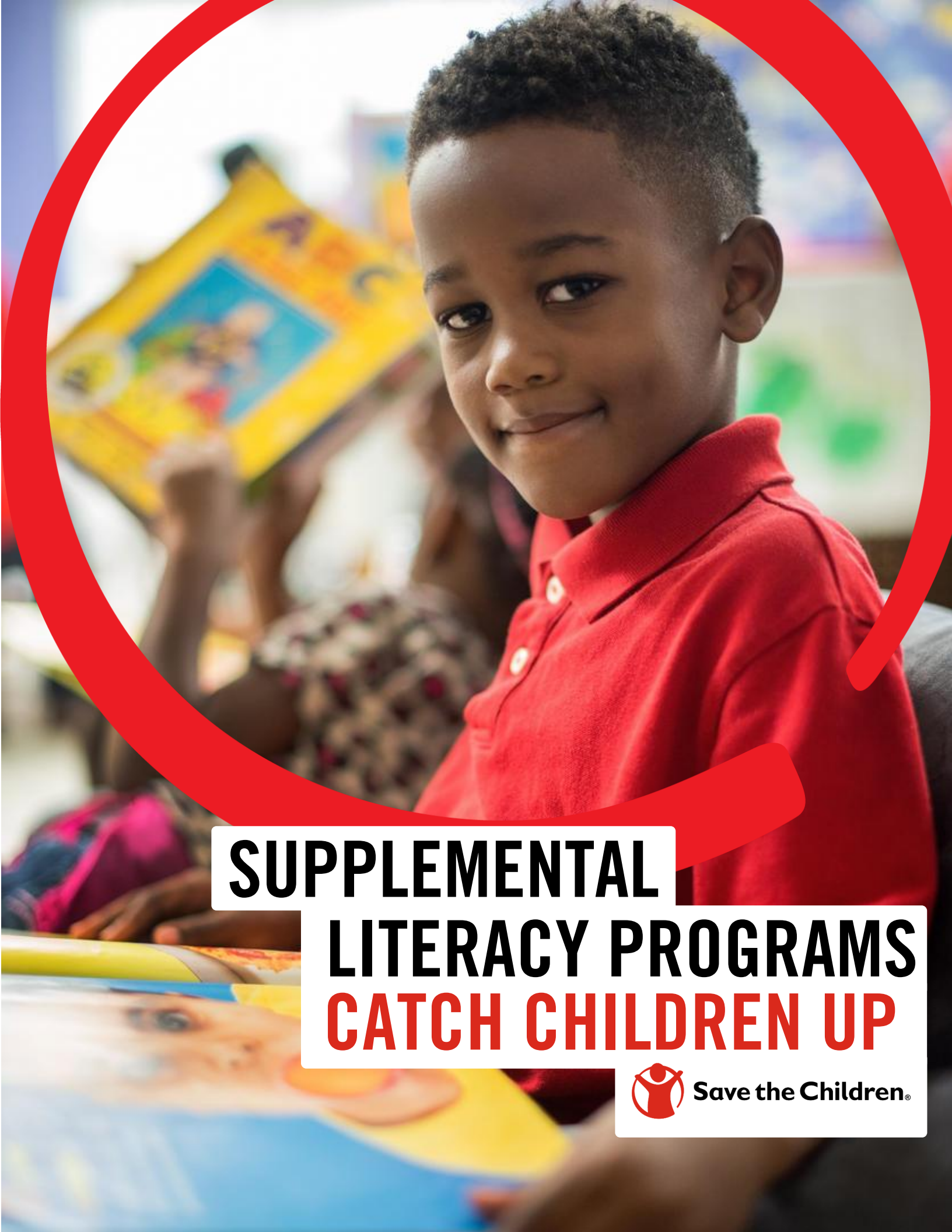

SUPPLEMENTAL LITERACY PROGRAMS CATCH CHILDREN UP

Save the Children®

Supplemental Literacy Programs Catch Children Up: Effective Strategies to Reduce Achievement Gap Faced by Students in Rural America, A Randomized Controlled Trial Study

Nicole Smith, M.A., Senior Specialist Monitoring and Evaluation, U.S. Programs and Advocacy, Save the Children
Patrick V. Iannone, MS Ed., State Director, U.S. Programs and Advocacy, Save the Children
Julie Meredith, Ph.D., Senior Research Associate, Policy Studies Associates
Anita I. Drever, Ph.D. Director of Monitoring and Evaluation, U.S. Programs and Advocacy, Save the Children

ABSTRACT

Children who are not reading on grade level by the end of third grade are at greater risk of dropping out of high school than their peers who have reached this essential milestone. Children who are living in poverty *and* not reading on grade level by the end of third are at even greater risk. Children in rural America[1] are especially likely to be both poor and below grade level.

Save the Children offers in-school literacy programming to elementary school students in kindergarten through third grade living in high-poverty, rural communities in order to help them overcome these obstacles and acquire the literacy skills that will position them for lifelong success. The purpose of this study was to determine, through the implementation of a rigorous, randomized controlled trial study design at nine program sites in Mississippi, whether Save the Children's daily, in-school literacy programming produces greater test score gains for students reading below grade level than participation in schools' language arts curriculum alone.

The study results indicate Save the Children's programming produces statistically significant literacy and reading comprehension test score gains for K-3rd grade students in Mississippi who began the school year reading below grade level. The results indicate that in-school literacy programming featuring Read-Alouds and Guided Independent Reading Practice in daily, half-hour one-on-one and small group tutoring sessions may help children living in low-income communities in rural America acquire the literacy skills they need to succeed throughout their education careers and beyond.

Published by Save the Children
December 2018

[1] We classify counties as rural if they are not part of Metropolitan Statistical Areas. The United States Office of Management and Budget (OMB) considers a county to be part of a Metropolitan Statistical Area if the county contains an urban core with a population of a least 50,000, or is adjacent to a county with a core of that size and has close social and economic ties with that core. For more information, see https://www.census.gov/programs-surveys/metro-micro/about.html.

INTRODUCTION

If a child is not reading on grade level by the end of third grade, that child is four times more likely to leave high school[2] without a diploma than a child who has achieved this milestone. A child not reading on grade level by the end of third grade who has also lived for a year or more in poverty is at even greater risk for not completing high school.[3] Children's early literacy skills are the foundation for their learning in all subject areas, not just language arts. Early deficits can affect lifelong success.

Children living in rural parts of the United States are especially likely to face the double jeopardy of growing up in poverty and not having had the opportunity to develop their basic literacy skills in elementary school. 23.5% of children living outside metropolitan areas in the United States were living in poverty in 2016. By contrast, 18.8% of children in metropolitan areas were living in poverty in 2016, the most recent year for which Census data were available at the time of the writing of this report.[4] In 2015, 69.5% of children living in small towns had 4th grade reading scores that were below proficient as compared to 60.4% of fourth graders living in suburban communities.[5]

Save the Children provides multifaceted support to vulnerable children ages 0 to 8 and their families living in rural parts of the United States that will help give those children a more equitable opportunity to acquire the literacy, numeracy and socio-emotional skills by the end of third grade that will position them for lifelong success. Save the Children supports rural children ages 0 to 5 and their families with home visits, book bag exchanges and kindergarten readiness playgroups as tactics to accelerate rural children's early learning. Once children are old enough to attend school, Save the Children provides kindergarten transition supports to children and families; summer and out-of-school-time enrichment-focused programming to counter the "summer slide" in the academic skills of low-income children, and in-school literacy supports to children who are not yet reading on grade level. This report summarizes the findings of a rigorous, randomized control trial study that examined the effectiveness of the in-school literacy supports in Save the Children's suite of 0-8 programming.

The randomized controlled trial was conducted in nine elementary schools in seven rural school districts implementing Save the Children's in-school literacy programming in Mississippi[6]. Students reading below grade level were randomly assigned to treatment and control groups at the beginning of the school year. The study's authors were able to compare the test scores of students who participated in 90 or more days of supplemental, in-school literacy programming focused on reading practice through Read-Alouds and Guided Independent Reading Practice with the test scores of similar students at the same schools who did not participate in the programming. The analysis results confirmed that the students who participated in the programming had significantly higher early literacy and reading comprehension test scores than the students who did not participate in the programming. These results indicate that in-school programs targeted at students reading below grade level that focus on Read-Alouds and Guided Independent Reading Practice help reduce the achievement gap between students in non-metropolitan and metropolitan areas of the United States.

---

[2] Statistic was calculated using data on high school graduation by age 19.
[3] Donald J. Hernandez, *Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation.* (Baltimore, Maryland: The Annie E. Casey Foundation. 2011).
[4] Save the Children. *U.S. Complement to the End of Childhood Report 2018: Growing Up Rural in America.* (Fairfield, Connecticut: 2018). Page 4.
[5] National Center for Education Statistics. "Rural Education in America." Accessed December 21, 2018. https://nces.ed.gov/surveys/ruraled/tables/b.2.a.-1r.asp.
[6] In one of the nine Mississippi schools, Save the Children also provided afterschool programming and some children from both the treatment and the control groups were enrolled.

PROGRAM OVERVIEW

Schools in low-income rural communities often lack supports for struggling readers. Small rural districts are less likely to have curriculum specialists to help teachers incorporate state-of-the-art pedagogical practices.[7] They are also less likely to be able to offer supplemental education services to help children who need additional support in the classroom.[8] Save the Children's in-school literacy programming for kindergarten through third graders therefore fills an important gap by providing supplemental literacy supports to students who would likely not otherwise receive them. During the 2016–2017 school year, Save the Children provided in-school and afterschool programming supports to over 9,000 students in high poverty, rural regions of 10 U.S. states.

Within Save the Children's in-school literacy programs, there are two major programmatic approaches: one that tutors use with kindergarten and first graders, and another that tutors use with second and third graders. Save the Children's **Emergent Reader** programming for kindergarten and first graders focuses on beginning reading skills. Children practice daily reading easy texts aloud and participate in hands-on active learning in small groups to support increased growth in phonemic awareness, letter recognition, sound-symbol correspondence and beginning sight words. Tutors monitor students' reading performance through individual module assessments targeting these skills. Tutors administer the module assessments weekly to gauge skill mastery and to inform students' subsequent placement in the different modules over time.

Save the Children's **Developing Reader** programming for second and third graders also incorporates Guided Independent Reading Practice with books aligned to students' interests and skill levels. Tutors monitor children's progress by having them regularly complete Renaissance Accelerated Reader online book quizzes, which help tutors support students in selecting developmentally appropriate books as skill levels improve. Programming may also include small-group tutoring sessions that focus on phonics, sight words, and vocabulary development.

Both the Emergent and Developing Reader approaches are five days per week, 30 minutes per day. Children are pulled out of their regular classes (but not literacy classes) to attend the program in groups of 5-10. See Appendix A for more detailed descriptions of Save the Children's Emergent and Developing Reader programs.

METHODS

Sample Characteristics

The study was conducted in a subset of the schools with Save the Children in-school literacy programming: nine elementary schools in seven rural Mississippi school districts. The nine schools in the study sample comprised all the locations where Save the Children was implementing in-school programming within the state of Mississippi during the 2016–2017 school year. At each of the schools in the study sample, more than 90% of the students enrolled were participating in the Free and Reduced Price Meals program, and Save the Children programming had been implemented at each of the sites for a minimum of seven years. At one of the sites, Save the Children was also implementing an after school program that students in both the treatment and control groups attended and therefore should not have influenced the results of this study. Save the Children's regional and national offices closely monitor program quality at all its program sites through site visits and oversight of student assessment data, providing additional guidance when sites do not achieve quality benchmarks.

---

[7] Michel McNeil, "Rural Areas Perceive Policy Tilt," *Education Week*, August 28, 2009, 1 and 22.
[8] Leslie M. Anderson and Katrina G. Laguarda. "Case Studies of Supplemental Services Under the *No Child Left Behind Act:* Findings *from 2003-04."* (Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development. 2005).

Study Design

The authors employed a clustered, randomized controlled trial study design to collect the data that would allow them to test whether in-school literacy program participation improved children's literacy and reading comprehension test scores. The randomized controlled trial design is the gold standard for collecting data to test the effectiveness of education interventions as it enables researchers to mitigate selection bias as well as any possible differences between the treatment and control groups on unmeasured factors. Weaker studies—where, for example, the performance of students whose families signed them up for a program are compared to the performance of students whose families did not—are subject to selection bias. If at the conclusion of the program, the test scores of the students who participated in the program are higher than the test scores of the student that did not, researchers have no convincing basis for determining whether these differences were due to program effects, or were due to the fact that families who are already more deeply engaged in supporting their students' success are also more likely to enroll their children in additional programming. By assigning children at random to treatment and control groups, that weakness is eliminated: no one can argue that whatever influenced a child's enrollment might also influence their likelihood of success in the program.

To determine eligibility for both the in-school literacy program and the study, all kindergarten and first grade students in the study sample completed Renaissance's STAR Early Literacy diagnostic assessment (SEL) and all second and third graders took Renaissance's online STAR Reading assessment. Students receiving test scores below the cutoff points[9] for assignment to Tier 2 and Tier 3 reading supports based on the Federal Response to Intervention (RTI) guidelines, but who had not received Individual Education Plans, were eligible for participation in both the program and the study. Because the number of children who were reading below grade level greatly exceeded staff capacity, not all children who needed literacy support could participate in the in-school literacy program. Random assignment of eligible students to intervention and control groups therefore met both programmatic and evaluation needs.

The study population of students was drawn from all of the grades at each school participating in the supplemental literacy program. Approximately 110 students at each of the nine schools (1,000 students total) were assigned in equal proportions to the treatment and control groups. Roughly the same number of students were selected for the study from each grade participating in the program at each school. (See Appendix B for a more detailed description of the random assignment process.) Students were included in the final analysis if they completed the pre- and post-tests within the appropriate timeframe, attended 90 or more days, and had student records that included information on their gender.

Within each school, the Save the Children literacy program takes place outside of the children's regular classroom, which means children not participating in the program cannot overhear or benefit from it unintentionally. Children attending the program leave their classroom at a time when the school's regular curriculum is not addressing literacy. Therefore, children in the program receive the same literacy education provided by the school as children not in the program, and the Save the Children literacy program is truly supplemental to the school curriculum. Differences in pre-test scores between the program participants and the comparison control group were not statistically significant and indicate the treatment and control groups were comparable (See Table 1).

[9] The cutoff point for second and third graders was a normal curve equivalent score of 50 or lower. The cutoff point for kindergarten and first graders was a scaled score of 674 or lower.

Table 1: Mean Pre-Test Scores for the Treatment and Control Groups by Grade Level Groupings

	Save the Children program participants	Comparison group non-participants
Kindergarten–1st grade	**N=285**	**N=262**
Average pre-test SEL scaled score (standard deviation)*	527.4 (94.9)	526.5 (90.4)
2nd–3rd grade	**N=155**	**N=160**
Average pre-test STAR scaled score (standard deviation)*	183.9 (83.6)	177.5 (77.5)

* Includes pre-test scores only for children who also completed post-tests

Outcome Measurement

The nine Mississippi schools participating in the study administered school-wide early literacy and reading comprehension testing at the start of the school year. Students in kindergarten and first grade completed Renaissance's STAR Early Literacy diagnostic assessment (SEL). The twenty-minute online adaptive diagnostic assessment measures students' vocabulary, phonics, language and numeracy skills and their understanding of literacy concepts such as words and letters. Students who completed the assessment received a scaled score that ranged from 300-900. Renaissance computes scaled scores based on the number of correct responses and the difficulty of the items. The same scaled score range is used for all test takers, which enables consistent comparison and analysis across grade levels.

Students in second and third grade completed Renaissance's online STAR Reading assessment, which measures students' skills in a variety of literacy domains. Students who completed the assessment received scaled scores between 0 and 1400, as well as normed scores in percentiles and Normal Curve Equivalents (NCEs). The SEL and STAR Reading post-test assessments were administered at least 90 days after the pre-test.

Statistical Analysis

The central research question posed for this evaluation is whether kindergarten through third graders reading below grade level who participate in Save the Children's in-school literacy programming achieve greater literacy gains than their counterparts who do not participate in the programming. The authors used hierarchical linear modeling (HLM) to measure whether students randomly selected to receive the in-school literacy supports achieved greater gains in their literacy and reading comprehension test scores (See Appendix C for detailed model descriptions). This statistical approach nests students in schools, and analytical approach to account for: the sampling and assignment strategy which selected participant and comparison students from the same school population; and the common experiences shared by students in the same school that could influence their test score gains, regardless of program participation.

Using HLM also allowed the authors to control for additional factors at the individual-level that could impact individual students' achievement (e.g., prior test scores, grade) and at the school-level such as average school wide STAR or SEL test scores. For example, differences in average school-wide test scores could be indicative of factors within the school that impact all children, both participants in the Save the Children program and non-participating students. At the individual-level, children's prior test performance and grade could impact the magnitude of their SEL or STAR gains, again independent of their participation in Save the Children programming.

RESULTS

The authors found positive, statistically significant results for the effect of children's participation in Save the Children's in-school literacy programs. Both the kindergarten–1st grade model, which measured the program's influence on early literacy, and the second–third grade model, which measured the program's influence on reading comprehension, yielded statistically significant differences in the treatment and control groups' post-test results. Kindergarten and first-graders in Save the Children programming gained, on average, 20 more points on the SEL post-test compared with students who did not receive services (γ_{20}=20.174, *se*=6.528, $p \leq 0.01$). Second- and third-grade participants in Save the Children programming also made greater gains in comparison to second- and third-grade non-participants—gaining, on average, 17 more points on the STAR test between pre- and post-assessments (γ_{20}=17.081, *se*=8.507, $p \leq 0.05$) than their non-participating peers (see Table 2). (See Appendix C for the full results of HLM models and more information about the model building process).

Table 2: Average Pre- and Post-Test Score Change for the Treatment vs. Control Groups by Grade Level

	Save the Children program participants	
Grade Level	**Coefficient**	**Standard error**
Kindergarten–1st grade		
Average SEL pre-test and post-test score change for Save the Children participants compared with non-participants	20.2***	6.5
2nd–3rd grade		
Average SEL pre-test and post-test score change for Save the Children participants compared with non-participants	17.1*	8.5

Statistical significance: * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Other findings of interest included a statistically significant effect of grade in the kindergarten–first grade models, which suggested the program may have produced stronger results for first grade students than for kindergarten students. In order to more closely examine the grade level differences, the authors ran additional analyses which indicate that while there was a positive effect of participation in Save the Children programming for children in both kindergarten and first grade, the effect was greater – and only statistically significant for – children in first grade (See Appendix C for further discussion). There was, however, no statistically significant difference in the pre-test and post-test score change for second graders as compared to third graders. Gender was also not significant in the second to third grade model.

IMPLICATIONS

This study provides strong evidence for the effectiveness of Save the Children's in-school literacy intervention. By randomly assigning students to treatment groups that participated in the daily in-school literacy programming and control groups that did not, the authors were able to ensure that the only difference between the two groups was their participation in the program. And as a result, differences in the outcome could be more confidently attributed to students' participation rather than other unmeasured factors. In addition, the study was conducted in one of the more challenging public education contexts in the United States: Mississippi was ranked 44th in the nation in 2017 for its

fourth grade reading scores.[10] The intervention's success there suggests that in-school literacy programming featuring Read-Alouds and Guided Independent Reading Practice in daily, half-hour one-on-one and small group tutoring sessions may help close achievement gaps in other states with large rural populations.

Save the Children's In-School Literacy Program is, however, only one of a suite of programs Save the Children offers children and their families in some of the most marginalized rural communities in the United States. Future research will examine the degree to which the full suite of programming – including in-school literacy programming for K-3rd grade children – influences 3rd grade reading scores. Save the Children hypothesizes that by offering year-round learning supports to children from birth to age eight, they are less likely to fall behind early in their educational careers, positioning them to succeed over the course of their entire school careers and beyond.

[10] "The Nation's Report Card," National Assessment of Educational Progress (NAEP), accessed December 21, 2018. https://www.nationsreportcard.gov/.

APPENDICES

APPENDIX A: Save the Children's Emergent and Developing Reader Programs

Save the Children offers a variety of literacy activities through its in-school literacy program:

- For children in Kindergarten and first grade who are reading below grade level, literacy activities that target beginning reading skills are available. Emergent Reader literacy activities include extended read-aloud with developmentally appropriate follow-up activities; reading together time where easy text is practiced orally by students daily; and hands-on active learning to support increased growth in phonemic awareness, letter recognition, sound-symbol correspondence and beginning sight words.

- For students in grades 2-3 reading below grade level, these activities include Guided Independent Reading Practice, which is supported by Read Alouds with related fluency-building activities. Technology is used to maximize children's learning through monitoring progress with regular Renaissance Accelerated reader book quizzes. Small-group tutoring sessions – targeting phonics, sight words, and vocabulary development – may also be incorporated to support children's reading skill development.

Study participants in grades K-1 received Emergent Reader literacy activities daily for 30-minutes. Daily activities included a Read Aloud with fluency building and vocabulary learning experiences incorporating grade-level books and materials. In addition, study participants in grades K-1 utilized hands-on literacy activities that promote active learning in small groups. Activities targeted specific reading skills such as phonemic awareness, letter recognition, sound-symbol correspondence, and beginning sight words.

For students in grades K-1, reading performance was monitored through individual module assessments that target phonemic awareness, letter recognition, sound-symbol correspondence, and beginning sight words. Module assessments were administered weekly to determine skill mastery, and subsequent student placement in different modules over time.

Study participants in the second and third grades received literacy activities that featured Read Alouds with related fluency building and vocabulary activities one week a month on successive days for 30 minutes a day in small groups of 5-10. Read Alouds were implemented using the following procedures:

1. The teacher introduced the book in an enthusiastic way to the students.
    - The teacher mentioned the title, author and, perhaps, the illustrator.
    - The teacher provided student-friendly definitions for vocabulary words that needed to be discussed before reading.

2. The teacher read the book aloud to the students.
    - The teacher read with expression at an appropriate speed, varying the pace to pause for emphasis and to allow time for students to think about what's happening or what might come next.
    - The teacher paused at planned stopping points to highlight key vocabulary terms.

3. After reading, the teacher spent a short amount of time discussing the book.
    - The teacher asked open-ended questions that required more than a yes or no response, or initiated a high-level discussion related to the book.

4. In addition, related vocabulary and fluency-building activities were incorporated to support Read Alouds.
    - Fluency-building primarily consisted of choral reading, while vocabulary learning games were used to highlight key terms from Read Aloud texts.

Study participants in grades 2-3 also received daily experiences in Guided Independent Reading Practice for 3 weeks a month on successive days for 30 minutes a day in small groups of 5-10. Guided Independent Reading Practice was implemented using the following procedures:

1. Students self-selected an appropriate book.
    - Based on the results of a beginning-of-program STAR assessment, each student's reading range was identified. This range of book levels was explained to the student, and they were shown how to select books of interest that have been labelled within their range.

2. Students read the book.
    - The teacher guided students as they read, suggesting word-level and discourse-level strategies as needed (i.e., to make sure students were able to decode challenging words and comprehend what was read).

3. Teacher checked for comprehension.
    - The teacher asked the students to summarize the text or asked him/her to describe what was happening in each picture.

4. Students took a Renaissance Accelerated Reader (AR) quiz.
    - Students took the quiz in a timely fashion (no more than one day after completing their book). Students were instructed to read each question and all response choices carefully before selecting their answers. Immediate feedback was provided on student performance through the AR reading software program.

For students in grades 2-3, reading performance was monitored on a daily basis using Renaissance's Accelerated Reader (AR) software program. When using AR, children choose books they want to read. To help them select successfully, every book has an assigned book level, interest level and point value. Book level indicates the readability of the text, while interest level relates to content. Point values are assigned to books based on the book's length. The intent is for children to read books that are appropriate to their reading level and interest.

APPENDIX B: Random Assignment Procedure

The study's authors processed the randomized selection and assignment. Those implementing the literacy program in the schools provided the pretest reports but were not involved in the selection or assignment. Once all inclusion and exclusion criteria were applied to the sampling frame, individual child results were split into groups by grade within each school. Each child was assigned a unique sequential identification number and a random integer set generator was used to create lists of random number sets in the schools' grades ranges. As described above in the Study Design section, equal numbers of children were selected from each school. However, as each school offered programming to different grade ranges (most served K-3, but some served K-2 or K-1), the number of children selected from each grade varied by the grades served at their school. For schools serving K-3, the first 16 random numbers and the corresponding child identification numbers were assigned to the intervention group receiving the literacy program. The subsequent 16 random numbers and the corresponding child identification numbers were assigned to the control group. The same process was used for schools serving fewer grades, specifically study groups with 22 children were generated for schools serving K-2 and 33 children per study group for schools serving K-1. If schools were able to provide programming to more children than were assigned to the intervention group, children who were reading below grade level at the beginning of the year but who were not selected for either the control or intervention group may have participated in the program but were not included in this study.

APPENDIX C: Hierarchical Linear Modeling (HLM) Descriptions & Results

Kindergarten & First Grade Model: Emergent Reader

Descriptive Statistics

Table C-1: Characteristics of Participant and Non-Participant Groups

	Participants in Save the Children program (N=285)	Non-participant comparison students (N=262)
Grade		
Percent of children enrolled in Kindergarten (N=309)	28.0%	28.5%
Percent of children enrolled in 1st grade (N=238)	24.1	19.4
Average pre-test SEL scaled score (standard deviation)	527.4 (94.9)	526.5 (90.4)

Model Building

To build the models for analysis, we first reviewed the information available for the participant and non-participant groups. Data for race and gender were not available for a large proportion (83 percent) of the non-participant comparison students. Therefore, we opted not to include these variables in our models, because doing so would have excluded most of our comparison group. We did, however, have the grade in which children were enrolled for both participants and non-participants, and opted to include this covariate.

Next, we added a level-2 covariate, the school mean of pre-test STAR Early Literacy (SEL) scaled score, to the model as a predictor of the level-1 intercept and, in a second model, as a predictor the level-1 intercept of pre-test SEL scale score. We tested these models against one another and against the model with no level-2 covariates. These tests suggested that the model using school mean as a predictor of the level-1 intercept was the best fitting model.

Finally, we tested models that fixed or allowed the level-2 slopes to randomly vary. Using a likelihood ratio test of model deviance statistics, we determined that a model that allowed level-2 slopes to vary fit better than a model with fixed level-2 slopes, and the best fitting model fixed the level-2 slope predicting treatment group while allowing the slopes of grade and pre-test SEL scale score to vary. The final model appears below:

Level 1:

SEL scaled score (post-test) = $\beta_0 + \beta_1$ [Grade] + β_2 [Treatment] + β_3(SEL scaled score [pre-test]) + r

Level 2:

$\beta_0 = \gamma_{00} + \gamma_{01}$ [Site mean pre-test SEL scaled score] + μ_1

$\beta_1 = \gamma_{10} + \mu_1$

$\beta_2 = {}_{20}$

$\beta_3 = {}_{30} + \mu_3$

Assumptions Checking

To check the assumptions of the multilevel model, we first examined the level-1 and level-2 residuals. These tests indicated that the model met the assumptions of normal distribution of level-1 variance, though there did appear to be some deviation from multivariate normality at level-2. The lack of multivariate normality is, perhaps, not surprising given the relatively few covariates in the model; deviation from normality may, for example, indicate that there are additional factors influencing the results that were not include as covariates in the model. Further, we have nine level-2 schools, and with additional schools, the distribution could more closely approximate normal. Typically, we would report the results using robust standard errors to account for this slight deviation from multivariate normality; in this case, however, we do not for two reasons. First, robust standard errors are appropriate only for datasets with a moderate to large number of level-2 units, and with nine level-2 units these data do not meet that criterion. Second, comparing the results between the normal and robust standard errors, the differences appear to be primarily in the effect of prior achievement, not a focal variable of this analysis.

Finally, we checked for homogeneity of level-1 variance and found that the model met this assumption.

Results

We find significant positive results for the effect of children's participation in Save the Children's literacy intervention, with an average increase in the SEL scale score for participating kindergarten and first-grade children of 20 more points between pre-test and post-test compared with non-participants (γ_{20}=20.174, *se*=6.528, $p \leq 0.01$, see Table C-2).

An examination of the raw differences in the SEL scale scores between pre-test and post-test SEL scale scores by grade suggests that there may have been a greater increase in SEL scale scores across both participant and non-participant groups for children in kindergarten (non-participants: 208 points v. participants: 219 points) and a smaller overall increase in SEL scale scores in first grade, but a greater difference between children in the non-participant and participant groups (non-participants: 114 points v. participants: 149 points, see Table C-3).

Table C-2. Results of the Analysis of SEL Scale Scores, All Grades

Fixed effects	Coeff. (se)			
For intercept, β_0				
Intercept, γ_{00}	723.592 (14.756)***			
Site-mean pre-test SEL scaled score, γ_{01}	-1.226 (0.958)			
For GRADE slope, β_1				
Intercept, γ_{10}	-54.859 (20.706)*			
For TREATMENT GROUP slope, β_2				
Intercept, γ_{20}	20.174 (6.528)**			
For PRE-TEST SEL SCALE SCORE slope, β_3				
Intercept, γ_{30}	0.742 (0.085)***			
Random effects	Variance component	*sd*	*df*	χ
Intercept, τ_{00}	1582.603***	39.782	7	59.490
Grade slope, τ_{10}	3075.325***	55.456	8	33.953
Pre-test SEL scaled score slope, τ_{30}	0.209***	0.044	8	21.598

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Table C-3. Comparison of Mean SEL Scaled Scores, By Grade

	Non-participants			Participants		
	Pre-test	Post-test	Difference	Pre-test	Post-test	Difference
Kindergarten	480.5 (74.9)	688.4 (111.1)	207.9	468.0 (81.7)	687.5 (96.7)	218.5
Grade 1	594.2 (65.0)	708.1 (83.3)	113.9	595.1 (56.5)	744.0 (78.9)	148.9

To more closely examine the differences at each grade, we ran additional analyses which indicate that while there was a positive effect of participation in Save the Children programming for children in both kindergarten and first grade, the effect was greater and significant for children in first grade. For kindergarten, participation in the literacy intervention was associated with an SEL scale score increase of approximately 9 more points than non-participating kindergarteners (γ_{10}=8.983, *se*=9.237, *p*=0.334, see Table C-4). The impact for first graders was an increase of nearly 36 more in SEL scale score between the pre- and post-assessment than their non-participating peers, and the effect of participation was significant (γ_{10}=35.664, *se*=9.143, $p \leq 0.001$, see Table C-5).

Table C-4. Results of the Analysis of SEL Scaled Scores, Kindergarten

Fixed effects	Coeff. (se)			
For intercept, β_0				
Intercept, γ_{00}	690.953 (15.518)***			
Site-mean pre-test SEL scale score, γ_{01}	-0.439 (1.854)			
For TREATMENT GROUP slope, β_1				
Intercept, γ_{10}	8.983 (9.237)			
For PRE-TEST SEL SCALE SCORE slope, β_2				
Intercept, γ_{20}	0.81 (0.098)***			
Random effects	Variance component	*sd*	*df*	χ
Intercept, τ_{00}	1722.087***	41.498	7	68.145
Pre-test SEL scale score slope, τ_{20}	0.221*	0.049	8	18.300

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Table C-5. Results of the Analysis of SEL Scaled Scores, First Grade

Fixed effects	Coeff. (se)			
For intercept, β_0				
Intercept, γ_{00}	709.799 (9.030)***			
Site-mean pre-test SEL scale score, γ_{01}	0.243 (1.013)			
For TREATMENT GROUP slope, β_1				
Intercept, γ_{10}	35.664 (9.143)***			
For PRE-TEST SEL SCALE SCORE slope, β_2				
Intercept, γ_{20}	0.666 (0.075)***			
Random effects	Variance component	*sd*	*df*	χ
Intercept, τ_{00}	288.481**	16.985	7	18.400

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Second & Third Grade Model: Developing Reader

Descriptive Statistics

Table C-6. Characteristics of Participant and Non-Participant Groups

	Participants in Save the Children program (N=155)	Non-participant comparison students (N=160)
Grade		
Percent of children enrolled in 2nd grade (N=144)	18.4%	27.3%
Percent of children enrolled in 3rd grade (N=171)	32.4	21.9
Gender	(N=151)	(N=152)
Male (N=156)	50.6%	49.4%
Female (N=147)	49.0	51.0
Average pre-test STAR scaled score (standard deviation)	183.9 (83.6)	177.5 (77.5)

Model Building

We first reviewed the information available about the students comprising to participant and non-participant groups to identify potential covariates for the analysis. Grade level was available for all participants and non-participants, and we opted to include this covariate in our model. Gender was available for almost all students (94.2 percent) and testing during model development showed it decreased model deviance. Data for children's race/ethnicity were not available for a substantial portion of the population (approximately 37 percent). Although missing race/ethnicity data were distributed equally across the participant and non-participant groups, we opted not to include race in our models, because doing so would have excluded approximately one-third of both participants and non-participants. No other variables describing the characteristics of participating students were available for the analysis.

Next, we added a level-2 covariate, the school mean of pre-test STAR Reading scaled scores, to the model as a predictor of the level-1 intercept and, in a second model, as a predictor the level-1 intercept of pre-test STAR scaled scores. We tested these models against one another and against the model with no level-2 covariates. These tests suggested that the model using school mean pre-test STAR scaled scores as a predictor of the level-1 intercept was the best fitting model.

Finally, we tested models that fixed or allowed the level-2 slopes to randomly vary. Using a likelihood ratio test of model deviance statistics, we determined that a model with fixed level-2 slopes yielded the best fitting model. The final model appears below:

Level 1:
SEL scaled score (post-test) = $\beta_0 + \beta_1$ [Grade] +β_2 [Treatment] + β_3(Gender) + β_4(Pre-test STAR scaled score) + r

Level 2:

$\beta_0 = \gamma_{00} + \gamma_{01}$ [Site mean pre-test STAR scaled score]
$\beta_1 = \gamma_{10}$
$\beta_2 = \gamma_{20}$
$\beta_3 = \gamma_{30}$
$\beta_4 = \gamma_{40}$

Assumptions Checking

To check the assumptions of the multilevel model, we first examined the level-1 and level-2 residuals. These tests indicated that the model met the assumptions of normal distribution of level-1 variance, though there did appear to be some deviation from multivariate normality at level-2. The lack of multivariate normality is, perhaps, not surprising given the relatively few covariates in the model; deviation from normality may, for example, indicate that there are additional factors influencing the results that were not include as covariates in the model. Further, we have only seven level-2 schools, and, with additional schools, the distribution could more closely approximate normal. Typically, we would report the results using robust standard errors to account for this slight deviation from multivariate normality; in this case, however, we do not. Robust standard errors are appropriate only for datasets with a moderate to large number of level-2 units, and with seven level-2 units these data do not meet that criterion.

Finally, we checked for homogeneity of level-1 variance and found that the model met this assumption.

Results

We find significant positive results for the effect of children's participation in Save the Children's literacy intervention, with participating children gaining an average of 17 more scale score points on the STAR between pre-test and post-test compared with non-participating students (γ_{20}=17.081, se=8.507, $p \leq 0.05$, see Table C-7).

The negative effect of grade, which was significant in our analyses of changes in performance on the SEL assessment for children in kindergarten and first grade, is not significant for the second and third graders included in these analyses (see Table C-8 for a comparison of pre- and post-test scores by grade). Additional analyses suggest that there may be a slightly larger effect of participation in the literacy program for children in third grade, though these effects are not significant, possibly due to the relatively small number of children and schools available for these individual grade analyses.

Table C-7. Results of the Analysis of STAR Reading Scaled Scores

Fixed effects		Coeff. (se)		
For intercept, β_0				
Intercept, γ_{00}		324.968 (29.036)***		
Site (School?)-mean pre-test STAR scaled score, γ_{01}		0.322 (0.423)		
For GRADE slope, β_1				
Intercept, γ_{10}		-19.592 (11.148)		
For TREATMENT GROUP slope, β_2				
Intercept, γ_{20}		17.081 (8.507)*		
For GENDER slope, β_3				
Intercept, γ_{30}		1.33 (8.373)		
For PRE-TEST STAR SCALED SCORE slope, β_4				
Intercept, γ_{40}		1.045 (0.067)***		
Random effects	Variance component	*sd*	*df*	χ
Intercept, τ_{00}	351.546***	18.750	5	19.890

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Table C-8. Comparison of Mean STAR Reading Scaled Scores, by Grade

	Non-participants			Participants		
	Pre-test	Post-test	Difference	Pre-test	Post-test	Difference
Grade 2	138.5 (50.8)	241.7 (101.8)	103.2	131.3 (51.1)	243.4 (101.8)	111.7
Grade 3	226.1 (77.9)	313.8 (105.0)	88.7	245.0 (71.7)	357.3 (92.8)	112.3