in the Wisconsin Reading Corps program started far below Fall benchmark compared to the average student in the Minnesota Reading Corps study, Kindergarten students were still able to reach benchmark within a single semester and first grade students made meaningful gains compared to their control group peers. Therefore, we are able to conclude the effectiveness of the Wisconsin replication site of the Reading Corps program on these age groups.

An equally consequential result of this study for the Reading Corps program is the documentation of their successful replication of the program model in Wisconsin. While we were unable to conduct in-depth site visits to the Wisconsin program to confirm fidelity to the Reading Corps model, the program appears to include several similar elements that we found in our previous *Process Assessment* result in a successful

program.[23] The Wisconsin replication program appears to have established a similar program-based infrastructure and resources, including critical partnerships with local nonprofit organizations, the implementing school district, and schools; the use of objective data based decisions to determine student eligibility, track learning progress, and make adjustments to interventions when needed; specialized interventions accurately implemented and closely monitored by a multi-layered supervisory structure of Internal and Master Coaches; and comprehensive, ongoing training to and supervision of tutors serving without possessing any specialized prerequisite technical skill. Therefore, the combination of Reading Corps program elements that previously resulted in positive impacts on student literacy outcomes in Minnesota have been shown to be a robust and effective model for the development of other successful reading intervention programs for K-1 students.

Also, the implications of this evaluation for research are critical. Much of the literature on the scaling of effective program models is theoretical in nature and mainly offers lessons learned and insights based on experience. Reading Corps has not only been shown through rigorous research to be an effective program in its original location, Minnesota, but now includes evidence of a successful replication of its program model in another state, Wisconsin. These current study findings coupled with the findings from the previous *Impact Evaluation* and *Process Assessment*, which identified those program elements that make the Reading Corps successful, add important and useful information to the literature on effective replications of reading interventions.

---

[23] Hafford, C., Markovitz, C., et al. (October 2014). *Process Assessment of the Minnesota Reading Corps PreK Program: Phase II.* (Prepared under contract to the Corporation for National and Community Service). Chicago, IL: NORC at the University of Chicago.

Appendix A: Logic Model

Exhibit A.1. Reading Corps Logic Model (for Impact Evaluation)

Problem

Students not school ready or reading below grade level

Program and School Inputs/ Resources

Selection of sites based on degree of student need & internal capacity to partner effectively

School selection of PreK and K-3 students for Tier 2 & 3 interventions based on literacy assessment of current literacy skills

School incorporation of web-based data management system to track & monitor student progress with literacy intervention

Research-based core literacy curriculum

Recruitment, screening & placement of tutor by program and school staff

Training of tutors & Literacy Coaches in literacy interventions, assessment, data-driven decision-making & program rules (N=800)

Identification & assignment of dedicated Literacy Coach at school to support & monitor tutors

Activities

Continuous improvement feedback loop among Master Coach, Internal Coach, tutors to ensure implementation fidelity

Coaching
Ongoing consultation with Master Coach

Supervision & coaching of tutors by Internal Coach
- Monthly meeting with tutors to review student progress & set individualized plans for each student
- Ongoing observation & coaching to refine each tutors' delivery of interventions
- Fidelity assessment of tutor delivery of RC interventions

Tutoring of Pre-K students by tutors
(N=20 per tutor per year)
- Support classroom instruction (1:1, small group, class)
- Enrich literacy environment
- Chart progress of targeted students
- Conduct monthly progress monitoring of targeted students
- Assess all students 3x per year on literacy skills
- Deliver 1:1 & small group literacy interventions to students needing extra assistance

Tutoring of K-3 students by tutors
(N=20-40 per tutor per year)
- Assess students identified by site staff as needing assistance (i.e., closest to target performance)
- Deliver 1:1 tutoring to selected students (20 minutes/day, 5 days/week)
- Chart weekly student progress
- Conduct weekly progress assessments & 3x per year benchmarking
- Exit students deemed to be on track & add new students

School Practices
- Implement core curricula
- Implement RC (supplemental intervention)
- Schedule tutoring sessions
- Designate Internal Coach
- Support implementation fidelity of RC model
- Support coaching, assessment & benchmarking practices
- Integrate tutor into school setting

Short-term Outcomes

Pre-K Students
- Number & % of students improving on 3-5 Fastbridge & Early Literacy Outcomes (ELO) Measures at subsequent benchmarking period: Fall & Spring

K-3 Students
- Number & % of students improving on literacy measures at subsequent benchmarking period: Fall, Winter, Spring
- Number & % of students successfully exiting from grade level intervention

System-level
- Enriched classroom literacy environment *(Pre-K only)*
- Expansion of early literacy goals in K-3
- Increased integration of assessment practices into K-3 school culture

Long-term Goals

Pre-K
Achievement of "kindergarten ready" targets on Fastbridge & ELO measures

K-3
Demonstrated proficiency on the state's 3rd grade reading test

Build internal capacity for RTI methods and literacy instruction

Expand literacy goals site-wide

Appendix B: Randomization Process

Appendix C: Baseline Analysis

The study team conducted baseline equivalence analyses, including chi-square tests of independence[24], t-test analyses of two populations[25], and Hedges' g[26] 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 both Kindergarten and first grade were comparable (see Tables C.1 and C.2). While some differences among smaller subgroups of demographic variables were found, the overall balance between the program and control groups for both Kindergarten and first grade student in Wisconsin is sound, indicating that the two groups were 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 Wisconsin Evaluation Participants, Kindergarten

Covariate	Chi-sq	p-value	Test	p-value2	Hedges g	Total N
Kindergarten						**78**
Gender						
Male	0.83	0.36	-0.91	0.36	-0.20	42
Female	0.83	0.36	0.91	0.36	0.20	36
Race/Ethnicity						
Black or African American	0.46	0.50	0.68	0.50	0.15	68
White	2.05	0.15	-1.43	0.15	-0.32	2
Hispanic/Latino	0.16	0.69	-0.40	0.69	-0.09	7
Asian	1.01	0.31	1.01	0.31	0.22	1
Multi-racial	—	—	—	—	—	0
American Indian, Alaskan Native Native Hawaiian	—	—	—	—	—	0
Student ELL status						
ELL-eligible	2.89	0.09	1.70	0.09	0.38	6
Not ELL-eligible	2.89	0.09	-1.70	0.09	-0.38	72
Fall Assessment Scores	—	—	-0.19	0.85	-0.04	78

---

[24] 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.

[25] 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.

[26] 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 Wisconsin Evaluation Participants, 1st Grade

Covariate	Chi-sq	p-value	Test	p-value2	Hedges g	Total N
First grade						**124**
Gender						
Male	0.03	0.86	-0.18	0.86	-0.03	61
Female	0.03	0.86	0.18	0.86	0.03	63
Race/Ethnicity						
Black or African American	0.48	0.49	0.69	0.49	0.12	101
White	0.00	1.00	0.00	1.00	0.00	2
Hispanic/Latino	0.44	0.51	-0.66	0.51	-0.12	10
Asian	0.00	1.00	0.00	1.00	0.00	10
Multi-racial	1.01	0.32	-1.00	0.32	-0.18	1
American Indian, Alaskan Native Native Hawaiian	—	—	—	—	—	0
Student ELL status						
ELL-eligible	0.90	0.34	0.95	0.34	0.17	11
Not ELL-eligible	0.90	0.34	-0.95	0.34	-0.17	113
Fall Assessment Scores	—	—	-0.12	0.90	-0.02	124

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 82 Kindergarten students were randomized into program or control conditions through the matched pair design. Of those randomized, one control student refused the study, three program students were removed from the program (e.g. low attendance, refusal to engage tutor, and became ineligible), and six withdrew from the school (two in the program group and four in the control group). 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 five extra students in the program condition and three extra students in the control condition due to incomplete pairs. As such, the analysis sample was reduced to 64 Kindergarten students, an attrition rate of 22%, 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 summarizes these findings.

Table D.1. Attrition in Kindergarten Sample

	Overall	Program	Control
Total Randomized	82	41	41
Attrition			
Refused	1	0	1
Removed from Program	3	3	0
Withdrawn	6	2	4
Pair member Attrition	8	4	4
Total Attrition	18	9	9
Analysis sample	64	32	32
Attrition Rate	22%	22%	22%

A total of 126 first grade students were randomized into program or control conditions through the matched pair design. Of those randomized, one program student was removed from the program (e.g. low attendance), and six withdrew (five in the program group and one 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 an extra student in the program condition and six extra students in the control condition due to incomplete pairs. As such, the analysis sample of 112 first grade students, an attrition rate of 11%, which given the lack differential attrition, is well within the norms outlined by the WWC.

Table D.2. Attrition in 1st Grade Sample

	Overall	Program	Control
Total Randomized	126	63	63
Attrition			
Refused	0	0	0
Removed from Program	1	1	0
Withdrawn	6	5	1
Pair member Attrition	7	1	6
Total Attrition	14	7	7
Analysis Sample	112	56	56
Attrition Rate	11%	11%	11%

Appendix E: Detailed Analysis Methods

The analysis for the impact evaluation employed a mixed model (defined below) to estimate the treatment effect for each outcome using data from specific 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

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

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

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

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 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.

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 measures 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.[27] 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 these variables were used in a regression model to predict the outcome values for program participants. Thus, the regression model is specified as

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

---

[27] 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.

where we report the standardized coefficients to present partial correlations.[28] Our model also includes control variables represented by $\sum_c \gamma_c X_{cij}$.

---

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

Appendix F: Results Tables

Table F.1 presents the detailed results from the K-1 analysis. The column labeled "Treatment effect" represents the β_1 parameter detailed in the analysis section, which represents the difference between the program and control groups' predicted means. The columns labeled "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.[29] The "Effect size" column presents the results from a model without controls that is converted into standard deviation units. The final columns are the predicted program and control means in the outcome metric and the total sample size.

Table F.1. Results from K-1 Outcome Analysis

Grade and Outcome	Treatment effect [a,b] (β_1)	p-value[a]	Effect size[c] $\left(\frac{\alpha_1}{\sigma}\right)$	Prog. Mean[d]	Control mean[d]	Total N
Kindergarten						
Letter Sounds fluency	6.460	0.035	0.549	27.183	20.723	64
	(3.065)					
First grade						
Nonsense Words fluency	8.735	0.004	0.462	43.814	35.079	112
	(3.032)					

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.

Table F.2 presents the detailed results from the K-1 dosage analysis. The column labeled "Partial correlation" represents the regression coefficient detailed in the analysis section and relates the dosage variable to the outcomes. The column labeled "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.

---

[29] Rejecting the null hypothesis is based on a range of data starting with the observed differences between groups.

Table F.2. Partial correlations between outcomes and dosage indicators

Grade, Outcome, and Covariate[a]	Partial correlation[b]	p-value
Kindergarten (Letter Sounds, N = 32)		
Number of sessions (λ_1)	-0.392	0.142
	(0.267)	
Average length of sessions (λ_2)	0.185	0.201
	(0.144)	
Attendance (λ_3)	0.138	0.590
	(0.257)	
First Grade (Nonsense Words, N = 56)		
Number of sessions (λ_1)	-0.526	0.011
	(0.208)	
Average length of sessions (λ_2)	0.188	0.101
	(0.115)	
Attendance (λ_3)	0.438	0.002
	(0.139)	

a: Model includes Fall assessment, gender, race, and ELL status control variables.
b: Variables coded into standard deviation units to produce partial correlations; Standard errors in parentheses