Massachusetts Early Literacy Tutoring Study

Summative Report

Katherine Gross, Daniel Hubbard, Ji Hyun Yang, Aleksandra Holod, Emily Ackman, and Emma Wilson

NOVEMBER 2023

AIR®
Advancing Evidence.
Improving Lives.

dese
MASSACHUSETTS
Department of Elementary and Secondary Education

Contents

Exhibits

Executive Summary

This summative evaluation of the Early Literacy Supplemental Services Tutoring program aimed to gather information on program models and predictors of student growth. The information in this evaluation came from data made available to the evaluation team for the period from September 2022 to June 2023.

- The evaluation analysis included 2,364 students, who received a total of 89,755 sessions of tutoring. Tutoring services were evaluated in 67 high-need elementary schools across 14 districts.
- Across all vendors, 84% of students included in the analysis of learning outcomes scored either well below or below benchmark on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) reading assessment at the beginning of tutoring services.
- Students served by all four tutoring vendors saw statistically significant gains in literacy outcomes, according to statistical models, with Ignite and Springboard performing particularly well.
- Within each DIBELS subscale, students with lower pretest scores on that subscale showed higher growth on that subscale at posttest, suggesting that tutoring services were effective in addressing students' greatest areas of weakness.
- Students who received tutoring services met expectations for typical grade level DIBELS score growth within the study observation window. On average, students' composite scores on the DIBELS increased by 18 points over a 3- to 4-month span, corresponding to about half of the growth expected on the DIBELS over a full 9-month school year, or students who participated in 14 weeks of tutoring gained 20 weeks of learning on DIBELS outcomes. Catapult students' scores increased by 14 points, on average; and Springboard, Ignite, and Literacy Lab students' scores increased by 19 points, on average.
- Between the baseline and follow-up DIBELS assessments, the proportion of Catapult students well below or below benchmark decreased by 15%; the proportion of Ignite students well below or below benchmark decreased by 38%; the proportion of Literacy Lab students well below or below benchmark decreased by 12%; and the proportion of Springboard students well below or below benchmark decreased by 38%.

Introduction

Overview of Early Literacy Tutoring

Only 56% of Massachusetts third graders were meeting proficiency targets for English language arts prior to the COVID-19 pandemic (Massachusetts Department of Elementary and Secondary Education [DESE], 2019). In addition, only 38% of the state's Black and Latinx students had reached proficiency in reading by third grade. School closures during the pandemic intensified the urgency of addressing students' literacy learning needs. The percentage of students meeting expectations decreased from 56% in 2019 to 51% in 2021 (DESE, n.d.). In response, DESE created the Early Literacy Supplemental Services Tutoring program to address the need for early literacy support, using funding from the Governor's Emergency Education Relief (GEER) Fund.

The Early Literacy Supplemental Services Tutoring program acknowledges that remote instruction during COVID-19–related school closures led to learning opportunity gaps. Some students now need additional support to meet grade-level reading benchmarks. A recent meta-analysis indicated that literacy tutoring is effective for early elementary-grade students, and even tutoring provided by nonprofessionals can have significant positive impacts (Nickow et al., 2020). DESE engaged several tutoring vendors who implemented different tutoring program models in elementary schools across districts in Massachusetts.

Overview of the Evaluation

DESE contracted with the American Institutes for Research® (AIR®) for a second year to conduct another formative and summative evaluation of the Early Literacy Supplemental Services Tutoring program, spanning the period from September 2022 to June 2023. This report focuses on findings from the summative evaluation, which incorporated quantitative data sources and analytic methods to examine the learning outcomes of participating students.

About the Study

The goal of the study was to evaluate how variations in tutoring service delivery relate to the literacy outcomes of a diverse group of students. The study used a mixed-methods evaluation design that incorporated both quantitative and qualitative data sources and analytic methods.

AIR conducted a formative evaluation in February 2023 to support improvements in the implementation and delivery of early literacy tutoring services. The formative evaluation used a qualitative design that incorporated findings from interviews and focus groups with program stakeholders—including tutoring program leaders; tutoring program coordinators embedded in districts or schools; principals; tutoring staff; and parents and guardians—as well as a tutor survey, student rosters, tutor rosters, and tutoring logs.

Research Questions

Exhibit 1 lists the research questions (RQs) that guided the summative evaluation. To answer these questions, the study team gathered and analyzed the data sources shown in Exhibit 1.

Exhibit 1. Research Questions and Data Sources

Research questions	Tutoring service logs	Tutoring leader interviews	Tutor interviews	Tutor survey	DIBELS assessment
1. Was there a significant difference between the impact of providers' program models on student outcomes?	✓	✓	✓	✓	✓
2. Which program factors best predicted student growth?	✓	✓	✓	✓	✓
3. Were there significant differences in how students developed specific skills that were impacted by program factors?	✓	✓	✓	✓	✓
4. Which student characteristics best predicted student growth?					✓
5. Were there significant differences in how students developed specific skills that were impacted by student characteristics?	✓	✓	✓	✓	✓
6. What significant interaction effects were found between any of the variables listed above?	✓	✓	✓	✓	✓

Note. DIBELS = Dynamic Indicators of Basic Early Literacy Skills reading assessment.

Methods

Participants

This evaluation focused on four tutoring programs: Catapult Learning (hereafter, Catapult); Ignite! Reading (hereafter, Ignite); Literacy Lab; and Springboard Collaborative (hereafter, Springboard). As part of being funded by the state, the four participating vendors participated in data collection activities, which the AIR study team designed to gather data to address the research questions. The study team focused on the students served by the four tutoring vendors between September 2022 and June 2023. This observation window only provided a

snapshot of the ongoing Early Literacy Supplemental Services Tutoring program, which is currently in its third year of operation. For this report, AIR evaluated all data provided by the four vendors from September 2022 through June 10, 2023. Data submitted to the study team after June 10 are not reflected in this report.

The summary of services sample (hereafter, "students served") consisted of participating students with both roster data and verifiable records of attendance in the tutoring log data. The analytic sample consisted of participating students with roster data, tutoring log data, baseline (pretest) DIBELS assessment data, and follow-up (posttest) DIBELS assessment data. (See Appendix A for a more detailed description of sample definitions, measures, and analytic methods.)

Student Rosters, Tutor Rosters, and Tutoring Logs. All four participating vendors shared rosters and logs with the AIR study team, which allowed the team to determine how much tutoring each student received and with which tutor. Tutor rosters included tutor names, school names, and email addresses. Student rosters included student names, demographic data, and names of students' schools and districts. The tutoring logs contained dates, tutor names, student names, and school names, providing a record of student attendance for each tutoring session.

Student Assessments. All four participating vendors shared student literacy assessments with the study team and provided both baseline (pretest) and follow-up (posttest) DIBELS assessment data for the students they served.[1] The AIR study team trained tutoring vendors and tutors to administer the DIBELS assessment to students. Tutoring vendors gathered DIBELS data from tutor-administered assessments and from school district data. The student assessments were collected at the start of tutoring to provide a measure of baseline achievement, and again at either the midyear or the end-of-year points, depending on the duration of the tutoring sessions in which students were enrolled, to capture students' literacy learning during their participation in tutoring services. Using the DIBELS assessment data, the study team calculated baseline (pretest) and follow-up (posttest) composite scores for students. Composite scores and benchmarks were calculated based on middle-of-year standards.[2,3]

Tutor Survey. The AIR study team administered a web-based survey to tutors that gathered data on the following topics: tutor training, content and structure of tutoring sessions,

[1] See the vendor-specific summaries for information regarding DIBELS assessment data collection.
[2] See https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2021-10/UO_Dibels8_Scoring_Guide_100121.pdf for more information.
[3] For cases with missing DIBELS subtest scores, the study team created a linear regression imputation model to estimate the missing scores within grade-level samples, using the available subtest scores as predictor variables. In the regression sample, 524 students (42%) had pretest composite scores calculated with imputed subtest data and 340 students (27%) had posttest composite scores calculated with imputed subtest data.

assessments administered to students, differentiation of instruction, coaching and supervision of tutors, and tutors' background. Survey participation was by email, using each tutor's contact information provided in the tutor rosters. Tutors were initially invited to participate on December 15 and received up to three reminder emails between January 10 and January 23 if they had not yet completed the survey. Participants who completed the tutor survey during the first administration received a $15 Target gift card after completing the survey. Tutors who did not complete the survey received several reminders regarding the opportunity to participate and receive the gift card.

The study team reopened the survey in the spring, on April 6, for new tutors. New tutors were added on a rolling basis, with up to three weekly reminder emails sent between April 6 and April 30. Participants who completed the survey during the second administration did not receive an incentive following survey completion. Across all four vendors, 105 tutors completed the survey, for a response rate of 31%. Exhibit 2 presents the survey sample size and response rate for each vendor, as well as for the full study population.

Exhibit 2. Survey Sample Sizes and Response Rates

Vendor	Surveys sent	Surveys completed	Response rate
Catapult	142	27	19%
Ignite	96	39	41%
Literacy Lab	18	11	61%
Springboard	79	28	35%
All	**335**	**105**	**31%**

Analytic Approach. The study team used a fuzzy string matching algorithm to probabilistically match the names of students served across the roster, tutoring log, and DIBELS assessment data files. We then conducted descriptive tabulations of these program data to determine the number of students and communities served, the demographic characteristics of participating students, and the magnitude of student learning gains. We followed this descriptive analysis with a series of regression analyses to answer questions about the associations between program factors, student demographic characteristics, and student learning outcomes. Students whose primary tutors did not complete the survey were included in the regression sample, with the survey variables set to zero and an indicator for "survey missing" set to 1. See Appendix A for more information on the variables, their sources, and their measurement properties.

Report Organization

Findings are presented in aggregate and then at the vendor level, followed by sections discussing the findings and the limitations of the study. The findings are organized by RQs.

Summary of Services

In this section, we report the total volume of services delivered to the participant sample of 2,364 students from September 2022 to June 2023 by the four tutoring vendors who participated in the evaluation. We describe the population of students served. We then present summaries for each vendor, describing their program models and the students they served.

Tutoring Services Delivered

The evaluation included a participant sample of 2,364 students, served in 67 schools and 14 districts across the Commonwealth between September 2022 and June 2023. Catapult accounted for 35% of students served, with supplementary services provided at 28 schools across five districts. Ignite accounted for 23% of students served, with supplementary services provided at 13 schools across six districts. Literacy Lab accounted for 9% of students served, with supplementary services provided at 11 schools across two districts. Lastly, Springboard accounted for 33% of students served, with supplementary services provided at 18 schools across four districts. It should be noted that some vendors provided concurrent supplemental services within the same districts during the observation window. For example, Catapult and Ignite both provided services to students in Worcester and Randolph, and Catapult and Springboard both provided services to students in Boston.

Collectively, according to tutoring log data, the four tutoring vendors provided 89,755 tutoring sessions in total during the course of the study. The length of individual tutoring sessions ranged from 15 minutes to 45 minutes, depending on vendors' service delivery models. Each student participated in an average of 37.97 tutoring sessions during the study observation window and an average of 2.43 tutoring sessions per week.

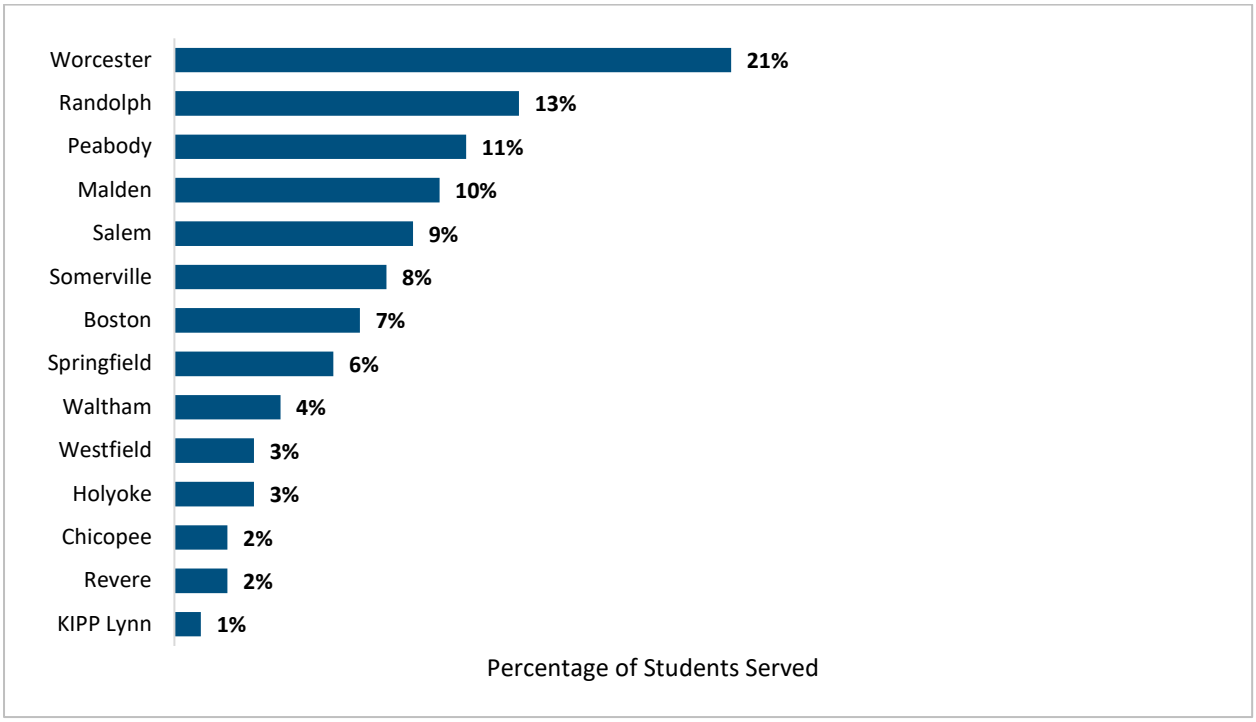
Exhibit 3 presents a snapshot of tutoring provided across the four vendors, including the students and communities served and the tutoring dosage.

Exhibit 3. Snapshot of Tutoring Provided Across the Four Vendors

Participating Students	Schools	Districts
2,364	67	14
Total Number of Tutoring Sessions	**Number of Tutoring Sessions Per Student**	**Number of Tutoring Sessions Per Student Per Week**
89,755	37.97	2.43

The greatest proportion of students included in the study (21%) attended school in Worcester, followed by Randolph (13%), Peabody (11%), Malden (10%), Salem (9%), Somerville (8%), Boston (7%), and Springfield (6%). Waltham (4%), Holyoke and Westfield (3% each), Chicopee and Revere (2% each), and KIPP Lynn (1%) were the least represented districts among students served. Exhibit 4 presents a snapshot of the home districts of students served.

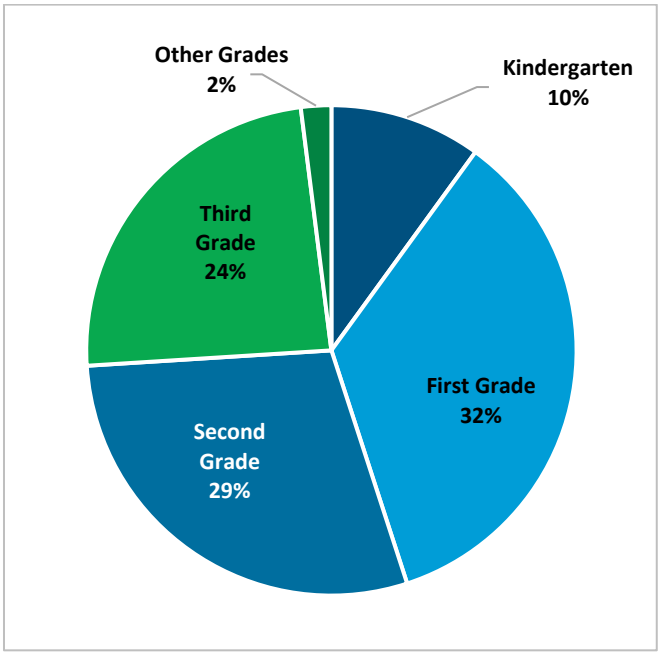
Exhibit 4. Home Districts of Students Served Across the Four Vendors



Note. *N* = 2,364. This count reflects the number of students served across all four vendors who were listed on vendor-provided rosters and matched to tutoring log attendance data.

According to student roster data, 10% of students served across the four vendors were in kindergarten; 35% were in first grade; 29% were in second grade; 24% were in third grade; and 2% were in grades outside the K–3 range, categorized as "Other" for the purposes of reporting. Exhibit 5 presents a snapshot of the grade levels of students served across all four vendors.

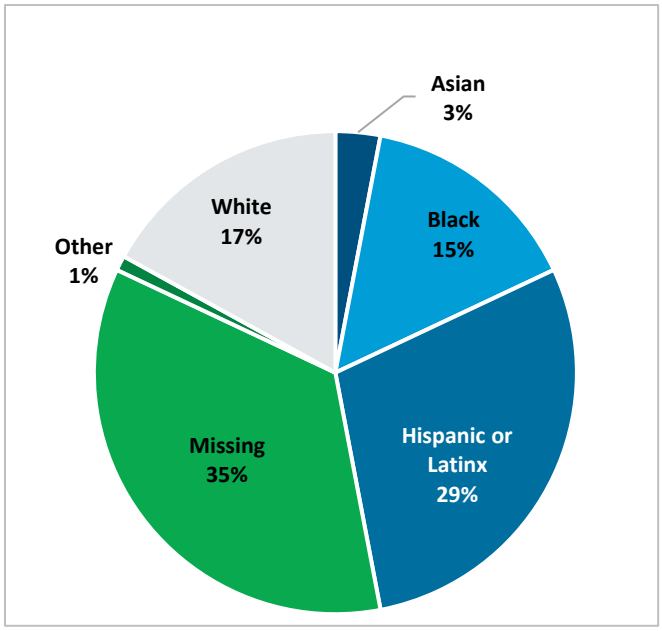
Exhibit 5. Grade Levels of Students Served Across the Four Vendors



Note. *N* = 2,364. This count reflects the number of students served across all four vendors who were listed on vendor-provided rosters and matched to tutoring log attendance data.

According to student roster data, 29% of students served across the four vendors were reported as Hispanic or Latinx, 17% were reported as White, 15% were reported as Black, 3% were reported as Asian, and 1% were reported as "Other." Less than 1% each of students served were reported as American Indian or Alaska Native, Native Hawaiian or Pacific Islander, and multiracial. Approximately 35% of students served did not have race and ethnicity data available for analysis. Exhibit 6 presents a snapshot of the demographic characteristics of students served across all four vendors.

Exhibit 6. Race/Ethnicity of Students Served Across the Four Vendors



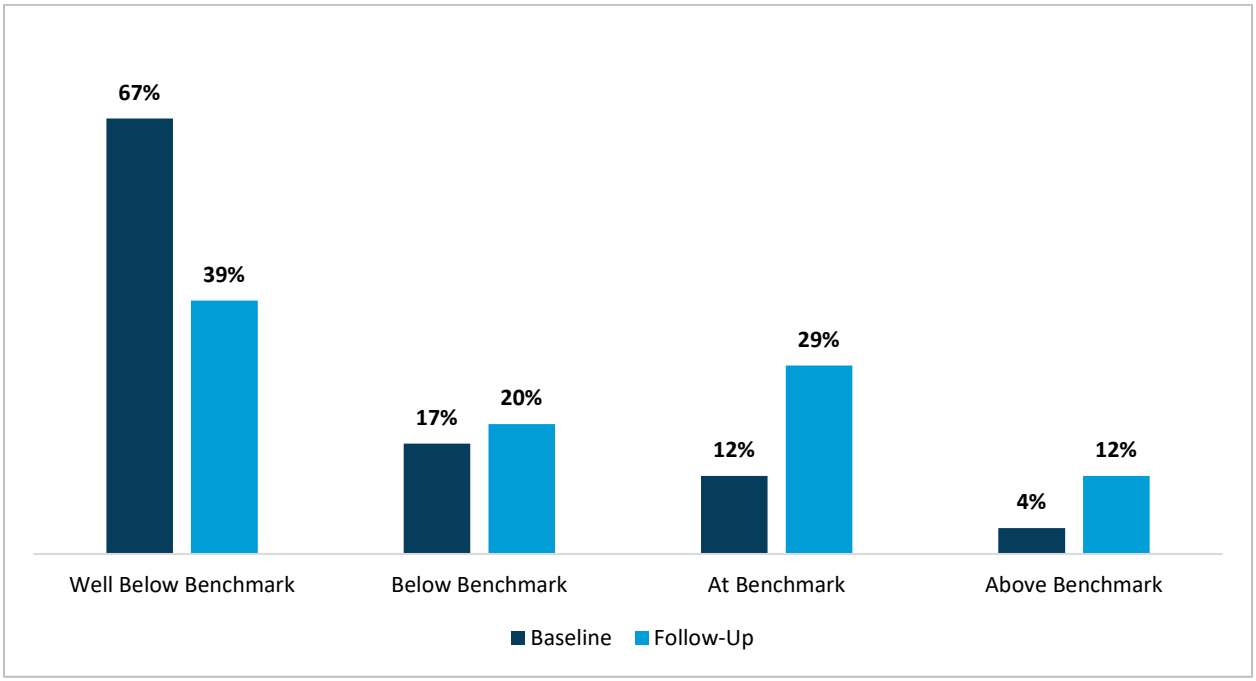
Note. *N* = 2,364. This count reflects the number of students served across all four vendors who were listed on vendor-provided rosters and matched to tutoring log attendance data.

Of the students included in our analysis of learning outcomes, 67% scored well below benchmark on the DIBELS assessment at the beginning of tutoring services, 17% scored below benchmark, 12% scored at benchmark, and 4% scored above benchmark. Students who score well below benchmark are classified as "at risk" for not meeting grade-level reading proficiency goals. Students who score below benchmark are classified as "some risk," students who score at benchmark are classified as "minimal risk," and students who score above benchmark are classified as "negligible risk."

Within the same sample of students, 39% scored well below benchmark on the follow-up midyear DIBELS assessment, 20% scored below benchmark, 29% scored at benchmark, and 12% scored above benchmark.

Exhibit 7 presents the baseline and follow-up DIBELS reading levels of students included in the analytic sample.

Exhibit 7. Baseline and Follow-Up DIBELS Reading Levels of Students Served Across the Four Vendors



Note. *N* = 1,259. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. This count reflects the number of students listed on the rosters provided by the four vendors, who were matched to tutoring log attendance data and had both pretest and posttest DIBELS composite scores available.

Catapult

Catapult (catapultlearning.com), a national tutoring company founded in 1976, uses a small-group instructional model that includes a diagnostic and prescriptive approach. Catapult tutors use formative assessments to identify students' literacy skills and needs, provide direct instruction to students, and incorporate structured lesson plans into tutoring sessions.[4] The Catapult implementation model is flexible because it allows sites to adapt how tutoring is implemented at their school to suit their needs. For example, sites can choose to offer tutoring during the school day, before school, or after school. In terms of tutors' backgrounds and qualifications, Catapult aims to hire credentialed teachers and paraprofessionals from the local community.

Catapult works closely with its partner schools, which identify and enroll students in the program. A Catapult supervisor coordinates the program for two or more districts, oversees management of the program at each participating school, and supervises all tutors in the participating districts. According to program staff, students served by Catapult received an estimated 40 minutes of tutoring per session.

[4] For more information about Catapult's instructional model, see https://catapultlearning.com/programs/instruction/.

Overview of Catapult Tutoring Services

Over the course of the study, Catapult tutors provided a total of 30,380 tutoring sessions to 828 students at 28 schools across five districts: Boston, Chicopee, Malden, Salem, and Worcester. Each student participated in an average of 36.69 tutoring sessions during the study observation window and an average of 2.17 tutoring sessions per week (Exhibit 8).

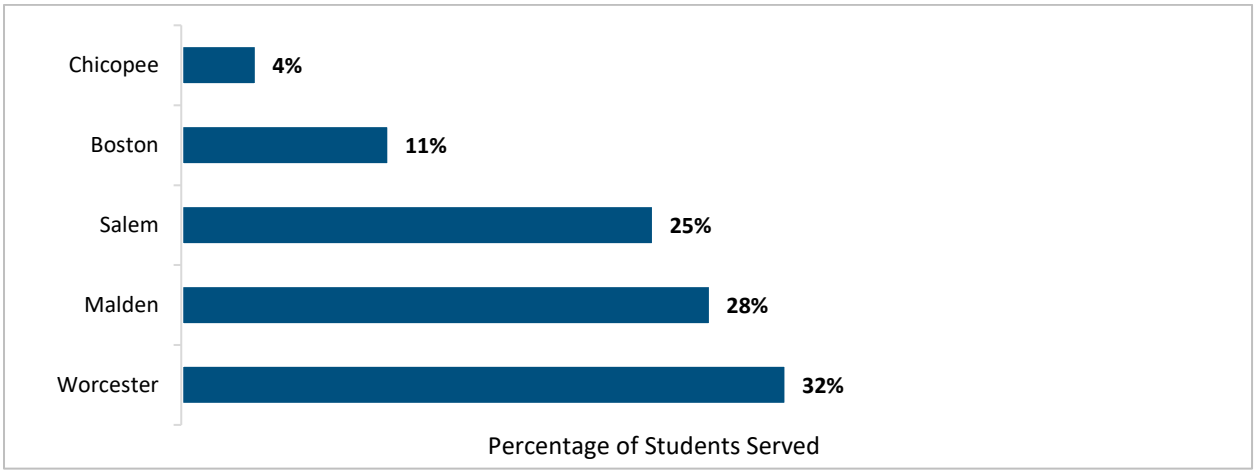
Exhibit 8. Snapshot of Catapult Tutoring Services Included in This Study

Participating Students	Schools	Districts
828	28	5
Total Number of Tutoring Sessions	**Number of Tutoring Sessions Per Student**	**Number of Tutoring Sessions Per Student Per Week**
30,380	36.69	2.17

Communities Served by Catapult

According to student roster and tutoring log data provided by Catapult, 32% of students included in the study attended school in Worcester, 28% attended school in Malden, 25% attended school in Salem, 11% attended school in Boston, and 4% attended school in Chicopee. Exhibit 9 provides a snapshot of the home districts of students served by Catapult.

Exhibit 9. Home Districts of Students Served by Catapult

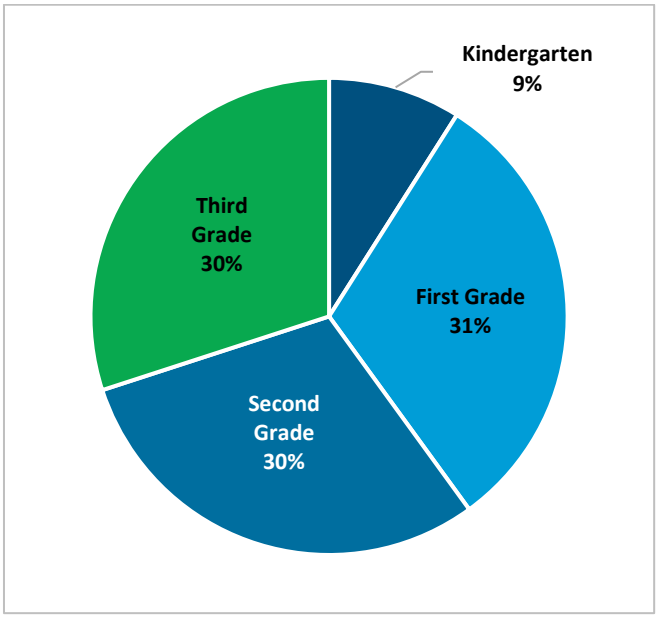


Note. *N* = 828. This count reflects the number of Catapult students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Characteristics of Students Served by Catapult

According to student roster and tutoring log data, 9% of students served by Catapult were in kindergarten, 31% were in first grade, 30% were in second grade, and 30% were in third grade. Less than 1% of students served by Catapult were in grades outside the K–3 range, categorized as "Other" for reporting purposes. Exhibit 10 presents a snapshot of the grade levels of students served by Catapult.

Exhibit 10. Grade Levels of Students Served by Catapult



Note. *N* = 828. This count reflects the number of Catapult students listed on the vendor-provided roster who were matched to tutoring log attendance data.

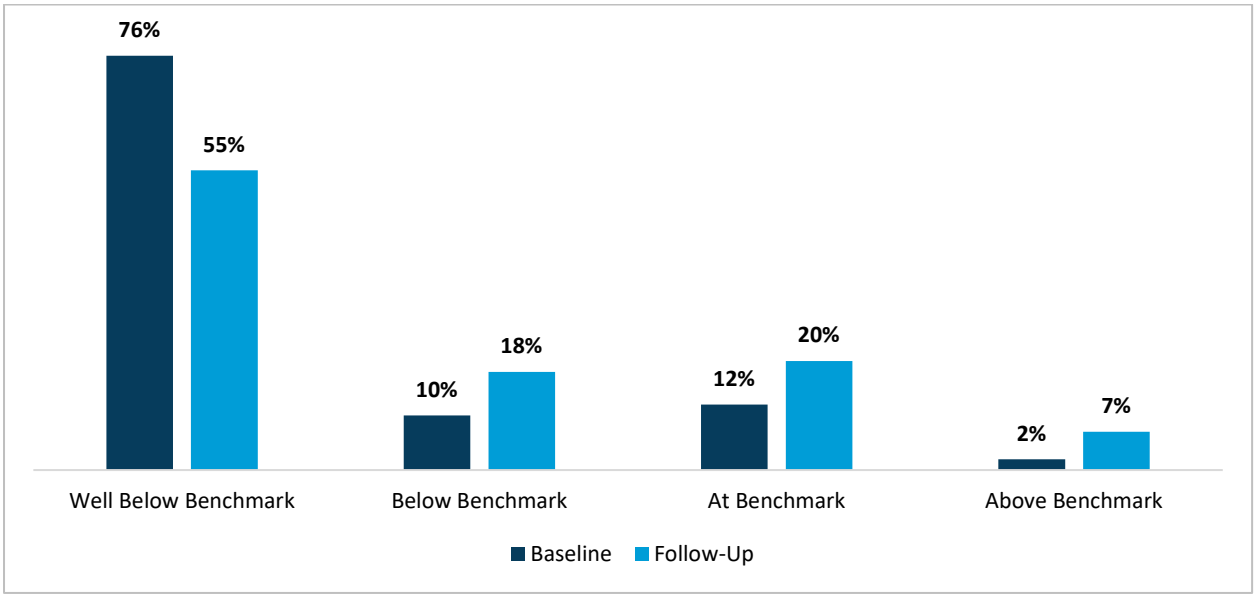
Catapult provided both baseline (pretest) and follow-up (posttest) DIBELS data for students served in all five districts: Boston, Chicopee, Malden, Salem, and Worcester. Of the 828 total students served by Catapult (students listed on the roster who received at least one tutoring session during the study observation window), 368 (44%) had sufficient data available to calculate both pretest and posttest DIBELS composite scores, and were therefore included in our analysis of learning outcomes.[5]

Of the Catapult students included in our analysis of learning outcomes, 76% scored well below benchmark on the baseline DIBELS assessment, 10% scored below benchmark, 12% scored at benchmark, and 2% scored above benchmark.

[5] DIBELS scoring requirements specify that complete subscale score data are required for composite score calculation. See https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2023-02/UO_Dibels_8_Scoring_Guide_2023.pdf for more information.

Within the same sample of Catapult students, 55% scored well below benchmark on the follow-up DIBELS assessment, 18% scored below benchmark, 20% scored at benchmark, and 7% scored above benchmark. Exhibit 11 presents the baseline and follow-up DIBELS reading levels of students served by Catapult.

Exhibit 11. Baseline and Follow-Up DIBELS Reading Levels of Students Served by Catapult



Note. *N* = 368. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. This count reflects the number of Catapult students listed on the vendor-provided roster who were matched to tutoring log attendance data and had both pretest and posttest DIBELS composite scores available. Approximately 56% of students served by Catapult did not have sufficient data available to calculate both pretest and posttest DIBELS composite scores.

According to available student roster data, 10% of students served by Catapult were reported as Hispanic or Latinx, 4% were reported as Black, and 1% were reported as White. Less than 1% of students were reported as Asian, and less than 1% were reported as multiracial. Eighty-five percent of students served by Catapult did not have race and ethnicity data available for analysis.

Ignite

Ignite (https://ignite-reading.com), a California-based tutoring company founded in 2020, focuses on high-fidelity implementation of a foundational reading skills curriculum. Ignite's mission is to ensure that every student is an independent reader at the start of second grade.

Ignite uses a virtual, one-on-one model that employs an evidence-based curriculum for tutoring. Ignite tutoring is provided for 15 minutes a day and occurs during the school day.

Ignite uses diagnostic tools to determine students' skill level and areas where they need more support. Ignite then works with schools to determine a program for students.

Ignite provides a 10-week certification period for their tutors, during which tutors receive 60 hours of professional learning and coaching. During the certification period, tutors work with four to six students so they can apply what they learn in real time. After training, tutors work with as many students as their schedule allows. Tutors are given feedback, including areas of growth, to strengthen lesson delivery.

Ignite gives full discretion to schools to decide which students would benefit from tutoring. The number of students participating in the program depends on a variety of factors, including tutors' schedules, the school schedule, and available resources at the school (e.g., computers, headphones).

Overview of Ignite Tutoring Services

During the study observation window, Ignite tutors provided a total of 32,832 tutoring sessions to 553 students at 13 schools across six districts: Lynn (KIPP Academy), Randolph, Revere, Waltham, Westfield, and Worcester. Each student participated in an average of 59.37 tutoring sessions in total during the study observation window and an average of 3.61 tutoring sessions per week (Exhibit 12).

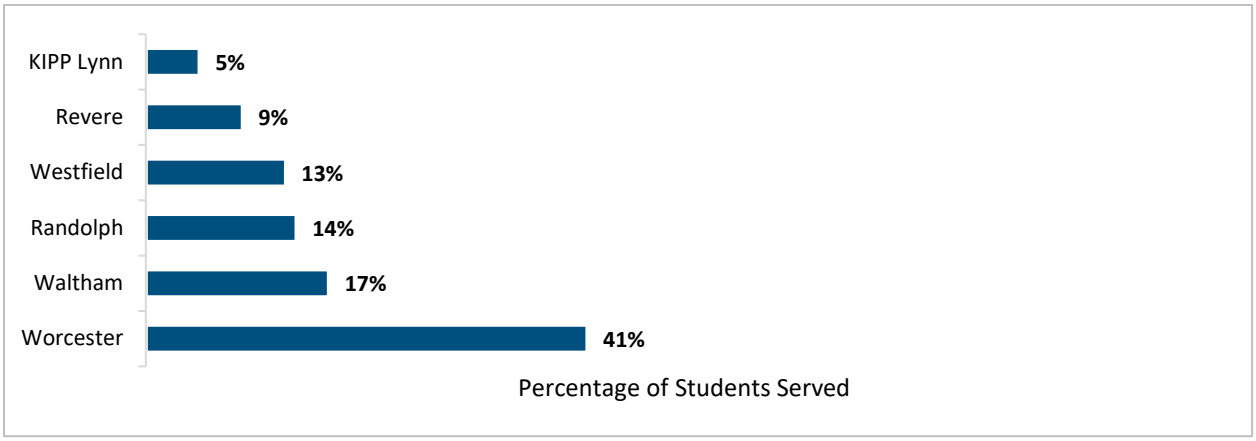
Exhibit 12. Snapshot of Ignite Tutoring Services Included in This Study

Participating Students	Schools	Districts
553	13	6
Total Number of Tutoring Sessions	**Number of Tutoring Sessions Per Student**	**Number of Tutoring Sessions Per Student Per Week**
32,832	59.37	3.61

Communities Served by Ignite

According to student roster and tutoring log data provided by Ignite, 41% of students included in the study attended school in Worcester, 17% attended school in Waltham, 14% attended school in Randolph, 13% attended school in Westfield, 9% attended school in Revere, and 5% attended school at KIPP Academy in Lynn. Exhibit 13 provides a snapshot of the home districts of students served by Ignite.

Exhibit 13. Home Districts of Students Served by Ignite

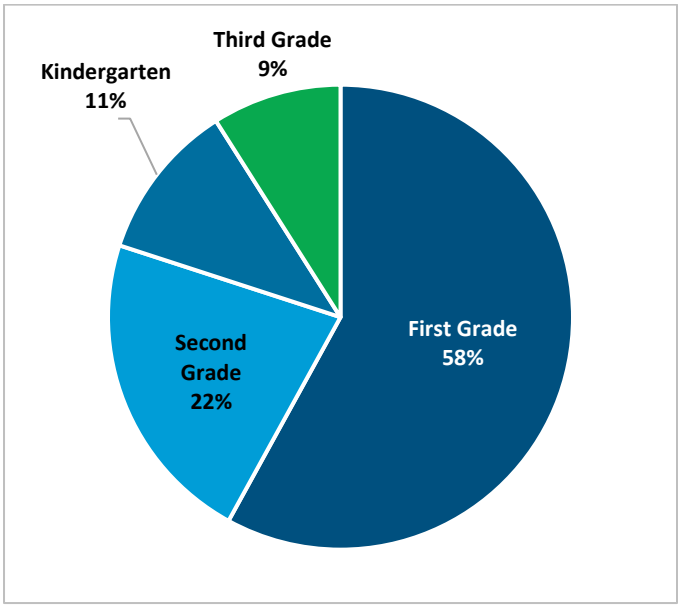


Note. *N* = 553. This count reflects the number of Ignite students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Characteristics of Students Served by Ignite

According to student roster and tutoring log data, 58% of students served by Ignite were in first grade, 22% were in second grade, 11% were in kindergarten, and 9% were in third grade. Exhibit 14 presents a snapshot of the grade levels of students served by Ignite.

Exhibit 14. Grade Levels of Students Served by Ignite



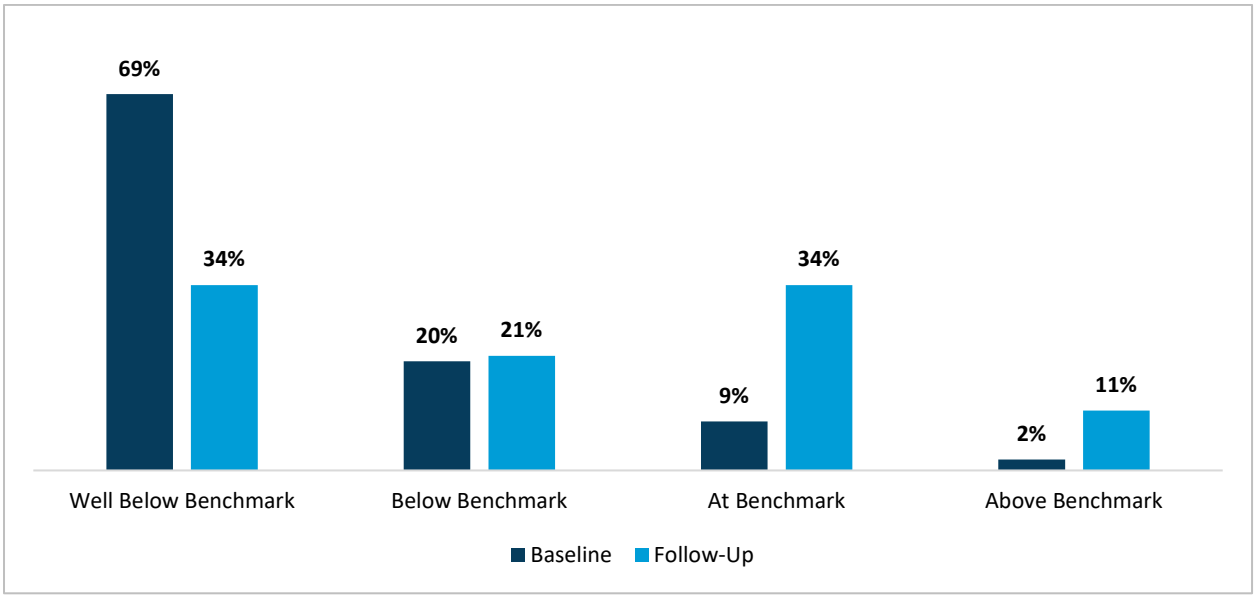
Note. *N* = 553. This count reflects the number of Ignite students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Ignite provided both baseline (pretest) and follow-up (posttest) DIBELS data for students in all six districts: KIPP Lynn, Randolph, Revere, Waltham, Westfield, and Worcester. Of the 553 total students served by Ignite (students listed on the roster who received at least one tutoring session during the study observation window), 353 students (64%) had sufficient data available to calculate both pretest and posttest DIBELS composite scores and were therefore included in our analysis of learning outcomes.[6]

Of the Ignite students included in our analysis of learning outcomes, 69% scored well below benchmark on the baseline DIBELS assessment, 20% scored below benchmark, 9% scored at benchmark, and 2% scored above benchmark.

Within the same sample of Ignite students, 34% scored well below benchmark on the follow-up DIBELS assessment, 21% scored below benchmark, 34% scored at benchmark, and 11% scored above benchmark. Exhibit 15 presents the baseline and follow-up DIBELS reading levels of students served by Ignite.

Exhibit 15. Baseline and Follow-Up DIBELS Reading Levels of Students Served by Ignite

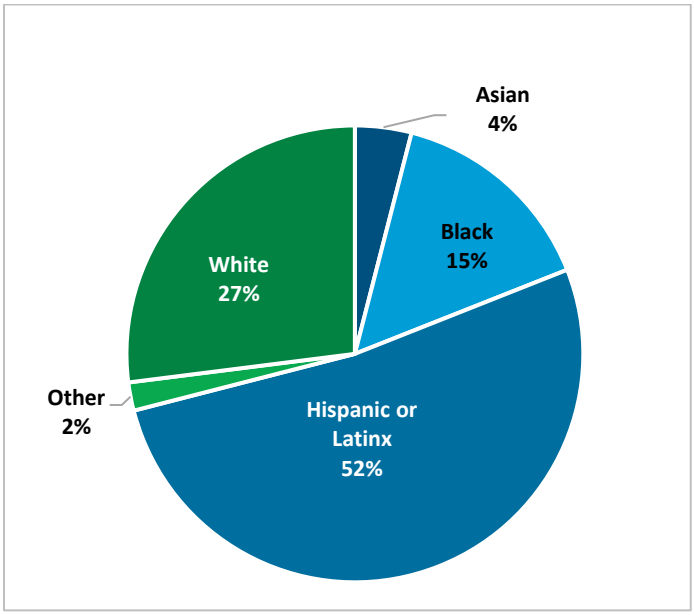


Note. *N* = 353. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. This count reflects the number of Ignite students listed on the vendor-provided roster who were matched to tutoring log attendance data and had both pretest and posttest DIBELS composite scores available. Approximately 36% of students served by Ignite did not have sufficient data available to calculate both pretest and posttest DIBELS composite scores.

[6] DIBELS scoring requirements specify that complete subscale score data are required for composite score calculation. See https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2023-02/UO_Dibels_8_Scoring_Guide_2023.pdf for more information.

According to student roster data, 52% of students served by Ignite were reported as Hispanic or Latinx, 27% were reported as White, 15% were reported as Black, 4% were reported as Asian, and 2% were reported as "Other." Less than 1% each of students were reported as American Indian or Alaska Native, Native Hawaiian or Pacific Islander, and multiracial. Exhibit 16 presents a snapshot of the demographic characteristics of students served by Ignite.

Exhibit 16. Race/Ethnicity of Students Served by Ignite



Note. *N* = 553. This count reflects the number of Ignite students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Literacy Lab

Literacy Lab (https://theliteracylab.org/) offers individualized reading instruction to children, with the goal of closing the literacy gap. Literacy Lab tutors, who are trained as part of an AmeriCorps program, focus on preventing reading failure and ensuring that all students are reading at grade level by third grade. Literacy Lab tutors work full time in elementary schools. After training, tutors provide daily, one-on-one, 20-minute intervention sessions with students in kindergarten through third grade. Tutors provide targeted reading skill practice in the areas of phonemic awareness, phonics, and fluency. Tutoring sessions occur during the school day, outside the student's classroom-based, core-reading instruction time.

Literacy Lab works closely with its partner schools, which identify and enroll students in the program. Each school provides a staff person—a literacy specialist, a coach, or an interventionist—to work as a program coordinator. The person in this role handles

communication with Literacy Lab, schedules students for the tutoring service, and serves as an on-site supervisor for the tutors at the school.

Literacy Lab gives full discretion to schools to decide which students would benefit from tutoring. Schools use a variety of metrics to prioritize students to receive tutoring services, including Standardized Test for the Assessment of Reading (STAR) and DIBELS data, students' access to existing school-based literacy services and support, and teacher recommendations regarding students' skills and prior education.

Overview of Literacy Lab Tutoring Services

During the study observation window, Literacy Lab tutors provided a total of 10,584 tutoring sessions to 203 students at 11 schools across two districts: Holyoke and Springfield. Each student participated in an average of 52.14 tutoring sessions in total during the study observation window and an average of 2.56 tutoring sessions per week (Exhibit 17).

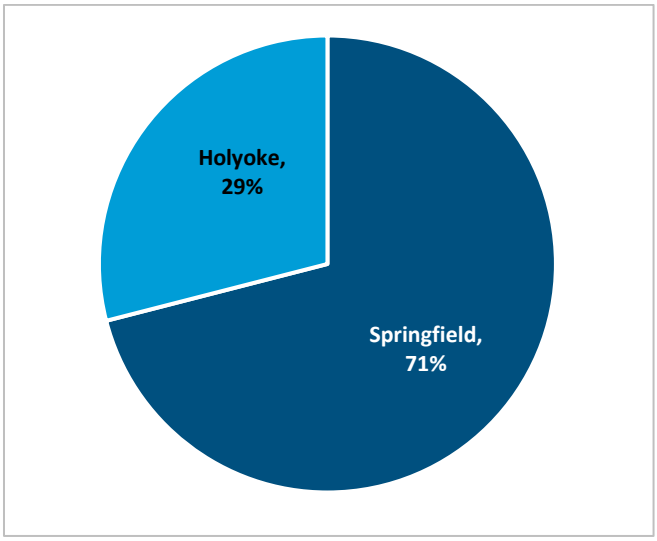
Exhibit 17. Snapshot of Literacy Lab Tutoring Services Included in This Study

Participating Students	Schools	Districts
203	11	2
Total Number of Tutoring Sessions	**Number of Tutoring Sessions Per Student**	**Number of Tutoring Sessions Per Student Per Week**
10,584	52.14	2.56

Communities Served by Literacy Lab

According to student roster and tutoring log data provided by Literacy Lab, 71% of students included in the study attended school in Springfield and 29% attended school in Holyoke. Exhibit 18 provides a snapshot of the home districts of students served by Literacy Lab.

Exhibit 18. Home Districts of Students Served by Literacy Lab

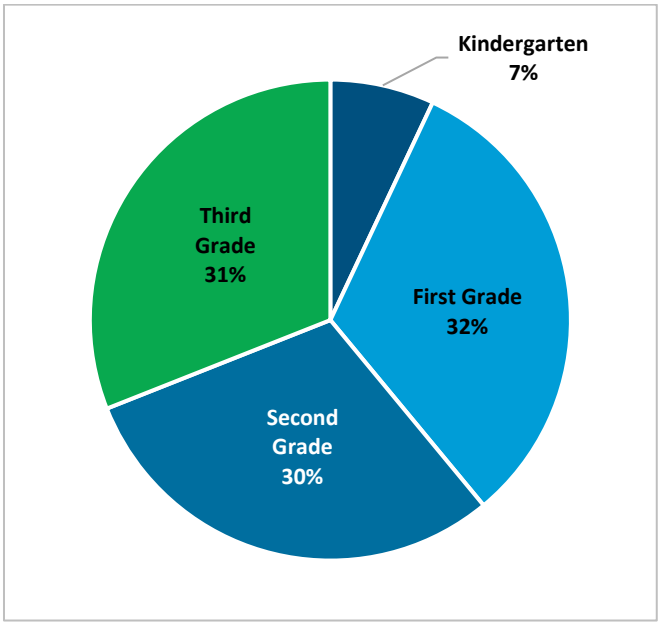


Note. *N* = 203. This count reflects the number of Literacy Lab students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Characteristics of Students Served by Literacy Lab

According to student roster and tutoring log data, 7% of students served by Literacy Lab were in kindergarten, 32% were in first grade, 30% were in second grade, and 31% were in third grade. Exhibit 19 presents a snapshot of the grade levels of students served by Literacy Lab.

Exhibit 19. Grade Levels of Students Served by Literacy Lab

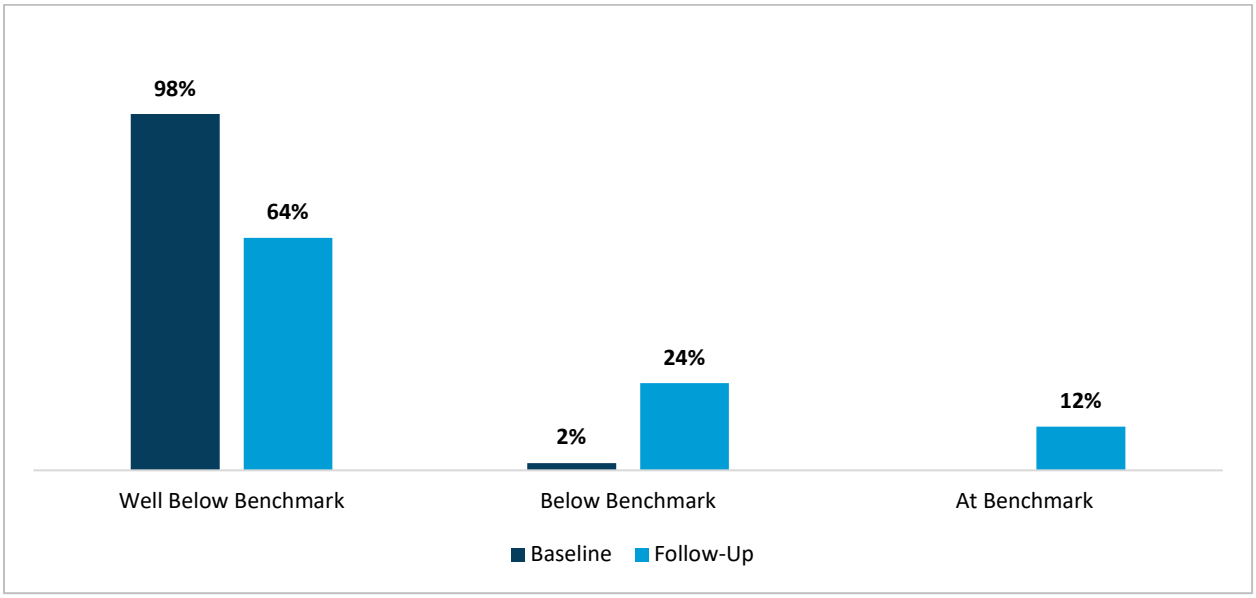


Note. *N* = 203. This count reflects the number of Literacy Lab students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Literacy Lab provided baseline (pretest) and follow-up (posttest) DIBELS data for students served in both Holyoke and Springfield. However, students served at Balliet Elementary in Springfield were missing pretest DIBELS data, and students served at Boland Elementary and Indian Orchard Elementary in Springfield were missing posttest DIBELS data. Of the 203 total students served by Literacy Lab (students listed on the roster who received at least one tutoring session during the study observation window), 58 students (29%) had sufficient data available to calculate both pretest and posttest DIBELS composite scores and were therefore included in our analysis of learning outcomes.[7] Of the Literacy Lab students included in our analysis of learning outcomes, 98% scored well below benchmark on the baseline DIBELS assessment and 2% scored below benchmark.

Within the same sample of Literacy Lab students, 64% scored well below benchmark on the follow-up DIBELS assessment, 24% scored below benchmark, 12% scored at benchmark. Exhibit 20 presents the baseline and follow-up DIBELS reading levels of students served by Literacy Lab.

Exhibit 20. Baseline and Follow-Up DIBELS Reading Levels of Students Served by Literacy Lab

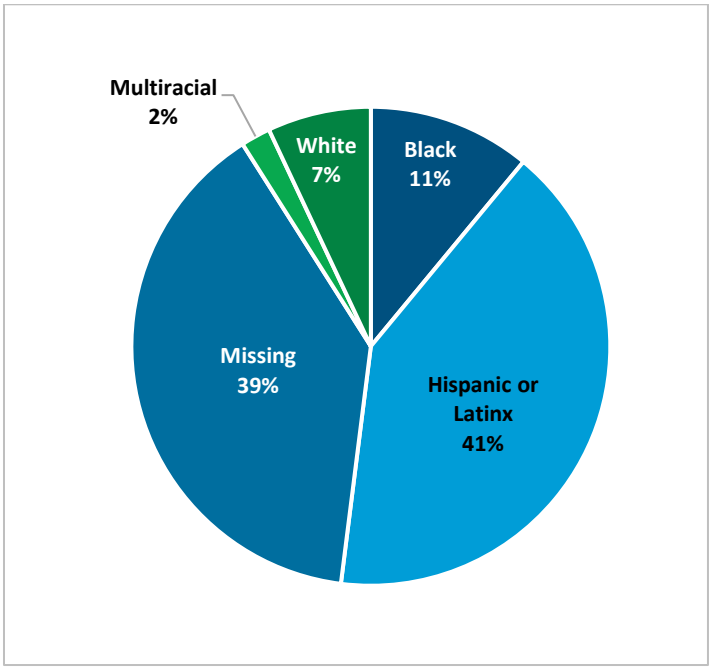


Note. *N* = 58. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. This count reflects the number of Literacy Lab students listed on the vendor-provided roster who were matched to tutoring log attendance data and had both pretest and posttest DIBELS composite scores available. Approximately 71% of students served by Literacy Lab did not have sufficient data available to calculate both pretest and posttest DIBELS composite scores.

[7] DIBELS scoring requirements specify that complete subscale score data are required for composite score calculation. See https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2023-02/UO_Dibels_8_Scoring_Guide_2023.pdf for more information. Note that the analytic sample of students served by Literacy Lab is not fully representative of the total population of students served, and the above findings should be interpreted with caution.

According to student roster data, 41% of students served by Literacy Lab were reported as Hispanic or Latinx, 11% were reported as Black, 7% were reported as White, and 2% were reported as multiracial. Less than 1% of students were reported as American Indian or Alaska Native. Thirty-nine percent of students served by Literacy Lab did not have race and ethnicity data available for analysis. Exhibit 21 presents a snapshot of the demographic characteristics of students served by Literacy Lab.

Exhibit 21. Race/Ethnicity of Students Served by Literacy Lab



Note. *N* = 203. This count reflects the number of Literacy Lab students listed on the vendor-provided roster who were matched to tutoring log attendance data.

Springboard

Springboard (www.springboardcollaborative.org), a national tutoring company founded in 2011, employs a tutoring model that combines a classroom curriculum with family engagement. The Springboard model follows a 10-week cycle that covers six unique steps to improve literacy. Its curriculum emphasizes phonemic awareness, phonics, fluency, vocabulary, and comprehension. Family engagement includes virtual workshops where tutors discuss literacy skills and concepts, along with reading strategies to use at home. Springboard tutors provide services both during the school day and after school. According to tutoring attendance log data, students served by Springboard received an average of 45 minutes per session.

When partnering with districts, Springboard identifies one or more program coordinators at program launch. The program coordinator is a district or school employee who serves as a

liaison with Springboard. The program coordinator oversees the program, monitors tutors, and helps recruit teachers and school staff to become tutors. The program coordinator is responsible for day-to-day communication with families. Springboard hires school personnel to serve in the role of tutor—giving priority to teachers—and hires other school staff, such as reading specialists, interventionists, and paraprofessionals.

Overview of Springboard Tutoring Services

During the study observation window, Springboard tutors provided a total of 15,959 tutoring sessions to 780 students at 18 schools across four districts: Boston, Peabody, Randolph, and Somerville. Each student participated in an average of 20.46 tutoring sessions in total during the study observation window and an average of 1.83 tutoring sessions per week (Exhibit 22).

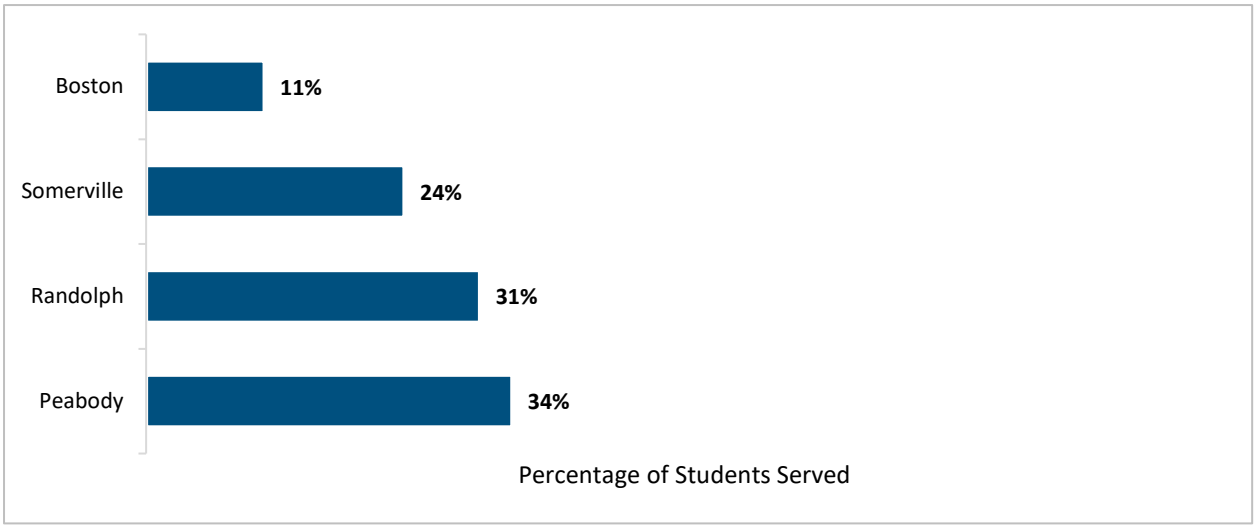
Exhibit 22. Snapshot of Springboard Tutoring Services Included in This Study

Participating Students	Schools	Districts
780	18	4
Total Number of Tutoring Sessions	**Number of Tutoring Sessions Per Student**	**Number of Tutoring Sessions Per Student Per Week**
15,959	20.46	1.83

Communities Served by Springboard

According to student roster and tutoring log data provided by Springboard, 34% of students included in the study attended school in Peabody, 31% attended school in Randolph, 24% attended school in Somerville, and 11% attended school in Boston. Exhibit 23 provides a snapshot of the home districts of students served by Springboard.

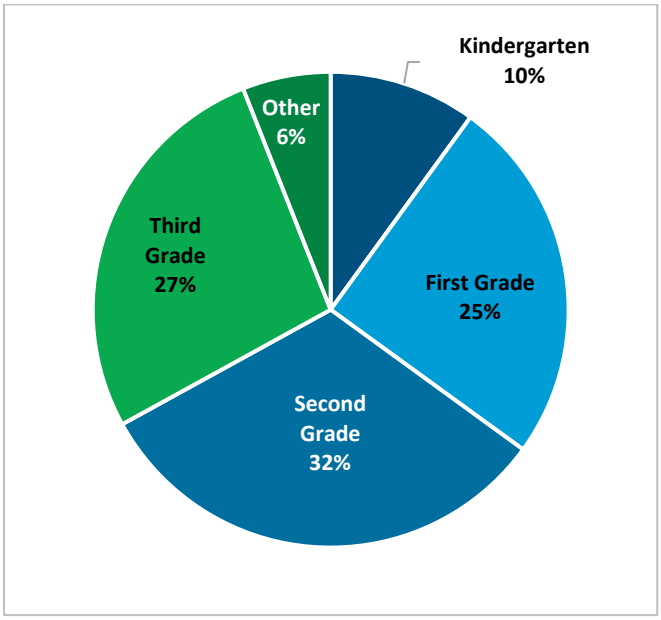
Exhibit 23. Home Districts of Students Served by Springboard



Note. *N* = 780. This count reflects the number of Springboard students listed on the vendor-provided roster who were matched to tutoring log attendance data.

According to student roster and tutoring log data, 10% of students served by Springboard were in kindergarten; 25% were in first grade; 32% were in second grade; 27% were in third grade; and 6% were in grades outside the K–3 range, categorized as "Other" for reporting purposes. Exhibit 24 presents a snapshot of the grade levels of students served by Springboard.

Exhibit 24. Grade Levels of Students Served by Springboard



Note. *N* = 780. This count reflects the number of Springboard students listed on the vendor-provided roster who were matched to tutoring log attendance data.

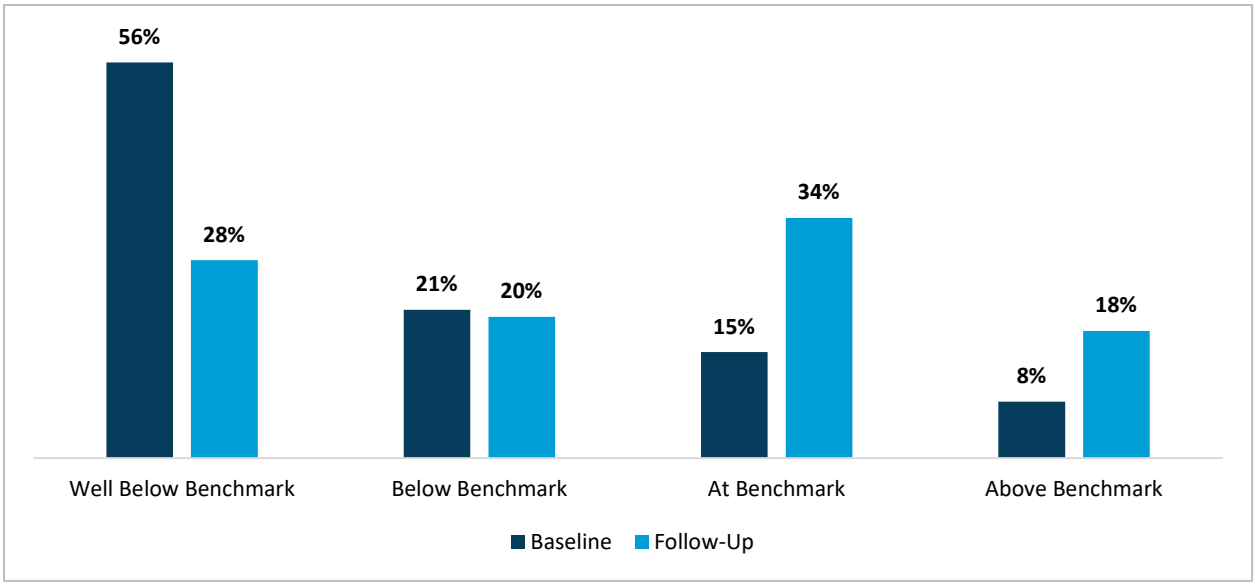
Springboard provided both baseline (pretest) and follow-up (posttest) DIBELS data for students served in all four districts: Boston, Peabody, Randolph, and Somerville. However, students served during the second tutoring session in Randolph (February through April) were missing posttest DIBELS data as of June 10. Additionally, the pretest and posttest DIBELS data provided for students at James F. Condon School in Boston and Prospect Hill Academy in Somerville were missing most of the required subtest scores. Of the 780 total students served by Springboard (students listed on the roster who received at least one tutoring session during the study observation window), 480 students (62%) had sufficient data available to calculate both pretest and posttest DIBELS composite scores and were therefore included in our analysis of learning outcomes.[8]

Of the Springboard students included in our analysis of learning outcomes, 56% scored well below benchmark on the baseline DIBELS assessment, 21% scored below benchmark, 15% scored at benchmark, and 8% scored above benchmark.

Within the same sample of Springboard students, 28% scored well below benchmark on the follow-up DIBELS assessment, 20% scored below benchmark, 34% scored at benchmark, and 18% scored above benchmark. Exhibit 25 presents the baseline and follow-up DIBELS reading levels of students served by Springboard.

[8] DIBELS scoring requirements specify that complete subscale score data are required for composite score calculation. See https://dibels.uoregon.edu/sites/dibels1.uoregon.edu/files/2023-02/UO_Dibels_8_Scoring_Guide_2023.pdf for more information.

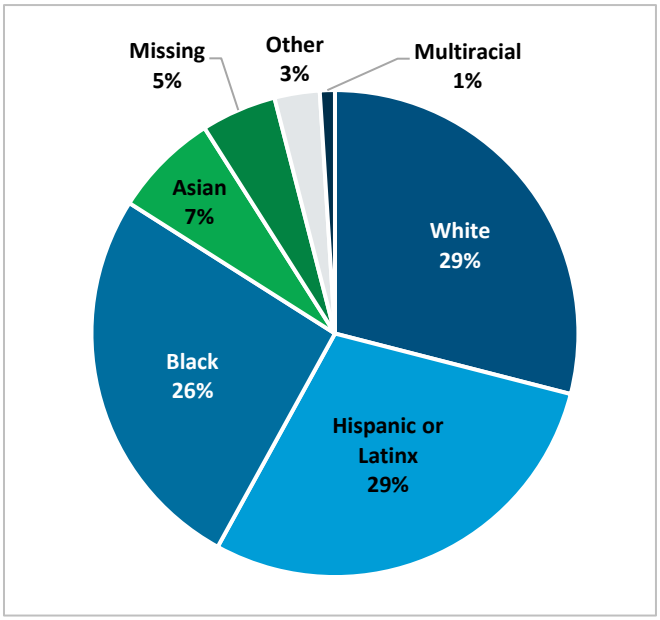
Exhibit 25. Baseline and Follow-Up DIBELS Reading Levels of Students Served by Springboard



Note. *N* = 480. DIBELS = Dynamic Indicators of Basic Early Literacy Skills. This count reflects the number of Springboard students listed on the vendor-provided roster who were matched to tutoring log attendance data and had both pretest and posttest DIBELS composite scores available. Approximately 38% of students served by Springboard did not have sufficient data available to calculate both pretest and posttest DIBELS composite scores.

According to student roster data, 29% of students served by Springboard were reported as Hispanic or Latinx, 29% were reported as White, 26% were reported as Black, 7% were reported as Asian, 3% were reported as "Other," and 1% were reported as multiracial. Five percent of students served by Springboard did not have race and ethnicity data available for analysis. Exhibit 26 presents a snapshot of the demographic characteristics of students served by Springboard.

Exhibit 26. Race/Ethnicity of Students Served by Springboard



Note. *N* = 780. This count reflects the number of Springboard students listed on the vendor-provided roster who were matched to tutoring log attendance data.

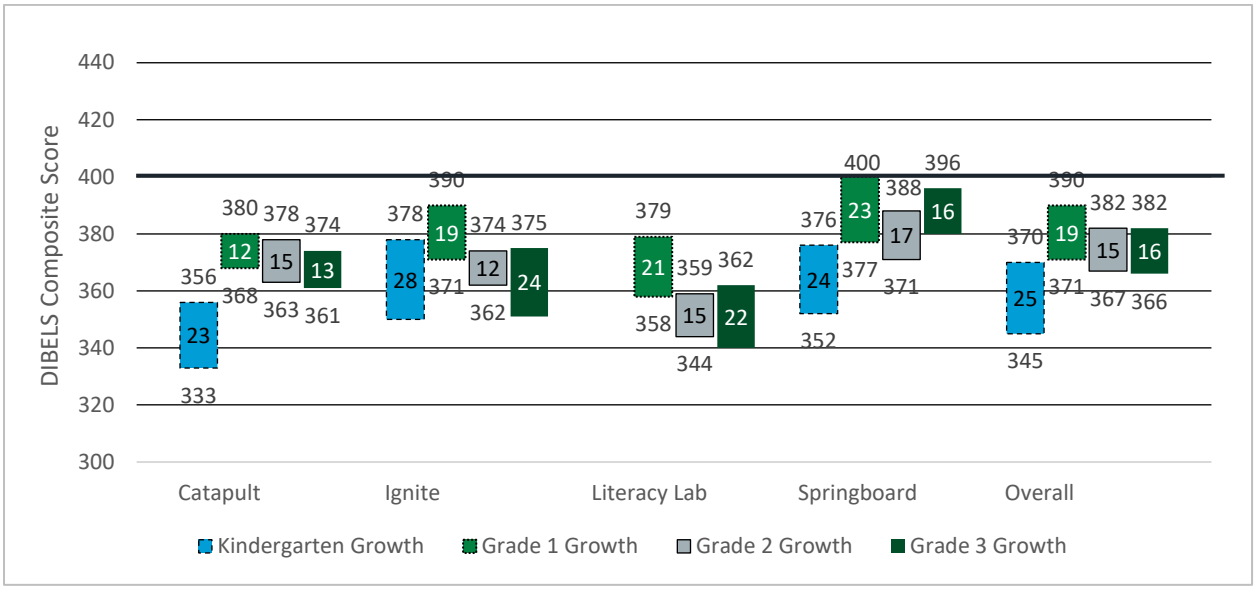
Summary of Student Learning Outcomes

In this section, we summarize student learning outcomes. We begin with descriptive analyses and then proceed to regression analyses that tested for associations across program features, student characteristics, and learning outcomes. On average and across the four vendors, students' composite scores on the DIBELS increased by 18 points (*SD* = 19). Catapult students' scores increased by 14 points, on average (*SD* = 18); Ignite students' scores increased by 19 points, on average (*SD* = 20); Literacy Lab students' scores increased by 19 points, on average (*SD* = 14); and Springboard students' scores increased by 19 points, on average (*SD* = 19).

Exhibit 27 presents the average DIBELS pretest scores, posttest scores, and score growth for each combination of tutoring vendor and grade level. The numbers below the bars are the average pretest scores, the numbers in the bars are the average growth, and the numbers above the bars are the average posttest scores. The clusters of bars correspond to vendors and the colors of the bars correspond to grade levels. For instance, the light blue bar on the far left presents results for kindergartners at Catapult tutoring sites. The average pretest score among kindergartners at Catapult sites is 333 points and the average posttest score among kindergartners at Catapult sites is 356 points, corresponding to an average growth of 23 points. A student reading at their expected grade level would earn a score around 400 at the midpoint of the school year; the nationwide standard deviation of test scores is 40 points at any given

moment in time. There were too few kindergartners at Literacy Lab sites with data for their scores to be presented separately.

Exhibit 27. Average Student Growth on DIBELS Composite Scores



Note. DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

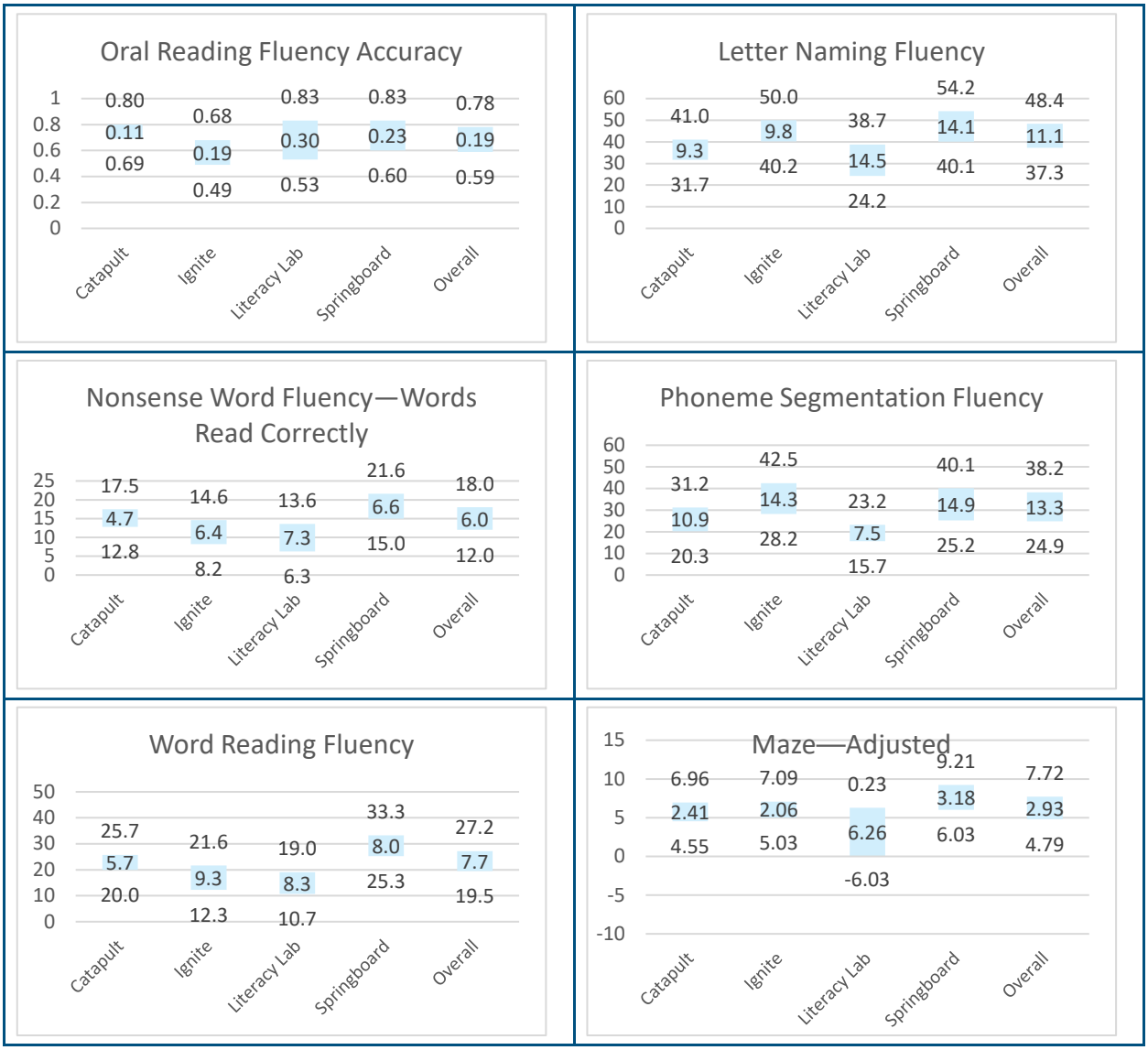
Approximately 68% of students with complete data had baseline DIBELS scores well below benchmark, while 16% of students scored below benchmark but not well below it. At posttest, 40% scored well below benchmark, while 20% scored below benchmark but not well below it. At baseline, 16% of students had DIBELS composite scores at or above benchmark; this increased to 41% at posttest. Out of the students who scored well below benchmark at baseline, 43% moved up at least one category at posttest (while 57% remained well below the benchmark), and 20% met the benchmark at posttest; of the students who scored below benchmark but not well below it at baseline, 76% scored at or above benchmark at posttest.

Exhibit 28 presents student growth on the DIBELS subscales; as in Exhibit 27, each bar includes the average pretest score on the given subscale for students working with a given vendor at the bottom and the average posttest score at the top, and the length of the bar represents the average growth on that subscale between pretest and posttest.[9] All subscale scores are expressed in terms of raw subscale points. In all six main subscales, students working with all

[9] These subscales include Oral Reading Fluency, in which students are asked to read a provided passage aloud for a minute; Letter Naming Fluency, in which students are presented with upper-case and lower-case letters in a sequence and asked to name as many as possible; Nonsense Word Fluency, in which students are presented with pronounceable combinations of letters and asked to sound out the "words" they form; Phoneme Segmentation Fluency, in which students are presented with words and asked to break them into phonemes; Word Reading Fluency, in which students are presented with a list of common words and asked to read them aloud; and Maze, in which students are presented with a passage in which every seventh word is removed and replaced with a menu of three options, from which students are expected to choose the word that fits the place in the passage.

four vendors scored higher on average at posttest than they did at pretest. As a percentage of pretest scores, the gains were particularly high in Nonsense Word Fluency and Phoneme Segmentation Fluency. In four of the six subscales, average growth among students at Catapult was the lowest of all four vendors, though this difference was not always significant.

Exhibit 28. Average Student Growth in DIBELS Scores, by Subscale



Note. DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

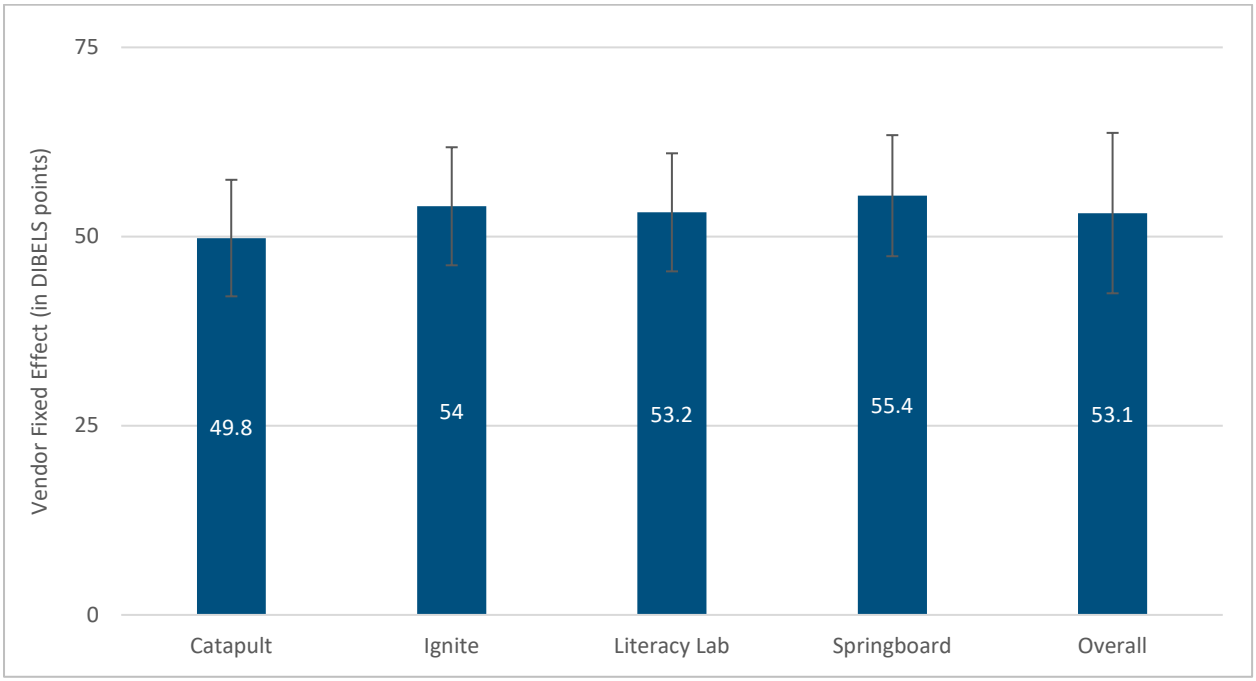
Associations Between Program Characteristics and Student Outcomes

RQ1: Was there a significant difference between the impact of providers' program models on student outcomes?

Student growth was greater among students enrolled in tutoring from Ignite and Springboard than students enrolled in tutoring from Catapult (see Exhibit 29); growth among students enrolled in tutoring from Literacy Lab was not statistically different from growth among students enrolled in tutoring from any of the other vendors. Fixed effects regression models were used to examine student growth from DIBELS pretest to posttest. Fixed effects regression models are a measure of the impact of a program after controlling for other factors. In this study, these models indicated that students gained 50 points on their DIBELS composite scores during Catapult's tutoring, 54 points during Ignite's tutoring, 53 points during Literacy Lab's tutoring, and 55 points during Springboard's tutoring, on average. The statistical models included controls for students' demographics and *excluded* specific details of service delivery (e.g., number of tutoring sessions, whether the tutor was a teacher). This approach allowed us to isolate the association between student learning and each vendor's program model as a whole package.

While this is a common analytic approach for comparing programs, it is important to keep in mind that these estimates assume values of zero on all covariates. A score of 0 on the DIBELS is impossible; the lowest observed DIBELS pretest score is 314. This modeling technique allowed us to compare the vendors with each other, and in doing so, we found that the effect of Ignite and Springboard's tutoring was significantly greater than the effect of Catapult's tutoring at all conventional levels of statistical significance. However, this does not mean that Catapult's tutoring was ineffective. The effect estimated in this model for Catapult was statistically significantly different from zero, implying that Catapult students' performance also improved appreciably during the period of time in which they received tutoring.

Exhibit 29. Vendor-Specific Effects on Student Outcomes



RQ2: Which program factors best predicted student growth?

Exhibit 30 presents the relationships between various aspects of program delivery and student DIBELS composite score growth. When all of these characteristics are included in a single regression model, only having a certified teacher as a tutor is significantly related to student growth. However, many of the characteristics of tutoring were strongly associated with single vendors or highly correlated with each other, which made it difficult to single out the impact of any single characteristic. For instance, Springboard tutors were more likely to be teachers and received slightly more training on average, but tutors who were teachers received much more training than tutors who were not teachers.

Exhibit 30. Relationships Between Program Factors and Student Growth

Program factor	Coefficient (standard error)
Number of students in the tutoring group	-2.11 (1.37)
Tutor is a certified teacher	5.56* (2.58)
Number of sessions per week	0.444 (0.690)
Number of hours of training tutor received	-0.067 (0.119)
Tutor received feedback often	-0.674 (2.40)
Tutoring takes place outside of school hours	-0.190 (1.37)
Number of hours of tutoring received	0.019 (0.039)

* $p < .05$.

RQ3: Were there significant differences in how students developed specific skills that were impacted by program factors?

To understand the potential differences in how students in each program developed specific skills, we conducted regression analyses that used each of the DIBELS subscale scores as outcomes. Most program factors were not significantly related to any of the subscale scores. Students whose tutors received more training had lower growth in Letter Naming Fluency but higher growth in Maze. Students whose tutors were educators other than teachers had higher growth in Nonsense Word Fluency and Maze, while students whose tutors were teachers had higher growth in Word Reading Fluency and Maze. Students who worked in smaller groups had greater growth in Word Reading Fluency. Students whose tutoring took place after school had higher growth in Word Reading Fluency but lower growth in Oral Reading Fluency. Students whose tutors had more hours of training had greater growth on the Maze, as did students exposed to more total hours of tutoring. As with the associations between program factors and overall DIBELS score growth, interpretation of these results is complicated by the extent to which program factors are distinctive aspects of particular vendors, making it difficult to distinguish the effect of, for instance, having a smaller group from the effect of being enrolled in tutoring from Catapult.

Associations Between Student Characteristics and Student Outcomes

RQ4: Which student characteristics best predicted student growth?

Exhibit 31 presents the associations between DIBELS score growth and pretest scores, student grade levels, and race/ethnicity groups. Score growth was slightly higher for students with lower baseline scores, and it was higher for students in second grade and lower for students in kindergarten. As Catapult did not report information on students' race and ethnicity, the associations between racial and ethnic groups and test score growth were derived entirely from students at Ignite, Literacy Lab, and Springboard; according to these data, Black students and Asian students exhibited more growth than the reference group, which consists of White students and students whose race/ethnicity is not reported. Furthermore, because race and ethnicity information was not available for Catapult students, and all of the coefficients on race and ethnicity variables were positive, it is possible that these estimated relationships between racial and ethnic categories and score growth actually represent the higher growth achieved by students working with the other vendors, as discussed in RQ2.

Exhibit 31. Relationships Between Student Characteristics and Student Growth

Student characteristic	Coefficient (standard error)
DIBELS pretest score	-0.081* (0.020)
Student is in kindergarten	4.10* (1.84)
Student is in first grade	0.00 (reference category)
Student is in second grade	-2.70* (1.36)
Student is in third grade	-2.17 (1.46)
Student is Black	4.69* (1.54)
Student is Hispanic/Latinx	1.20 (1.26)
Student is Asian	10.18* (2.69)
Student is White or did not report race/ethnicity	0.00 (reference category)

*$p < .05$.

RQ5: Were there significant differences in how students developed specific skills that were impacted by student characteristics?

The regression specifications with DIBELS subscale scores included pretest scores in the same subscales, along with grade and race/ethnicity variables, as covariates. Students with lower pretest scores had higher growth in all subscales, including Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Word Reading Fluency, Oral Reading Fluency, and Maze. Black students had higher growth in Nonsense Word Fluency and Maze; no other differences by race or ethnicity were statistically significant. Kindergartners had lower growth in Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, and Word Reading Fluency. Second graders had higher growth in Word Reading Fluency and Oral Reading Fluency, and third graders had higher growth in Oral Reading Fluency.

What Works, for Whom, and Under Which Conditions?

RQ6: What significant interaction effects were found between any of the variables in RQs 1–5?

We tested interactions that combined DIBELS pretest scores with indicators for whether a student was enrolled with Catapult, Ignite, Literacy Lab, or Springboard. If the coefficient on this interaction term is negative and significant for a particular vendor, it implies that the vendor is making progress in helping lower achieving students catch up to grade level. However, a coefficient that is positive and significant implies that the vendor is helping higher achieving students experience faster growth, while students with less-developed reading backgrounds fall farther behind. This coefficient was negative and significant for Catapult and positive and significant for Springboard, implying that students who started out further behind

were showing faster growth with Catapult, while students who started out performing well were showing faster growth with Springboard. This coefficient was not statistically significant for Ignite or Literacy Lab, implying that there was not a significant relationship between pretest scores and student growth for students served by those two vendors.

To determine whether different program characteristics were more or less helpful for particular types of students, we reestimated the main model separately with each of the following interaction terms added one by one:

- The number of hours of training received by the tutor, interacted with whether the tutor was a teacher or another type of educator
- The number of students per tutor, interacted with whether the tutor was a teacher or another type of educator
- The number of students per tutor, interacted with indicators for each grade level
- Whether the tutor and student were both Black or both Hispanic or Latinx. (There were insufficient data to support the inclusion of an interaction term for the tutor and student both being Asian.)

While there are theoretical justifications for each of these interactions being relevant to DIBELS growth, only some were statistically significant. Larger amounts of training were associated with lower growth for students whose tutors were teachers and for students whose tutors were other educators. The directions of causation for these results are not obvious; it could be that educators require less training to be able to implement the intervention than non-educators do and that the training adds to the educators' burden, making them less effective, or it could be that vendors are providing more training to tutors who appear to need it most while allowing tutors to work with less training if they already appear capable. All other interactions were not significantly related to DIBELS growth.

Limitations

Several limitations of this study should be kept in mind when interpreting its findings. Without a comparison group that receives only regular classroom instruction, we cannot determine how much student learning is a result of the tutoring services versus regular classroom instruction. Even among tutored students, while variation in program characteristics within vendors was often minimal, variation across vendors was substantial. For instance, all Springboard students received tutoring after school, but the average student with a Springboard tutor received 8 more hours of tutoring than the average student with a Catapult tutor. As such, program characteristics were often highly correlated with each other, making it very difficult to distinguish the effects of different aspects of vendors' program delivery.

Additionally, data were only available for a subset of students served by the four tutoring vendors. Several sites did not administer the DIBELS posttest or provide the data about its results in time for their students to be included in the analysis. The analytic sample only included students with DIBELS pretest and posttest scores (either for all subscales or with enough data to impute missing ones) who could be matched to vendors' student rosters and attendance logs. Other missing data, including demographic data or survey responses, could be replaced with constant values and missing value indicators. However, students in schools that did not administer DIBELS in a timely fashion, students whose tutors did not record attendance data, and students who were absent on the day of testing were excluded. As a result, the analytic sample was smaller than the population and not entirely representative. This limits the statistical power of the analyses and the extent to which findings can be extrapolated to other settings where tutoring took place.

Conclusion

This summative evaluation aimed to identify program and student characteristics among students who participated in the Early Literacy Supplemental Services Tutoring program, determine whether these students' performance improved during participation, and determine the extent to which program and student characteristics were associated with student growth.

- **Students with available data improved their literacy skills while enrolled in tutoring.** Students served by all four tutoring vendors saw statistically significant gains in student learning, according to statistical models, with Ignite and Springboard performing particularly well. It was difficult to tease apart the value of specific program components or approaches because of the level of standardization in each vendor's approach.
- **Vendors focused on improving the skills where students needed the most work.** Within each DIBELS subscale, students with lower pretest scores on that subscale had higher growth on that subscale at posttest than students with higher pretest scores on the same subscale had, even if the relationship between composite pretest scores and composite score growth was not statistically significant. This suggests that tutoring was effective in addressing students' greatest areas of weakness.

References

Massachusetts Department of Elementary and Secondary Education (DESE). (2019). *Next generation MCAS achievement results*. https://profiles.doe.mass.edu/statereport/nextgenmcas.aspx

Nickow, A., Oreopoulos, P., & Quan, V. (2020). *The impressive effects of tutoring on preK–12 learning: A systematic review and meta-analysis of the experimental evidence*. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w27476/w27476.pdf

Appendix A. Variable Sources and Measurement Properties

Variable	Source	Measurement properties
Number of hours of tutoring	Tutoring logs	Continuous: The number of sessions attended multiplied by the length of the session, in hours (e.g., a student attending three 45-minute sessions would be coded as 2.25)
Frequency of tutoring	Tutoring logs	Continuous: The number of sessions divided by the number of weeks between the first and last sessions
Student–tutor race match	Tutor survey, student rosters (missing for Catapult)	Binary: 1 if the student and the tutor were the same race/ethnicity, 0 otherwise
Total training hours received	Tutor survey	Integer: The two questions about tutor training (summer and school year) would be coded to number of hours and summed (e.g., if a tutor received "more than 4, up to 8 hours" of training in the summer and "more than 2, up to 4" in the school year, they would be coded as 5 + 3 = 8)
Tutor is a certified teacher	Tutor survey	Binary: 1 if the tutor was a certified teacher and 0 otherwise
Tutor is an educator but not a teacher	Tutor survey	Binary: 1 if the tutor listed any district employment besides teaching, 0 if the tutor was a teacher or not employed by a district
Feedback received	Tutor survey	Integer: 3 if the tutor received feedback every session, 2 if the tutor received feedback often, 1 if the tutor received feedback sometimes, 0 if the tutor never received feedback or was not observed
Vendor fixed effects	Student rosters	Binary: Indicator for each tutoring vendor
Student race/ethnicity	Student rosters (missing for Catapult)	Binary: Indicator for each racial/ethnic group (Asian, Black, Hispanic/Latinx, White)
DIBELS pretest score	DIBELS data	Continuous
Tutoring during or after the school day	Interviews/document review	Binary: 1 if after school, 0 if during school
Tutor–student ratio	Tutoring logs, survey	Continuous: Number of students served by student's tutor Binary: 1 for group, 0 for individual
Student grade level	Student rosters	Categorical: Separate indicators for each value

About the American Institutes for Research

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit institution that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

MASSACHUSETTS
Department of Elementary and Secondary Education

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

21715_11/23