springboard COLLABORATIVE

June 2020

Springboard Summer Reading Program Evaluation Report

McClanahan Associates
Evaluation for Progress

Kelly Piccinino, Ph.D.
Sarah K. Pepper, Ph.D.
Hannah Salomon, M.S.Ed.
Sara Greenfield
Wendy McClanahan, Ph.D.

impactED
University of Pennsylvania

Claire Robertson-Kraft, Ph.D.

TABLE OF CONTENTS

SPRINGBOARD SUMMER

Springboard Collaborative (Springboard) aims to reduce the literacy gap by closing the gap between home and school. Springboard coaches teachers and family members to help their children cultivate reading habits through in-person programming and technology to help them read on grade-level by fourth grade.

Springboard Summer is an intensive 5-week summer program that combines daily reading instruction geared towards rising kindergarten through third grade students they refer to as "scholars," weekly workshops training parents to teach reading at home, a rigorous coaching cycle for teachers, and an incentive structure that awards learning tools to families in proportion to scholar reading gains. The purpose of Springboard Summer is to reverse summer (reading) slide[1] while helping schools build internal capacity in four ways:

- Developing children as readers by increasing high-quality instruction at home and school in order to help students reach ambitious reading goals.
- Developing parents as teachers by training them to be effective one-on-one reading coaches at home.
- Developing teachers as instructors through training, Professional Learning Communities, and coaching on data-driven instruction and family engagement.
- Developing schools' leadership pipelines by coaching teacher leaders through management experiences.

STUDY OVERVIEW

Over the past eight years, through examinations of school-administered reading assessments, Springboard has demonstrated internally that the average Springboard Summer scholar improves in reading during the program. Several partner school districts have compared Springboard scholars to their non-participating counterparts and found that, on average, Springboard scholars grow faster in reading than their peers. This promising evidence prompted Springboard to seek an external evaluation, with a focus on comparing Springboard Summer scholars to their non-participating counterparts.

Building on Springboard's prior internal and district-led evaluations, this external evaluation is the next step on the path to validating Springboard's efforts at improving student literacy outcomes. This evaluation is a valuable opportunity to learn about Springboard Summer scholar outcomes by examining district-administered reading assessment data with select districts (those implementing Springboard Summer with efficacy[2]). Additionally, it uses a rigorous methodology - propensity score matching, which controls for differences between students who get Springboard Summer versus those who do not. For these reasons, this study is an important extension of prior work. It serves as an external test of the Springboard Summer model, specifically focusing on districts where the program is implemented as intended, reducing the chance of findings being a result of poor implementation.

[1] "Summer slide" refers to the "tendency for students, especially those from low-income families, to lose some of the achievement gains they made during the previous school year." Source: https://www.cde.state.co.us/cdelib/summerslide.
[2] McCormick, M. (2019, August). Linking fidelity of implementation to outcomes in real-world settings. MDRC | Building knowledge to improve social policy. https://www.mdrc.org/sites/default/files/Linking_Fidelity.pdf.

This study's main research objectives are to determine:

1. **How participation in Springboard Summer is associated with reading growth.**
2. **How Springboard Summer scholars' reading achievement compares to that of non-participating Springboard Summer students.**

HIGH-LEVEL FINDINGS

Analyses found that, on average, Springboard Summer scholars experienced reading growth:

- Across all grades, on average, scholars who participated in Springboard Summer showed improvement on their reading assessment score from the end of the school year before Springboard Summer to the start of the following school year.
  - The largest gains were for scholars who started below grade level.
  - Given high scholar and parent attendance, the evaluation team could not discriminate between the effect of high and low attendance on reading growth.
- Springboard Summer scholars show larger improvements in reading scores, on average, when compared to similar students who did not participate. The quality of evidence from this study satisfies the Every Student Succeeds Act (ESSA) Tier 2 standards.[3] (See a fuller discussion of ESSA standards as they relate to this study on page 13.)

REPORT OVERVIEW

This report shares the findings from this external evaluation so that Springboard can use them to inform internal data-driven decisions, identify areas where programmatic changes may be needed, and to generate findings about implementation and outcomes that can be shared with internal and external stakeholders. The report is broken out into the following sections:

- **Study Methodology** – This section of the report describes the sample, data collection, measures, and the rationale for the analytic approach.
- **Findings** – This section of the report presents key findings from the analyses.
- **Discussion** – The final section of the report discusses strengths and limitations of the evaluation and highlights recommendations related to the research findings.
- **Appendices** - The appendices include figures and tables that supplement the main findings (district selection matrix, propensity score weighting, regression models, and Springboard scholar data compared to non-participating Springboard Summer students by district), sensitivity analyses findings related to the sample's largest subpopulation consisting of Public Prep scholars, and stakeholder survey constructs.

[3] Tier 2 standards require: 1) a quasi-experimental study design, 2) statistically significant positive effects of programming without any negative findings, 3) adequate sample size (at minimum 350 students), and 4) setting requirements of more than one district or school.

STUDY METHODOLOGY

This is a quasi-experimental study exploring: 1) the association between Springboard Summer and reading growth, and 2) Springboard Summer scholars' reading achievement compared to that of non-participating Springboard Summer students. The evaluation team gathered student data from select school districts who implement Springboard Summer programming with high fidelity to the program model and combined this with programmatic Springboard Summer data, yielding a sample of 673 rising kindergarten through 4th grade Springboard scholars from five districts participating in Springboard Summer in either 2018 or 2019.

STUDENT SAMPLE

To prepare for the project, the evaluation team conducted a document review and inventoried all relevant Springboard documents and data. The document review highlighted important variations among Springboard districts and sites that were relevant to consider for evaluation purposes. Springboard districts varied by: a) type of reading assessment used, b) how sites selected Springboard Summer scholars, c) the demographics they targeted, and d) the perceived[4] quality of Springboard Summer implementation. Accordingly, the evaluation team used a criteria matrix in order to hone in on a group of districts that would be most aligned to the parameters of this study. Furthermore, the number of districts included was limited by the project budget.

Below, a list of the factors are provided that this study considered when selecting districts for the study sample (in order of consideration):

1. ***Fidelity to the Springboard Summer Program Model.*** Fidelity to the Springboard Summer program model was the most important factor considered when selecting the study sample because the study goal was to explore Springboard Summer's effects on reading growth when *implemented as designed*. First, Springboard Collaborative nominated strong implementers. The evaluation team followed up with program staff (e.g., site/cluster leaders) to confirm each site's fidelity to the model with no more than minor local program adaptations.[5]

   The districts in this study implemented the following program elements in 2018 and 2019:

   a. 5 weeks of literacy programming (or equivalent of) during summer break;
   b. Springboard Summer's 3-pronged approach;
      i. Student facing program elements designed to increase reading levels and achievement;
         1. 5-component literacy block (read aloud, shared reading, word work, differentiated reading & literacy centers, and writing period);
         2. achievement incentives for scholars;
      ii. Program elements designed to bolster family involvement;
         1. initial home visits;
         2. weekly family workshops (training parents to teach reading at home);

[4] Since this study did not include an implementation evaluation, we relied on retrospective staff perceptions of program implementation quality.
[5] Home visit adaptations include alternate meeting locations when teachers did not feel safe going into a home or families didn't feel comfortable with an official person coming into the home. In 2019, one location ran for four weeks but an hour longer each day to provide the equivalent numbers of hours of a five-week program. Some locations adapted the schedule of family workshops and offered make-up sessions to accommodate parents. Teachers were given flexibility regarding the order they were expected to implement the literacy block on a given day. One location used a different computer program (MyOn and Lexia vs. Raz-Kids) for guided reading groups for the purposes of variety since students use Raz-Kids during the school year.

iii. Program elements designed to build capacity for teacher instruction;
    1. teacher training;
    2. coaching supports;

c. Target rising kindergartners through rising fourth graders.

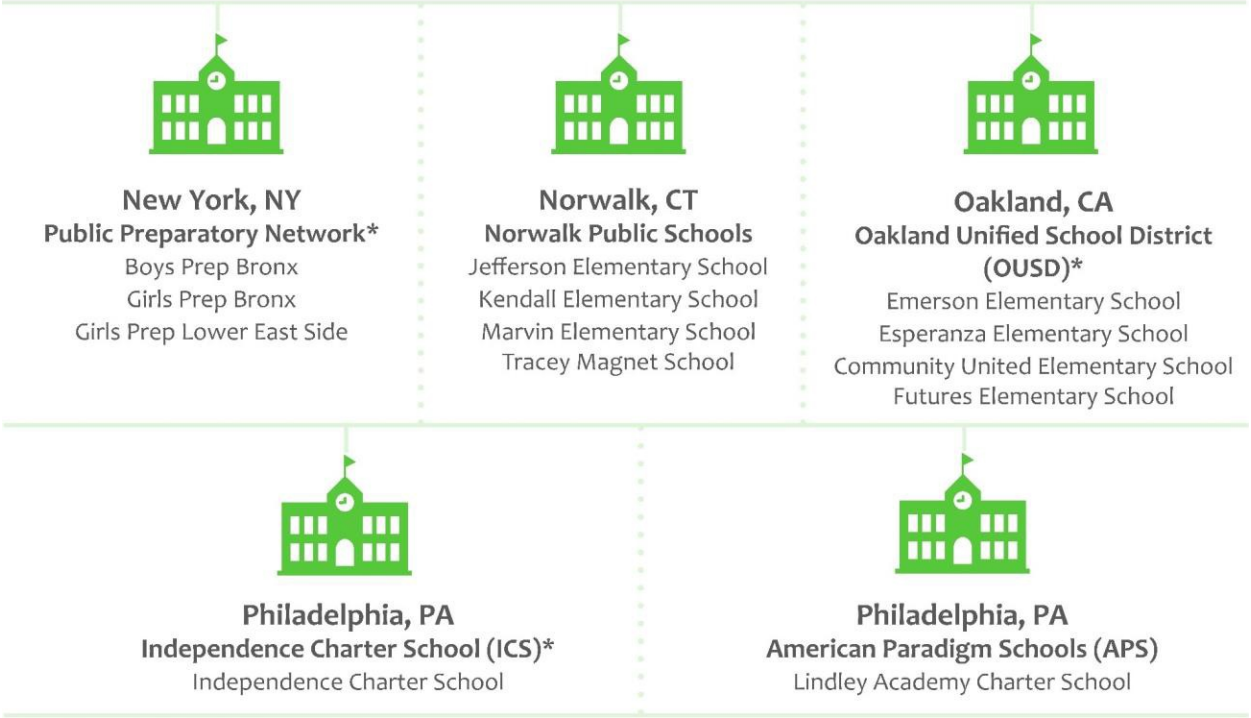
2. ***Reading Assessment Type.*** The evaluation team worked with Springboard to identify which reading assessments were most aligned to Springboard Summer's program model and intended outcomes. As a result, districts using assessments that do not readily measure incremental progress, such as i-Ready, were excluded from the sample.

3. ***Feasibility of Securing School District Data.*** After considering program fidelity and reading assessment alignment, the evaluation team researched district regulations for data-sharing processes and connected with district gatekeepers to inquire about the feasibility of securing school district data within the study timeline[6] for all remaining Springboard Summer implementing districts. Some school districts conveyed upfront that they did not have the ability to participate in a study such as this one. They reported inadequate staffing to provide needed data and/or an inability to accommodate the required timeline. Other districts initially agreed to participate but then were unable to meet the timeline.

Six districts were selected through the initial selection criteria and agreed to participate in the study (see Appendix A, Section I for detail on districts excluded and reasons why). Ultimately, only five districts were included in the study since one district who had initially agreed to fulfill the data request for this study was not able to complete the data request due to capacity constraints and study timeline. Figure 1 below, cites each district included in the study and their respective school sites within each district. The evaluation team worked closely with school district leadership to establish data-sharing agreements with the selected school districts.

[6] The study timeline required district data to be provided to the evaluation team prior to April 1st, 2020 in order to meet analysis and reporting goals.

FIGURE 1. Selected Springboard Summer Districts/Sites



*Public Prep, OUSD, and ICS provided data on non-participants.

The treatment sample is composed of 673 scholars who were rising kindergarteners through fourth graders at the time of their participation in Springboard Summer (2018 or 2019) and had district reading assessment data from both the end of the school year (EOY) before their participation in Springboard Summer, as well as from the beginning of the school year (BOY) which followed their participation in Springboard Summer. Each of the five districts sharing data used a different reading assessment, and over one-half of the student sample with available paired assessments that straddle student participation in Springboard Summer are from the Public Prep district. Figure 2 represents the sample of Springboard Summer scholars by district.

FIGURE 2. Springboard Summer Scholars by District

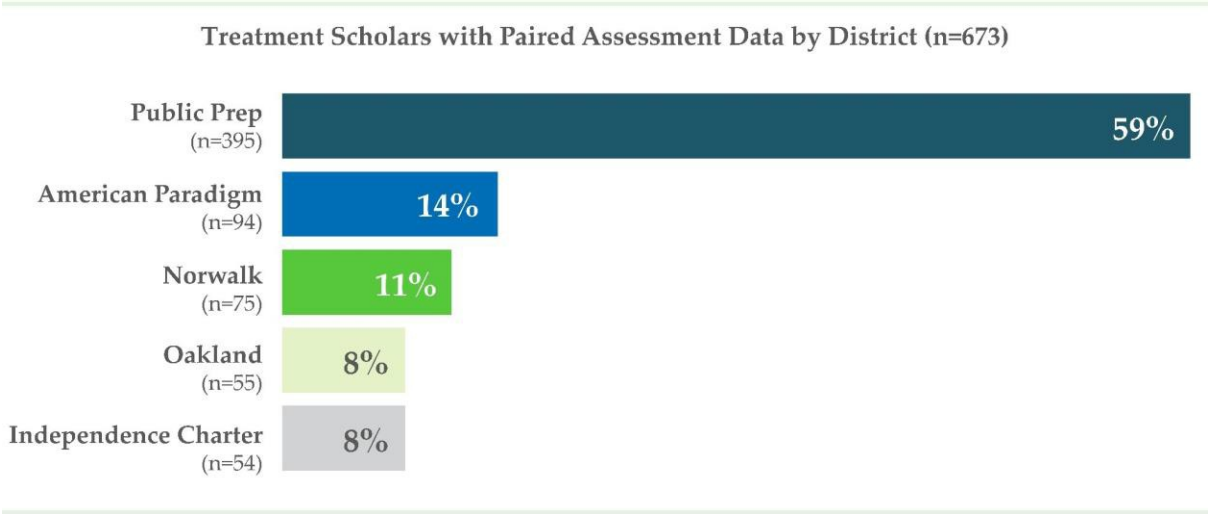


Table 1 describes Springboard Summer scholar demographics by district and in total (overall) for the study sample. Overall by grade, there are only a small number of rising kindergarten scholars with paired data available, so subgroup analysis for kindergarten is underpowered; the remainder of the sample is approximately evenly divided between rising first, second, third, and fourth grade scholars. A little under half (47%) of the overall sample was male. The four poverty level categories show that scholars in the study sample tend to live in zip code areas with higher concentrations of poverty (about a third, or 32%, are within a zip code where 35-52% are classified as living below poverty). Academic data from the sample indicate that 71% of Springboard Summer scholars in the study started summer programming below grade level (30% started more than one grade below reading level and 41% started within one grade below reading level). While 21% of the sample scholars were English Language Learners (ELLs), only 3% took the district reading level assessments in Spanish. Those with an IEP made up 23% of the sample of Springboard Summer scholars.

TABLE 1. Distribution of Springboard Summer Scholars in Analysis Sample by Grade & District

Variables	American Paradigm (n = 94)	Independence Charter (n = 54)	Norwalk (n = 75)	Oakland (n =55)	Public Prep (n = 395)	Overall (n = 673)
Scholar Demographics						
Kindergarten	1%	2%	0%	0%	6%	4%
1st Grade	37%	35%	0%	0%	32%	27%
2nd Grade	21%	39%	25%	0%	27%	25%
3rd Grade	21%	0%	44%	62%	20%	25%
4th Grade	19%	24%	31%	38%	15%	20%
Gender: Male	49%	56%	61%	49%	42%	47%
Percent of Zip code below poverty: 0 - <15%	0%	15%	100%	9%	2%	14%
Percent of Zip code below poverty: 15 - <28%	3%	43%	0%	29%	29%	23%
Percent of Zip code below poverty: 28 - < 35%	83%	26%	0%	62%	21%	31%
Percent of Zip code below poverty: 35 - 52%	14%	17%	0%	0%	48%	32%
Scholar Academics						
Starting Reading Level Relative to Grade Level: Start more than 1 GL below	23%	19%	72%	93%	15%	30%
Starting Reading Level Relative to Grade Level: Start within 1 GL below	55%	38%	25%	4%	46%	41%

TABLE 1. Distribution of Springboard Summer Scholars in Analysis Sample by Grade & District (CONT.)

Variables	American Paradigm (n = 94)	Independence Charter (n = 54)	Norwalk (n = 75)	Oakland (n =55)	Public Prep (n = 395)	Overall (n = 673)
Starting Reading Level Relative to Grade Level: Start at/above GL	23%	43%	3%	4%	38%	29%
Percent English Language Learners	5%	28%	57%	71%	10%	21%
Percent with IEP	23%	24%	29%	9%	23%	23%
Percent Tested in Spanish	0%	43%	0%	0%	0%	3%

DATA COLLECTION

This study relies on two main data sources: Springboard Summer program data provided by Springboard, and student records data provided by select school districts. The following section outlines how each type of data were collected.

Springboard Summer Program Data

Springboard provided the evaluation team the following scholar-level Springboard Summer programmatic data for 2018 and 2019:

- Springboard Summer Program Data
  - Daily program attendance
  - Family workshop participation
  - Family home visit participation
  - Scholar summer reading goals
  - Incentives each scholar earned at Springboard Summer
  - Springboard's internal reading assessment scores
- Scholar Demographic Data
  - Birthday
  - Race/ethnicity
  - School ID
  - School name
  - Grade

The evaluation team used the internal Springboard school data to confirm the district data and ensure accurate data on Springboard scholars from the districts were received.

District Data

The evaluation team requested the following student-level data (for Springboard Summer scholars and within school non-participating Springboard Summer students - for comparison purposes) from each district for the spring and fall bookending the 2018 and 2019 summers:

- Demographics
  - Teacher name
  - School name
  - Grade
  - School attendance (individual-level attendance - excused, unexcused - spanning across both years and for some, it would be just one summer)
  - Gender
  - Ethnicity
  - Zip code
  - English Language Learner (ELL) status
  - Special education services (i.e., Individualized Education Program (IEP))

- District Reading Assessment Data
  - District reading assessment type (e.g., STAR)
  - Reading assessment raw scores: BOY, middle of the school year (MOY), and EOY component scores (oral reading fluency, reading comprehension, language usage, vocabulary, etc.)
  - Reading proficiency status (e.g., below grade level, on grade level, above grade level, if available)
  - District-level score conversions (i.e., stanine, standardized, number of months reading growth, or GLE, if available)

Of note, the evaluation team requested reading assessment data for multiple time points from each school district[7]; however, not all districts were able to provide student data for each time point. Ultimately, the evaluation team selected two data points for analysis: EOY (prior to participation in Springboard Summer), and BOY (following participation in Springboard Summer), as districts in the sample were able to provide these data at minimum. At the school level, the percentage of student's eligible for free or reduced-price lunch (FRPL) for each school the Springboard Summer and comparison students attended during the school year were also requested. Because these data were not shared uniformly across districts, the evaluation team opted to use student residential zip code to create a proxy for student socioeconomic status (SES). The U.S. Census Bureau 2018 American Community Survey[8] provides estimates of the percentage of people living in poverty by zip code and these data were linked to each student.

The evaluation team secured Institutional Review Board (IRB) approval required by many school districts as a prerequisite for establishing data-sharing agreements. As school districts have different data-sharing guidelines, regulations and processes, the data requests were unique to each district. The evaluation team developed data request manuals for each district tailored to the specific requests. Early in the data request process, the evaluation team connected with school district personnel to walk them through the data request in an effort to field data-related questions upfront and expedite the turnaround. The evaluation team worked with each district to set up a data transfer protocol that allowed districts to maintain de-identified data, and allowed the evaluation team to maintain the key variables of interest (i.e., student demographic data and district reading assessment data). Since identifiable educational data are protected by FERPA regulations, the evaluation team specifically

[7] The evaluation team originally requested reading assessment data from four (4) time points: middle of the school year (MOY) prior to participation in Springboard Summer (T1); end of the school year (EOY) prior to participation in Springboard Summer (T2); beginning of the school year (BOY) following participation in Springboard (T3); and MOY following participation in Springboard Summer (T4).
[8] Source: https://data.census.gov/cedsci/table?q=Income%20and%20Poverty&tid=ACSST1Y2018.S1701&vintage=2018&t=Income%20and%20Poverty.

requested that districts create an extra variable to de-identify Springboard Summer program participants, based on the names of the participants provided to districts using internal Springboard data. Despite providing districts with data request manuals, the data received from each school district was very different from one another, both in format and content.[9] Furthermore, two of five districts[10] were ultimately unable to provide whole-school student-level data - to be used for comparison purposes.

MEASURES

To answer the study research questions, the evaluation team examined both reading growth among Springboard Summer scholars and reading growth comparing Springboard Summer scholars to non-participating Springboard Summer students.

The dependent/outcome variable is the number of months of reading growth from the pre- to post-program assessments administered by each school district. The districts in the treatment sample shared data from the following assessment types to track student reading progress:

- Norwalk: DIBELS (Oral Reading & Fluency)
- ICS: DRA, EDL, STAR Reading
- Public Prep: STEP
- APS: STAR Reading, STAR Early Literacy
- OUSD: SRI

At ICS, students were assessed using three different tools: DRA, EDL (the Spanish version of DRA) and STAR. The evaluation team included ICS students in the analysis sample only where their two assessments were conducted using the same language with the same tool. Although APS also used two different instruments for reading assessment, both tools are created by the same company and include direct conversions to equate scores from the two instruments within their technical manual. Accordingly, the evaluation team used these conversions rather than eliminating students from the sample when they were assessed first using STAR Early Literacy and then using STAR Reading.

In an ideal approach, the evaluation team would use the same or similar reading assessments across all districts, and would be able to document shifts in scholars' performance following intervention on a normal curve equivalent (NCE) which compares students to a norm of the expected distribution of students across all abilities within a given grade level. Springboard would be deemed successful if it shifted scholars who participated to a higher position on the curve for their grade. While ideal, the evaluation team could not follow this approach, however, because some of the assessments are not norm-referenced.[11] An alternative would be to use scores for comparison youth within each grade to create internally normed data to use as a reference point. Because we didn't have enough non-participating Springboard Summer students within each grade from two of the three districts that provided data on non-participants, the evaluation team could not follow this alternate approach across all of the districts. Instead, because each of these assessments has a unique scale, the evaluation team converted the assessment-specific scores to a common metric of Grade

[9] Different types of assessments were used by each of the five districts. The evaluation team also requested any available conversion data (i.e., stanine, standardized, reading growth in months, GLE).

[10] An APS manager had originally granted approval for sharing whole-school data for Lindley, but during the data collection process Lindley's chief administrator denied access to this data. At the time we were in-taking Springboard Summer scholar data, Norwalk expressed internal capacity issues and data access complications (with the vendor who collected the data in those years) as reasons for not providing whole-school data.

[11] Norm-referenced measures compare a student's scores or performance to those of their peers, the norm group.

Level Equivalents (GLE)[12] using conversion charts so that all data could be analyzed jointly. Despite not being best practice, GLE were used for the dual reason that the evaluation team 1) did not have exclusively norm-referenced tests and 2) because there was not enough comparison data to generate our own internally normed standardizations. To address these limitations, the evaluation team did a series of sensitivity tests with the districts where we did have enough comparison data to generate our own internally normed standardizations - Public Prep.

Another metric limitation relates to how districts shared back Springboard program attendance data. Due to concerns related to student anonymity, one district provided attendance data classified into two groups: scholars who attended 80 percent or more of the scheduled days of Springboard Summer programming and scholars who attended fewer days of programming. Since more fine-grained measures of daily attendance were not provided by all five districts the evaluation team used attendance data at this level.

ANALYSIS

The goal of this study is to determine **how participation in Springboard Summer is associated with reading growth.** To answer this question, the evaluation team explores five distinct sub-questions which are outlined below along with the analysis strategies employed to answer each one.

- Is there any **improvement in reading** for Springboard Summer scholars?
- Does scholars' **improvement vary by starting position relative to grade level**?
- How is **participation dosage associated** with scholars' **reading growth** over a summer?
- What influence does **family workshop participation** have on scholars' **reading growth**?
- How does **reading growth among Springboard Summer scholars compare** to **students within the same district who are not invited or choose not to participate** in Springboard Summer?

RQ1: Is there any improvement in reading for Springboard Summer scholars? For this research question the evaluation team examines if there is a difference between the EOY GLE prior to participation in Springboard Summer and the BOY GLE following participation in Springboard Summer. The evaluation team use a paired t-test to assess whether any differences in GLE from EOY to BOY are statistically significant.

RQ2: Does scholars' improvement vary by starting position relative to grade level? To address this research question, students are classified into one of three categories according to their EOY district assessment: at/above grade level, within one grade below grade level, or more than

[12] The GLE score is expressed as a decimal number with the number on the left representing the year of the grade level and the number to the right of the decimal representing the month. For example, when a student has a GLE score of 3.2 in Track My Progress, this means that the student scored as well as the average student in the second month of third grade.

one grade below grade level. Within each of these categories, the evaluation team examines if there is a difference between the EOY GLE prior to participation in Springboard Summer and the BOY GLE following participation in Springboard Summer. The evaluation team use a paired t-test to assess whether any differences in GLE from EOY to BOY are statistically significant. Kindergarten students are not included in this analysis because there is not sufficient variation in their starting point relative to grade level as all are starting from relatively comparable positions.

RQ3: How is participation dosage associated with scholars' reading growth over a summer? Springboard's internal studies indicate that participation dosage tends to have a positive relationship with reading growth. Since more fine-grained measures of daily attendance were not provided by all five districts, the evaluation team examines the relationship within the current sample of treatment students by first using data from Springboard Summer to classify participants into one of two groups: scholars who attended 80 percent or more of the scheduled days of programming and scholars who attended fewer days of programming. Here, the evaluation team calculates the dependent variable as the change in GLE from the EOY to Springboard Summer to the BOY after Springboard Summer for each student. Then, within each grade, t-tests for a difference of means assess if scholars who meet the target threshold of attending at least 80 percent of the scheduled Springboard Summer sessions show greater average change in reading as measured by grade level equivalents than students who do not meet the target threshold.

In addition, recognizing that the t-tests did not allow the evaluation team to control for other characteristics that varied across students in the treatment sample, and that these differences may have masked an actual effect of meeting the attendance threshold, the evaluation team estimated a series of regression models within each grade predicting the change in grade level equivalent on the reading assessment. Meeting the attendance threshold was the predictor of interest, and the evaluation team controlled for starting point relative to grade level, gender, having an IEP, being an English language learner, testing in Spanish, participation year and district.

Finally, although a continuous representation of Springboard Summer attendance could not be linked to district assessment data within all districts, the evaluation team were able to test whether a continuous representation of attendance rather than the dichotomous representation of meeting the 80 percent attendance threshold or not was related to change in GLE among students within the Public Prep district.

RQ4: What influence does family workshop participation have on scholars' reading growth? A core component of Springboard Summer is family engagement. Each Springboard Summer site offers four to five family workshops (with four main workshops and one potential make-up). Springboard believes that attending 80 percent of the family workshops is sufficient for engagement and program success. Similarly to the analyses testing the relationship between Springboard Summer scholar attendance and reading growth, the evaluation team explored the relationship between family workshop participation and reading growth within the current sample of treatment students by first using data from Springboard Summer to classify participants into one of two groups: students whose families attended 80 percent or more of the scheduled family workshops, and students whose families attended fewer. The evaluation team use the same dependent variable, the change in GLE from the EOY prior to Springboard Summer to the BOY after Springboard Summer for each student, and within each grade, use t-tests for a difference of means to assess whether scholars whose families meet the target threshold of attending at least 80 percent of the scheduled family workshops show greater average change in reading as measured by grade level equivalents than scholars whose families do not meet the target threshold.

Recognizing that the t-tests did not allow the evaluation team to control for other characteristics that varied across students in the treatment sample, and that these differences may have masked an actual effect of meeting the family workshop attendance threshold, the evaluation team estimated a series of regression models within each grade predicting the change in GLE on the reading assessment. Meeting the family workshop attendance threshold was the predictor of interest, and the evaluation team controlled for starting point relative to grade level, gender, having an IEP, being an English language learner, testing in Spanish, participation year, and district.

RQ5: How does reading growth among Springboard Summer scholars compare to students within the same district who are not invited or choose not to participate in Springboard Summer? Since participants were not randomly assigned to Springboard Summer programming, the evaluation team use a form of propensity score matching to identify the comparison group that is most similar to the group of students who participated in Springboard Summer. A propensity score is an estimate of the probability that an individual would receive treatment given his/her characteristics and experiences. This score is used to create comparability among groups of Springboard Summer scholars and students who did not participate in Springboard Summer along a single numerical dimension, to the extent that differences between the groups are captured in the available data. Within each grade level, propensity score matching with replacement is used across the following measured variables to identify the subset of non-treatment students with available assessment data who have the greatest propensity to have received treatment: reading score at EOY, English language learner status, IEP status, gender, percent of the residential zip code living below poverty, testing in Spanish, and school district. The evaluation team required exact matches on the final two variables.

Within each grade, the resulting matched comparison group is statistically equivalent to the treatment group, as measured on each of the seven characteristics used to calculate the propensity score. Because the comparison group constructed in this way has baseline equivalence to the treatment group, the evaluation team use t-tests for a difference of means to assess whether students who participate in Springboard Summer show greater average change in reading than students in the matched comparison group. In all cases, weights calculated by SAS within the propensity score matching procedure are applied in all subsequent t-tests.

While classification of research studies by quality of evidence into ESSA tiers is not completely straightforward, using the criteria outlined by Neild et al. (2019), this study meets ESSA Tier 2 for Moderate Evidence.[13] The study is quasi-experimental. For second, third, and fourth grades, it meets the standard for baseline equivalence outright with no required statistical adjustments, as the difference between the treatment and matched comparison group on the reading score at EOY is less than .05 times the standard deviation of the score for the treatment group. For first grade, the study similarly meets the Tier 2 standards because the difference between the treatment and matched comparison group on the reading score at EOY is less than .25 times the standard deviation of the score for the treatment group, and the evaluation team use the acceptable statistical adjustment of simple mean scores. The study does not meet the standards for Tier 2 for kindergarten, however, because of a provider or administrative unit confound due to the fact that all kindergarten matched treatment and comparison students are from a single school district, Public Prep. The study demonstrates a statistically significant positive effect of Springboard Summer on

[13] Neild, R.C., Wilson, S.J., & McClanahan, W. (2019). Afterschool programs: A review of evidence under the Every Student Succeeds Act. Philadelphia: Research for Action.

reading levels, and does not include any strong negative findings. The study also meets the sample size and setting requirements for Tier 2.

EVALUATION FINDINGS

HIGH-LEVEL FINDINGS

Analyses found positive results for Springboard Summer scholars:

- Across all grades, on average, scholars who participated in Springboard Summer showed improvement on their reading assessment score from the end of the school year before Springboard Summer to the start of the following school year.
    - The largest gains were for scholars who started below grade level.
    - Given high scholar and parent attendance, the evaluation team could not discriminate between the effect of high and low attendance on reading growth.

- Springboard Summer scholars show larger improvements in reading scores, on average, when compared to similar students who did not participate. The quality of evidence from this study satisfies the Every Student Succeeds Act (ESSA) Tier 2 standards.

DETAILED FINDINGS

The first group (research questions 1-4) of findings examine reading growth as measured by the district administered reading assessments for Springboard Summer scholars only, with all sample districts. Later, the evaluation team examine how they compare to non-participants (research question 5). Since Public Prep made up a large proportion of the study sample, Public Prep's findings are briefly discussed in the main body of the report for each research question, specifically to see if they differed from the all-district analysis. Full results from the Public Prep-only sensitivity analyses can be found in Appendix B.

Research Question 1: Is there any improvement in reading for Springboard Summer scholars?

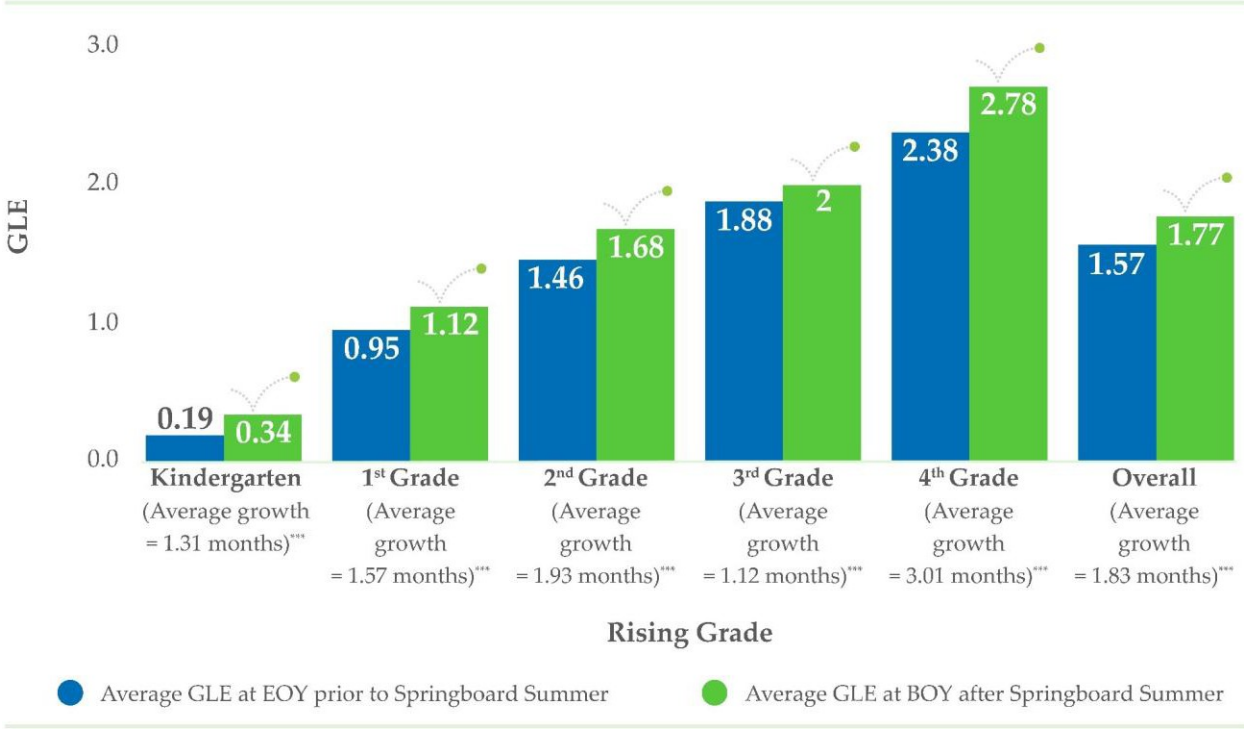
All Springboard Summer Scholars

Overall, findings indicate that there is improvement in reading for scholars who participate in Springboard Summer. Figure 3 presents the results of overall and within grade paired t-tests of the difference in mean grade level equivalent on the reading assessment at the EOY before Springboard Summer participation (in blue) and at the BOY following Springboard Summer participation (in green).

Within each grade, and overall, the average change is positive and statistically significant, suggesting that, on average, scholars who participated in Springboard Summer performed more favorably on the district administered reading assessment at the beginning of the school year than they had on the district administered reading assessment at the end of the school year before Springboard Summer.

Figure 3 shows, **on average, scholars showed about 1.8 months of reading growth (ranging between approximately 1 month for third graders and 3 months for fourth graders).**[14]

FIGURE 3. Average Change in District Reading Assessment from EOY to BOY after Springboard Summer Participation



Public Prep Only

When only scholars from Public Prep are considered, the findings are similar. **On average, scholars who participated in Springboard Summer at Public Prep increased their reading level between the end of school year and beginning of year district assessments.** See Appendix B, Section I for full Public Prep results.

Research Question 2: Does scholars' improvement vary by starting point relative to grade level?

All Springboard Summer Scholars

The second research question explored is, do scholars who start Springboard Summer already reading on grade level *and* do scholars who start Springboard Summer reading below grade level *both* show improvement, on average, after participating in the program? **The next few figures demonstrate the answer is mixed: in some cases all scholars show statistically significant reading growth, and in other cases, only scholars who are behind grade level show it.**

[14] The numbers in the bars represent a grade level equivalent score. These differences between the green and blue bars are translated into months of reading growth as follows: (green GLE - blue GLE) * 9 months, so for the overall bars: (1.77 - 1.57)*9 = 1.80. This differs slightly from the 1.83 in the chart label because the numbers in the bars were not rounded for the original calculations.

Scholars are classified according to their EOY district assessment as at/above grade level, within one grade level below grade level, or more than one grade below grade level. The following figures present the results of paired t-tests within each of these categories within grade to assess if scholars show similar average change in reading as measured by grade level equivalents whether they start at/above grade level, within one grade level below, or farther below grade level. Kindergarten scholars are not included in this analysis because there is not sufficient variation in their starting point relative to grade level as all are starting from relatively comparable positions.

Figure 4 shows **for rising first graders, both scholars who started at or above grade level and those who joined the program reading below grade level, on average, experienced improvement in their reading level** following participation in Springboard Summer.

FIGURE 4. Change in Reading Assessment by Starting Level Relative to Grade Level for *Rising First Grade Scholars*

Public Prep Only

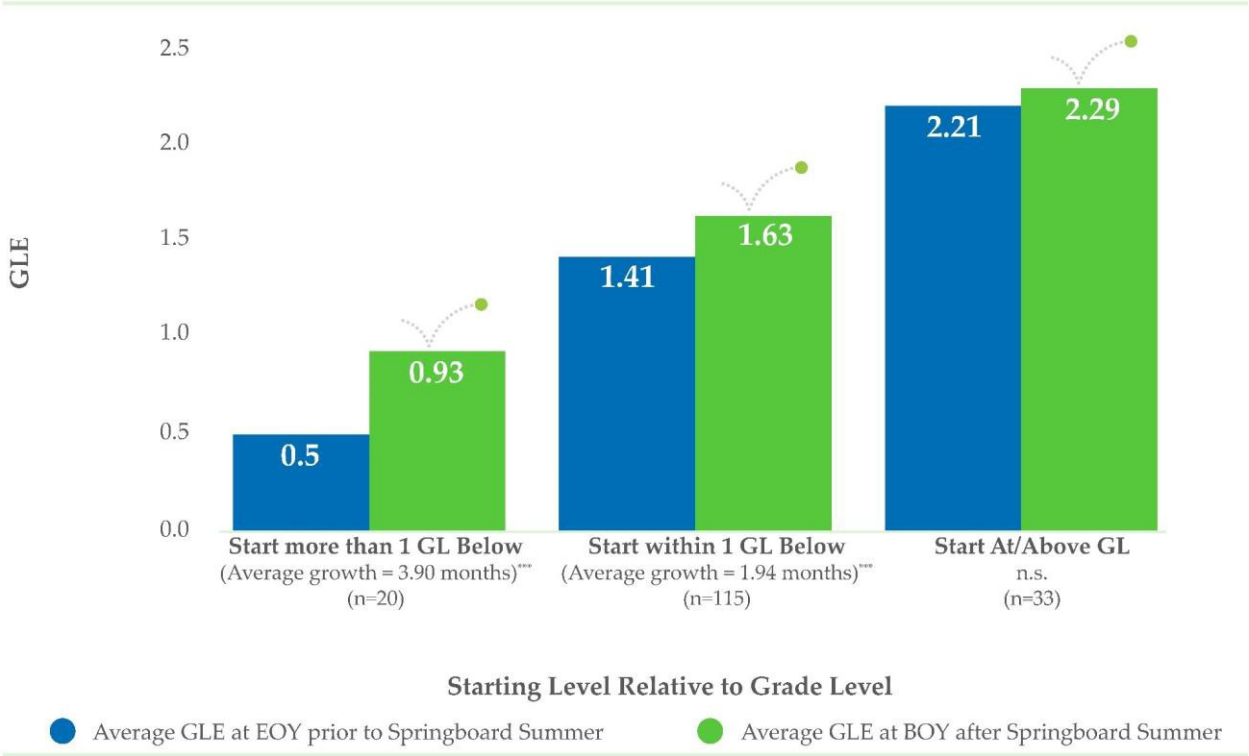
Rising first grade scholars who participated in Springboard Summer within the Public Prep district, on average, experienced improvement in their reading level following participation in the program whether they started at or above grade level as well as if they joined the program reading below grade level. See Appendix B, Section II for full Public Prep results.

All Springboard Summer Scholars

Figure 5 demonstrates that **rising second grade scholars who began the program reading more than a year below their grade level and those who began the program reading within a year below their grade level, on average, experienced improvements in their reading ability** when assessed at the start of the year following their participation in Springboard Summer. On average, scholars who began the program already reading at or above grade level did not experience

significant changes in their reading level; however, the number of scholars in this category is small, which makes it hard to detect statistically significant change.

FIGURE 5. Change in Reading Assessment by Starting Level Relative to Grade Level for *Rising Second Grade Scholars*



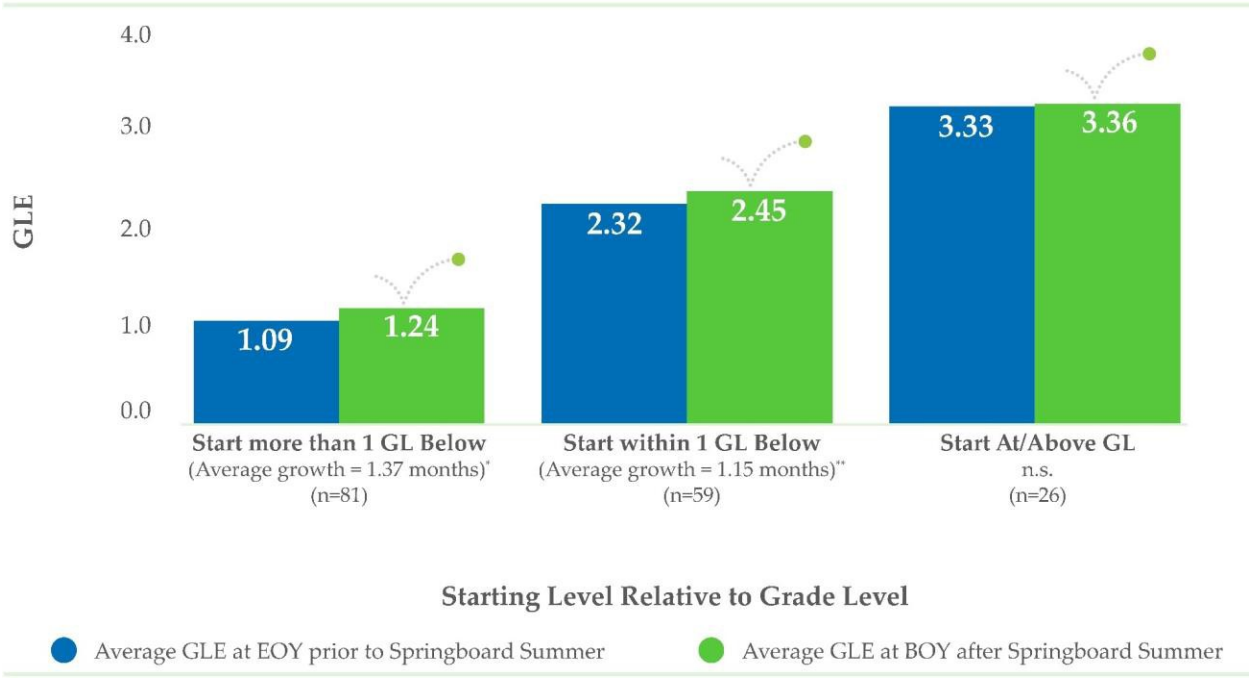
Public Prep Only

For Public Prep only, very few rising second grade scholars began the program reading more than a year below their grade level. The data suggest that these scholars experienced substantial growth, but with a sample of only three students, the change reaches only borderline significance. **Scholars who began the program reading within a year below their grade level and those who began reading at or above grade level, on average, both experienced improvements in their reading ability when assessed at the start of the year following their participation in Springboard Summer.** See Appendix B, Section II for full Public Prep results.

All Springboard Summer Scholars

Figure 6 shows that **rising third grade scholars who began the program reading more than a year below their grade level and those who began the program reading within a year below their grade level, on average, experienced improvements in their reading ability** when assessed at the start of the year following their participation in Springboard Summer. On average, scholars who began the program already reading at or above grade level did not experience statistically significant changes in their reading level.

FIGURE 6. Change in Reading Assessment by Starting Level Relative to Grade Level for *Rising Third Grade Scholars*



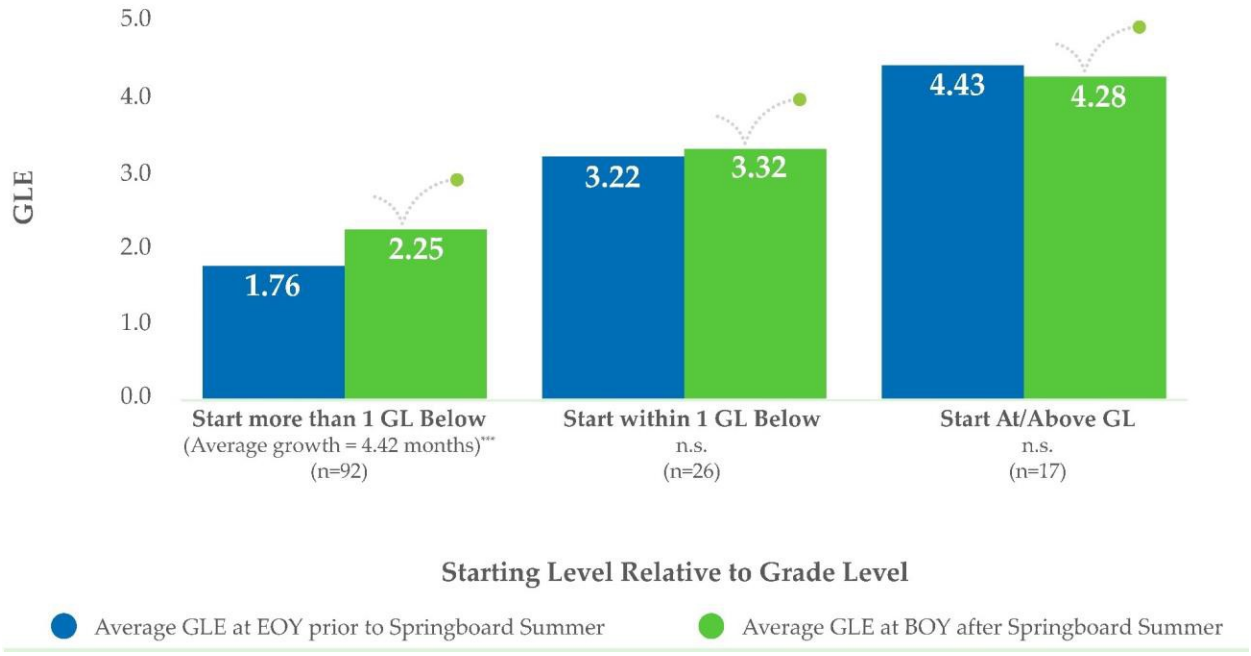
Public Prep Only

Rising third grade scholars at Public Prep experienced improvements in their reading ability when assessed at the start of the year following their participation in Springboard Summer whether they began the program reading at or above grade level, within one grade level below, or more than one grade level below. See Appendix B, Section II for full Public Prep results.

All Springboard Summer Scholars

Figure 7 shows that **rising fourth grade scholars who began the program reading more than a year below their grade level, on average, experienced improvements in their reading ability when assessed at the start of the year following their participation in Springboard Summer**. On average, scholars who began the program reading within one year below grade level and those already reading at or above grade level did not experience statistically significant changes in their reading level.

FIGURE 7. Change in Reading Assessment by Starting Level Relative to Grade Level for *Rising Fourth Grade Scholars*



Public Prep Only

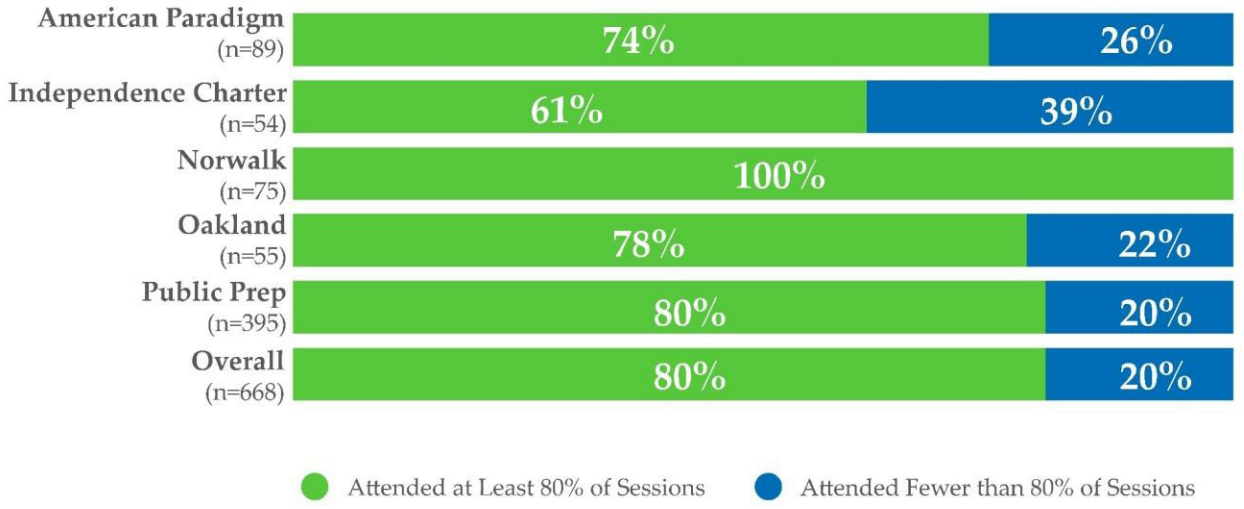
On average, **rising fourth grade scholars at Public Prep who entered the summer reading below grade level, experienced improvements in their reading ability when assessed at the start of the year following their participation in Springboard Summer.** On average, scholars who began the program already reading at or above grade level did not experience significant changes in their reading level. See Appendix B, Section II for full Public Prep results.

Research Question 3: How is participation dosage associated with scholars' reading growth over a summer?

All Springboard Summer Scholars

Across the treatment sample, attendance at Springboard Summer is high. One school district was particularly concerned with scholar anonymity and would only share Springboard Summer participation data back as a dichotomous cut (80 percent of sessions attended or more versus less than 80 percent of sessions attended). Because of this, the evaluation team were unable to explore attendance as a continuous representation across all districts. Since 80 percent of the sample with matched assessment data has met 80 percent or more of scheduled sessions, the sample has little variation in scholar program attendance.

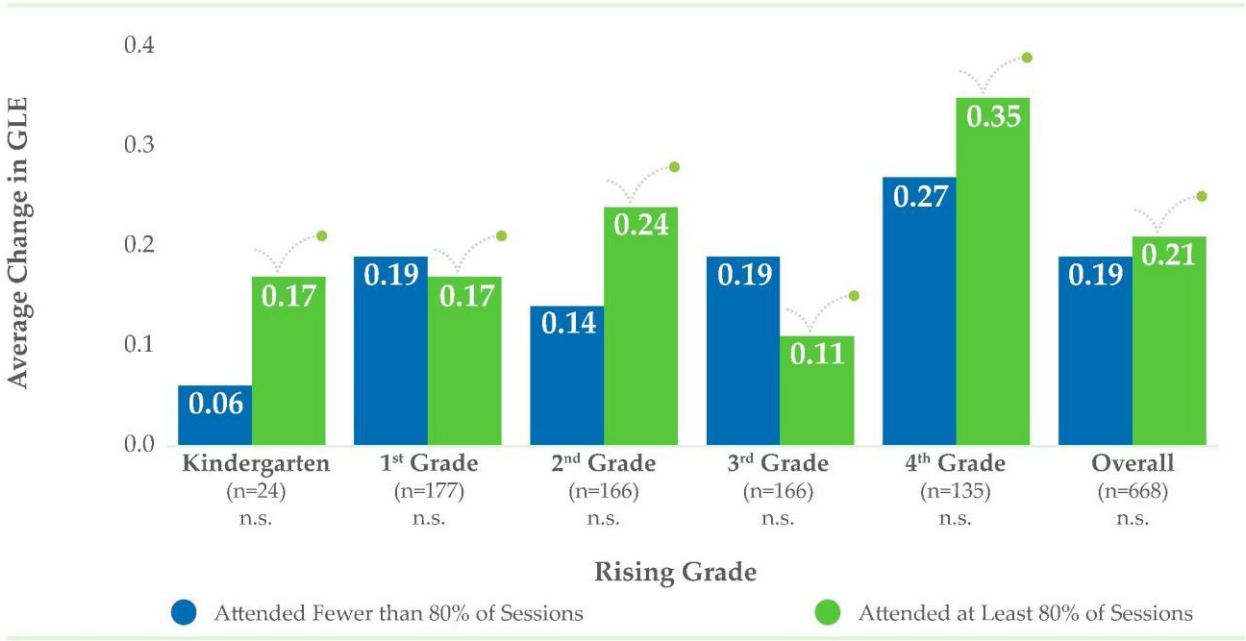
FIGURE 8. Percentage of Scholars that Attended 80% or More Sessions (by District)



Within each grade, t-tests for a difference in means assess if scholars show similar average change in reading as measured by grade level equivalents whether or not they meet the threshold of attending at least 80 percent of the scheduled Springboard Summer sessions. As noted in the previous chart, 80 percent of scholars do meet the threshold. Only 133 scholars do not meet it. This includes 6 rising kindergarten scholars, 47 rising first grade scholars, 29 rising second scholars, 23 rising third grade scholars, and 28 rising fourth grade scholars.

Figure 9 below, presents the findings by grade level. Both overall and within each grade, **the average change in GLE was not significantly different for scholars who meet the Springboard Summer participation threshold**. While high attendance is indicative of a strong program, and beneficial to participating scholars, it has one drawback in that it results in minimal variations in attendance across the sample making it more difficult to detect statistically significant differences in reading growth by attendance level. This does not mean that scholar attendance is not useful. It may be that there is just too little variation in scholar attendance within the sample to be able to discriminate between effects for high and low attenders.

FIGURE 9. Average Change in Reading Level within Grade by Scholar Attendance at Springboard Summer Sessions



The t-tests did not allow us to control for other characteristics that varied across scholars in the treatment sample, and these differences may have masked an actual effect of meeting the attendance threshold. To explore this, the evaluation team estimated a series of regression models within each grade predicting the change in grade level equivalent on the reading assessment. Meeting the attendance threshold was the predictor of interest, and the evaluation team controlled for starting point relative to grade level, gender, having an IEP, being an English language learner, testing in Spanish, participation year, and district. Within each grade, the evaluation team retained control variables only if they added predictive value to the model. Detailed results of these analyses are presented in Appendix A, Section II, and tell the same story as the t-tests. Even after controlling for other characteristics of scholars in the treatment sample, there is not a significant effect on change in grade level equivalent in reading associated with meeting the Springboard Summer attendance target.

Public Prep Only

Both overall and within each grade, the average change in GLE is not significantly different between Public Prep scholars who meet the Springboard Summer attendance threshold as compared to those who do not. See Appendix B, Section III for full Public Prep results.

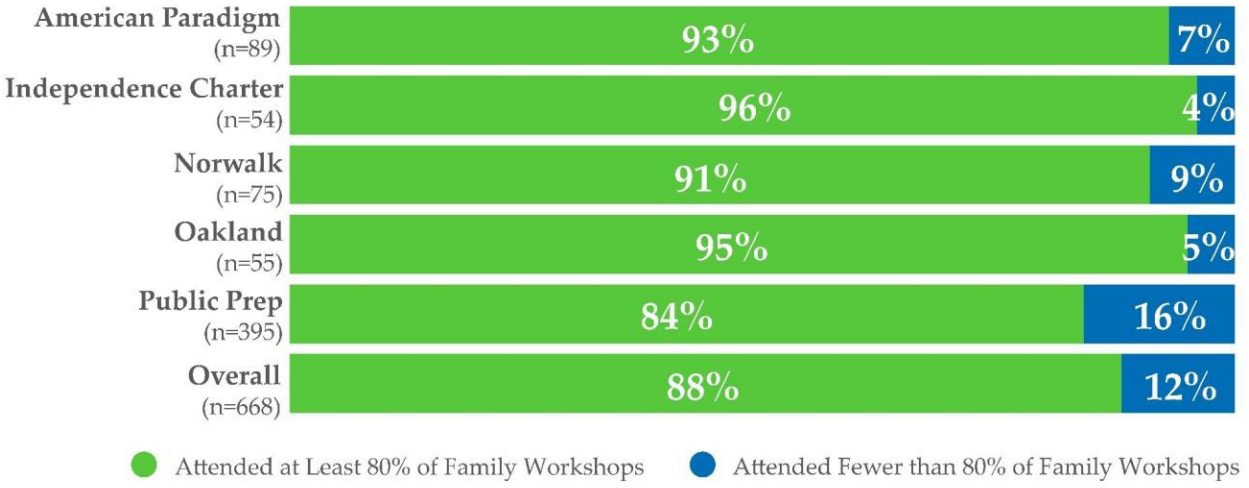
The evaluation team also explored Springboard Summer attendance on a continuous scale for Public Prep scholars (as opposed to meeting the 80 percent threshold). For the subset of treatment students who participated in Springboard Summer within the Public Prep district, **a continuous representation of Springboard attendance is not significantly related to change in GLE in reading within any of the five grade levels.** Reduced regression models exploring the relationship between continuous representation of Springboard Summer attendance and change in reading level by grade for Public Prep can be found in Appendix B, Section III.

Research Question 4: What influence does family workshop participation have on scholars' reading growth?

All Springboard Summer Scholars

Across the treatment sample, attendance at Springboard Summer family workshops is high. Springboard has set a family participation target of attending 80 percent or more of the scheduled family workshops, and as Figure 10 shows, 88 percent of the sample with paired assessment data has met this target.

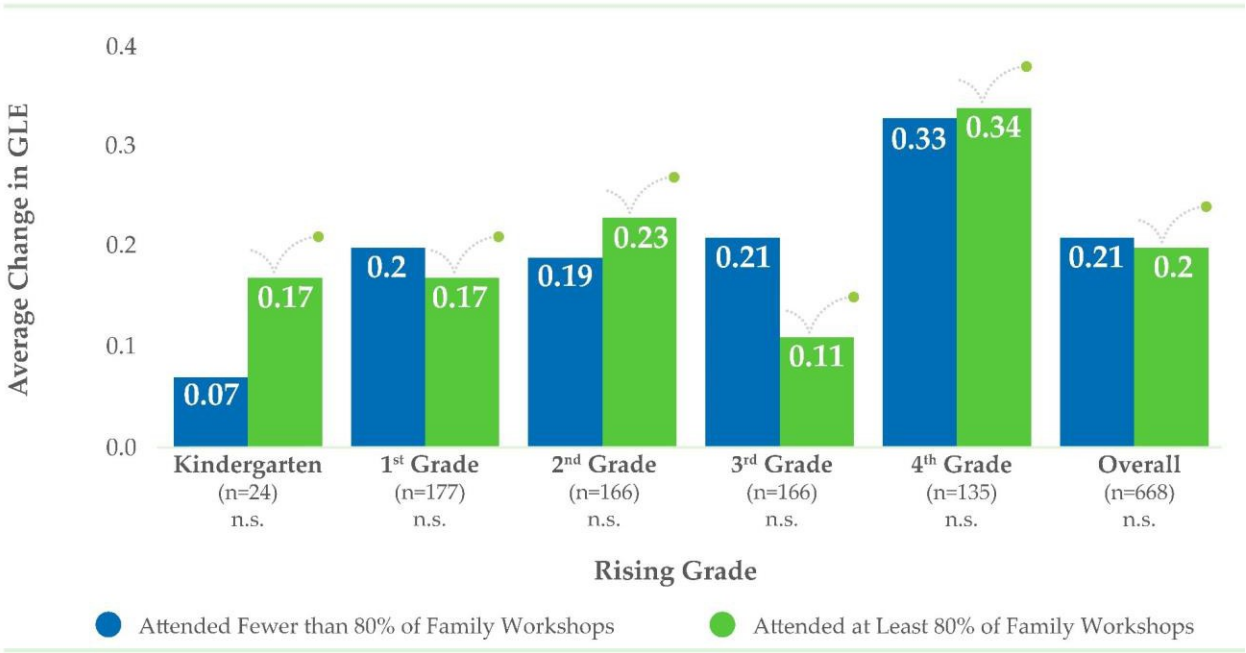
FIGURE 10. Percentage of Scholars Whose Families Attended 80% or More of the Family Workshops (by District)



Within each grade, t-tests for a difference of means assess if scholars show similar average change in reading as measured by grade level equivalents whether or not their families met the target threshold of attending at least 80 percent of the scheduled Springboard Summer family workshops. As noted in the previous chart, the families of most scholars meet this target. Only 64 scholars do not meet the family workshop attendance target. This includes 5 rising kindergarten scholars, 26 rising first grade scholars, 17 rising second grade scholars, 20 rising third grade scholars, and 12 rising fourth grade scholars.

Figure 11 below, shows that both overall and within each grade, **the average change in GLE is not significantly different for scholars who meet the family workshop attendance threshold as compared to those who do not.**

FIGURE 11. Average Change in Reading Level within Grade by Attendance at Springboard Summer Family Workshops



As was the case when exploring the effect of Springboard Summer scholar attendance on reading change, the evaluation team recognize that the t-tests did not allow us to control for other characteristics that varied across scholars in the treatment sample, and that these differences may have masked an actual effect of meeting the family workshop attendance threshold. To explore this, the evaluation team estimated a series of regression models within each grade predicting the change in grade level equivalent on the reading assessment. Meeting the family workshop attendance threshold was the predictor of interest, and the evaluation team controlled for starting reading level relative to grade level, gender, having an IEP, being an English language learner, testing in Spanish, participation year, and district. Within each grade, control variables were retained only if they added predictive value to the model. Results are presented in Appendix A, Section III.

Even after controlling for other characteristics of scholars in the treatment sample, there is not a significant effect on change in GLE in reading associated with meeting the Springboard family workshop attendance target. ***This does not mean that the family workshops are not useful.*** High workshop attendance is characteristic of Springboard programming, and for this study the evaluation team chose strong implementers of the Springboard Summer program model. While useful for students, minimal variations in family workshop attendance across the sample makes it difficult to detect statistically significant differences in reading growth by attendance level.

Public Prep Only

Findings for only Public Prep Springboard Summer scholars were similar. Within each grade, **the average change in GLE for Public Prep scholars who participated in Springboard Summer is not significantly different for scholars who meet the Springboard family workshop attendance threshold as compared to those who do not,** although all differences are in the same

direction suggesting more favorable results from participation. **When all grades are considered jointly, the difference reaches statistical significance.** Again, the evaluation team followed this analysis by controlling for other characteristics of scholars in the treatment sample. **Even after controlling for other characteristics of scholars in the treatment sample, there is not a significant effect on change in GLE in reading associated with meeting the Springboard family workshop attendance target.** See the Appendix B, Section IV for full Public Prep results.

Research Question 5: How does reading growth among Springboard Summer scholars compare to students within the same district who are not invited or choose not to participate in Springboard Summer?

Through this point, we have only been exploring the experience of Springboard Summer scholars. Next, the evaluation team compare them to students who were not offered the option to or chose not to participate in Springboard Summer.

We received data on non-participating Springboard Summer students from three school districts:

- Public Prep
- ICS
- OUSD

Appendix A, Section IV presents the number of Springboard scholars and comparison students by district.

Within each grade, the treatment and full sample of potential comparison students are significantly different on one or more characteristics. The differences in the starting reading level are most concerning. Appendix A, Section IV presents these differences. Through propensity score matching the evaluation team identify the subset of potential comparison students who have baseline equivalence to the treatment scholars and use this sample for analysis to compare reading growth for Springboard Summer scholars and the similar students within their district who did not participate. These analyses are followed by two variations to test the sensitivity of results across various scenarios.

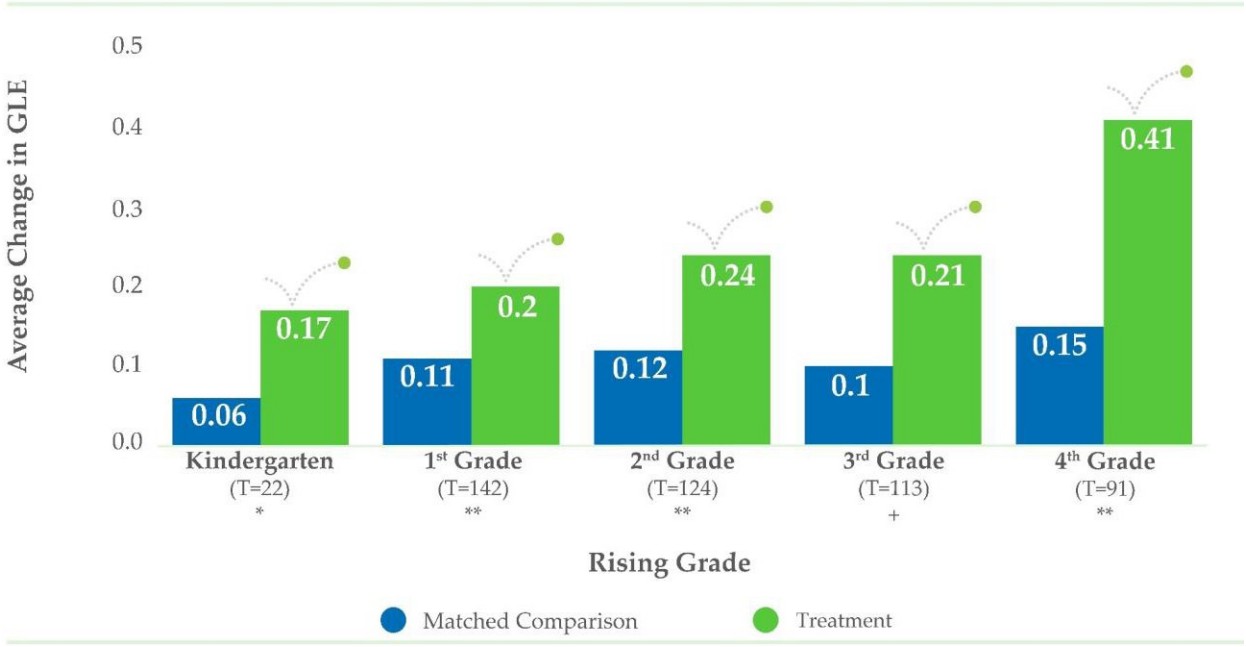
All Springboard Summer Scholars: In-district Matches

Within each grade level, the evaluation team use propensity score matching with replacement across the following measured variables to identify the subset of non-treatment students with available assessment data who have the greatest propensity to have received treatment: reading score at EOY, English language learner status, IEP status, gender, percent of the residential zip code living below poverty, testing in Spanish, and school district. The evaluation team required exact matches on the final two variables. This process identifies the subset of potential comparison students who are most like the Springboard Summer scholars at EOY, and these students are then defined as the matched comparison group. Because the evaluation team required matches to be in-district, Springboard Summer scholars at Norwalk and APS, neither of which provided data for comparison students, are not included in this analysis. Appendix A, Section IV presents the resulting treatment and weighted matched comparison sample.[15]

[15] After propensity score matching, the resulting weighted matched comparison group is not statistically significantly different from the treatment group on any of the matching characteristics.

Once the matched comparison sample was identified, the evaluation team explored whether this group performs differently on the district assessments for scholars who participated in Springboard Summer. Figure 12 shows the results of these comparisons and indicates that **the treatment sample shows larger improvements, on average, in their reading score when compared to students in the matched comparison group. For rising kindergarten, first, second, and fourth grade scholars, these differences are large enough to be considered statistically significant**, while for rising third grade scholars they are marginally significant.

FIGURE 12. Average Change in Reading Level within Grade by Treatment Status when Restricted to In-District Matches



Our data were imperfect: the evaluation team did not have comparison data from every district we requested it from, each district used a different reading assessment, and more than one-half of the data is from Public Prep. Due to this, and the fact that we could not convert to NCE across all districts, the evaluation team conducted several different variations of the analysis to test the sensitivity of the results across various scenarios. The findings are presented in the following sections.

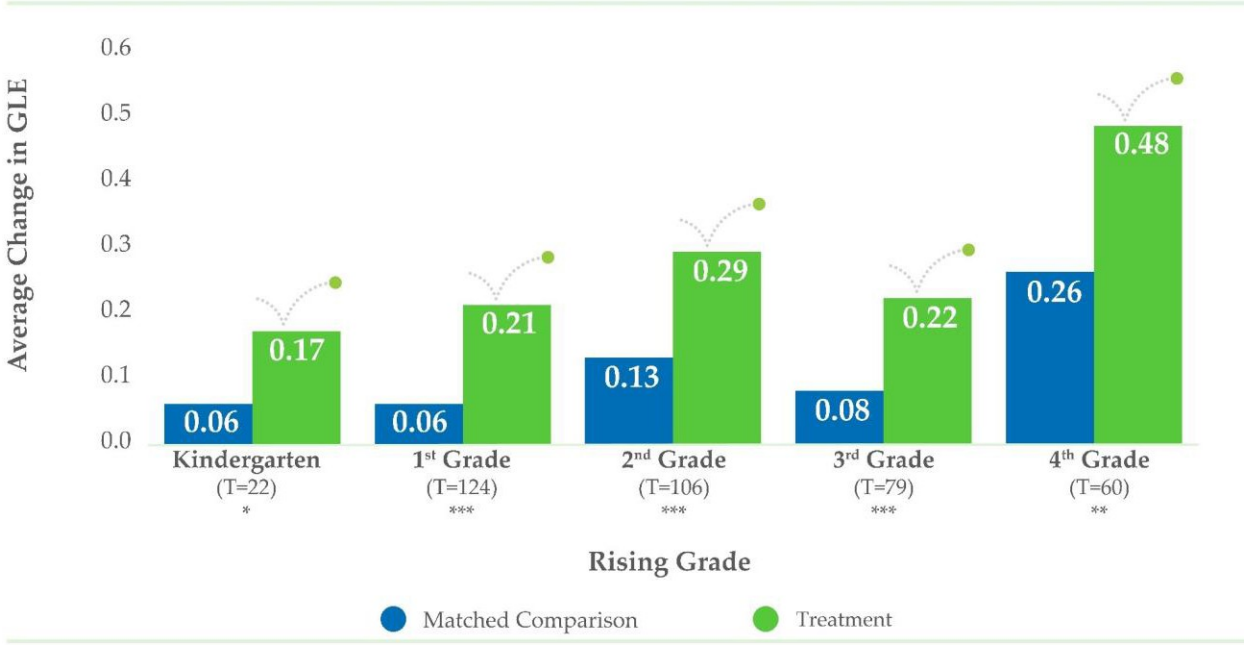
Sensitivity Analysis 1: Public Prep GLE

Since Public Prep made up a large portion of the study sample, the evaluation team examined if we found the same results when including only the Public Prep scholars and non-participants in the analysis sample. The evaluation team began the analysis with propensity score matching with replacement to identify the appropriate matched comparison sample. See Appendix A, Section IV for a comparison of the Public Prep treatment and matched comparison sample on the baseline reading assessment and measured demographic characteristics within each grade.[16]

[16] After propensity score matching, the resulting weighted matched comparison group was statistically the same as the treatment group on each of the matching characteristics.

In Figure 13, we see similar findings - **when the sample is restricted to scholars in the Public Prep district, within each grade, the treatment sample shows larger improvements in their reading score than students in the matched comparison group**.

FIGURE 13. Sensitivity Analysis 1: Average Change in Reading Level within Grade by Treatment Status | Public Prep Only



Sensitivity Analysis 2: Public Prep Standardized Scores

Public Prep uses the STEP assessment which is not norm-referenced. In order to test the robustness of our findings, the evaluation team mimic a norm-referenced analysis by using the full range of assessment scores for non-participating Springboard Summer Public Prep students within each grade to calculate standardized scores. To calculate the standardized dependent variable, we calculate the average score and standard deviation within a grade for all students who did not participate in Springboard Summer. Then, we use this mean and standard deviation to calculate the standardized score for the two time points for both comparison students and treatment scholars using the following equation:

> Standardized score = (Student STEP score at time point - Average STEP score for comparison student in that grade at that time point)/Standard Deviation of STEP scores for comparison student in that grade at that time point.

Using this formula, the resulting average standardized score at each time point for comparison students is 0. Students who score below that have a negative standardized score, and students who score above that average for students in their grade, would have a positive standardized score.

If students all learn at the same rate, their relative positions on the standardized scale would be consistent across time. In contrast, if scholars who participated in Springboard Summer learned more, their position on the standardized scale would move to a more positive position, or to a less

negative position, if they were improving but still performing below the average non-participating student in their grade.

Figure 14 on the following page shows within grade comparisons between treatment and matched comparison students at Public Prep on these standardized reading scores, rather than the GLE used in the previous analyses. To understand the figure below, first look at the rising second grade students. From the green bar, we see that rising second grade scholars who participated in Springboard Summer improved their standardized assessment score 0.35 standard deviations on average, while the matched comparison group student who did not participate in Springboard Summer improved over the same time period by only 0.03 standard deviations, on average.

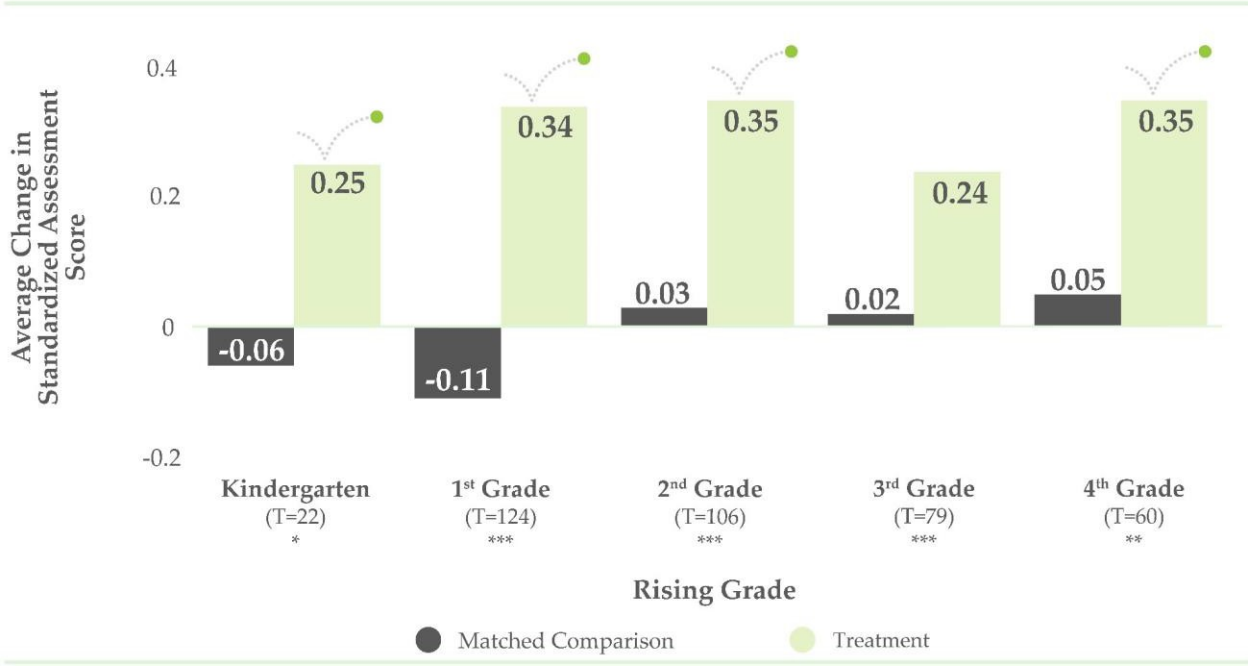
As shown in the first line of Table T11 in Appendix B, Section VI, both groups were performing worse than the average non-participating second grader at the end of the school year, with the average second grade treatment scholar scoring 0.41 standard deviations below the average, and the average matched comparison second grade student scoring 0.43 standard deviations below average. As shown in Figure 14, after the summer however, the average non-participating rising second grader in the matched comparison group changed their relative position very little, improving only 0.03 standard deviations from -0.43 to -0.40 standard deviations. The rising second grade scholars who participated in Springboard Summer, however, saw an improvement of 0.35 standard deviations from -0.41 to -0.06 standard deviations which, while still slightly below the average for non-participants in the grade level (which is always 0 by definition), indicates a shift to a level of performance much closer to the average second grader in the full comparison group.

See Appendix B, Section VI for a comparison of the treatment and matched comparison sample that demonstrates baseline equivalence of the two samples on the baseline reading assessment and the measured demographic characteristics.[17]

Within each grade, when scholars who participated in Springboard Summer at Public Prep are compared with matched comparison students who did not, **scholars showed significantly higher growth between the end of school year assessment before Springboard Summer to the start of school year assessment after Springboard Summer than their non-participating peers.**

[17] After propensity score matching, the resulting weighted matched comparison group was statistically the same as the treatment group on each of the matching characteristics.

FIGURE 14. Sensitivity Analysis 2: Average Change in Standardized Reading Level within Grade by Treatment Status | Public Prep Only



Findings of Public Prep standardized scores analyses for research questions 1-4 can be found in Appendix B, Section VI.

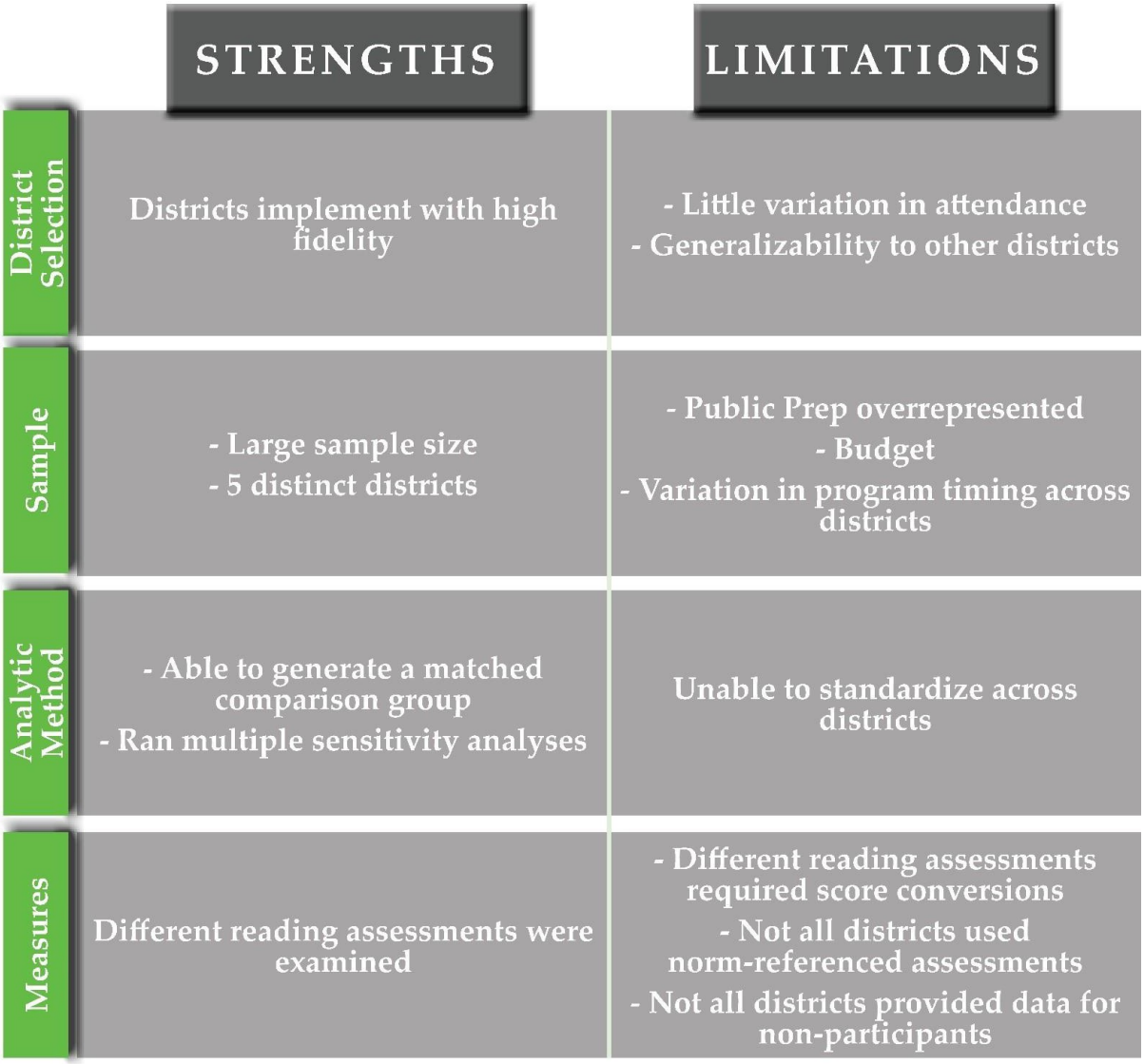
EVALUATION DISCUSSION

This study shows that Springboard Summer scholars from districts implementing Springboard Summer with high fidelity in 2018 and 2019 show reading growth - across all grades Kindergarten through 4th. The largest gains were for scholars who started below grade level. Furthermore, reading growth is larger for scholars than matched comparison peers who were not invited to or chose not to participate in Springboard Summer in 2018 and 2019.

STUDY STRENGTHS AND LIMITATIONS

In this section some notable strengths of our study are discussed, and although the evaluation team worked to counter many study limitations, we discuss a few caveats to the study and design approach. Figure 15 below, summarizes key strengths and limitations broken out by some study facets. The section to follow discusses each of these in more detail.

FIGURE 15. Study Strengths and Limitations



District Selection. The evaluation team was purposeful in selecting districts that implemented Springboard Summer with high fidelity and this is a major strength of this study. This is an external test of the Springboard Summer model, specifically focusing on districts where the program is implemented as intended, reducing the chance of findings being a result of poor implementation. Because we wanted to select school districts with high fidelity to the Springboard Summer program model, our sample reported very high program and family workshop attendance rates, making it more difficult to examine how the variation in attendance rates affects reading outcomes for youth. The selection of strong implementers also means that the results of this study may not be generalizable to other districts where there is lower implementation fidelity.

Sample. The number of districts included in this study is limited by the project budget, study timeline, and district capacity. Working with districts in the sample to secure the appropriate data was time intensive. It required the evaluation team to collaborate with districts in establishing data-

sharing agreements, to convey data request expectations (sometimes multiple times), and to follow-up in order to secure the necessary data (oftentimes needing to reconnect with districts multiple times following the initial transfer of data). Ultimately, the sample comprised 673 rising Kindergartners through 4th graders across five distinct districts, which was a large enough sample size to detect both significant improvements in reading across time as well as differences in reading growth for scholars and comparison students who did not participate in Springboard Summer. The sample does overrepresent the Public Prep school district though, as they made up a generous portion of it (58.7%). Additionally, districts in the sample varied in terms of when district-administered assessments were given and when Springboard Summer occurred. The evaluation team recognize our estimates may be influenced by the timing of the program in relation to the summer months (i.e., whether programming took place at the beginning, middle, or end of the summer months). If scholars participated in Springboard Summer early in the summer (at a time point close to the end of year reading assessment) they may benefit for the remainder of the summer (e.g., because they are motivated to continue to work on their reading for the remainder of the summer once programming is over or because their parents, who have participated in Springboard Summer family workshops, are better equipped to continue to reinforce reading education post-program). On the other hand, if programming takes place toward the end of the summer, there could be risk for summer slip during the gap in time between EOY district reading assessments and programming. Exploring how the timing of programming influenced estimates is beyond the scope of this research.

Analytic Method. The evaluation team explored several research designs, including a natural experiment and regression discontinuity design (RDD). However, the aforementioned designs require specific types of selection methods. For example, a natural experiment would have provided the most rigorous design, but would have required that scholars were selected via lottery (random selection) only. An RDD requires a cutoff point (based on assessment/scoring), for which the treatment group can consist of *only* participants whose pretest scores were below the cutoff point and the control group consists of *only* participants whose pretest scores were above the cutoff point; while many Springboard Summer programs targeted students who had lower assessment scores (for which an RDD would have been a viable option), several of these program sites then opened enrollment to all students, which eliminated the RDD option. Propensity score matching is the best fit and most feasible design given these limitations. A major strength of our analysis is that we were able to generate a matched comparison group. However, it is important to note that the quality of the matches between scholars and potential comparison group members is dictated by the availability of data on potential confounders. In this analysis we had access to important and obvious confounders, most notably baseline reading score. However, it's possible that we were not able to account for other important confounders because data on them was not available to us. Another major strength related to analysis is that the evaluation team conducted a series of sensitivity tests with the district where we did have enough comparison data to generate our own internally normed standardizations (Public Prep).

Measures. A strength to this study is that it examines multiple types of reading assessments. As a metric limitation, each district used a different reading assessment (i.e., DIBELS, STEP, STAR, SRI, DRA), some of which are not norm-referenced.[18] The non-norm-referenced assessments did not lend themselves to easy conversions to standardize the scores across assessments for reading growth analysis. Because each of these assessments has a unique scale, we converted the assessment-specific

[18] Norm-referenced measures compare a student's scores or performance to those of their peers, the norm group.

scores to a common metric of GLE. Despite not being best practice, we use GLE for the dual reason that we did not have norm referenced tests and because we did not have enough comparison data to generate our own internally normed standardizations. Other limitations associated with our measures relate to the data we *don't* have. A limitation of this study is that we do not have any information about the summer experiences of students who do not participate in Springboard Summer. They may be receiving additional reading supports other than Springboard (i.e., a different summer reading program, receiving more reading support at home). In other words, we don't know *what* our comparison represents - whether our comparison shows the benefits of Springboard over no summer reading intervention or whether it shows Springboard's benefits over other reading supports that are available to students over the summer, or a combination of both. Additionally, not all school districts in the study ended up having the capacity to provide comparison data (non-Springboard student data) from each school for our comparison analysis.

RECOMMENDATIONS

In the following section, the evaluation team presents what we learned about how Springboard can strengthen their data collection processes moving forward.

Systematically document program implementation. The evaluation team relied on retrospective accounts of program staff impressions of implementation fidelity. Going forward, Springboard should systematically document how Springboard Summer programming is implemented in each school site. All data of interest should be recorded by those directly implementing the program (i.e. Springboard Summer instructors). Data might include implementation components such as number of family workshops offered (distinguishing between required and make-up sessions, if structure differs), internal reading assessment administration date(s) (any and all), number of coaching support hours, etc. Staff perceptions of the quality in internal reading assessment administration (like the retrospective accounts staff provided early in this study) as well as their perceptions of reading assessment accuracy (how well the reading assessments map onto the local Springboard Summer program) could also be assessed on an on-going basis. Springboard should ensure that program benchmarks and data tracking expectations are made readily available and are well-recognized within each Springboard Summer site.

Formalize data-sharing partnerships with school districts. Springboard may want to consider formalizing data sharing partnerships with participating school districts. It would be advantageous to Springboard to identify champions who are willing to share data for the purpose of furthering programmatic improvements by evaluation, and developing strong relationships with these school districts and individuals who are gatekeepers for data within the districts. This will provide an opportunity for Springboard to routinely align school district data with their Springboard scholars for future evaluative purposes.

Track more nuanced implementation measures. Springboard Summer strives for high attendance from Springboard Summer scholars and high engagement from Springboard Summer families. For this study, we saw little variation in program session and family workshop attendance, because we chose sites with strong implementation and fidelity to Springboard Summer's model. Springboard could begin tracking student program participation, student daily engagement, and family engagement in more nuanced ways to more deeply examine how different measures tend to affect reading outcomes for Springboard Summer scholars.

Consider including Springboard Summer stakeholders in evaluation. During this study, Springboard expressed interest in learning more about district-administered surveys in participating Springboard Summer districts (see Appendix C for more detail). Springboard should take teacher experiences into consideration when evaluating their summer program. Tools for gathering data on teacher experiences may include either systematic surveys for all teachers and/or semi-structured interviews (may be beneficial for collecting data from a sample of Springboard teachers). This data may help gather more robust context and provide insight on how nuances experienced or implemented by teachers impact outcomes for Springboard scholars. In the future, Springboard could develop and administer a Springboard Summer scholar survey and/or a parent survey to help better understand how scholar and parent experiences impact program outcomes. Where possible, these Springboard Summer surveys could be aligned to district-administered student and parent surveys to gauge and report out on outcomes of interest to partnering districts to showcase Springboard Summer scholar outcomes. Additionally, questions related to implementation fidelity (student and parent perceptions of how well their expectations of program implementation align to the Springboard Summer model) and program quality (student and parent perceptions of their own successes and challenges related to programming, such as scholar/staff relationships, variety in programming, scholar and family engagement, program climate, etc.).

Continue to consider best practices in equating varied district reading assessments. Converting district reading assessment data across all participating Springboard Summer locations remains a challenge. Springboard should continue thinking about how best to equate all reading assessments in some standardized way to systematically capture growth across districts that are using different assessments to measure reading level.

Consider targeting students below grade level for Springboard Summer programming. This study shows that scholars from districts implementing Springboard Summer with high fidelity in 2018 and 2019 show reading growth - with the largest gains for scholars who started below grade level. Springboard may maximize their impact by targeting students who are reading below grade level at the end of the school year before the program. Districts in the sample reported using waiting lists to varying degrees. Strategies to strategically target students may include reserving spots for these students and proposing more systematic processes across districts implementing Springboard Summer. This will ensure students who are most likely to benefit, and more importantly in the most need, will be served by Springboard Summer.

APPENDICES

APPENDIX A

Section I

Six districts were selected through the initial selection criteria and agreed to participate in the study. Ultimately, only five districts were included in the study because one was unable to fulfill the data request due to capacity constraints and study timeline. The following Appendix table here, in Section I provides detail on district exclusions and the rationale for why those districts were excluded from being selected for this study.

TABLE T1. District/ Site Selection Matrix

District	Springboard Summer 2018 Programming		Springboard Summer 2019 Programming		Across 2018/2019 Programming Years	
	Exclusion Reason(s)		Exclusion Reason(s)		Exclusion Reason(s)	
	Assessment alignment (i-Ready)	Did not meet implementation criteria	Assessment alignment (i-Ready)	Did not meet implementation criteria	Unresponsive/ limited capacity	Turnaround Time
Alum Rock	✓	✓	✓	✓		
American Paradigm: Lindley*						
Baltimore County Public Schools (BCPS)		✓				✓[19]
DCPS Independent				✓		
DC Public Schools		✓		✓		✓[20]
Eagle Academy Public Charter					✓	
Fresno Unified SD	✓		✓			
Global Community CS					✓	
Great Oaks				✓		
Hilltop Independent				✓		

[19] BCPS required internal IRB processes.
[20] DCPS required a letter of support for the study (either from a district-level or school-level authority) and a lengthy turnaround time.

TABLE T1. District/ Site Selection Matrix (CONT.)

District	Springboard Summer 2018 Programming		Springboard Summer 2019 Programming		Across 2018/2019 Programming Years	
	Exclusion Reason(s)		Exclusion Reason(s)		Exclusion Reason(s)	
	Assessment alignment (i-Ready)	Did not meet implementation criteria	Assessment alignment (i-Ready)	Did not meet implementation criteria	Unresponsive/ limited capacity	Turnaround Time
Independence Charter School (summer site only)*						
Kuumba Academy					✓	
New York City DOE						✓[21]
New York City DOE (non-Standard)				✓		
Norwalk Public Schools*						
Oakland Unified SD (non-Standard)		✓		✓		
Oakland Unified SD (Standard) (excluding Fruitvale)*						
Oakland Unified SD (Independent)				✓		
PAVE Schools					✓	
Public Prep*						
Salinas					✓	
San Francisco Unified SD		✓		✓		
School District of Philadelphia					✓	
Stamford Public Schools				✓		
Stockton					✓	

*Note: Districts/sites included in the study are also included in the table above and are noted with an asterisk and highlighted in blue in the far left column.

[21] NYC DOE required internal IRB processes.

Section II

Section II provides supplementary statistics from the findings presented in the main narrative of the report for Research Question 3: How is participation dosage associated with scholars' reading growth over a summer?

The regression results tell the same story as the t-tests. Even after controlling for other characteristics of students in the treatment sample, there is not a significant effect on change in grade level equivalent in reading associated with meeting the Springboard attendance target. This does not mean that attendance is not useful. It may be that there is just too little variation in attendance within the sample, most of whom are meeting the target, to be able to discriminate between effects for high and low attenders.

TABLE T2. Reduced Regression Models Exploring Relationship between Level of Springboard Attendance and Change in Reading Level by Grade[22]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Attend 80% or more of Springboard Sessions	.1285 (.0845) n.s.	-.0161 (.0541) n.s.	-.0288 (.0667) n.s.	.0309 (.1020) n.s.	.1437 (.1498) n.s.
Starting Reading Level Relative to Grade Level:[23] > 1 year below grade level	---	---	.3156 (.1097)**	.2005 (.0903)*	.4833 (.1531)**
At/above grade level	---	-.1439 (.0573)*	-.1295 (.0643)+	-.1661 (.1036) n.s.	-.3575 (.2158) n.s.
Has IEP	---	---	-.1862 (.0612)**	---	---
Data from 2019 Springboard Summer	---	---	---	.2004 (.0842)*	---
Used Spanish Assessments	---	.3708 (.1506)*	---	---	---
English Language Learner	---	---	---	---	-.5980 (.1737)***
Male	---	---	---	---	-.2679 (.1202)*
District:[24] APS	-.0680 (.1820) n.s.	-.2547 (.0710)***	-.4500 (.0786)***	-.1970 (.1130)+	-.0696 (.1866) n.s.
ICS	-.3965 (.1710)*	-.2372 (.1228)+	-.3750 (.0757)***	---	.2144 (.2183) n.s.
Norwalk	---	---	-.2082	-.6431	.0582

[22] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Also tested but not included in the final model for any grade was the percentage of residential zip code below poverty.
[23] Omitted/comparison category is starting within one year below grade level.
[24] Public Prep is the omitted/comparison category.

			(.1142)+	(.1134)***	(.1920) n.s.

TABLE T2. Reduced Regression Models Exploring Relationship between Level of Springboard Attendance and Change in Reading Level by Grade (CONT.)

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Oakland	---	---	---	-.2868 (.1151)*	-.0254 (.2031) n.s.
Intercept	.0680 (.0743) n.s.	.3277 (.0662)***	.3877 (.0701)***	.1018 (.1172) n.s.	.1954 (.1885) n.s.
Adjusted R^2	.1720	.0739	.2889	.1453	.2175
n	24	177	166	166	135

Section III

Section III provides supplementary statistics from the findings presented in the main narrative of the report for Research Question 4: What influence does family workshop participation have on scholars' reading growth? The table in this Appendix presents the reduced regression models exploring the relationship between level of attendance at Springboard family workshops and change in reading level by grade.

The regression results tell the same story as the t-tests. Even after controlling for other characteristics of students in the treatment sample, there is not a significant effect on change in grade level equivalent in reading associated with meeting the Springboard family workshop attendance target. This does not mean that the family workshops are not useful. It may be that there is just too little variation in attendance within the sample, most of whom are meeting the family workshop attendance target, to be able to discriminate between effects for high and low attenders.

TABLE T3. Reduced Regression Models Exploring Relationship between Level of Attendance at Springboard Family Workshops and Change in Reading Level by Grade[25]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Attend 80% or more of Springboard Family Workshops	.1285 (.0845) n.s.	-.0133 (.0685) n.s.	.1270 (.0804) n.s.	-.0822 (.1053) n.s.	.1253 (.2044) n.s.
Starting Reading Level Relative to Grade Level:[26] > 1 year below grade level	---	---	.2969 (.1090)**	.2057 (.0903)*	.4893 (.1533)**
At/above grade level	---	-.1434 (.0573)*	-.1147 (.0630)+	-.1631 (.1029) n.s.	-.3473 (.2161) n.s.
Has IEP	---	---	-.1824 (.0608)**	---	---

[25] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Also tested but not included in the final model for any grade was gender.
[26] Omitted/comparison category is starting within one year below grade level.

TABLE T3. Reduced Regression Models Exploring Relationship between Level of Attendance at Springboard Family Workshops and Change in Reading Level by Grade (CONT.)

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Data from 2019 Springboard Summer	---	---	.0941 (.0530)+	.1893 (.0843)*	---
Used Spanish Assessments	---	.3711 (.1509)*	---	---	---
English Language Learner	---	---	---	---	-.5864 (.1741)**
Male	---	---	---	---	-.2763 (.1200)*
District:[27]					
APS	-.1965 (.1710) n.s.	-.2536 (.0716)***	-.4316 (.0784)***	-.1939 (.1127)+	-.1059 (.1833) n.s.
ICS	-.3965 (.1710)*	-.2340 (.1226)+	-.3539 (.0743)***	---	.1756 (.2162) n.s.
Norwalk	---	---	-.2424 (.1131)*	-.6334 (.1118)***	.0731 (.1914) n.s.
Oakland	---	---	---	-.2846 (.1146)*	-.0417 (.2064) n.s.
Intercept	.0680 (.0743) n.s.	.3264 (.0769)***	.1996 (.0880)*	.2023 (.1198) n.s.	.1992 (.2306) n.s.
Adjusted R^2	.1720	.0737	.3046	.1481	.2141
n	24	177	166	166	135

Section IV

Section IV provides supplementary statistics from the findings presented in the main narrative of the report for Research Question 5: How does reading growth among Springboard Summer scholars compare to students within the same district who are not invited or choose not to participate in Springboard Summer?

Table T4 presents the number of Springboard Summer scholars and comparison students within each district that provided comparison data. It also notes the number in each group that are identified through propensity score matching for inclusion in the matched sample for analysis comparing Springboard Summer scholars to comparison students.

[27] Public Prep is the omitted/comparison category.

TABLE T4. Springboard Students and Comparison Sample by District

District	Number of Springboard Scholars	Number of Matched Scholars	Number of Potential Comparison Students	Number of Matched Comparison Students
Public Prep	395	392	649	227
ICS	54	45	268	39
OUSD	55	55	252	45

Within each grade, the treatment and potential comparison youth are significantly different on one or more characteristics. The differences in the starting reading level are most concerning.

TABLE T5. Comparison of Treatment and *Potential* Comparison Sample on Demographic Characteristics and Baseline Reading Assessment by Grade

Variable	Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	Treatment	Potential Comparison	Treatment	Potential Comparison	Treatment	Potential Comparison	Treatment	Potential Comparison	Treatment	Potential Comparison
Starting GLE	0.19	0.06+	0.95	1.13***	1.46	1.81***	1.88	2.00	2.38	3.12***
% ELL	4.2%	1.3%	10.6%	3.4%**	19.6%	9.8%**	31.3%	40.9%*	25.9%	13.1%**
% IEP	20.8%	6.7%	23.3%	13.7%*	19.6%	15.0%	24.7%	10.6%***	24.4%	17.1%+
% Male	33.3%	38.7%	51.1%	44.5%	45.8%	44.9%	46.4%	38.6%	45.2%	34.4%*
% of Zip Code Below Poverty	33.0%	34.2%	30.7%	29.1%*	28.9%	28.6%	26.1%	27.1%	27.2%	26.3%
% Tested in Spanish	0%	0%	6.7%	22.7%***	6.6%	19.6%***	0%	0%	0%	0%
n	24	75	180	292	168	194	166	303	135	305

All Springboard Summer Scholars: Within District Matches

After propensity score matching, the resulting weighted matched comparison group is not statistically different from the treatment group on any of the matching characteristics.

TABLE T6. Comparison of Treatment and *Matched* Comparison Sample on Demographic Characteristics and Baseline Reading Assessment by Grade When Restricted to In-District Matches

Variable	Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison
Starting GLE	0.17	0.21	1.03	1.06	1.56	1.56	1.92	1.92	2.30	2.31
% ELL	4.6%	0%	9.2%	7.0%	16.1%	13.7%	28.3%	27.4%	20.9%	18.7%
% IEP	18.2%	22.7%	21.1%	27.5%	19.4%	13.7%	21.2%	24.8%	23.1%	22.0%
% Male	31.8%	40.9%	50.7%	50.7%	46.0%	54.8%	40.7%	33.6%	35.2%	39.6%
% of Zip Code Below Poverty	33.7%	35.0%	31.3%	30.9%	31.9%	31.1%	30.0%	31.0%	30.5%	31.2%
% Tested in Spanish	0%	0%	7.0%	7.0%	8.1%	8.1%	0%	0%	0%	0%
n	22	15[28]	142[29]	77[30]	124[31]	72[32]	113	78[33]	91[34]	69

[28] The included comparison youth are weighted to match the treatment sample.
[29] Three treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.
[30] The included comparison youth are weighted to match the treatment sample.
[31] Five treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.
[32] The included comparison youth are weighted to match the treatment sample.
[33] The included comparison youth are weighted to match the treatment sample.
[34] Three treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.

APPENDIX B

Public Prep made up a large portion of the study sample. Therefore, the evaluation team did some separate analysis for Public Prep students only in order to confirm that they aligned with the overall findings. Appendix B presents the findings from each Public Prep analysis.

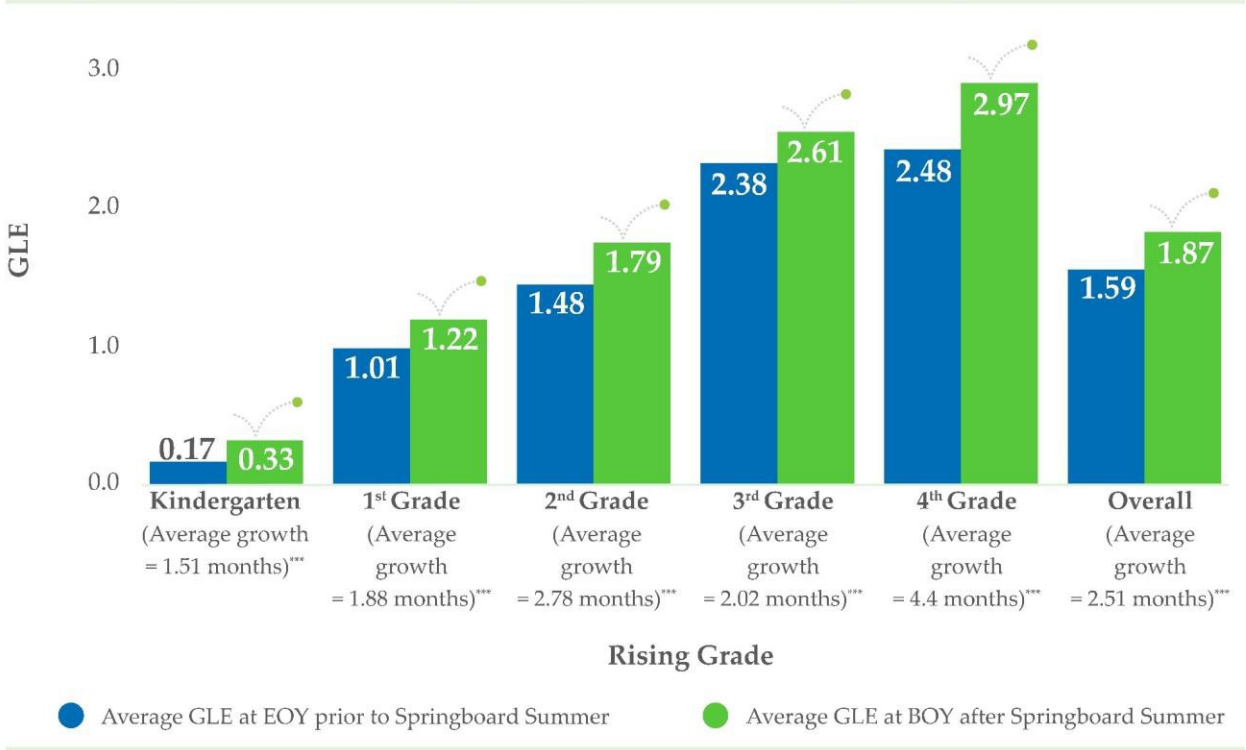
Section I

Appendix B, Section I presents the results of Public Prep students who participated in Springboard Summer for Research Question 1: Is there any improvement in reading for Springboard Summer scholars? Here, we examine overall growth for Public Prep students.

Public Prep: Overall Growth

When only scholars from Public Prep are considered, the findings are similar. Figure F1 shows, within each grade, and overall, the average change is positive and statistically significant, suggesting that, **on average, scholars who participated in Springboard Summer at Public Prep performed more favorably on the district administered reading assessment at the beginning of the school year following Springboard Summer than they had on the district administered reading assessment at the end of the school year before Springboard Summer.**

FIGURE F1. Average Change in District Reading Assessment from EOY to BOY after Springboard Summer Participation | Public Prep Only



Section II

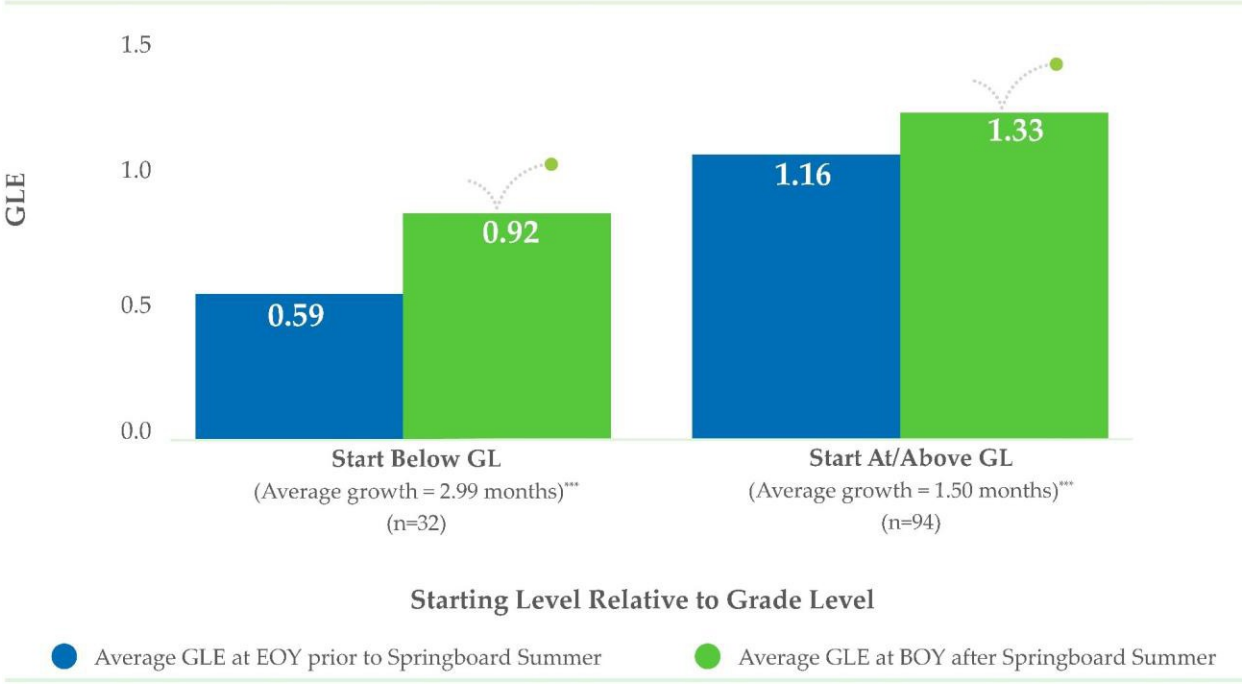
Section II presents Public Prep's results in changes in reading assessment by starting level relative to grade level for rising kindergarten through rising fourth grade Public Prep students who participated

in Springboard Summer, answering Research Question 2: Does scholars' improvement vary by starting point relative to grade level?

Public Prep Scholars: Growth amongst First Graders

Figure F2 below, shows that **first graders who participated in Springboard Summer within the Public Prep district, on average, experienced improvement in their reading level following participation in the program whether they started at or above grade level as well as if they joined the program reading below grade level.**

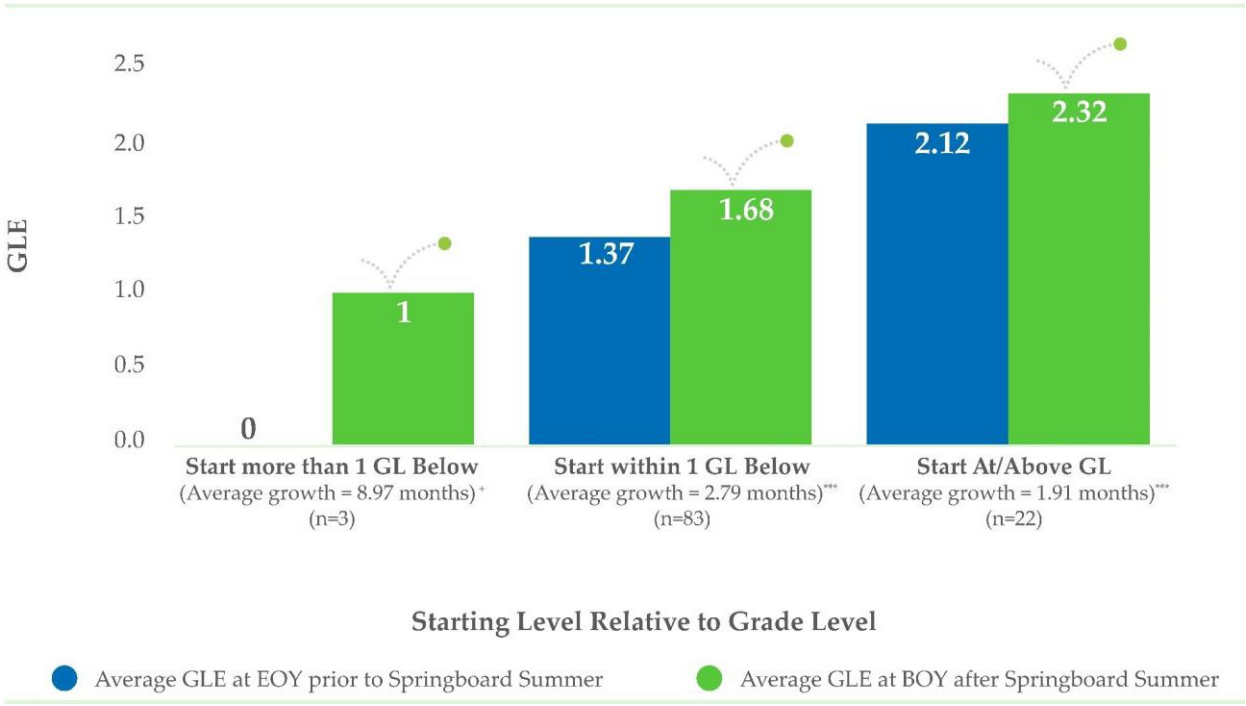
FIGURE F2. Change in Reading Assessment by Starting Level Relative to Grade Level for Rising First Grade Scholars | Public Prep Only



Public Prep Scholars: Reading Growth amongst Second Graders

Figure F3 below, shows Public Prep data for rising second graders. Very few second grade scholars who participated in Springboard Summer at Public Prep began the program reading more than a year below their grade level. The data suggest that these students experienced substantial growth, but with a sample of only three students, the change reaches only borderline significance. **Students who began Springboard Summer reading within a year below their grade level and those who began reading at or above grade level, on average, both experienced improvements in their reading ability when assessed at the start of the year following their participation in Springboard Summer. For those reading at or above grade level, this is different than when all districts were considered jointly.**

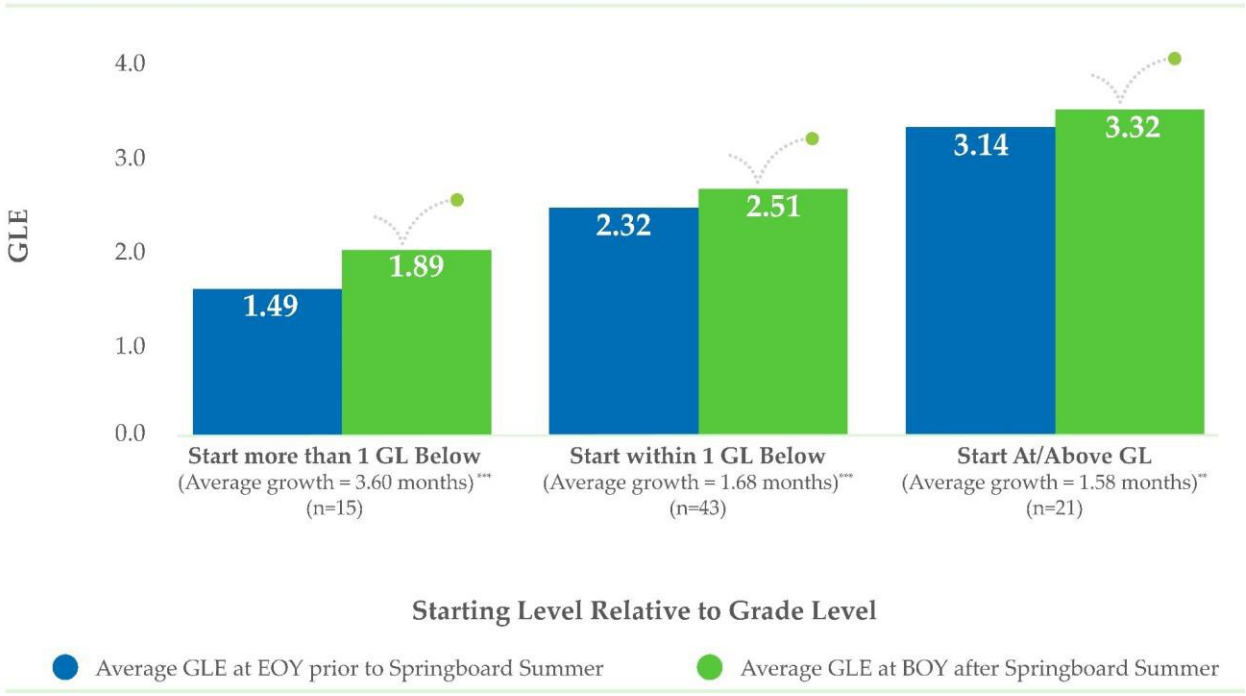
FIGURE F3. Change in Reading Assessment by Starting Level Relative to Grade Level for Rising Second Grade Scholars | Public Prep Only



Public Prep Scholars: Reading Growth amongst Third Graders

Figure F4 below, shows that on average, **third grade scholars participating in Springboard Summer at Public Prep experienced improvements in their reading ability when assessed at the start of the year following their participation in Springboard Summer whether they began the program reading at or above grade level, within one grade level below, or more than one grade level below.**

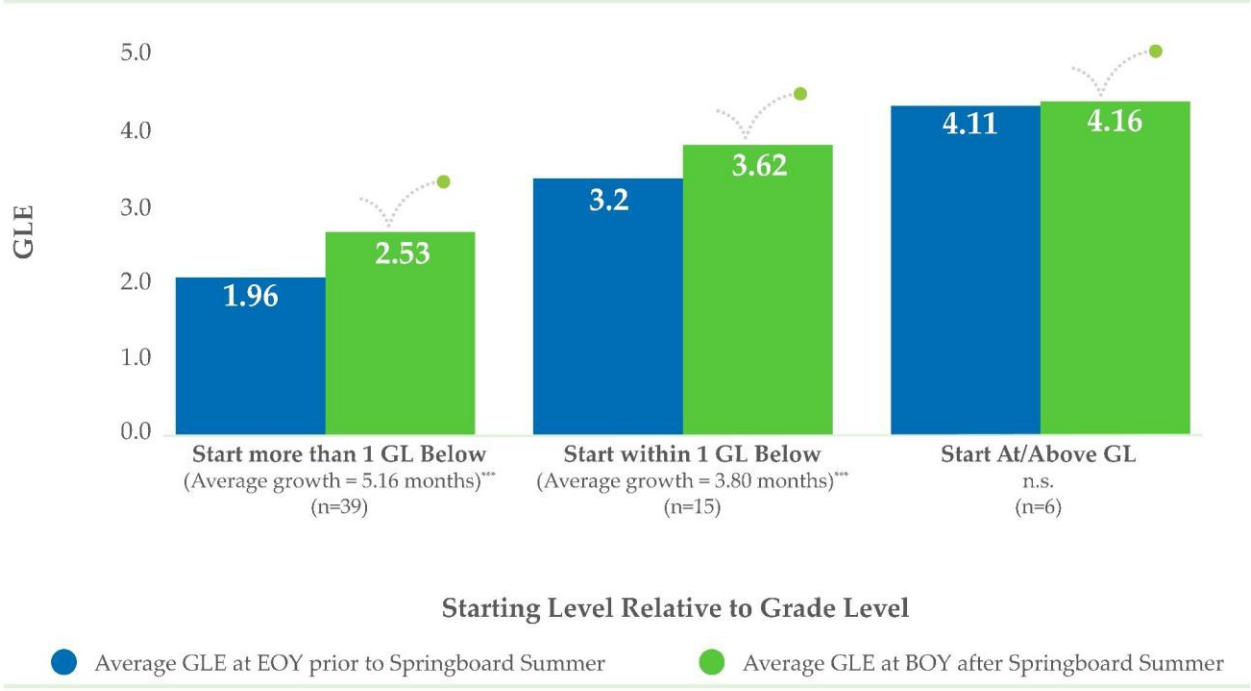
FIGURE F4. Change in Reading Assessment by Starting Level Relative to Grade Level for Rising Third Grade Scholars | Public Prep Only



Public Prep Scholars: Reading Growth amongst Fourth Graders

Figure F5 below, shows that on average, **fourth grade scholars participating in Springboard Summer at Public Prep who entered the summer reading below grade level, experienced improvements in their reading ability** when assessed at the start of the year following their participation in Springboard Summer. On average, students who began the program already reading at or above grade level did not experience significant changes in their reading level.

FIGURE F5. Change in Reading Assessment by Starting Level Relative to Grade Level for Rising Fourth Grade Scholars | Public Prep Only



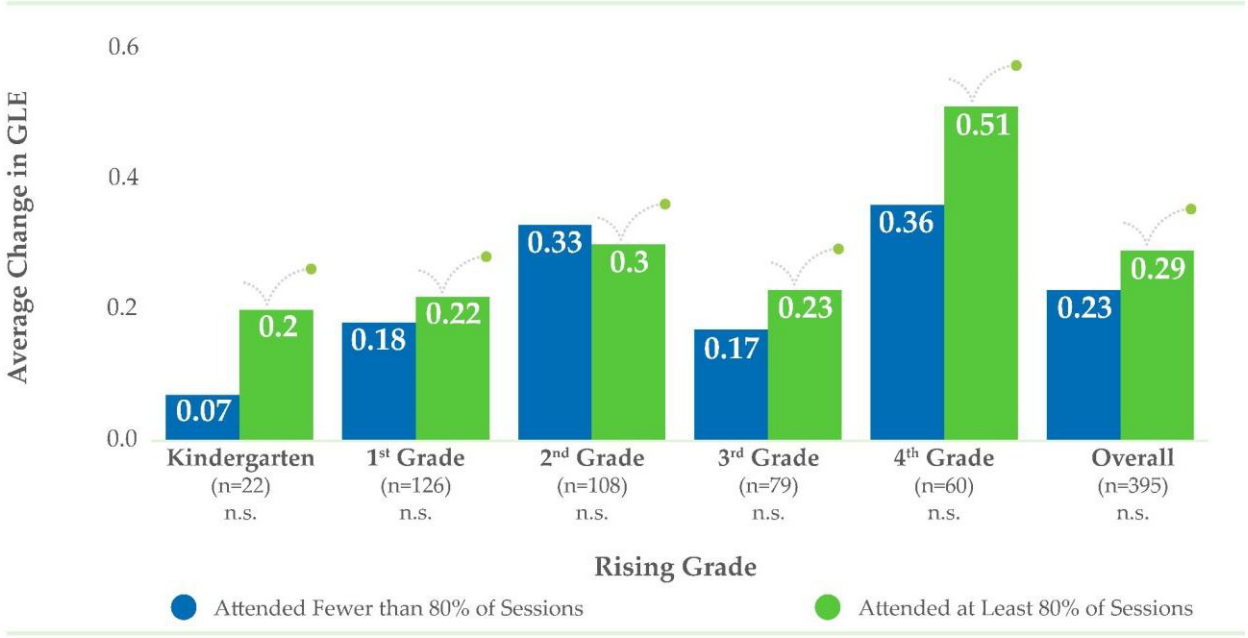
Section III

The following section presents the results of Public Prep scholars who participated in Springboard Summer for Research Question 3: How is participation dosage associated with scholars' reading growth over a summer?

Public Prep Scholars: Program Attendance

Figure F6 below, shows similar results for Public Prep compared to all Springboard scholars. **Both overall and within each grade, the average change in GLE is not significantly different between Public Prep scholars who meet the Springboard attendance threshold as compared to those who do not.**

FIGURE F6. Average Change in Reading Level within Grade by Scholar Attendance at Springboard Summer Sessions | Public Prep Only



Again, the t-tests did not allow the evaluation team to control for other characteristics that varied across students in the treatment sample, and these differences may have masked an actual effect of meeting the attendance threshold. To explore this, the evaluation team estimated a series of regression models within each grade predicting the change in grade level equivalent on the reading assessment. Meeting the attendance threshold was the predictor of interest, and the evaluation team controlled for starting point relative to grade level, gender, having an IEP, being an English language learner, testing in Spanish, participation year and district. Within each grade, we retained control variables only if they added predictive value to the model. Detailed results are presented in Table T7 below, and tell the same story as the t-tests. For the subset of treatment scholars who participated in Springboard Summer within the Public Prep district, there is not a significant effect on change in grade level equivalent in reading associated with meeting the Springboard attendance target even after controlling for other characteristics of students in the treatment sample.

TABLE T7. Reduced Regression Models Exploring Relationship between Level of Springboard Attendance and Change in Reading Level by Grade, Public Prep Only[35]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Attend 80% or more of Springboard Sessions	.1285 (.0845) n.s.	.0457 (.0465) n.s.	-.0072 (.0747) n.s.	.0603 (.0747) n.s.	.1320 (.1438) n.s.
Starting Reading Level Relative to Grade Level:[36] > 1 year below grade level	---	---	.6767 (.1584)***	.2254 (.0661)**	.1069 (.1293) n.s.

[35] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Gender was tested but not included in the final model for any grade.
[36] Omitted/comparison category is starting within one year below grade level.

TABLE T7. Reduced Regression Models Exploring Relationship between Level of Springboard Attendance and Change in Reading Level by Grade, Public Prep Only (CONT.)

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
At/above grade level	---	-.1681 (.0471)***	-.1604 (.0675)*	-.0252 (.0585) n.s.	-.3937 (.2074)+
Has IEP	---	---	-.2089 (.0662)**	---	---
Data from 2019 Springboard Summer	---	---	---	.1607 (.0496)**	---
English Language Learner	---	-.1314 (.0703)+	---	---	---
% of Residential Zip Code Below Poverty	---	---	-.6445 (.3426)+	---	1.7822 (.9055)+
Intercept	.0680 (.0743) n.s.	.3133 (.0543)***	.5812 (.1271)***	.0541 (.0767) n.s.	-.2637 (.3087) n.s.
Adjusted R^2	.0587	.0980	.2290	.1946	.1508
n	22	126	108	79	60

The evaluation team also explored Springboard Summer attendance on a continuous scale for Public Prep scholars (as opposed to meeting the 80 percent threshold). For the subset of treatment students who participated in Springboard Summer within the Public Prep district, **a continuous representation of Springboard attendance is not significantly related to change in grade level equivalent in reading within any of the five grade levels.** Reduced regression models exploring the relationship between continuous representation of Springboard attendance and change in reading level by grade for Public Prep can be found in Table T8, below.

TABLE T8. Reduced Regression Models Exploring Relationship between Continuous Representation of Springboard Attendance and Change in Reading Level by Grade, Public Prep Only[37]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
% of Springboard sessions attended	.0022 (.0026) n.s.	.0016 (.0014) n.s.	.0018 (.0021) n.s.	.0017 (.0022) n.s.	.0049 (.0042) n.s.
Starting Reading Level Relative to Grade Level:[38] > 1 year below grade level	---	---	.6821 (.1573)***	.2256 (.0662)**	.0983 (.1290) n.s.

[37] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Gender was tested but not included in the final model for any grade.
[38] Omitted/comparison category is starting within one year below grade level.

TABLE T8. Reduced Regression Models Exploring Relationship between Continuous Representation of Springboard Attendance and Change in Reading Level by Grade, Public Prep Only (CONT.)

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
At/above grade level	---	-.1682 (.0469)***	-.1482 (.0674)*	-.0264 (.0588) n.s.	-.3758 (.2032)+
Has IEP	---	---	-.2082 (.0659)**	---	---
Data from 2019 Springboard Summer	---	---	---	.1668 (.0509)**	---
English Language Learner	---	-.1364 (.0702)+	---	---	---
% of Residential Zip Code Below Poverty	---	---	-.7088 (.3452)*	---	1.8101 (.8835)*
Intercept	-.0211 (.2273) n.s.	.2056 (.1277) n.s.	.4331 (.2035)*	-.0465 (.2069) n.s.	-.5967 (.4312) n.s.
Adjusted R^2	-.0142	.1011	.2344	.1937	.1587
n	22	126	108	79	60

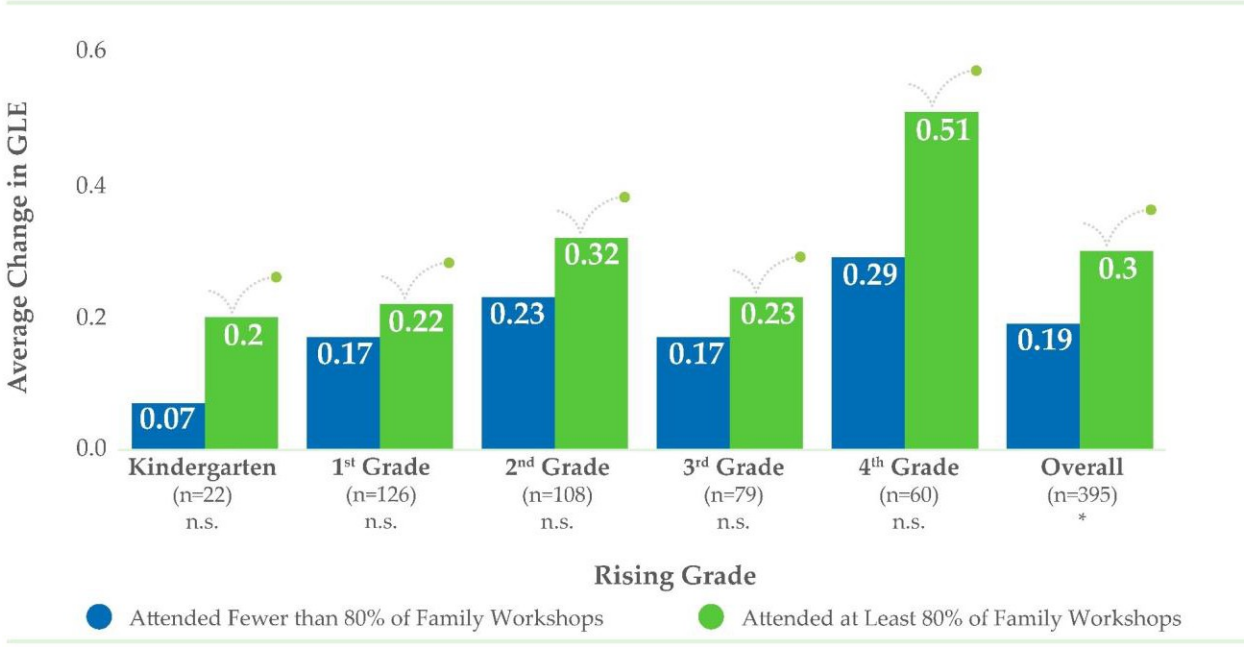
Section IV

The following section presents the results of Public Prep scholars who participated in Springboard Summer for Research Question 4: What influence does family workshop participation have on scholars' reading growth?

Public Prep: Family Workshop Participation

Findings for only Public Prep Springboard Summer scholars were similar. In Figure F7 below, within each grade, **the average change in GLE for Public Prep scholars who participated in Springboard Summer is not significantly different for students who meet the Springboard family workshop attendance threshold as compared to those who do not,** although all differences are in the same direction suggesting more favorable results from participation. **When all grades are considered jointly, the difference reaches statistical significance.** Again, the evaluation team followed this analysis by controlling for other characteristics of scholars in the treatment sample.

FIGURE F7. Average Change in Reading Level within Grade by Attendance at Springboard Summer Family Workshops | Public Prep Only



Even after controlling for other characteristics of scholars in the treatment sample, there is not a significant effect on change in grade level equivalent in reading associated with meeting the Springboard family workshop attendance target. See Table T9, below.

TABLE T9. Reduced Regression Models Exploring Relationship between Level of Attendance at Springboard Family Workshops and Change in Reading Level by Grade, Public Prep Only[39]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Attend 80% or more of Springboard Family Workshops	.1285 (.0845) n.s.	.0592 (.0526) n.s.	.0777 (.0796) n.s.	.0871 (.0670) n.s.	.2905 (.1588)+
Starting Reading Level Relative to Grade Level:[40] > 1 year below grade level	---	---	.6707 (.1572)***	.2239 (.0656)**	.0877 (.1270) n.s.
At/above grade level	---	-.1687 (.0470)***	-.1549 (.0662)*	-.0193 (.0580) n.s.	-.4150 (.2016)*
Has IEP	---	---	-.2148 (.0661)**	---	---
Data from 2019 Springboard Summer	---	---	---	.1702 (.0502)**	---

[39] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Gender was tested but not included in the final model for any grade.
[40] Omitted/comparison category is starting within one year below grade level.

TABLE T9. Reduced Regression Models Exploring Relationship between Level of Attendance at Springboard Family Workshops and Change in Reading Level by Grade, Public Prep Only (CONT.)

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
English Language Learner	---	-.1345 (.0704)+	---	---	---
% of Residential Zip Code Below Poverty	---	---	-.6364 (.3380)+	---	2.1038 (.8533)*
Intercept	.0680 (.0743) n.s.	.2994 (.0597)***	.5043 (.1372)***	.0268 (.0760) n.s.	-.5057 (.3338) n.s.
Adjusted R^2	.0587	.1002	.2361	.2041	.1873
n	22	126	108	79	60

Section V

Section V provides supplementary statistics from the findings presented in the main narrative of the report for Sensitivity Analysis 1: Public Prep GLE. This first sensitivity analysis corresponds to Research Question 5: How does reading growth among Springboard Summer scholars compare to students within the same district who are not invited or choose not to participate in Springboard Summer?

After propensity score matching, the resulting weighted matched comparison group is statistically the same as the treatment group on each of the matching characteristics.

TABLE T10. Comparison of Treatment and Matched Comparison Sample on Demographic Characteristics and Baseline Reading Assessment by Grade, Public Prep Only

Variable	Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison
Starting GLE	0.17	0.21	1.01	1.06	1.52	1.52	2.38	2.40	2.48	2.48
% ELL	4.6%	0%	8.1%	7.3%	14.2%	16.0%	10.1%	8.9%	3.3%	1.7%
% IEP	18.2%	22.7%	20.2%	22.6%	20.8%	16.0%	25.3%	25.3%	33.3%	30.0%
% Male	31.8%	40.9%	48.4%	46.0%	47.2%	47.2%	35.4%	36.7%	31.7%	30.0%
% of Zip Code Below Poverty	33.7%	35.0%	31.7%	31.2%	32.8%	33.1%	32.2%	30.7%	34.3%	34.6%
% Tested in Spanish	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

TABLE T10. Comparison of Treatment and Matched Comparison Sample on Demographic Characteristics and Baseline Reading Assessment by Grade, Public Prep Only (CONT.)

Variable	Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison
n	22	15[41]	124[42]	64[43]	106[44]	62[45]	79	50[46]	60	44[47]

Section VI

Section VI provides supplementary statistics from the findings presented in the main narrative of the report for Sensitivity Analysis 2: Public Prep Standardized Scores.

After propensity score matching, the resulting weighted matched comparison group is the same as the treatment group on each of the matching characteristics.

TABLE T11. Comparison of Treatment and Matched Comparison Sample on Demographic Characteristics and Baseline Reading Assessment by Grade, Public Prep Only

Variable	Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison
Starting Standardized Reading Score	.34	.48	-.12	.06	-.41	-.43	-.26	-.20	-.48	-.51
% ELL	5.0%	0%	8.1%	8.1%	14.2%	17.0%	10.1%	8.4%	3.3%	1.7%
% IEP	18.2%	22.7%	20.2%	21.8%	20.8%	15.1%	25.3%	25.3%	33.3%	31.7%
% Male	31.8%	40.9%	48.4%	46.8%	47.2%	45.3%	35.4%	38.0%	31.7%	30.0%
% of Zip Code Below Poverty	33.7%	35.0%	31.7%	31.2%	32.8%	33.3%	32.3%	30.9%	34.3%	34.5%
% Tested in Spanish	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

[41] The included comparison youth are weighted to match the treatment sample.
[42] Two treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.
[43] The included comparison youth are weighted to match the treatment sample.
[44] Two treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.
[45] The included comparison youth are weighted to match the treatment sample.
[46] The included comparison youth are weighted to match the treatment sample.
[47] The included comparison youth are weighted to match the treatment sample.

TABLE T11. Comparison of Treatment and Matched Comparison Sample on Demographic Characteristics and Baseline Reading Assessment by Grade, Public Prep Only (CONT.)

Variable	Kindergarten		1st Grade		2nd Grade		3rd Grade		4th Grade	
	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison	Treatment	Matched Comparison
n	22	15[48]	124[49]	64[50]	106[51]	64[52]	79	51[53]	60	43[54]

The remainder of this section corresponds to research questions 1-4. Additional results are presented in this appendix.

Public Prep: Standardized Score Analysis for RQ1: Is there any improvement in reading for youth who participate in Springboard?

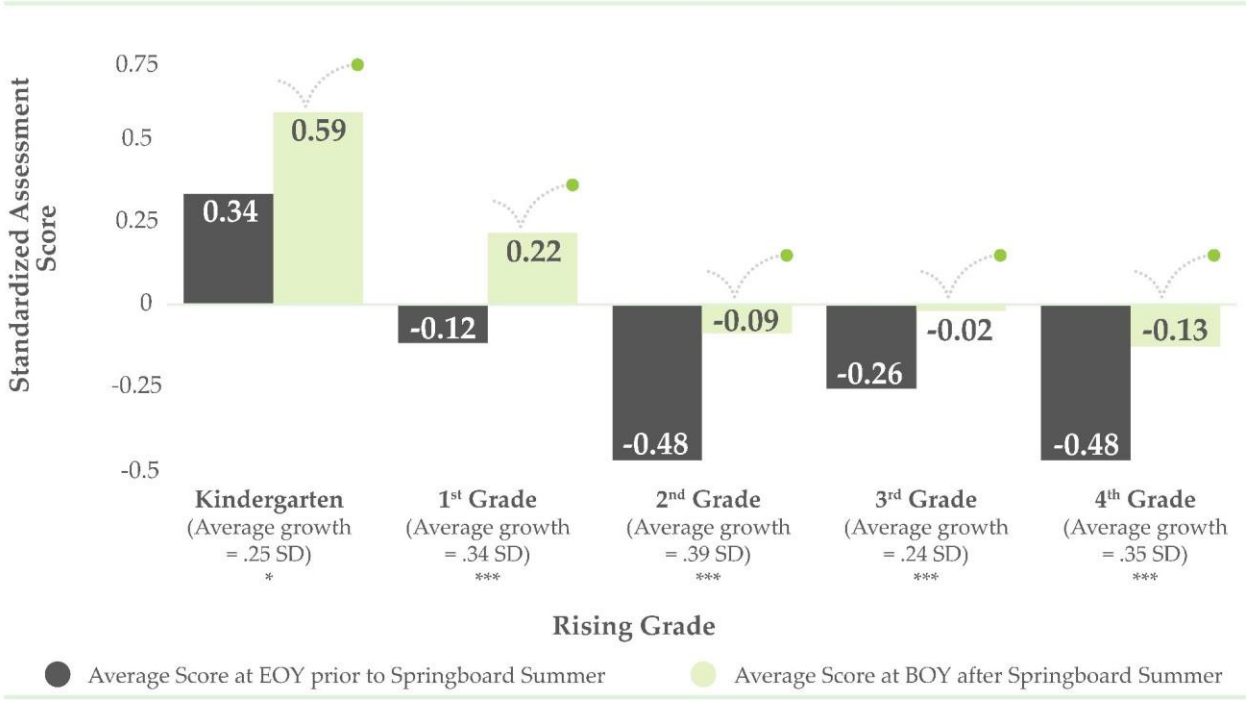
The following sequence of analyses use internally normed standardized scores as the dependent variable in place of GLE.[55]

Public Prep Standardized Scores Sensitivity Analysis

Figure F8 below, shows a similar story for 1st through 4th grade scholars who participate in Springboard Summer. In each of these grades, scholars who participate in Springboard Summer score below the average for Public Prep students in their grade at the end of the year district assessment. After participating in Springboard Summer, first grade scholars have improved their relative standing, and score better than the average comparison student in their grade. Second, third, and fourth grade scholars participating in Springboard Summer, while still testing below the average comparison student, have improved their relative position. Kindergarten scholars who participated in Springboard Summer tested more favorably than their non-participating peers before Springboard Summer and continued to improve their relative position.

[48] The included comparison youth are weighted to match the treatment sample.
[49] Two treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.
[50] The included comparison youth are weighted to match the treatment sample.
[51] Two treatment youth are excluded from the analysis sample due to lack of suitable matched comparisons.
[52] The included comparison youth are weighted to match the treatment sample.
[53] The included comparison youth are weighted to match the treatment sample.
[54] The included comparison youth are weighted to match the treatment sample.
[55] The construction of the dependent variable is explained within the body of the report which addresses Sensitivity Analysis 2.

FIGURE F8. Average Change in Standardized District Reading Assessment Score from EOY to BOY after Springboard Summer Participation | Public Prep Only

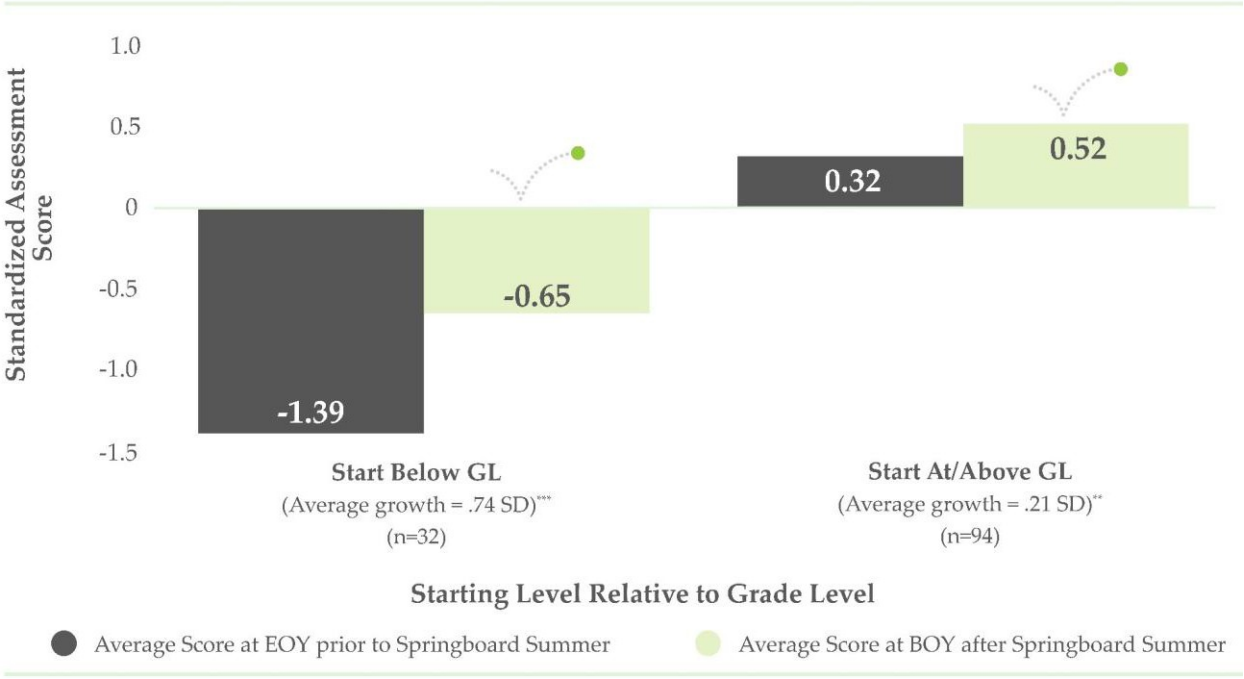


Public Prep: Standardized Score Analysis for RQ2: Does student growth vary by starting position relative to grade level?

Public Prep Standardized Scores Sensitivity Analysis: First Grade

As Figure F9 below shows, within first grade, Public Prep scholars who participated in Springboard Summer improved their reading ability relative to students who did not participate in Springboard Summer irrespective of whether they entered the program reading below or at/above their grade level.

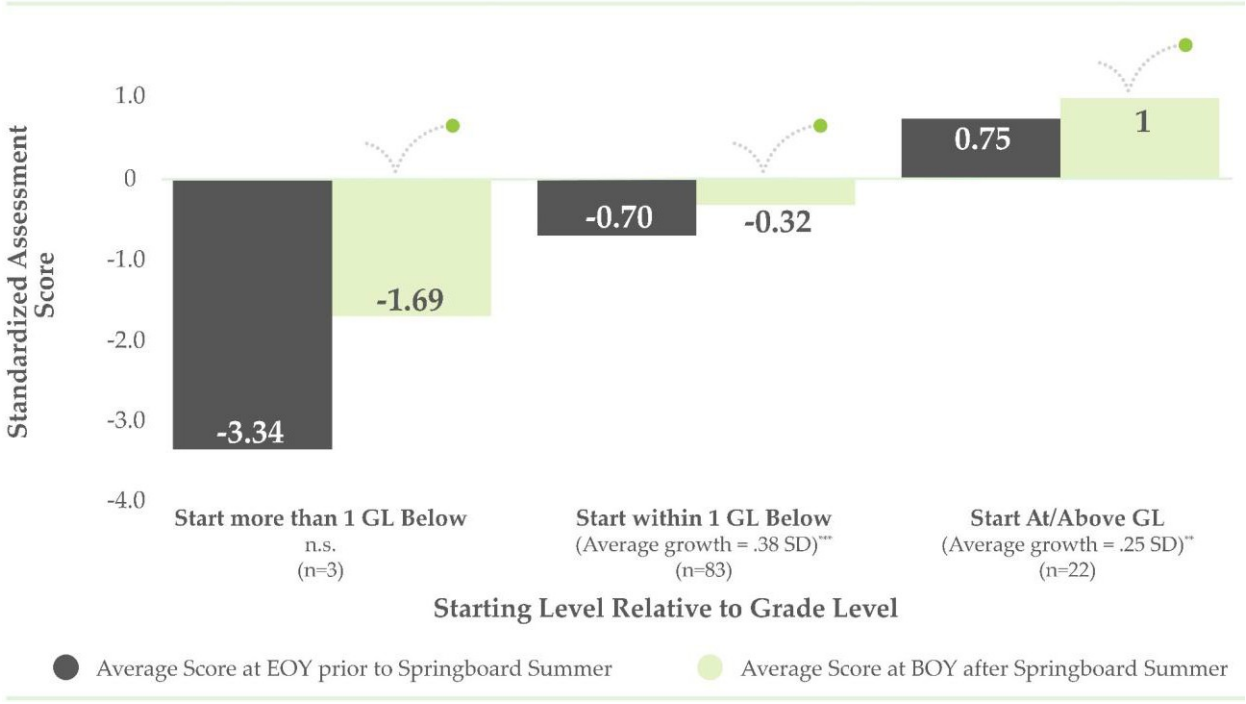
FIGURE F9. Average Change in Standardized District Reading Assessment Score by Starting Level for Rising First Grade Students | Public Prep Only



Public Prep Standardized Scores Sensitivity Analysis: Second Grade

Within second grade, as shown in Figure F10, Public Prep scholars who participated in Springboard Summer improved their reading ability relative to students who did not participate in Springboard Summer when they entered the program reading within one grade level below, at or above their grade level. For scholars who entered the program reading more than one grade below grade level, the improvement is not statistically significant, but this is due in large part to the small size of this group which only includes three students.

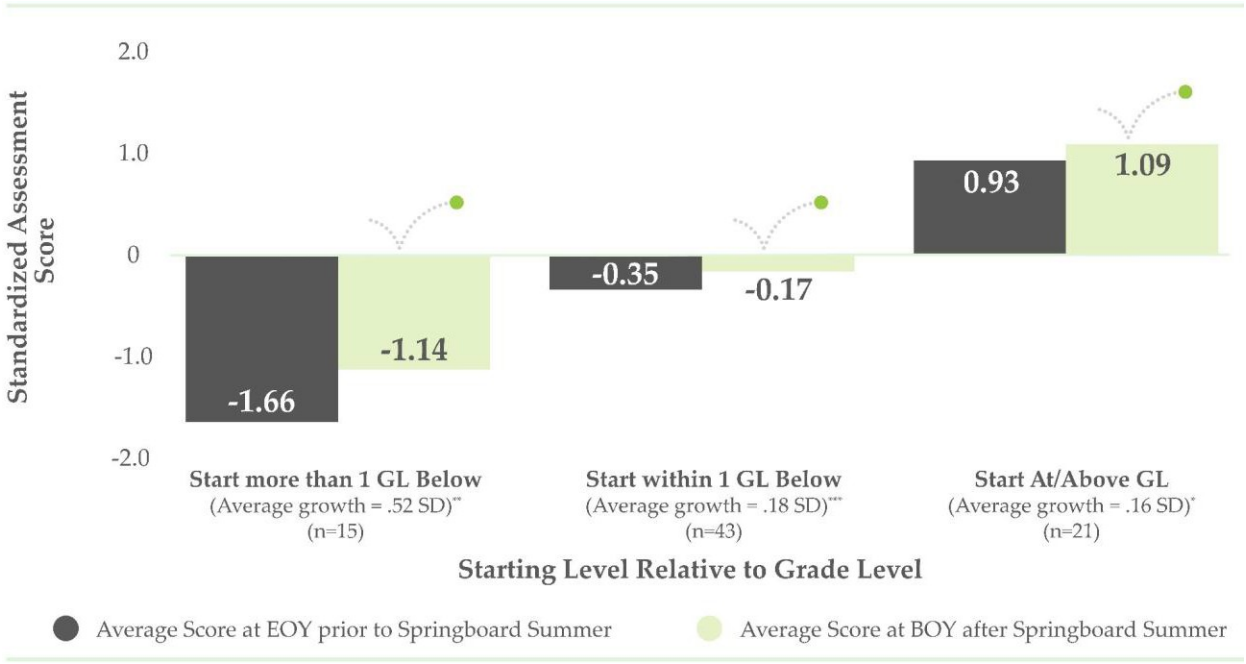
FIGURE F10. Average Change in Standardized District Reading Assessment Score by Starting Level for Rising Second Grade Students | Public Prep Only



Public Prep Standardized Scores Sensitivity Analysis: Third Grade

Within third grade, Public Prep scholars (see Figure F11 below) who participated in Springboard Summer improved their reading ability relative to students who did not participate in Springboard Summer irrespective of whether they entered the program reading far below, below, or at/above their grade level.

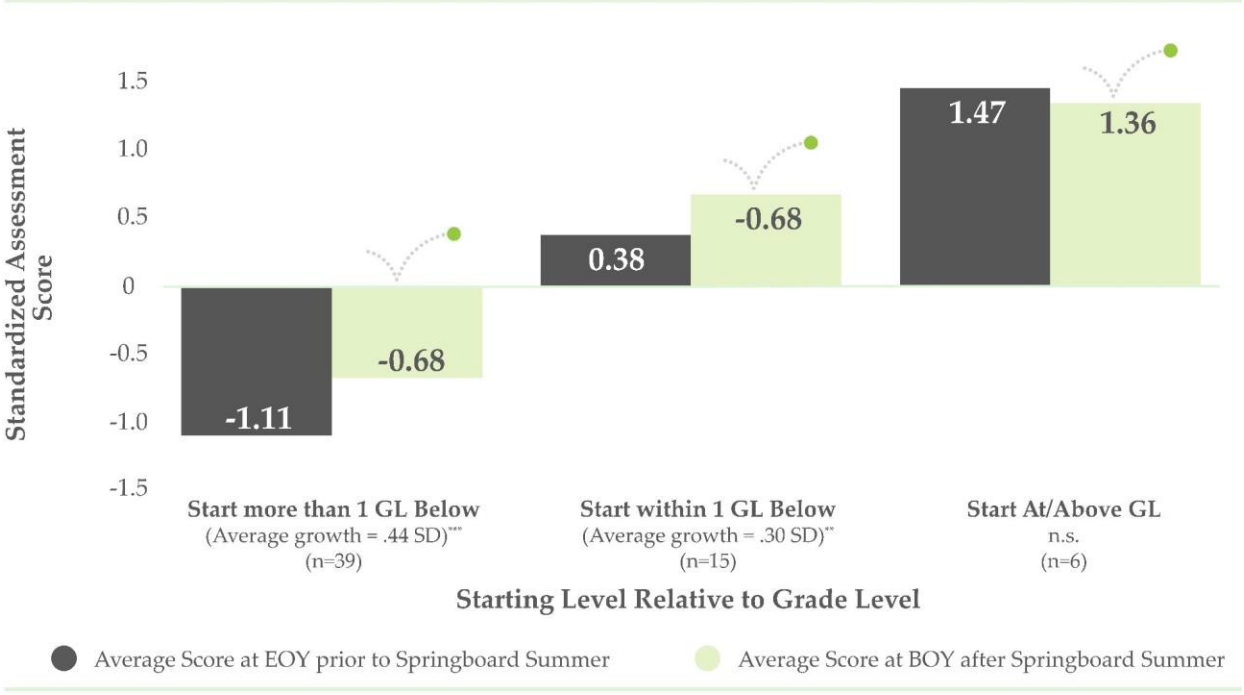
FIGURE F11. Average Change in Standardized District Reading Assessment Score by Starting Level for Rising Third Grade Students | Public Prep Only



Public Prep Standardized Scores Sensitivity Analysis: Fourth Grade

As Figure F12 shows, within fourth grade, Public Prep scholars who participated in Springboard Summer improved their reading ability relative to students who did not participate in Springboard Summer when they entered the program reading below grade level. For scholars who entered the program reading at or above grade level, the change in reading level is not statistically significant.

FIGURE F12. Average Change in Standardized District Reading Assessment Score by Starting Level for Rising Fourth Grade Students | Public Prep Only



Public Prep standardized scores for RQ3: How is participation dosage associated with reading growth over a summer?

Public Prep: Scholar Attendance

For students in 1st through 4th grades (see Figure F13 below), the differences in relative reading improvement are not significantly different between Springboard Summer participants at Public Prep who meet the attendance threshold as compared to those who do not. Except in second grade however, the differences are in the expected direction with higher attenders faring better. Among kindergarten students, the difference is marginally significant. It is likely that there is not sufficient variation in attendance across the treatment sample to adequately address the research question because attendance was uniformly quite high.

FIGURE F13. Average Change in Standardized Reading Level within Grade by Scholar Attendance at Springboard Summer Sessions | Public Prep Only

For the subset of treatment scholars who participated in Springboard Summer within the Public Prep district, the regression results tell the same story as the t-tests. Even after controlling for other characteristics of students in the treatment sample, there is not a significant effect on change in standardized reading level associated with meeting the Springboard attendance target for first through fourth grade students. For kindergarten students, the effect is marginally significant.

TABLE T12. Reduced Regression Models Exploring Relationship between Level of Springboard Attendance and Change in Standardized Reading Level by Grade, Public Prep Only[56]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Attend 80% or more of Springboard Sessions	.4330 (.2470) +	.1335 (.1335) n.s.	-.0181 (.1483) n.s.	.0957 (.1164) n.s.	.1666 (.1754) n.s.
Starting Reading Level Relative to Grade Level:[57] > 1 year below grade level	---	---	1.2464 (.3145)***	.3560 (.1030)***	.0784 (.1578) n.s.
At/above grade level	---	-.5371 (.1351)***	-.2562 (.1339)+	-.0422 (.0911) n.s.	-.4496 (.2531)*
Has IEP	---	---	-.4252 (.1313)**	---	---
Data from 2019 Springboard Summer	---	---	---	.2511 (.0773)**	---
English Language Learner	---	-.3782 (.2018)+	---	---	---
% of Residential Zip Code Below Poverty	---	---	-1.3063 (.6800)+	---	2.1778 (1.1052) +
Intercept	-.0842 (.2171) n.s.	.6818 (.1560)***	.9316 (.2524)***	-.0253 (.1195) n.s.	-.5386 (.3768) n.s.
Adjusted R^2	.0898	.1176	.2081	.1989	.1242
n	22	126	108	79	60

Public Prep standardized scores for RQ4: What influence does family workshop participation have on reading growth?

Public Prep: Family Workshop Attendance

For students in first through fourth grades, the differences in relative reading improvement are not significantly different between Springboard Summer scholars at Public Prep who meet the family workshop attendance threshold as compared to those who do not. Among kindergarten students, the difference is marginally significant. It is likely that there is not sufficient variation in family workshop attendance across the treatment sample to adequately address the research question because attendance was uniformly quite high with only a small number of families not meeting the threshold within each grade.

[56] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Gender was tested but not included in the final model for any grade.
[57] Omitted/comparison category is starting within one year below grade level.

FIGURE F14. Average Change in Standardized Reading Level within Grade by Attendance at Springboard Summer Family Workshops | Public Prep Only

For Public Prep students, the within grade regression results tell a similar story to the t-tests. For first through third grade students, even after controlling for other characteristics of students in the treatment sample, there is not a significant effect on change in standardized reading level associated with meeting the Springboard family workshop attendance target. For both kindergarten and fourth grade students the effect is marginally significant and in the expected direction suggesting greater improvement in reading where families attend at least 80 percent of the Springboard family workshops.

TABLE T13. Reduced Regression Models Exploring Relationship between Level of Attendance at Springboard Family Workshops and Change in Standardized Reading Level by Grade, Public Prep Only[58]

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Attend 80% or more of Springboard Family Workshops	.4330 (.2470)+	.1700 (.1511) n.s.	.1416 (.1581) n.s.	.1346 (.1090) n.s.	.3643 (.1936)+
Starting Reading Level Relative to Grade Level:[59] > 1 year below grade level	---	---	1.2365 (.3123)***	.3539 (.1090)**	.0544 (.1548) n.s.
At/above grade level	---	-.5388 (.1348)***	-.2454 (.1315)+	-.0329 (.0903) n.s.	-.4759 (.2458)+

[58] The final presented model for each grade includes only the Springboard Summer attendance variable together with the subset of variables that had a significant relationship to change in reading level within the grade. Gender was tested but not included in the final model for any grade.
[59] Omitted/comparison category is starting within one year below grade level.

TABLE T13. Reduced Regression Models Exploring Relationship between Level of Attendance at Springboard Family Workshops and Change in Standardized Reading Level by Grade, Public Prep Only (CONT.)

Independent Variables	Kindergarten	1st Grade	2nd Grade	3rd Grade	4th Grade
Has IEP	---	---	-.4359 (.1313)**	---	---
Data from 2019 Springboard Summer	---	---	---	.2657 (.0782)**	---
English Language Learner	---	-.3866 (.2022)+	---	---	---
% of Residential Zip Code Below Poverty	---	---	-1.2945 (.6713) +	---	2.5830 (1.0402)*
Intercept	-.0842 (.2171) n.s.	.6437 (.1714)***	.7881 (.2725)**	-.0653 (.1184) n.s.	-.8419 (.4069)*
Adjusted R^2	.0898	.1195	.2142	.2079	.1636
n	22	126	108	79	60

APPENDIX C

Upon Springboard request, the evaluation team reached out to districts we received data from to learn about whether they administer student, teacher and parent/caregiver surveys. The table below shows what constructs (main topics) the surveys cover.

TABLE T14. District Survey Construct Crosswalk by Springboard Summer Sample Districts

Springboard Summer District[60]	Parent Survey				Teacher/Staff Survey					Student Survey			
	Main Parent Survey Constructs				Main Teacher Survey Constructs					Main Student Survey Constructs			
	Climate/ Engagement	Instruction	Leader-ship	Community Ties/ Engagement	Climate/ Engagement	Instruction	Leader-ship	Com-munity Ties/ Engagement	Profes-sional Capacity	Climate/ Engage-ment	Instruction/ Academic/ Social-Emotional Learning	Leader-ship/ Positive Youth Develop-ment	Health/ Well-Being
American Paradigm Schools	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Independence Charter School[61]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Public Prep[62]	✓	✓	✓	✓									
Oakland Unified School District[63]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[60] Norwalk Public Schools: The survey data that is displayed on the NPS data dashboard is the result of the Comprehensive School Climate Inventory, a survey that was used districtwide for several years. This year the district transitioned to Panorama Surveys. The survey window closed early, when their students and teachers transitioned to distance learning. As a result, low participation levels will hinder reliable results. At this time, they do not have access to survey questions online.

[61] ICS administers the School District of Philadelphia's surveys. It has a Teacher Survey, Student Survey, and Family Survey Component. ICS's local surveys are generally developed to specifically gauge an aspect of the school. For example, this year, they have a survey for Offsite Recess (First year they had to do this) and Distance Learning.

[62] Public Prep doesn't administer surveys internally, but their families do participate in the NYC School Survey. However, network schools may administer their own student and staff surveys. Results for the NYC School Survey can be found here: BP, GPBX, GPLES

[63] Source: https://calschls.org/survey-administration/downloads/