A Randomized Controlled Trial Study of the Once Early-Reading Program in Multiple School Districts: Randomization Brief Summary

Ashley A. Grant, PhD, Amanda J. Neitzel, PhD, Steven M. Ross, PhD

March 2026

JOHNS HOPKINS
UNIVERSITY

A Randomized Controlled Trial Study of the Once Early-Reading Program in Multiple School Districts: Randomization Brief Summary

Ashley A. Grant, PhD, Amanda J. Neitzel, PhD, Steven M. Ross, PhD

Center for Research and Reform in Education
Johns Hopkins University School of Education
2800 N. Charles St
Baltimore, MD 21218
https://education.jhu.edu/crre

INTRODUCTION

Overview of Once Early-Reading Program

Once Early-Reading Program is a tutoring model designed to accelerate early literacy development by leveraging existing school paraprofessionals. Unlike many tutoring programs that rely on external providers or volunteers, Once trains paraprofessionals to deliver daily 30-minute, one-on-one reading sessions, thereby exceeding high-impact tutoring standards (which typically involve three sessions per week). Due to varying start dates, the number of weeks of tutoring will vary between 25 and 35.

Overview of the Evaluation

Once partnered with the Center for Research and Reform in Education (CRRE) to conduct an evaluation of Once's Early-Reading Program in a project funded by Accelerate and Once. The evaluation of the Once Early-Reading Program involves a multisite randomized control trial (RCT) designed to examine the impact of the program on kindergarten students' literacy outcomes in eight elementary schools across three districts during the 2025-26 school year.

The present study uses a randomized control trial to examine these research questions:

Outcome Questions
1. Does the Once Early-Reading Program, delivered by trained paraprofessionals, improve early reading outcomes for kindergarten students compared to standard (business-as-usual) instruction?
2. Do the impacts of the Once program on student literacy outcomes differ by student demographic subgroups (e.g., gender, race/ethnicity, socioeconomic status, English language classification, special education classification, free or reduced-price lunch status, prior achievement)?
3. Are student literacy outcomes associated with implementation factors (e.g., tutoring dosage, session attendance, engagement levels)?

Descriptive Questions
4. What are the demographic and academic characteristics of the kindergarten students participating in the study and the schools where the Once Early-Reading Program is implemented?
5. What are the characteristics of teachers and paraprofessionals delivering instruction in the participating schools?

Implementation Questions
6. What is the scheduled tutoring dosage (sessions, minutes, lessons) for students assigned to the Once program, and what is the actual dosage achieved?
7. How do tutoring dosage and student engagement vary by student characteristics

(e.g., gender, race/ethnicity, English learner status, special education classification, socioeconomic status), or school?

8. What is the nature of literacy instruction provided to control group students (the counterfactual condition), and how does it differ from the instructional experiences of treatment group students (treatment contrast)?

Cost & Cost-Effectiveness Questions

9. What are the total and per-student costs of implementing the Once Early-Reading Program in participating schools?

10. What is the cost-effectiveness of the program, expressed as the cost per standard deviation of literacy improvement achieved?

This randomization brief documents the results from the randomization process of the study, which is relevant for the impact analyses related to research questions 1-3. In the sections that follow, we provide information about how students were randomized to groups and assess the baseline equivalence between groups. Impact findings will appear in future reports.

METHOD

Research Design

During the 2025-26 school year, 8 schools in Indiana and Missouri are taking part in a randomized control trial to evaluate the impact of the Once Early-Reading Program on literacy outcomes for students. This study is designed to meet the Every Student Succeeds Act (ESSA) Tier 1 standards for "strong" evidence, as well as the What Works Clearinghouse (WWC) standards "without reservations." At all schools, students were identified who were the lowest achieving students in kindergarten. These students were randomly assigned within schools to participate in the Once Early-Reading Program, and thus to receive daily, in-person, one-on-one reading tutoring from paraprofessionals in the students' elementary schools (treatment), or to continue with their business-as-usual classroom instruction and literacy supports (control). Following an intent to treat (ITT) approach, students were included in their assigned group, regardless of whether or not they received tutoring.

Within each of the eight schools participating, one paraprofessional tutors the treatment kindergarten students, while the control students in each school continue with business-as-usual instruction. The study will estimate the program's impacts on student literacy outcomes, including early reading assessments already in use within each district (e.g., i-Ready, Acadience Reading), measured from the beginning to the end of the school year.

In addition to examining confirmatory impacts on student achievement, the study will address exploratory questions about dosage and subgroup effects. A teacher survey

will document the nature of literacy instruction in control classrooms, providing evidence on the treatment-control contrast. Finally, a cost analysis will estimate total and per-student program costs, as well as the cost-effectiveness of the model in improving early literacy skills. Findings will inform both policymakers and practitioners seeking scalable, sustainable approaches to delivering high-impact tutoring.

Participants

This evaluation includes 442 kindergarten students recruited by the grantee in 8 elementary schools across 3 districts within two states (Indiana and Missouri). Details about study participants are presented below.

3 districts	8 schools	8 tutors	442 Kindergarten students
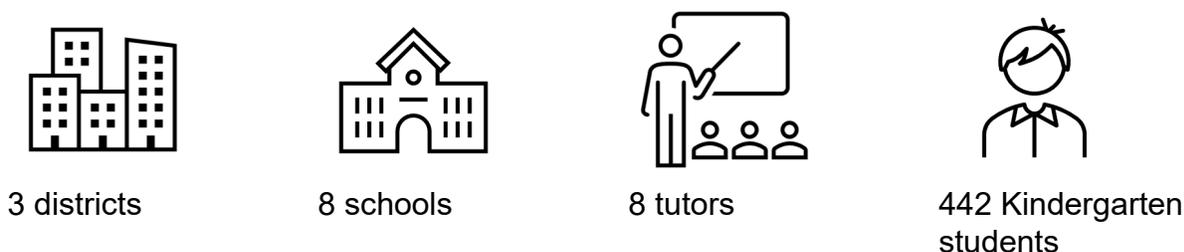
Figure 1
Demographic Snapshot of Student Participants[1]

Race / Ethnicity

- 33% Black
- 33% Hispanic
- 27% White
- 24% Other/Mixed

Characteristics

English learners
Female

0%   20%   40%   60%   80%   100%
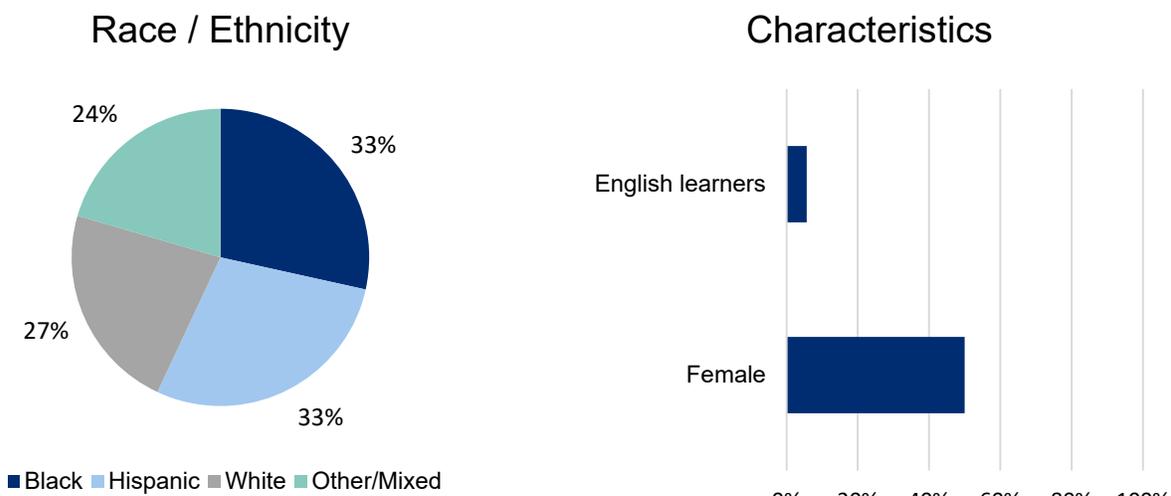
The 442 students who were deemed eligible to participate were randomized to either

---

[1] Data from districts.

the Once Early-Reading Program (n = 184) or the business-as-usual control condition (n = 258). All of these students had baseline data and were included in the analytic sample for baseline analyses. Demographics of the analytic sample are presented in Table 1.

Table 1
Demographics of Analytic Sample

		All students %	Treatment %	Comparison %
Gender	Female	50.2	51.1	49.6
Race/Ethnicity	White	26.5	34.1*	21.0
	Black	33.4	33.0	33.7
	Hispanic	33.4	27.5	37.7*
	Other or mixed race/ethnicity	24.0	22.5	25.0
ELL		5.7	6.0	5.4
Grade	Kindergarten	100	100	100
	n	442	184	258

Notes. 1. ELL = English language learners. 2. Data on student socioeconomic status and special education status was not included due to large amount of missing data on these characteristics between districts (78% and 33% respectively). 3. *p* < .05.

Overall, treatment and control groups were similar across background characteristics. However, there were statistically significant differences (*p* < .05) between groups on race/ethnicity: treatment students were more likely to be White than control students and control students were more likely to be Hispanic than treatment students.

Measures

To assess the results of the randomization process, the districts provided administrative and test data from the baseline time point, fall 2025.

Early-literacy achievement. The confirmatory primary outcome will be each student's end-of-year (EOY) reading/early-literacy score on the district-adopted assessment, measured in spring of 2025–26 (district testing window, typically Apr–Jun). Therefore, baseline measures of literacy achievement were used as the pretest measure for reading, to assess student literacy achievement prior to any intervention. Two different assessments of early literacy were used among the districts involved in the study (see below). To account for variation across assessments, baseline scores were standardized (z-scores), with dummy variables included for assessment type. These student achievement data were collected by Once from the school district.

- **i-Ready Reading Diagnostic.** The i-Ready Diagnostic assessment is a test that measures multiple foundational reading skills. Computer-adaptive, the assessment adapts its difficulty based on the student's responses, which helps to

accurately determine their reading levels and specific learning needs. This assessment is designed to be used with grades K-8, though in kindergarten, the assessment focuses on early literacy skills. Scores are vertically scaled and nationally normed across grades, ranging from 0-800. This test can be given multiple times a year to show student growth. The overall scale score from fall 2025 was used for baseline equivalence analyses.

- **Acadience (formerly DIBELS Next).** The Acadience Reading assessment was designed for formative assessment and measures basic early literacy skills. It is administered individually through brief, efficient measures, in order to provide teachers with immediate feedback on student performance. It is given up to three times a year, during the fall, winter, and spring, to students in grades K-1. The reading composite score from fall 2025 was used for baseline equivalence analyses.

Covariates. The program collected additional data from school districts, including demographics (race/ethnicity, gender, English learner status, special education status, and economic disadvantage) and classroom or teacher assignment.

Procedure

After recruitment of districts and schools, each school identified the lowest achieving students in kindergarten for participation in the study. This list of students was shared with Once.

Randomization Process. After eligible students were identified, they were randomized at the student-level to the tutoring treatment group or the business-as-usual control group. Randomization was conducted on a rolling basis as schools enrolled by program personnel at Once using the *IndependentRandomizer* app developed by Amanda J. Neitzel. This app allows for transparent and reproducible randomization while maintaining data security. Once then shared the deidentified results with the evaluation team for evaluation and confirmation.

In this study, randomization proportions were determined with a tiered approach based on the number of eligible students. This more complex approach helped to ensure consistent treatment assignment proportions across schools while also providing flexibility to schools to account for differences in school size, student availability, and intervention capacity. We used a three-pronged approach designed to maximize study power while accommodating variation in school-level student eligibility and tutoring capacity.

1) First, in standard cases, 20 students were randomly assigned to tutoring and the remainder to the business-as-usual control condition. This approach resulted in a larger control group, as shown in Table 2. This "whole school" approach was used in two of the eight schools.

2) Second, in schools where fewer than 50 students were available, we used pair-matched randomization to create up to 25 student pairs. Within each pair, one student was assigned to tutoring and one to control, ensuring balance even with smaller samples. This approach, labeled "pairs", was used in four of the eight schools.

3) Third, in schools with more than 40 eligible students, we randomly selected students for tutoring in accordance with the school's tutoring capacity, potentially selecting more than 20 students for participation in the program to maximize services. Typically, this meant assigning 24 students to tutoring, with the remaining students assigned to control. This "whole school plus" approach was used in two of the eight schools.

Table 2 shows the results of the random assignment of students overall. It also describes the approach to stratification used in each district, along with randomization results by district.

Table 2
Results of the Student-Level Random Assignment

Number of strata	Students
Treatment students *n*	184
Control students *n*	258
Total *N*	442

In the next section, we assessed the success of randomization at achieving baseline equivalence between the treatment and control groups.

Analytical Approach

All statistical analyses were completed using *Stata v.18.0.* To estimate the impact of Once Early-Reading Program on student literacy outcomes, we will use ordinary least squares (OLS) regression models with covariate adjustment. We will analyze impacts at the student level given that students' treatment condition was determined at the student level for the RCT. Further, randomization of students was conducted within strata within each school, so models will include blocking variables from the random assignment process. This approach provides unbiased estimates of treatment impacts while accounting for stratification used in the randomization.

Baseline Equivalence Assessment. We assessed baseline equivalence on students' early literacy as measured by the i-Ready Diagnostic or Acadience reading assessments. To adjust for the differences in scoring across these assessments, we standardized students' scores to create z-scores. We did this by subtracting students' individual scores from the district mean on a particular assessment and then divided that by the standard deviation of the district mean score, effectively translating students'

scores into standard deviation units. We estimated the baseline treatment-control difference using a multiple linear regression model with the same structural components we will use in the final impact model (including the treatment group indicator and random error terms but without other covariates) but where the pre-test is the dependent variable instead of an independent variable. To account for the differences in variation between schools, we use cluster robust standard errors. The coefficient for the treatment indicator thus estimates the treatment-control difference on the pre-test measure.

$$Pretest_i = \beta_0 + \beta_1 Treat_i + \beta_k RandomizationBlocks_i + \beta_l AssessmentType_i + r_i$$

where $\textbf{\textit{Pretest}}_i$ is the standardized pretest scale score (e.g. student achievement in reading) for student i, $\boldsymbol{\beta_0}$ is the covariate-adjusted grand mean test score for the control group; $\boldsymbol{\beta_1}$ is the average treatment effect, or difference between treatment and control groups on the pretest; $\textbf{\textit{Treat}}_i$ is the binary treatment indicator for student i, $\boldsymbol{\beta_k}$ is a vector of regression coefficients for blocking variables used in randomization, $\textbf{\textit{RandomizationBlocks}}_i$ is a vector of blocking variables used in randomization, including the strata within each school in which students were randomized, $\boldsymbol{\beta_l}$ is a vector of regression coefficients for the assessment type, and $\textbf{\textit{AssessmentType}}_i$ is a vector of dummy variables denoting the specific assessment used, and r_i is the student-level residual. Dummy variables for assessment type were included to control for structural differences in scoring and scaling. This approach ensures comparability while preserving the integrity of site-specific data.

The full analysis will additionally incorporate student demographic covariates, such as English language learner status, free and reduced lunch status, gender, race/ethnicity, and special education status.

Table 3 shows the results from baseline equivalence analyses of the standardized pretest score on early literacy in Fall 2025. In the full sample, treatment and control students demonstrated baseline equivalence, with a standardized mean difference of 0.07 SDs, well below the What Works Clearinghouse (2022) threshold of 0.25 standard deviations.

Table 2
Baseline Equivalence on Standardized Early Literacy Scores for Treatment and Control Students, Overall and by District

Analytic Sample	Treatment			Comparison			Treatment-Comparison Difference
	n	*M*	*SD*	*n*	*M*	*SD*	*Effect Size*
All	184	-0.420	1.034	258	-0.487	0.972	.067

Note. 1. Effect sizes are experimental minus control means divided by the control group standard deviation.

Revised Power Analysis. Power analyses were conducted using PowerUp (Dong et al., 2013). Table 4 shows the original power analysis with the anticipated sample size of 400 and the specifications (see below) resulted in an MDES of 0.236 SDs.

Table 4
Power Analysis Parameters and Resulting MDES

Variable	Original power analysis	Updated
Cluster size or Sample size	400	442
Alpha	0.05	0.05
#-tailed test	2	2
Power	0.8	0.8
%treatment	0.66 (160T/240C)	0.42 (184/258)
Covariates R^2	0.4	0.62
MDES	.236	.165

Note.

The updated power analysis, given the final sample size of 442, revealed the potential to detect an even smaller program effect size (0.165), providing further confidence in our final analysis plan's ability to detect evidence of program effectiveness.

DISCUSSION

The purpose of the present randomization brief was to document the randomization process and the results from this process for the evaluation of the Once Early-Reading Program. A sample of 8 schools across 3 districts were successfully recruited for the evaluation. From these schools, 442 students were identified as eligible for participation and randomized at the student-level in strata to treatment (Once Early-Reading Program) or control (business as usual) conditions. This randomization process produced two equivalent groups (baseline difference *ES* = 0.07), and an updated power analysis of the final sample numbers produced an improved minimum detectable program impact of *ES* = 0.165. Thus, this brief establishes that the current evaluation of the Once Early-Reading Program is set up to successfully detect any impacts of the program absent attrition.

Limitations

At present, there have been few limitations in the randomization process and its results.

REFERENCES

Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24-67.

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