THE SCHOOL DISTRICT OF PHILADELPHIA

Fall 2021 Performance on the Star Assessments for Students who Attended 2021 Summer Programs offered by SDP

Historically, students have attended summer programming in SDP because they have failed one or more courses, are at risk of failure during the upcoming school year, could benefit from intense academic support, or require year-round schooling. Although the 2021 summer programming was expanded from prior summers with the goal of supporting all students who may want additional support to make up for potential learning losses due to more than a year of virtual or hybrid learning, it was still anticipated that students who enrolled in summer programming had overall average lower academic performance than their peers who did not enroll in summer programming. Key findings include:

- Comparing summer program attendees and a matched sample of non-attendees, a similar percentage performed At or Above the National Average (in the High Average or Above Average performance groups) on the fall 2021-22 Star Reading and Math assessments.
- A higher percentage of students who attended summer programs for 75%-100% days performed At or Above the National Average (in the High Average or Above Average performance groups) on the fall 2021-22 Star Reading and Math than students who attended summer programming for fewer days.

Molly Schlesinger,
Senior Research Associate

Roland Reyes,
Senior Statistician

Office of Research and Evaluation

September 2022

Table of Contents

Acknowledgements

The authors would like to acknowledge the hard work and contributions of our colleagues in the School District of Philadelphia who planned and implemented summer programming in 2021, including: Benaline Baluyot, Miulcaeli Batista, Sonya Berry, Charlotte Brickhouse, Aja Carpenter, Diane Castelbuono, Sheree Ellison, Nyshawana Francis-Thompson, Khaled Ismail, La Tanya Miller, Janicka Newbill, Alexandra Robinson-Rogers, Malika Savoy Brooks, and Malicha White. The authors would like to acknowledge the support and contributions of Joy Lesnick, Ebru Erdem, Jenny Hoedeman-Eiteljorg, and Giselle Saleet in the Office of Research and Evaluation.

Table of Figures

Introduction

Each year, the School District of Philadelphia (SDP) and its partner organizations offer a variety of summer programs to ensure that students, especially those most vulnerable to experiencing summer learning loss, have the opportunity to continue learning during the summer months.[1] In response to the digital and hybrid learning experiences of students from March 2020 – June 2021, summer 2021 programming offers were expanded from previous summers.

SDP's Office of Research and Evaluation partnered with the Office of Academic Supports (OAS) to conduct a program evaluation of the summer programs organized by OAS. OAS worked with several District offices, including the Office of Curriculum and Instruction (OCI), the Office of Specialized Services, the Office of Multilingual Curriculum and Programs, the Office of High School Supports, and the Office of Early Childhood Education to design, implement, and support numerous programs. In this report, we present findings from an analysis of how students performed on the Renaissance Star ELA and Math assessments and whether they attended summer programming in Summer 2021.

[1] For more information about summer learning loss, see: https://www.philasd.org/research/2020/06/30/a-four-year-summary-of-summer-learning-loss-changes-in-k-2-independent-reading-levels-from-june-to-november-2016-2019/

About Summer 2021 Programming

In 2021, the Office of Academic Support (OAS) at the School District of Philadelphia (SDP) organized a series of summer programs (Table 1).[2]

Table 1. Overview of the 2021 summer programs offered by SDP

Summer Program	Description of Students Served
Kindergarten Transition Program	Students entering kindergarten
Programming for Grades 1-8	Students entering 1st – 8th grade
English Learner Newcomer Program	English Learners entering 1st – 8th grade who benefit from sheltered English Learning classrooms
Summer Bridge	Students entering 9th grade
Summer Credit Recovery	Rising 10th, 11th, 12th, and graduating students who failed specific courses in 2020-21
Quarter 5 Grade Improvement	Rising 10th, 11th, 12th, and graduating students who failed specific courses in 2020-21
Extended School Year	Special Education students entering 1st – 12th+ grade whose IEPs require accommodations for summer support

Notes: See *2021 Summer Programs in SDP: Offerings, Attendance, Survey Results, and Recommendations* report for descriptions of summer programs, https://www.philasd.org/research/2021/12/08/2021-summer-programs-in-sdp-offerings-attendance-survey-results-and-recommendations/. Data for kindergarteners are not included in the analysis because they did not participate in the spring 2020-21 assessments. An EL Newcomer program for high school students was originally offered. Due to staffing shortages, high school EL Newcomers were enrolled in Quarter 5 courses with an EL support teacher.

About Star Assessments

SDP began using Star Assessments, a suite of tests developed by Renaissance Learning, to assess K-12 students' reading and math skills District-wide in the 2021-22 school year. SDP uses Star to assess student progress toward the School District of Philadelphia Board of Education Performance Goals and to track student progress throughout the school year.[3]

[2] For more information about summer 2021 programming, see: https://www.philasd.org/research/2021/12/08/2021-summer-programs-in-sdp-offerings-attendance-survey-results-and-recommendations

[3] For more about SDP's Goals and Guardrails see: https://www.philasd.org/schoolboard/goals-and-guardrails/ and https://www.philasd.org/era/goals-and-guardrails/

Students who Attended Summer Programming

SDP students enroll in summer programming for a variety of reasons, including basic needs support for students receiving Special Education services, credit recovery in order to graduate, intense academic support to catch up to grade-level peers, and general academic support to reduce summer slide. In 2021, students also had the option to enroll for academic and social enrichment. This represents a change from the typical summer programming, where District students only qualify if they require or would strongly benefit from intensive academic support or year-round schooling.

Qualifying for Summer Programs

Extended School Year

Extended School Year (ESY) is a recurring summer program mandated by the Individuals with Disabilities Education Act (IDEA) that provides support to students with Individualized Education Plans (IEPs) who require services that extend beyond the school year. Students who qualify for ESY are students who are enrolled in Special Education services during the school year and whose IEPs require the District to offer year-round schooling.[4] Not all students with IEPs or who receive Special Education services qualify for ESY; this programming is reserved for students with the most intensive needs or who require supports that cannot be provided by a student's household alone.[5] This suggests that students who attended ESY had more intense learning and performance difficulties than students with IEPs who did not qualify for ESY, and we would thus anticipate ESY student to have lower performance than non-ESY students on the Star Assessments.

Summer Credit Recovery and Quarter 5 Grade Improvement

High school students attended Summer Credit Recovery and Quarter 5 Grade Improvement because they failed one or more courses and were working to recover a credit or improve a grade. This, by definition, means that high school students who qualified for these programs had lower grades (e.g., failed at least one course) compared to high school students who did not qualify for summer programming because they did not fail any courses.[6] Given the connection between grades and within-year assessments, we would anticipate that, on average, students who failed a high school course would have lower Star performance than students who had not failed a course.

---

[4] The primary purpose of ESY is to maintain the progress the student made on their most critical goals and objectives during the regular school year. The ESY program is based on the student's individual needs and typically does not include every goal and objective in the student's IEP.

[5] Not all students with an IEP are required to take Star assessments. For more information on how students with IEPs are waived from District and state assessments, visit: https://www.philasd.org/research/2021/05/03/assessment-participation-and-reading-and-math-performance-of-k-8-students-receiving-special-education-services/

[6] There were additional high school students who failed a course during the 2020-21 school year and qualified for summer programming and chose not to attend Summer Credit Recovery and Quarter 5 Grade Improvement. Over 400 credits recovered and grades improved were from students who needed the credits in order to graduate during summer 2021.

Summer Bridge

Similarly, Summer Bridge was recommended for students entering 9th grade who would benefit from summer ELA and math courses to help ease their transition into high school, and students who were not at risk of struggling during the transition to high school were not recommended for Summer Bridge. Therefore, students encouraged to enroll in Summer Bridge had lower reading and math performance entering summer programming than their counterparts who were not recommended for Summer Bridge, and we would anticipate this to be reflected in their Star performance.

English Learner Newcomer Program

Students who qualified for the English Learner Newcomer Program had low English language proficiency according to the ACCESS assessment,[7] since the goal of the program was to support English language learning as much as it was to support general competency in ELA and math. Historically, English Learners (ELs) as a population have had lower average performance on District-wide[8] reading and State-wide[9] ELA assessments than non-EL students, typically due to their lower English proficiency and the assessments being administered in English. We would anticipate similar patterns to emerge on the 2021-22 Star data.

Programming for Grades 1-8

These programs concentrated on providing ELA and math instruction through an in-person, project-based learning approach that encouraged students to integrate and demonstrate their learning through projects. Elementary and middle school students may have attended summer programming for intensive or strategic intervention or enrichment because during summer 2021, the elementary summer learning program was open to all students who registered, regardless of performance during the school year.[10] This means that while all students attended in order to increase their reading and math skills, students in the same classroom were at different levels of academic proficiency, which we anticipate will be reflected in Star performance.

---

[7] For more information about the ACCESS for ELLs assessment, visit https://wida.wisc.edu/assess/access
[8] For more information on historical EL performance on Star and aimswebPlus, visit: https://www.philasd.org/research/2021/09/17/k-12-student-participation-and-performance-in-math-and-literacy-assessments-during-the-2020-21-mostly-virtual-school-year/
[9] For more information on historical EL performance on the PSSAs, visit: https://www.philasd.org/research/2021/10/06/four-year-analysis-of-2015-16-kindergarteners-aimswebplus-reading-and-pssa-performance-from-2015-16-to-2018-19/
[10] In previous summers, similar programs for rising elementary and middle school students were only offered to District students who required intensive intervention. Summer 2021 programs were offered to all students to compensate for the academic impacts of virtual and hybrid school. For more information on the summer 2020 program see: https://www.philasd.org/research/2021/12/08/2021-summer-programs-in-sdp-offerings-attendance-survey-results-and-recommendations/

Summary

In summary, students who attended 2021 summer programming may have faced additional challenges compared to students who did not qualify for summer programming. Due to these challenges, we would anticipate students who attended summer programming to have lower overall assessment performance than students who did not qualify for summer programming.

Research Questions

Two research questions are the focus of this report.[11]

1. Were there differences in the fall 2021-22 Star Reading and Star Math performance between students who attended SDP 2021 summer programming and a matched sample of students who did not? Were there differences by student characteristics?
2. Were there performance differences on the fall 2021-22 Star Reading and Star Math assessments for students who attended 75-100% of summer program days compared to students who attended fewer than 75%?

Additionally, it is critical to recognize that students may qualify and enroll for summer programming but not attend. We note that a large number of students who enrolled in summer programming did not attend. This report focuses on students who enrolled in summer programming *and* attended. Implications are noted in the *Limitations* section.

Assessment Data Used for this Analysis

Star

SDP began using Star Assessments, a suite of tests developed by Renaissance Learning, to assess K-12 students' reading and math skills in the 2021-22 school year. For students in grades 6-12, SDP initially used Star in 2019-20 to assess skill development aligned to state and Common Core standards. SDP switched to using Star District-wide in 2021-22 for all K-12 students to assess student progress toward the School District of Philadelphia Board of Education Performance Goals.[12]

Since the beginning of 2021-22 school year, the District has been administering various Star tests to students across all grade levels (K-12) four times a year. For reading, students are assessed using either Star Early Literacy (grades K-2) or Star Reading (grades 3-12).[13] Both are computer adaptive tests that may become more or less difficult depending on whether students answer the previous question correctly. These tests are designed to broadly assess students' skills across a number of

[11] See addendum to this report for analyses that account for spring 2020-21 performance: https://www.philasd.org/research/2022/08/29/fall-2021-performance-on-the-star-assessments-for-students-who-attended-2021-summer-programs-offered-by-sdp

[12] For more about SDP's Goals and Guardrails see: https://www.philasd.org/schoolboard/goals-and-guardrails/ and https://www.philasd.org/era/goals-and-guardrails/

[13] Students in grades K-5 are also administered Star Reading Curriculum-Based Measures, which are 60-second one-to-one assessments of basic literacy skills, but these tests are not used in this analysis.

literacy domains. For math, students are administered either the Star Math Curriculum-Based Measures (grades K-2) or Star Math (grades 3-12). Star Math Curriculum-Based Measures (CBMs) are a series of short, 60-second subtests designed to measure students' foundational math skills. Star Math, like Star Early Literacy and Star Reading, is a computer adaptive test designed to measure students' math skills across several math-related domains.

The domains that make up the Star Early Literacy/Star Reading assessments for each grade level are as follows:[14]

- Grades K-2: The Star Early Literacy assessment domains include **Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning,** and **Number and Operations**
- Grades 3-12: The Star Reading assessment domains include **Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning, Analyzing Literary Text, Understanding Author's Craft,** and **Analyzing Argument and Evaluating Text**

The subtests/domains that make up the Star Math CBMs/Star Math assessment for each grade level are as follows:

- Kindergarten: The Star Math CBMs include **Number Recognition (NR) and Quantity Comparison (QC)**
- Grade 1: The Star Math CBMs include **NR, QC, and Addition to 10 (A10)**
- Grade 2: The Star Math CBMs include **A10, Addition to 20 (A20), and Subtraction from 10 (S10)**
- Grade 3: The Star Math assessment domains include **Number and Operations, Algebra, Geometry and Measurements**, and **Data Analysis, Statistics, and Probability**

The assessment outcome used in this analysis is the normal curve equivalent (NCE) score in the fall 2021-22. Scores on Star Early Literacy, Star Reading, and Star Math summarize students' overall proficiency across the different domains assessed on each test.

Key Outcome Data Points

For the analyses in this report, we use a data point called the normal curve equivalent (NCE) score to analyze student performance. NCE is closely related to another common performance metric called national percentile rank. However, unlike national percentile rank, which is ordinal, NCE scores can be used in arithmetic operations, such as calculating an average, because it is an interval variable. This section provides more details on the similarities and differences between each score type.

[14] For more information see: https://www.philasd.org/era/star-information/#1618402180282-71187e13-0e42

National Percentile Rank

Percentile rank is a norm referenced performance measure that compares students' scaled scores to a nationally representative sample of grade-level peers. The percentile rank is useful for understanding student skill development in comparison to students of the same grade nationally. Based on the number of correct responses, each student is assigned a national percentile rank. Percentiles range from 1-99. For example, a national percentile rank in the 23rd percentile indicates that the student is performing better than 23% of the nationally-normed sample based on their number of correct responses. While percentile ranks are a familiar metric for most readers, they should not be used in arithmetic operations, such as averaging percentile ranks across multiple students in the same student group, because the intervals between percentile ranks are not the same across the percentile range. [15]

Normal Curve Equivalent

Normal curve equivalent (NCE) scores are another type of norm-referenced performance measure, and are one way to address the limitations of percentile ranks. Like percentile ranks, NCEs describe students' performance among a nationally normed sample. Like percentile ranks, NCEs are a type of norm-referenced performance measure that describes students' performance among a nationally normed sample. However, unlike percentile ranks, NCEs are interval-scaled so that the intervals between test scores that correspond to each NCE score are the same.[16] Therefore, NCEs can be used to calculate an average or to calculate differences between groups. NCEs are scaled with a mean of 50, a standard deviation of 21.06, and range from 1-99.[17]

An NCE score of 50 indicates performance at the average of the national norming sample while scores higher or lower than 50 indicate performance above or below the average of the national norming sample, respectively.[18] For example, a student who has an NCE score of 50 or above performed the same or better than 50% of students nationally; in comparison, a student with an NCE score below 50 performed worse than 50% of students nationally. For the purposes of this analysis, percentile ranks were converted to NCE scores. It is important to note that despite their similarities, NCE scores and percentile ranks do not align.[19] Therefore, NCE scores should not be interpreted in the same way as percentile ranks.

---

[15] For example, the test score interval between the 23rd percentile and 24th percentile and the interval between the 50th percentile and 51st percentile are not equivalent to each other. These unequal intervals make it so that arithmetic results based on percentile ranks will be difficult to interpret.

[16] For example, the interval between NCE scores of 23 and 24 is the same as the interval between NCE scores 50 and 51.

[17] In this report, they are derived from the percentile ranks using the equation NCE = 21.06*z-score + 50 (Lipsey et al., 2012), where the z-score comes from the percentile rank value.

[18] For more information see Lipsey, M. W., Puzio, K, Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, W., Anthony, K. S., Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms.* https://ies.ed.gov/ncser/pubs/20133000/

[19] For more information see Pennsylvania Department of Education. (2021). *Making sense of NCEs and standard errors.* https://www.education.pa.gov/Documents/K-12/Assessment%20and%20Accountability/PVAAS/Methodology/MakingSenseOfNCEsAndStandardErrors.pdf

Performance Groups

Based on their NCE data, students fall into one of four performance groupings (Table 2). These performance groups give us a more nuanced understanding of student performance while still allowing us to categorize students based on their performance. If our students had similar performance to the national sample, about 50% would have performed in the High Average or Above Average NCE performance groups and the other 50% of our students would have performed in the Below Average or Low Average NCE performance groups. While there are other ways of grouping students (by Tier level, for example), performance groups are useful for analyzing how students at various levels of proficiency are performing and improving over time.[20]

Table 2. Star and aimswebPlus assessment Normal Curve Equivalent (NCE) group names and NCE range

Grouping Name	NCE Score Range
Above Average	75-99
High Average	50-74
Low Average	25-49
Below Average	1-24

Identifying the Student Sample

Student Sample

About half of students who attended summer programs took fall 2021-22 Star assessments and could be matched with similar students who did not attend summer programs.

In summer 2021, 16,453 students were enrolled across eight SDP summer programs (Figure 1). Of that population 12,840 students attended at least one program for at least one day, or attended between 1%-100% of days their program was offered. Of those students, 6,074 attended their program (or at least one of their courses) for 75%-100% of days the program was offered. 6,663 students took both Early Literacy/Reading and Math Star assessments in fall 2021. Only students who took the Star assessments in fall 2021 are included in this analysis.

Of the students who attended summer programming and took the fall Star assessments, 6,565 students were matched with similar students who did not attend summer programming—this group is our analytic sample of students included in the analyses.[21]

[20] aimswebPlus NCE scores were converted from National Percentile Ranks (NPR).
[21] Spring 2020-21 performance data was not included in the matching process due to even smaller percentages of the summer program attendees having taken spring 2020-21 assessments. Analyses and comparisons accounting for spring 2020-21 data are included in the addendum: https://www.philasd.org/research/2022/08/29/fall-2021-performance-on-the-star-assessments-for-students-who-attended-2021-summer-programs-offered-by-sdp

Figure 1. The number of students who were enrolled in summer programs, attended 1%-100% of summer program days, took fall 2021 Star assessments, and could be matched with similar students who did not attend summer programming

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 21, 2021.

The matched sample of students who did not attend summer programming

Propensity score matching (PSM)[22] was used to match the sample of summer students with a similar group of students who did not attend summer programming. This latter group is referred to as the matched sample. PSM is a statistical approach that aims to make the two groups similar on the basis of selected characteristics. Our matching model used several demographic characteristics including gender, race/ethnicity, English Learner status, and economic disadvantage status to identify the matched sample. Students in the matched sample were also required to have attended the same grade and school in spring 2020-21 as the students who attended summer programming. After the matching process was complete, we evaluated the quality of the matches to determine that the two groups, students who attended summer programming and students who did not, were sufficiently balanced with respect to the demographic characteristics included in the matching model (see Appendix A for details on propensity score matching and determining the ideal matched sample).[19] This matched sample includes both students who would have qualified and not qualified for summer programming. Due to the year of virtual learning due to the COVID-19 pandemic, the population of students who qualified for summer programming in 2021 was less straightforward than the populations who qualified in prior summers.

[22] Austin, 2011; Rosenbaum & Rubin, 1983; Stuart, 2010

Demographic characteristics of the students who attended summer programming and the matched sample

Students in 10th grade (13%) made up the highest proportion of students in the analytic sample (summer program attendees and matched non-summer program attendees combined are the analytic sample),[23] conversely, 9th graders (4%) made up the smallest proportion in the analytical sample (Table 3).[24] The low proportion of 9th grade students is likely due to the minimal programs offered to rising 9th graders (i.e., Summer Bridge) compared to the courses offered to other high school students (e.g., Quarter 5, Credit Recovery).

Black/African American students made up 60% of the analytic sample, Hispanic/Latinx students made up 19%, White students made up 9%-10%, Asian students made up 8%, and Multi-Racial/Other students made up 4% of the analytic sample (see Appendix A, Table A1 to see the student demographic makeup of the SDP population overall and the sample of students who attended summer programming).

Female students made up between 42%-43% of the analytic sample when compared to male students who made up between 57%-58%. Economically disadvantaged students made up between 75%-78%, students with IEPs made up 33%-37%, and English Learners (ELs) made up 13% of the analytic sample.

It should be noted that the characteristics of students who attended summer programming do not reflect the overall District composition for certain student groups (see Appendix A and B for the percentages of students by demographic characteristics who attended summer programming, for the District overall during the same time frame, the matched sample included in this report, and overall District population). For example, 43% of the analytic sample are female students, while female students make up 49% of the District population overall (See Appendix B, Table B2).

[23] Due to the PSM matching, the percentages of students in each student group in the summer attendee population and non-summer attendee population in the analytic sample are the same; see Table 2.
[24] Repeating 12th graders (22 students) were excluded from the analysis.

Table 3. The demographic characteristics of the analytic sample of students who attended summer programs and a matched sample of students who did not attend summer programs

Student Characteristic	Attended Summer Program (n = 6,565)	Did Not Attend Summer Program (matched sample) (n = 6,565)	Total Number of Students in Analytic Sample (N = 13,130)	Percent of Total Students in Analytic Sample (N = 13,130)
2021-22 Grade Level				
1	527	527	1,054	8%
2	584	584	1,168	9%
3	552	552	1,104	8%
4	582	582	1,164	9%
5	625	625	1,250	10%
6	565	565	1,130	9%
7	487	487	974	7%
8	378	378	756	6%
9	270	270	540	4%
10	852	852	1,704	13%
11	689	689	1,378	10%
12	454	454	908	7%
Race/Ethnicity				
Asian	494	495	989	8%
Black/African American	3,944	3,928	7,872	60%
Hispanic/Latinx	1,223	1,269	2,492	19%
Multi-Racial/Other	281	244	525	4%
White	623	629	1,252	10%
Gender				
Female	2,780	2,806	5,586	43%
Male	3,785	3,759	7,544	57%
Socio-Economic Status				
Econ. Disadvantage	4,950	5,140	10,090	77%
Not Econ. Disadvantage	1,615	1,425	3,040	23%
Special Education Status				
Students with IEPs	2,409	2,164	4,573	35%
Students without IEPs	4,156	4,401	8,557	65%
English Learner Status				
English Learner	854	848	1,702	13%
Not an English Learner	5,711	5,717	11,428	87%

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021.

Note: For the percentage of each demographic group represented by attendees and non-attendees, see Appendix B, Table B1.

How to read this table: This table displays the number of students in the two groups (summer attendee vs non-attendee) and in each student demographic group. For example, 2,780 summer attendees were female, and 2,806 non-attendees were female. The total number of female students in the sample was 5,586, representing 43% of the overall analytic sample of 13,130 students.

Summer Programming Attendance

Summer programming attendance ranged from attending 1% of days to 100% of days the program was offered.[25] Over 50% of students in the analytic sample attended their summer program for 75%-100% of program days (Table 4).

Table 4. The percentage of students in the analytic sample who attended summer programming by the percentages of summer program days they attended

Summer Programming Attendance Category	Number of students in each attendance category	Percent of students in each attendance category
1% - 24% of days	1,378	21%
25% - 49% of days	753	11%
50% - 74% of days	980	15%
75% - 100% of days	3,454	53%

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021.

Findings

Were there differences in the fall 2021-22 Star Reading and Star Math performance between students who attended SDP 2021 summer programming and a matched sample of students who did not? Were there differences by student characteristics?

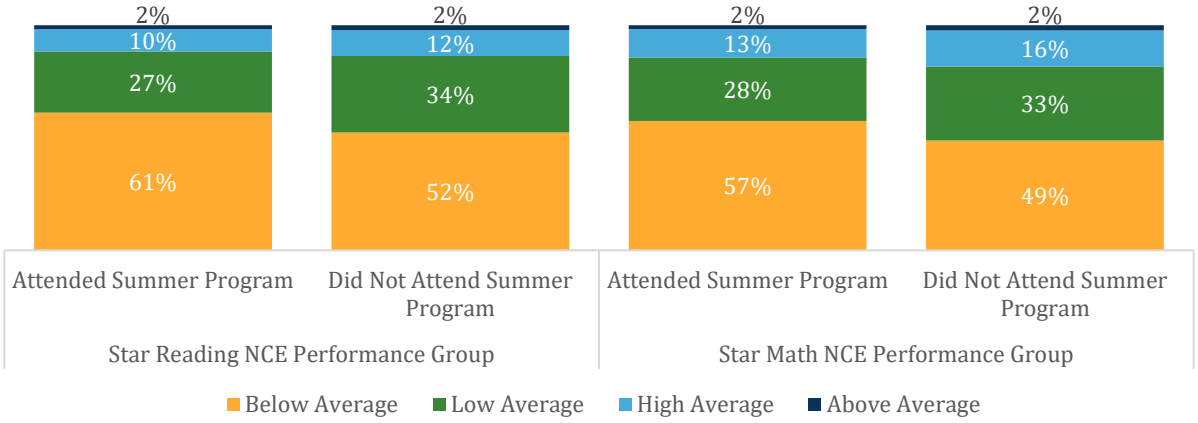
For the entire sample, a similar percentage of summer program attendees and non-attendees performed in the High Average and Above Average fall Star Reading and Math groups.

Of all students in the analytic sample, 12% of summer program attendees scored in the High Average or Above Average NCE performance groups on fall Star Reading (Figure 2) and 88% scored Below Average or Low Average. Students who did not attend a summer program had similar performance, with 14% scoring in the High Average or Above Average NCE performance groups and 86% scoring Below Average or Low Average.

For math, 15% of summer program attendees scored in the High Average or Above Average NCE performance groups on fall Star Math while 85% scored Below Average or Low Average. Here again, students who did not attend summer programs had similar performance, with 18% scoring High Average or Above Average and 82% scoring Below Average or Low Average.

[25] For more information about summer 2021 programming attendance, see https://www.philasd.org/research/2021/12/08/2021-summer-programs-in-sdp-offerings-attendance-survey-results-and-recommendations. Students who were enrolled and attended zero days are not included in this analysis.

Figure 2. The percentage of students in analytic sample who performed in each of the four NCE groups on fall Star Reading and Math



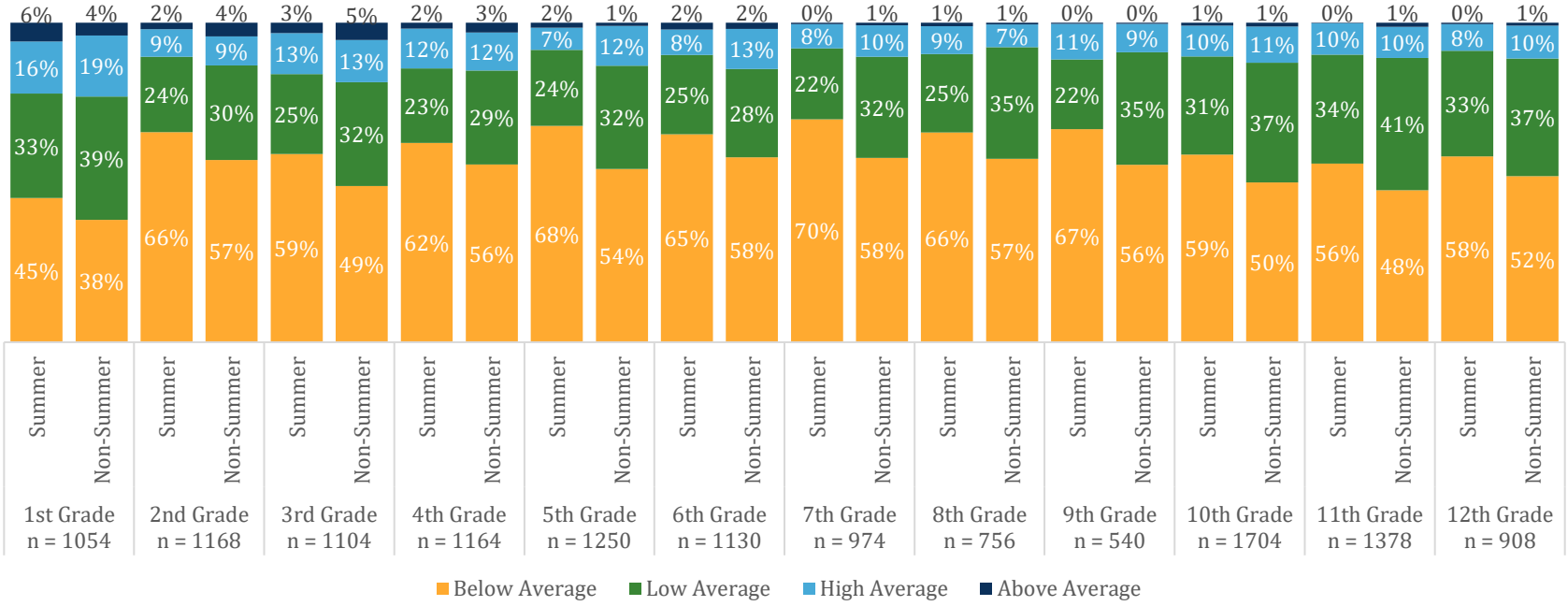
Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: Both outcomes include 6,565 students who attended summer programming and 6,565 who did not attend summer programming. See Appendix C, Table C1 for average NCE scores for each group.

A smaller percentage of 5th and 6th grade summer program attendees performed in the High Average and Above Average Star Reading group than non-attendees.

Across all grades, a slightly smaller percentage of summer program attendees scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups than non-attendees (Figure 3). Additionally, in the percentage of 5th and 6th graders who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups there was a four to five point difference between the percentage of summer program attendees (9-10%) and non-attendees (13-15%). The percentage of students in all other grades who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups differed by three points or fewer by summer program attendee status.

Figure 3. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Reading by 2021-22 grade level

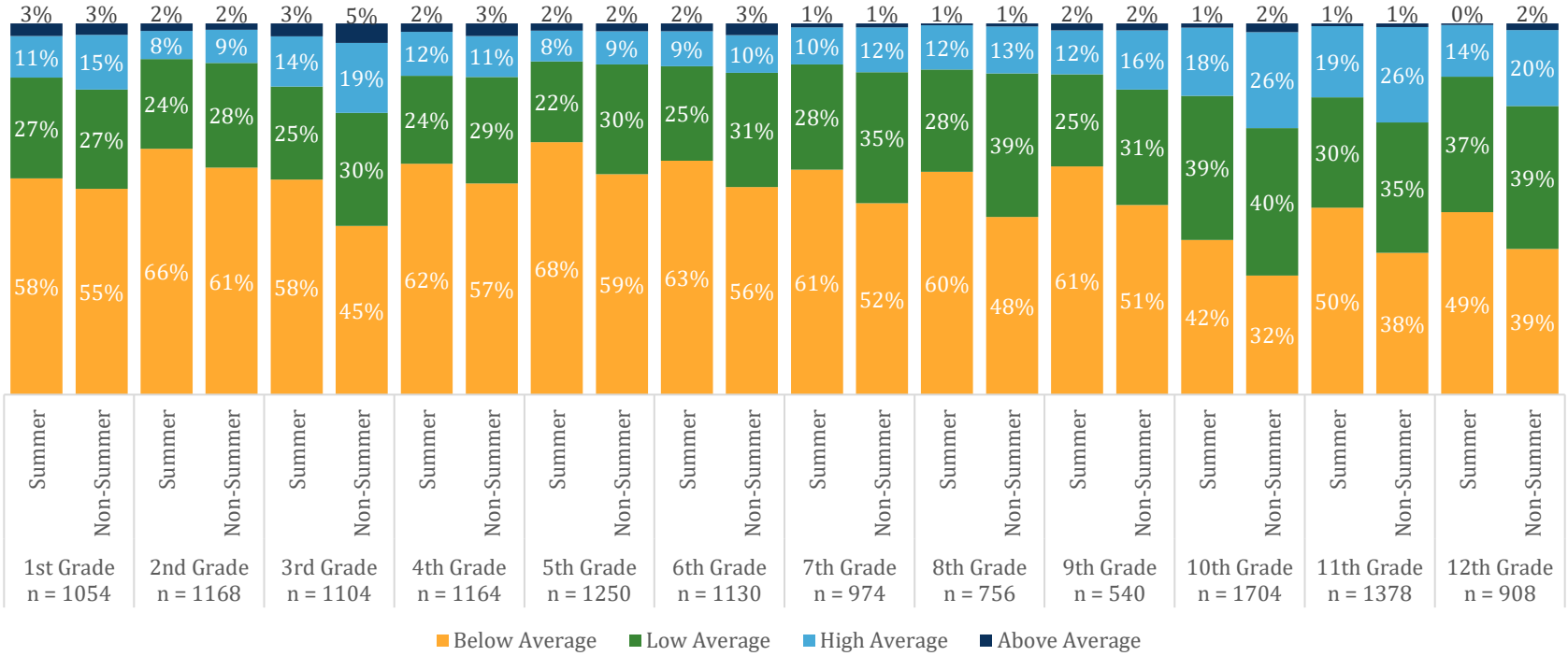


Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: Summer indicates students attended summer programming and non-summer indicates students did not attend summer programming. See Appendix C, Table C2 for average NCE scores. Due to rounding, bars with 0% labels can include as many as 24 students. The number of students in the *Summer* and *Non-Summer* columns is the n-count listed under the student group name divided by two.

Across all grade levels, a slightly smaller percentage of summer program attendees scored in the High Average (light blue) and Above Average (dark blue) fall Star Math NCE performance groups than non-attendees (Figure 4). There was a greater difference in the percentage points of 3rd, 10th, 11th, and 12th grade summer program attendees (14-20%) and non-attendees (22-28%) who scored in the High Average (light blue) and Above Average (dark blue) Star Math NCE performance groups than for students in any other grades.

Figure 4. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Math by 2021-22 grade level



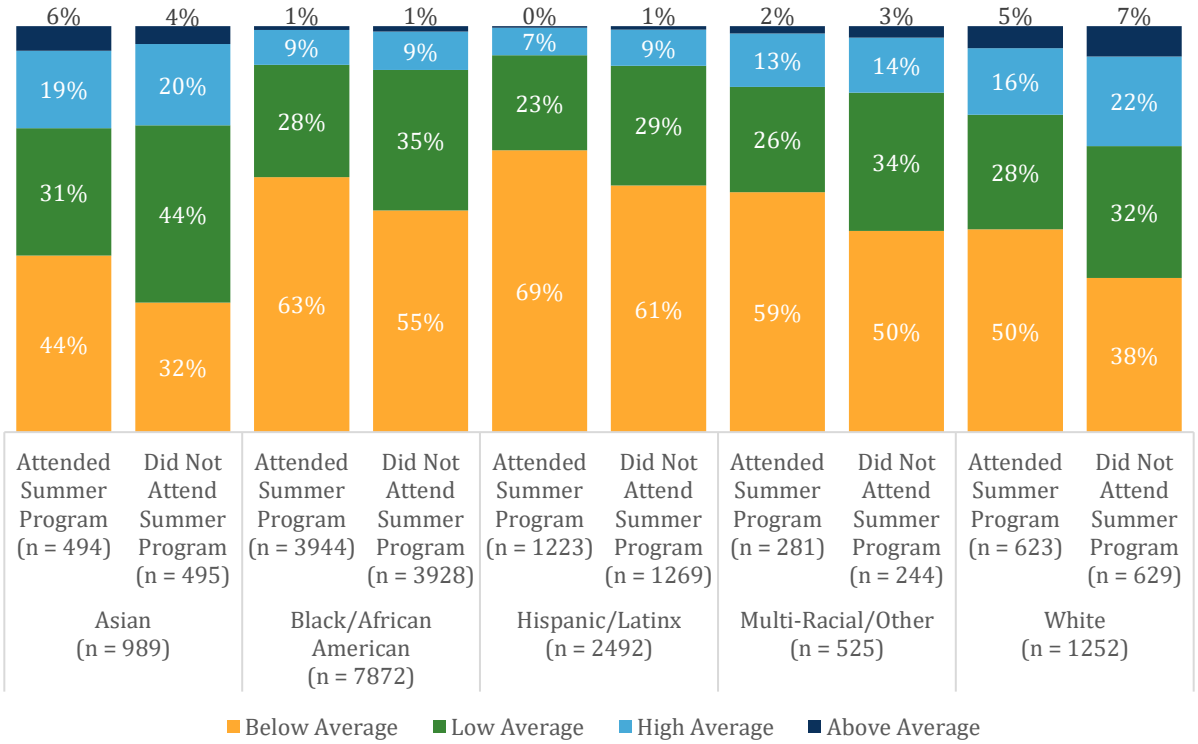
Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: Summer indicates students attended summer programming and non-summer indicates students did not attend summer programming. See Appendix C, Table C3 for average NCE scores. Due to rounding, bars with 0% labels can include as many as 24 students. The number of students in the *Summer* and *Non-Summer* columns is the n-count listed under the student group name divided by two.

A smaller or similar percentage of summer program attendees performed in the High Average and Above Average fall Star performance groups than non-summer program attendees across almost all racial/ethnic student groups.

Across almost all race/ethnicity student groups (Asian students did not follow this trend), a smaller percentage of summer program attendees scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups than non-attendees (Figure 5). The percentage of White students who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups was eight points lower for summer program attendees (21%) than non-attendees (29%). The percentage of students in all other race/ethnic student groups who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups differed by less than three points between summer program attendee status.

Figure 5. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Reading by student race/ethnicity

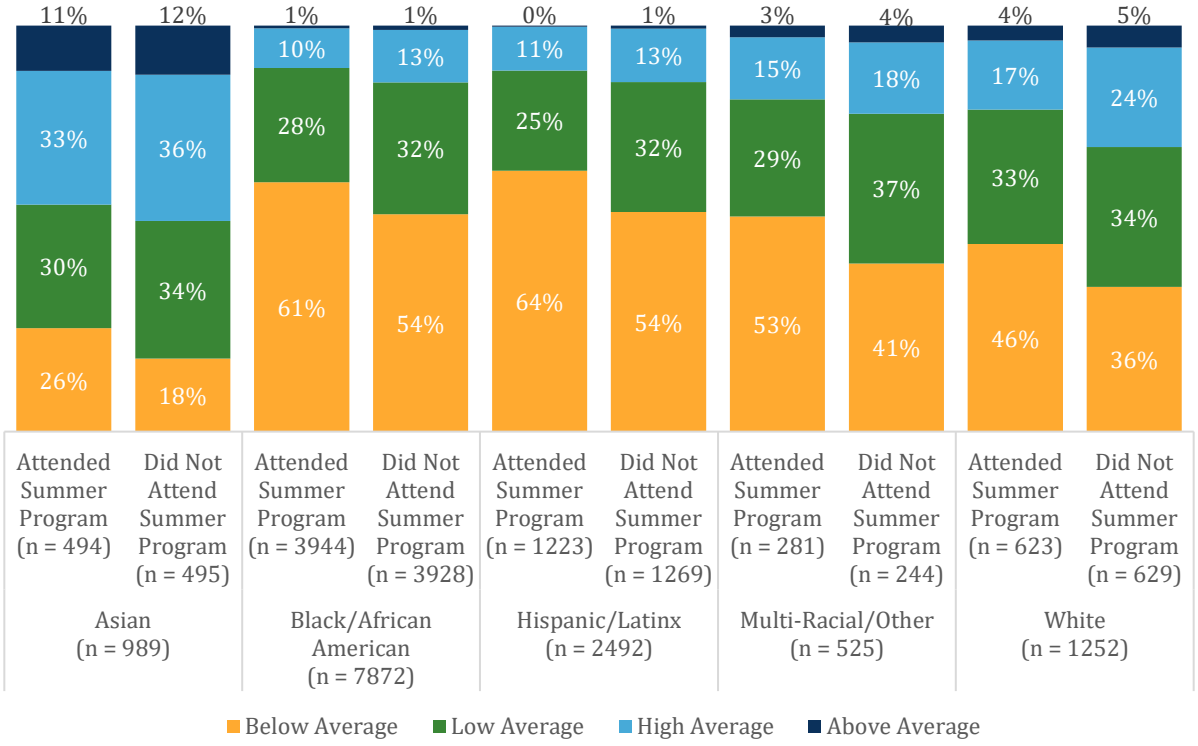


Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: See Appendix C, Table C4 for average NCE scores.

The percentage of students within race/ethnicity student groups who scored in the High Average (light blue) and Above Average (dark blue) fall Star Math NCE performance groups was smaller for summer program attendees than non-attendees (Figure 6). The percentage of White students who scored in the High Average (light blue) and Above Average (dark blue) fall Star Math NCE performance was eight points lower for summer program attendees (21%) than non-attendees (29%). Within all other racial/ethnic student groups, there was a four-percentage point or smaller difference in the proportion of summer program attendees and non-attendees within each Star Reading NCE performance level.

Figure 6. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Math by student race/ethnicity



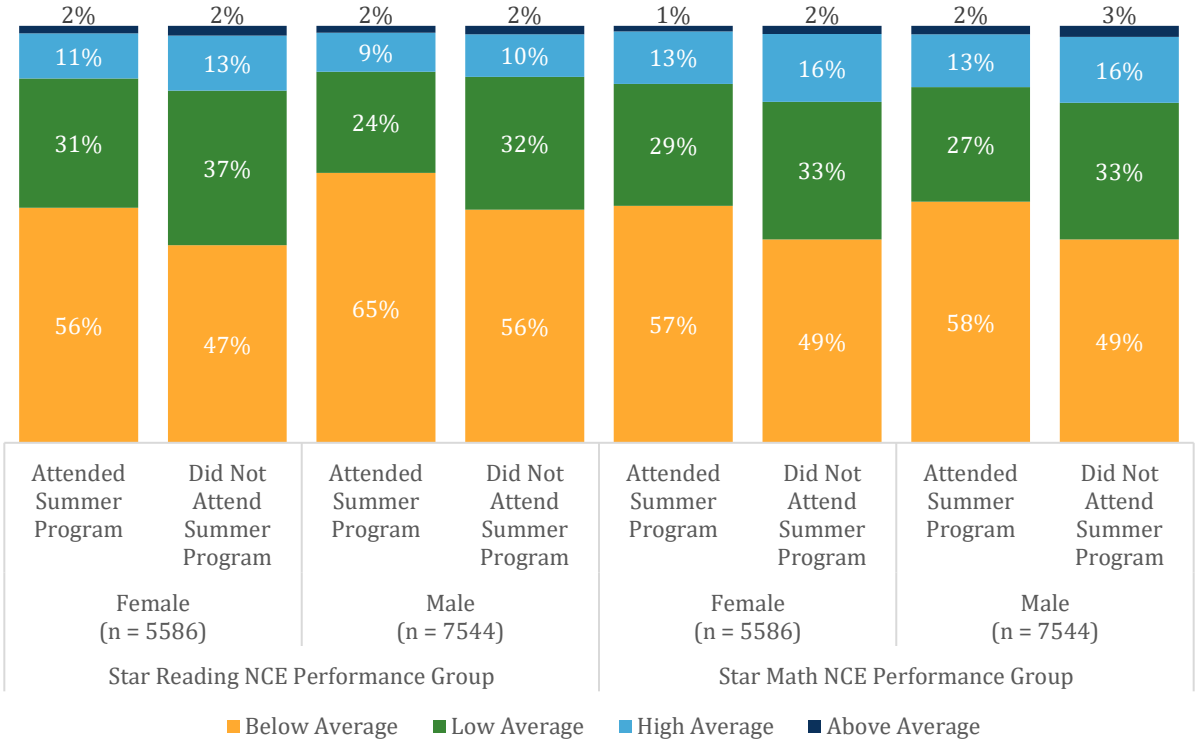
Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: See Appendix C, Table C4 for average NCE scores.

A slightly lower percentage of female and male summer program attendees performed in the High Average and Above Average Star Reading groups than non-attendees.

The percentage of both male and female students who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading and Math NCE performance groups was smaller for summer program attendees (11-15%) than non-attendees (12-19%) (Figure 7).

Figure 7. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Reading and Math by student gender



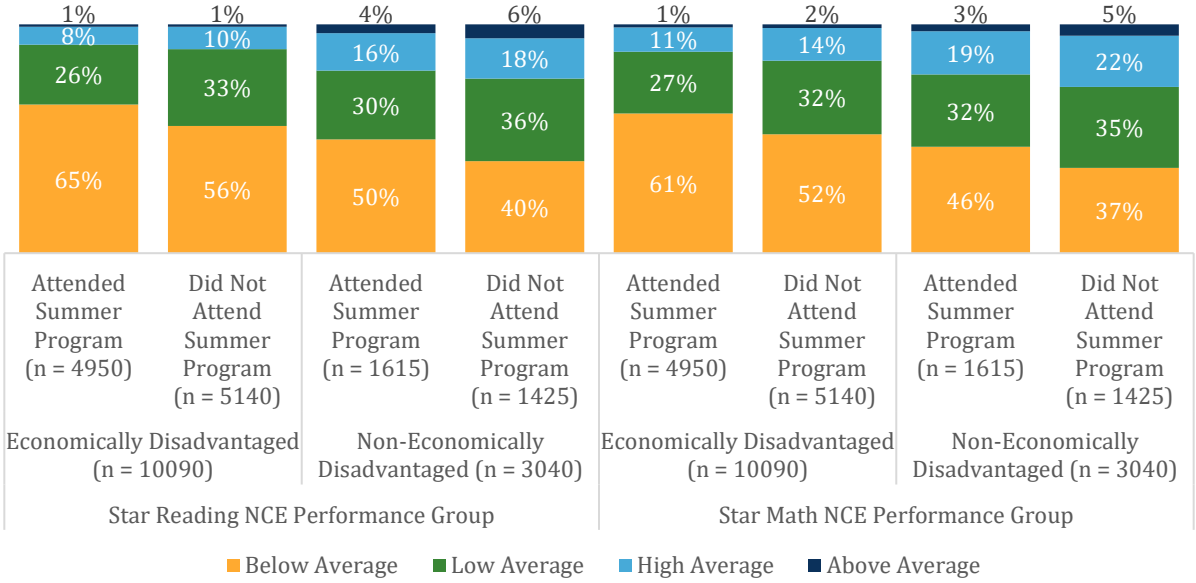
Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: See Appendix C, Table C5 for average NCE scores. The number of students in the *Summer* and *Non-Summer* columns is the n-count listed under the student group name divided by two.

A smaller percentage of economically disadvantaged students and non-economically disadvantaged summer program attendees performed in the High Average and Above Average fall Star Reading and Math groups than non-summer program attendees.

The percentage of economically disadvantaged students who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading and Math NCE performance groups differed by whether students were summer program attendees (9-12%) or non-attendees (11-16%). Similarly, the percentage of non-economically disadvantaged students scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading and Math NCE performance groups also differed by whether students were summer program attendees (20-22%) or non-attendees (24-27%) (Figure 8).

Figure 8. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Reading and Math by economically disadvantaged status



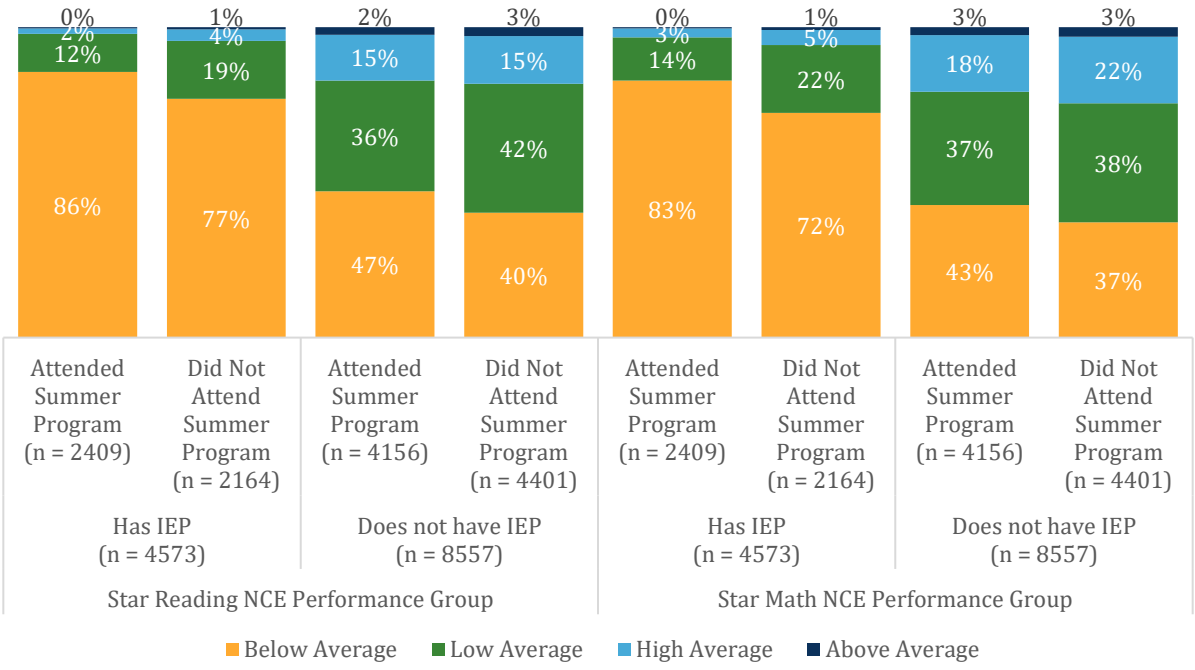
Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: See Appendix C, Table C6 for average NCE scores.

A smaller percentage of students with an IEP performed in the High Average and Above Average fall Star Reading and Math groups than students without an IEP, although the percentages did not differ much by whether students attended summer programming.

The percentage of students with an IEP who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading and Math NCE performance groups who attended summer programming (2-3%) was lower than students who did not attend summer programming (5-6%).

Figure 9. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Reading and Math by special education status



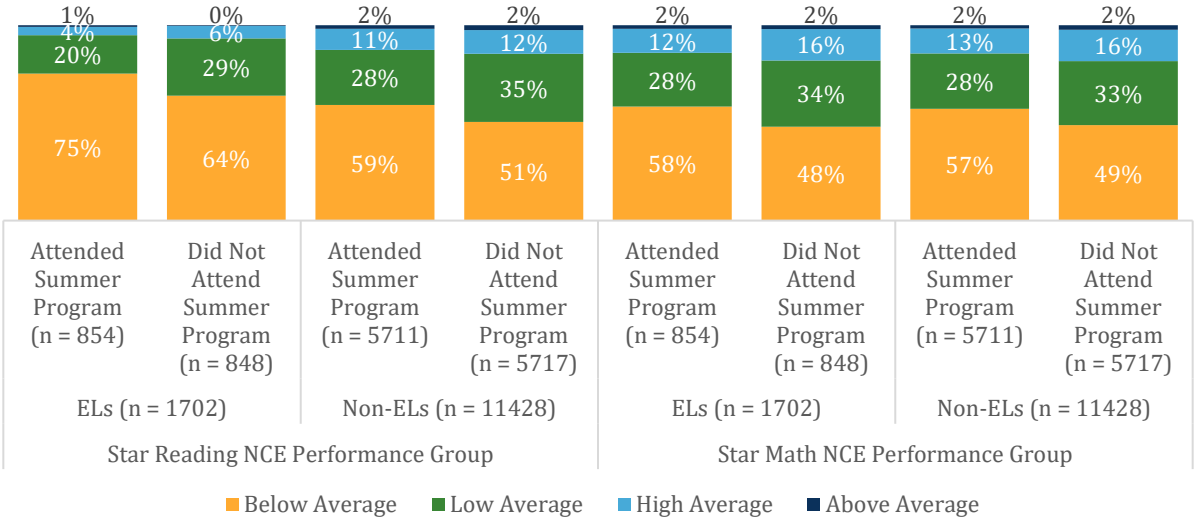
Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: See Appendix C, Table C7 for average NCE scores.

A similar percentage of English Learner and non-English Learner summer program attendees performed in the High Average and Above Average fall Star Reading and Math groups.

A similar percentage of English Learners (ELs) who attended and did not attend summer programming (5-6%) scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading NCE performance groups (Figure 10). However, the percentage of ELs who scored in the High Average (light blue) and Above Average (dark blue) fall Star Math NCE performance groups than ELs differed by whether they were summer program attendees (14%) or non-attendees (18%).

Figure 10. The percentage of students in the analytic sample who performed in the four NCE groups on fall Star Reading and Math by English Learner status



Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

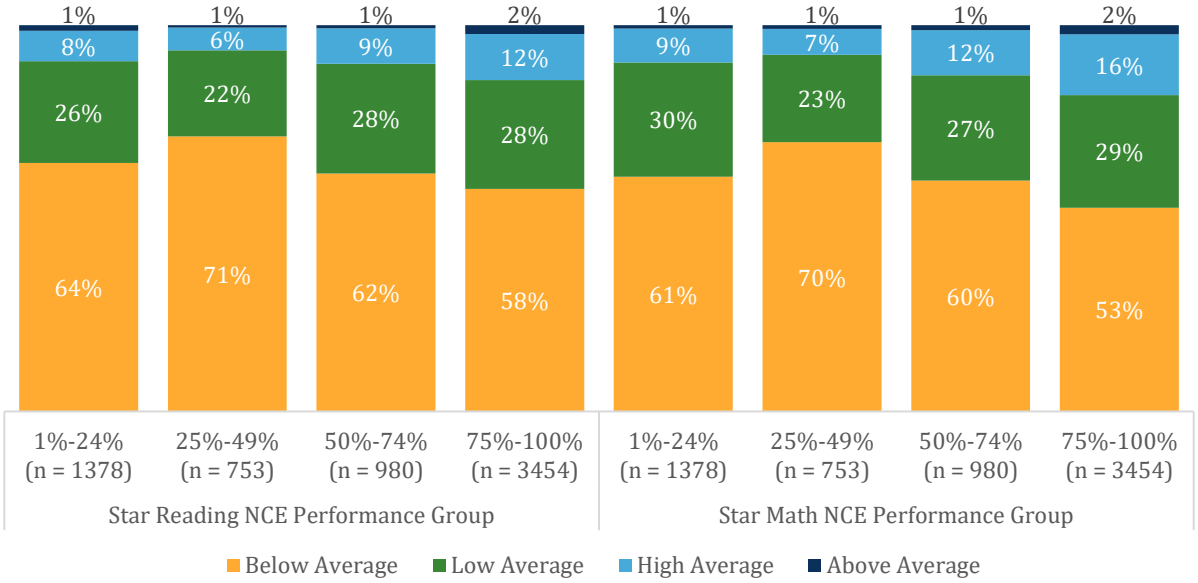
Note: See Appendix C, Table C8 for average NCE scores.

Were there performance differences on the fall 2021-22 Star Reading and Star Math assessments for students who attended 75-100% of summer program days compared to students who attended fewer than 75%?

Higher percentages of students who attended 75%-100% of summer program days scored in the High Average and Above Average fall Star Reading and Math groups than students who attended fewer than 75%-100% of summer program days.

Student performance on the Star Reading and Math are presented by the percentage of summer program days that students attended summer programming; this section only includes students who attended summer programming, not students in the matched sample (Figure 11). The percentage of students who scored in the High Average (light blue) and Above Average (dark blue) fall Star Reading and Math NCE groups was greater for students who attended summer programming for 75%-100% of days (14-18%) than students who attended fewer than 75%-100% of days (7-13%). Because the summer programs were between 10 to 19 days and five to six weeks, the relation between attendance levels and Star performance should be interpreted with caution. It is more likely that students who would go onto score in the High Average (light blue) and Above Average (dark blue) fall Star Reading and Math NCE groups would generally have higher summer attendance patterns than students with lower fall performance, rather than expecting attendance in a short program to promote high performance on Star Reading and Math in the fall.

Figure 11. The percentage of students in the analytic sample who attended summer programming and performed in the four NCE groups on fall Star Reading and Math within attendance group in their summer program



Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: See Appendix C, Table C9 for average NCE scores.

Summary of Results

Of the nearly 13,000 students who attended summer programming for at least one day, only about half of those students participated in fall 2021-22 Star Reading and Math assessments.[26] Students who attended summer programming—and took fall 2021-22 Star Reading and Math assessments—were matched with students who did not go to summer programming, but did attend the same grades and schools, and had similar demographics characteristics.

Star performance by whether students attended summer programming

When comparing the overall population of summer program attendees to their matched population of non-summer program attendees, there was only a two- to three-percentage point difference in the proportion of students who performed in the High Average and Above Average fall Star Reading and Math performance groups. This indicates that the students who attended summer programming had similar, albeit slightly lower, fall Star Reading and Math performance as a matched sample of students who did not attend summer programming. The notable aspect,

[26] About 25% of high school students who attended summer programming were taking Credit Recovery or Quarter 5 Grade Improvement courses in order to graduate during summer 2021. These students by definition would not have fall 2021-22 Star data because they graduated.

however, is that between 82%-88% of students in the analytic sample scored in the Low Average and Below Average fall Star Reading and Math performance groups. Because performance groups are based on NCE scores linked to national norms, this indicates that most students scored below the national average on fall Star assessments. Indeed, for each group in the analytic sample, the largest proportion of students (about half) was located in the lowest performance category for either Reading or Math (see Table C1).

Star performance by demographic student groups and whether students attended summer programming

When analyzing differences between summer program attendees and non-attendees within demographic student groups, the differences are larger than looking across all student groups. For example, there was an eight point difference between the percentage of White students who scored in the High Average and Above Average fall Star Reading and Math performance groups, favoring students who did not attend summer programming. Similarly, there was an 11 point difference between the percentage of non-economically disadvantaged students who scored in the High Average and Above Average fall Star Reading and Math performance groups, favoring students who did not attend summer programming.

With other student demographic groups, there were often differences between demographic groups, but minor differences between whether students attended or did not attend summer programming. For example, there was a 12 to 16 point difference between the percentage of students with an IEP and without an IEP who scored in High Average and Above Average fall Star Reading performance groups regardless of whether they attended summer programming.

Star performance by the percentage of summer programming days students attended

When looking into the performance of students who attended summer programming by the percentage of days that students were in attendance, an unexpected pattern emerged. In the High Average and Above Average fall Star Reading and Math performance groups, there was a difference of five to seven percentage points between students who attended summer programming for 75%-100% of days and students who attended fewer days—favoring those who attended more days.

Limitations

This analysis represents a population of students who attended 2021 summer programming. By default, this population of students is different from students who did not qualify for summer programming in some critical ways. Students typically attend summer programming because they require or would strongly benefit from intensive academic support, strategic support, or year-round schooling. At minimum, students who did not qualify for 2021 summer programming may not have faced the same type of social, academic, or economic challenges as students who qualified for summer programming. Due to the confluence of these challenges, we would anticipate students who qualified for summer programming to have lower overall assessment performance than students who did not qualify for summer programming, and therefore not be representative of the

District population overall (see Appendix B, Table B2 for comparisons between the analytic sample in this report and the overall District population in 2021-22).

As noted in the Introduction, this report focused on the subset of students who enrolled in *and* attended summer programs. However, there were a large number of District students who qualified and registered for the summer programs but did not attend. The reasons why these students did not attend vary widely, and typically come down to resources and not finding the summer programming worthwhile over other activities or commitments.[27] For example,

- Summer programming was five to six weeks and between 10 to 19 days (depending on the program). Some students were expected to provide childcare for their younger siblings, family members, or neighbors in order for their caregivers to work outside of the home during the summer (some continuing a practice from the school year), which would have overlapped with summer programming hours. Some students were expected to work outside of the house in order to provide for their family during the summer programming hours. Additionally, other students would not have been able to attend another summer program outside of the District that occurred during District summer programming hours. Especially for students who did not need to improve a grade or recover a credit in order to graduate, the time commitment and sacrifice of other summer activities or obligations may have outweighed the potential benefits of attending.
- In 2021, summer programming sites were scattered throughout the city, and students were not necessarily registered at the school closest to their home address or regular year-round school. Transportation was inconsistent or non-existent for some students living far from their assigned summer site due to transportation staffing shortages. In some cases, students could not get transportation, or the commute took over an hour each way.
- Students were burnt out or emotionally spent from over a year of virtual learning in 2020-21, and they were hesitant to continue virtual learning in the summer, reluctant to continue in an academic context for the summer when they ordinarily would have had a summer break to recharge, fearful of returning in person when COVID-19 was still prevalent and safety protocols were uncertain, or struggling with social-emotional health.
- Summer 2021 was unique in that it was the first large-scale effort to open numerous District schools to full capacity since the COVID-19 pandemic required schools move to virtual learning in March 2020. This effort was met with many complications and frustration amongst staff, families, and students alike. Students and families may have decided summer programming was more trouble than it was worth when faced with not knowing where students were registered on the first day of summer programming, believing schools were ill-prepared, a lack of confidence that students were learning, concerns about adherence to COVID-19 safety protocols, or schools that were understaffed.

Overall, students who qualified for summer programming but did not attend likely faced additional challenges compared to students who attended summer programming.

[27] For more information about summer 2021 programming, see https://www.philasd.org/research/2021/12/08/2021-summer-programs-in-sdp-offerings-attendance-survey-results-and-recommendations

Lastly, the PSM process of identifying a matched sample of students did not account for spring 2020-21 reading and math performance. Spring 2020-21 performance was excluded from the PSM process due to an overwhelming amount of missing data that would have reduced the analytic sample from about 13,000 to under 6,000 students, and would have further limited the generalizability of the sample to the population of students who had scores in both spring and fall, and who attended summer programming (or matched the characteristics of students who attended summer programming). The addendum to this report controls for spring 2021 performance on the Star reading and math assessments.

Appendix A: Propensity Score Matching

This report used propensity score matching (PSM; Austin, 2011; Rosenbaum & Rubin, 1983; Stuart, 2010) to study the effect of summer programming on 2021-22 fall academic screener performance. PSM is a method for examining the effect of a treatment (e.g., summer programming) on an outcome (e.g., fall academic performance) when random assignment is not feasible. PSM employs a statistical approach to form treatment and control groups that are balanced on measured characteristics (e.g., gender, race/ethnicity, socioeconomic status, etc.). After the groups are balanced, the average effect of the treatment can be estimated by comparing average performance on the outcomes between the two groups. By applying the PSM method, the influence of baseline differences between the treatment and control groups are reduced when estimating the treatment effect, although these differences may not necessarily be eliminated.[28] In this analysis, the treatment refers to summer programming and the outcomes are fall academic screener reading and math performance. The term treatment group is used to describe students who were enrolled in summer programming, and control group is used to describe students who were enrolled in the 2020-21 and 2021-22 school years but did not attend summer programming.[29]

The PSM procedure operates as follows. First, a sample of students is obtained that includes both students who attended summer programming (the treatment group) and students who did not attend summer programming (pool of control students). A statistical model is applied in order to estimate a *propensity score* for each student: The propensity score indicates the probability that a given student would have been assigned to the treatment group and it is based on a set of variables that are included in the propensity score model; variables in the model should be related to either the treatment, the outcome(s), or both (Austin, 2011; Rosenbaum & Rubin, 1983; Stuart, 2010). After propensity scores are estimated, students from the pool of control students are matched to students in the treatment group using a matching algorithm; we employed nearest neighbor matching. This approach matches students who have the smallest difference in their propensity scores, where a good match would be a pair of students, one in the treatment and one in the pool of controls, who have equal propensity scores (Austin, 2011; Stuart, 2010). In the matching phase, the researcher can also apply various matching restrictions in order to obtain the desired match quality (Jacovidis et al., 2017). For example, a caliper can be applied which only allows matches that are within a certain range of propensity scores, or exact matching can be required on a subset of variables. The matching process concludes when all students in the treatment group are matched to a student from the pool of controls, the latter henceforth termed the control group.[30] By matching on the propensity score, the students in the treatment and control groups are balanced on the

[28] A randomized controlled trial (RCT), where District students are randomly assigned into a treatment (attending summer programming) and a control group (not attending summer programming) could potentially account for all differences in baseline characteristics, on average (Ho et al., 2007). However, an RCT may not always be feasible, and PSM is one approach for approximating the RCT design when the option of randomization is not available (Austin, 2011; Ho et al., 2007; Stuart, 2011).

[29] Based on District data pulled from the Qlik Total Student Enrollment Yearly Report, accessed 10-10-2021.

[30] Depending on the matching requirements, not all students in the treatment group may have an acceptable match. This will result in students from the treatment group being removed from the sample.

characteristics included in the statistical model (Rosenbaum & Rubin, 1983; Stuart, 2010). After matching, match quality is evaluated using numerical and graphical summaries (Ho et al., 2011; Stuart, 2010).

The PSM approach assumes that all variables affecting enrollment in summer programming (treatment) and all variables related to fall academic screener scores (the outcomes) are included in the model used to estimate the propensity score (Austin 2011; Ho et al., 2007; Stuart, 2010). Key variables that are not included in the model will not be subject to the PSM procedure, therefore, the treatment and control groups may remain unbalanced with respect to those variables. As a result, existing baseline differences will persist when estimating the treatment effect.

In this study, the full sample was comprised of all students enrolled in the 2020-21 school year.[31] To be eligible for the final analytical sample (i.e., matched group of students), we required that 1) students had to have both Star reading and math scores within the 2021-22 fall screening window (if a student had multiple scores for a subject, the best score was used), and 2) students in the treatment group had to have attended summer programming for at least one day. Table A1 presents the demographic characteristics for both full sample and the group of eligible students.[32]

The group of eligible students well-represents the full population of students who did and did not attend summer programming (Table A1). This means that these students did not generally differ from students who were excluded from the sample due to not having taken progress monitoring assessments. For example, in both the full sample of students who attended summer programming and the group of eligible students who attended summer programming 60% of the sample was Black/African American, 75%-76% of the sample was economically disadvantaged, 13% were English Learners, and 37%-40% of the sample had an IEP.

[31] Based on data from Qlik Total Student Enrollment Yearly, data accessed on October 2021.

[32] This the PSM did not account for spring 2020-21 reading and math performance. Spring 2020-21 performance was excluded from the PSM process due to an overwhelming amount of missing data that would have reduced the analytic sample from about 13,000 to under 6,000 students, and would have further diminished the generalizability of the sample to the population of students with scores in both spring and fall, and attended summer programming (or matched the characteristics of students who attended summer programming). The addendum to this report controls for spring 2021 performance on the Star reading and math assessments.

Table A1. Demographic characteristics for the full sample and sample of eligible students

Student Group	Full sample		Eligible students	
	Summer = 0	Summer = 1	Summer = 0	Summer = 1
Number of students	203,562	15,442	75,327	6,663
Gender				
Female	49%	40%	50%	42%
Male	51%	60%	50%	57%
Race/Ethnicity				
Asian/Pacific Islander	7%	6%	11%	8%
Black/African American	52%	60%	45%	60%
Hispanic/Latinx	21%	21%	23%	19%
Multi-Racial/Other	14%	5%	5%	4%
White	6%	9%	16%	10%
Socio-Economic Status				
Economically Disadvantaged	40%	76%	69%	75%
Not Economically Disadvantaged	60%	24%	31%	25%
English Learner Status				
English Learner	10%	13%	14%	13%
Not an English Learner	90%	87%	86%	87%
Special Education Status				
Students with IEPs	16%	40%	12%	37%
Students without IEPs	84%	60%	88%	63%

The sample of eligible students was submitted to propensity score matching using the MatchIt package (Ho et al., 2011) in the statistical program R (R Core Team, 2020). Propensity scores were estimated using a logistic regression model that included student demographic characteristics. The model was:

$$logit(P(Y_i = 1|\boldsymbol{X}_i)) = \beta_0 + \beta_1 Gender_i + \beta_2 Race/Ethnicity_i + \beta_3 SES\ Status_i + \beta_4 English\ Learner_i + \beta_5 Special\ Education\ Status_i \quad (1)$$

In Equation 1, i indexes an individual student in the sample. The variables gender, race/ethnicity, Economically Disadvantaged status, English Learner status, and Special Education status are each categorical variables. The dependent variable, Y_i, is a binary treatment variable where 1 indicates student i attended summer programming for at least one day and 0 means the student was not enrolled in summer programming. The value, $P(Y_i = 1|\boldsymbol{X}_i)$, is the estimated propensity score indicating the probability that student i would have been assigned to the treatment group based on that student's characteristics (Arpino & Cannas, 2017). In the matching process, the logit of the propensity score, $logit(P(Y_i = 1|\boldsymbol{X}_i))$, was used (Austin, 2011; Stuart, 2010).

Next, the nearest neighbor method with 1:1 matching was applied to match the pool of control students with treatment students (Ho et al., 2011). We restricted matches to be exact on school and grade level. We did not apply a caliper (require matches to be within a specific propensity score range) in order to retain the maximum number of students possible in the treatment group (Jacovidis et al., 2017). Based on the matching specifications, a total of 76 students in the treatment

group were dropped from the resulting analytical sample due to the school by grade level restriction.

Table A2 shows the composition of the treatment and control groups before and after matching. To evaluate match quality, we examined the column labeled "Standardized Mean Difference" (SMD). The SMD is a measure of how different the treatment and control groups are on each variable (Ho et al., 2011; Stuart, 2010; Zhang et al., 2019). The SMD can be examined before and after matching, and variables with SMD values less than |.10| after matching are considered balanced (Ho et al., 2011; Zhang et al., 2019). Results met this criterion.

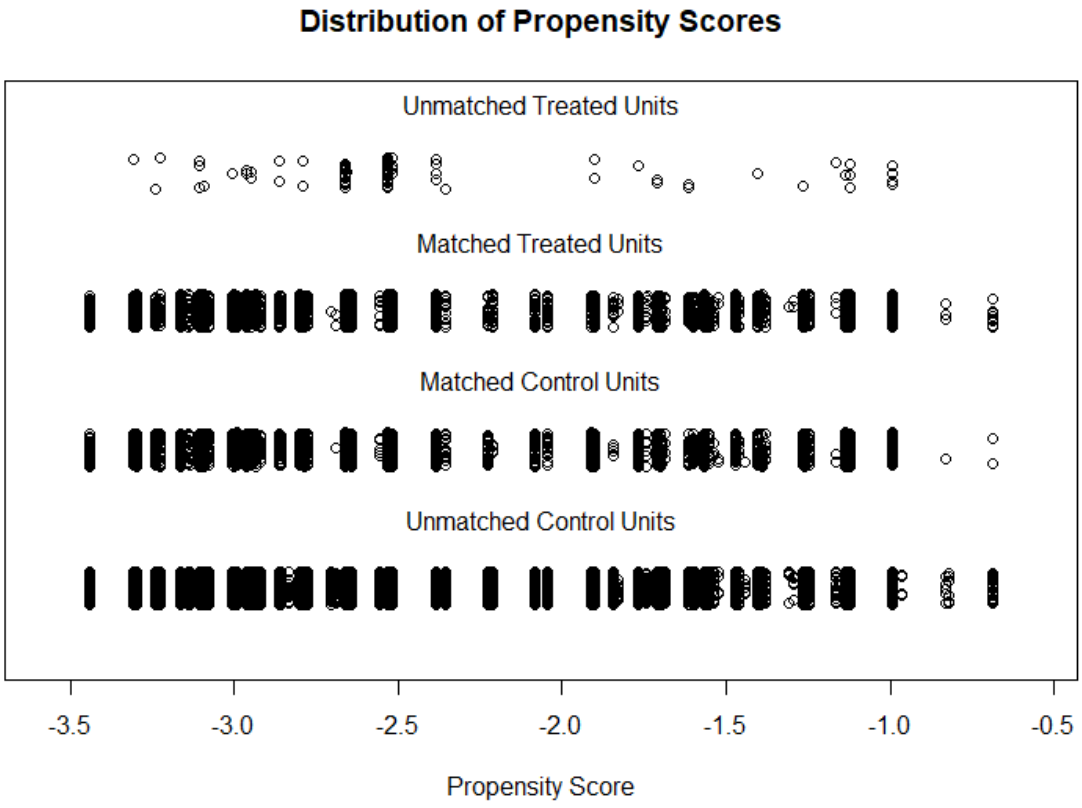
Table A2. Standardized mean difference for the student samples

Student Group	Eligible students			Analytical sample		
	Summer = 0	Summer = 1	Mean Difference	Summer = 0	Summer = 1	Mean Difference
Number of students	75,327	6,663		6,587	6,587	
Gender						
Female	50%	42%	-0.16	43%	42%	-0.01
Male	50%	57%	0.16	57%	58%	0.01
Race/Ethnicity						
Asian/Pacific Islander	11%	8%	-0.14	8%	8%	0.00
Black/African American	45%	60%	0.31	60%	60%	0.01
Hispanic/Latinx	23%	19%	-0.11	19%	19%	-0.02
Multi-Racial/Other	5%	4%	-0.02	4%	4%	0.03
White	16%	10%	-0.23	10%	10%	0.00
Socio-Economic Status						
Economically Disadvantaged	69%	75%	0.14	78%	75%	-0.06
Not Economically Disadvantaged	31%	25%	-0.14	22%	25%	0.06
English Learner Status						
English Learner	14%	13%	-0.02	13%	13%	0.00
Not an English Learner	86%	87%	0.02	87%	87%	0.00
Special Education Status						
Students with IEPs	12%	37%	0.51	33%	37%	0.08
Students without IEPs	88%	63%	-0.51	67%	63%	-0.08

Note: Standardized mean difference indicates the proportion difference for each variable. Values less than |.10| suggest the variables are balanced (Ho et al., 2011; Zhang et al., 2019).

In addition to the SMD, we also inspected the degree that the estimated propensity scores overlapped in the analytical sample (Stuart, 2010). Figure A1 presents a jitter plot that plots each student's estimated propensity score along the distribution of those scores. The two rows labeled "Matched Treated Units" and "Matched Control Units" are of primary interest. Using Figure A1 and concentrating on these two rows, the distribution of propensity scores across the two groups should overlap (Stuart, 2010). Results showed that scores displayed sufficient overlap.

Figure A1. Jitter plot displaying overlap between the treatment and control groups



Several limitations should be noted. The first pertains to the variables used in our propensity score model. Specifically, we used demographic characteristics (gender, race/ethnicity, Economically Disadvantaged status, English Learner status, and Special Education status) in the model. As noted previously, PSM assumes that all variables related to either the treatment (summer programming) or the outcomes (fall academic screener scores) are in the model, and those not included will remain unbalanced in the estimation of the treatment effect. As such, preexisting differences between the two groups will persist when estimating the treatment effect; thus, the resulting treatment effect estimates should be interpreted in light of variables omitted from the model. Second, our approach for selecting the sample of eligible students resulted in a large reduction in the treatment group (among the full sample of summer programming students, ~20.5% did not attend at least one day and ~47.3% did not have both fall reading and math scores). As such, the effect should also be interpreted in light of this change in sample size.

Appendix B: The Student Sample

Table B1. The demographic characteristics of analytic sample of students who attended summer programs and a matched sample of students who did not attend summer programs

Student Characteristic	Attended Summer Program	Did Not Attend Summer Program (matched sample)	Total Number of Students
Grade Level			
1	50%	50%	1054
2	50%	50%	1168
3	50%	50%	1104
4	50%	50%	1164
5	50%	50%	1250
6	50%	50%	1130
7	50%	50%	974
8	50%	50%	756
9	50%	50%	540
10	50%	50%	1704
11	50%	50%	1378
12	50%	50%	908
Race/Ethnicity			
Asian	50%	50%	989
Black/African American	50%	50%	7872
Hispanic/Latinx	49%	51%	2492
Multi-Racial/Other	54%	46%	525
White	50%	50%	1252
Gender			
Female	50%	50%	5586
Male	50%	50%	7544
Socio-Economic Status			
Economically Disadvantaged	49%	51%	10090
Not Economically Disadvantaged	53%	47%	3040
Special Education Status			
Students with IEPs	53%	47%	4573
Students without IEPs	49%	51%	8557
English Learner Status			
English Learner	50%	50%	1702
Not an English Learner	50%	50%	11428

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021.

How to read this table: This table displays the percentage of each demographic group represented by summer program attendees and non-attendees. For example, 49% of Hispanic/Latinx students in the overall sample were summer attendees, and 51% were non-summer attendees. The column on the right displays the total number of students in each student group by summing summer program attendees and non-attendees.

Table B2. The demographic characteristics of analytic sample and the overall District population

Student Characteristic	Analytic Sample	District Overall
Race/Ethnicity		
Asian	8%	10%
Black/African American	60%	47%
Hispanic/Latinx	19%	24%
Multi-Racial/Other	4%	4%
White	10%	15%
Gender		
Female	43%	49%
Male	57%	51%
Special Education Status		
Students with IEPs	35%	16%
Students without IEPs	65%	84%
English Learner Status		
English Learner	13%	15%
Not an English Learner	87%	85%

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021. District data retrieved from https://www.philasd.org/performance/programsservices/open-data/school-information/#district_enrollment

How to read this table: This table includes the percentages of students in the student groups in the analytic sample and the District overall. The *Analytic Sample* column includes the same percentages as displayed in Table 3. The *District Overall* column displays the percentages of students in the listed student groups at all District schools in the 2021-22 school year, and includes the sample of students in the *Analytic Sample* column.

Note: The *District Overall* column includes all students in the District, not only students who were qualified to attend summer programming or eligible to be matched with students who attended summer programming.

Appendix C: Average NCE Scores

Table C1. Average NCE scores on the fall Star Reading and Math assessments by whether students attended summer programming and by their respective NCE performance groups

Star Assessment	Group	Below Average	Low Average	High Average	Above Average
Reading	Attended Summer Program	7.7	36.2	58.5	82.2
	Did Not Attend Summer Program	8.9	36.0	58.3	83.2
Math	Attended Summer Program	8.5	36.2	59.4	83.1
	Did Not Attend Summer Program	9.9	36.5	59.4	83.8

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Note: Both outcomes include 6,565 students who attended summer programming and 6,565 who did not attend summer programming.

Table C2. Average NCE scores on the fall Star Reading assessments by whether students attended summer programming, grade levels, and by their respective NCE performance groups

Group	Grade Level	Below Average	Low Average	High Average	Above Average
Attended Summer Programming	1	10.4	35.9	59.8	83.3
	2	6.6	34.6	59.1	82.8
	3	7.1	37.4	57.8	82.3
	4	6.4	36.3	59.0	81.9
	5	6.3	36.7	58.1	81.5
	6	7.3	35.5	58.9	80.8
	7	7.1	35.7	59.6	84.6
	8	7.4	36.1	58.2	79.0
	9	5.8	36.3	59.2	86.9
	10	9.6	36.2	57.2	80.5
	11	8.8	36.7	57.4	78.2
	12	9.5	36.9	57.9	87.2
Did Not Attend Summer Program	1	10.9	35.2	59.0	82.6
	2	7.0	35.6	58.3	86.0
	3	8.7	37.0	59.7	82.7
	4	7.2	36.6	58.5	84.6
	5	8.3	35.6	59.3	80.4
	6	9.0	36.2	58.1	80.9
	7	8.2	35.6	58.7	81.6
	8	10.0	35.6	56.5	77.2
	9	7.6	37.4	59.2	79.6
	10	10.3	36.1	56.9	82.5
	11	9.9	36.4	56.7	85.8
	12	9.7	35.3	58.5	77.6

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C3. Average NCE scores on the fall Star Math assessments by whether students attended summer programming, grade levels, and by their respective NCE performance groups

Group	Grade Level	Below Average	Low Average	High Average	Above Average
Attended Summer Programming	1	8.8	37.3	61.3	84.7
	2	6.5	35.1	60.8	84.0
	3	7.9	36.8	61.2	82.0
	4	7.8	36.4	59.6	84.7
	5	7.5	35.3	59.3	82.7
	6	8.0	35.8	59.4	84.1
	7	7.7	35.2	59.8	79.6
	8	9.0	36.9	60.4	89.3
	9	7.0	35.7	60.3	88.4
	10	11.5	35.6	58.7	79.8
	11	10.9	37.1	57.9	78.6
	12	9.6	37.3	58.1	79.0
Did Not Attend Summer Program	1	9.1	36.9	60.0	84.6
	2	7.9	35.5	58.3	83.1
	3	10.1	36.6	61.4	83.1
	4	8.8	35.9	58.9	85.3
	5	8.9	36.1	58.0	85.1
	6	9.3	36.3	59.1	83.4
	7	8.7	36.0	60.6	82.3
	8	10.9	36.0	59.8	79.8
	9	10.1	37.4	58.9	84.8
	10	13.6	36.0	59.4	81.8
	11	11.9	37.3	59.4	86.9
	12	11.4	38.5	58.1	85.2

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C4. Average NCE scores on the fall Star Reading and Math assessments by whether students attended summer programming, student race/ethnicity, and by their respective NCE performance groups

Star Assessment	Race/ Ethnicity	Group	Below Average	Low Average	High Average	Above Average
Reading	Asian	Attended Summer Program	8.1	37.0	59.8	82.0
		Did Not Attend Summer Program	11.0	37.3	58.4	83.5
	Black/ African American	Attended Summer Program	8.0	36.1	57.8	81.4
		Did Not Attend Summer Program	8.9	35.8	57.5	81.9
	Hispanic/ Latinx	Attended Summer Program	6.9	36.3	58.6	86.0
		Did Not Attend Summer Program	8.1	35.8	58.0	82.6
	Multi-Racial/ Other	Attended Summer Program	6.1	36.7	58.3	85.3
		Did Not Attend Summer Program	8.2	35.6	59.5	85.6
	White	Attended Summer Program	8.4	36.2	59.6	82.4
		Did Not Attend Summer Program	9.9	36.3	60.5	84.2
Math	Asian	Attended Summer Program	9.7	38.3	61.2	86.0
		Did Not Attend Summer Program	10.4	38.7	61.3	84.6
	Black/ African American	Attended Summer Program	8.4	36.1	58.5	79.6
		Did Not Attend Summer Program	9.7	36.3	58.4	81.5
	Hispanic /Latinx	Attended Summer Program	8.5	35.9	59.1	78.8
		Did Not Attend Summer Program	9.9	36.1	60.0	81.7
	Multi-Racial/ Other	Attended Summer Program	8.4	34.7	60.6	82.4
		Did Not Attend Summer Program	9.0	35.5	61.1	84.9
	White	Attended Summer Program	9.2	36.2	60.1	81.1
		Did Not Attend Summer Program	11.9	37.0	59.5	85.6

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C5. Average NCE scores on the fall Star Reading and Math assessments by whether students attended summer programming, gender, and by their respective NCE performance groups

Star Assessment	Gender	Group	Below Average	Low Average	High Average	Above Average
Reading	Female	Attended Summer Program	8.7	36.4	58.8	83.8
		Did Not Attend Summer Program	9.7	35.9	58.0	84.1
	Male	Attended Summer Program	7.1	36.1	58.2	81.0
		Did Not Attend Summer Program	8.4	36.1	58.7	82.3
Math	Female	Attended Summer Program	9.0	36.6	59.1	83.4
		Did Not Attend Summer Program	10.0	36.6	59.2	83.6
	Male	Attended Summer Program	8.2	35.9	59.7	83.0
		Did Not Attend Summer Program	9.8	36.4	59.6	83.9

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C6. Average NCE scores on the fall Star Reading and Math assessments by whether students attended summer programming, economically disadvantaged status, and by their respective NCE performance groups

Star Assessment	Economically Disadvantaged status	Group	Below Average	Low Average	High Average	Above Average
Reading	Economically Disadvantaged	Attended Summer Program	7.6	35.9	58.0	80.8
		Did Not Attend Summer Program	8.8	35.8	57.7	81.9
	Non-Economically Disadvantaged	Attended Summer Program	8.3	37.1	59.2	83.4
		Did Not Attend Summer Program	9.3	36.8	59.7	83.9
Math	Economically Disadvantaged	Attended Summer Program	8.3	36.1	59.1	83.5
		Did Not Attend Summer Program	9.6	36.3	59.0	83.8
	Non-Economically Disadvantaged	Attended Summer Program	9.5	36.5	60.0	82.7
		Did Not Attend Summer Program	11.1	37.3	60.4	83.9

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C7. Average NCE scores on the fall Star Reading and Math assessments by whether students attended summer programming, special education status, and by their respective NCE performance groups

Star Assessment	Special Education Status	Group	Below Average	Low Average	High Average	Above Average
Reading	Has IEP	Attended Summer Program	5.0	34.6	57.0	81.8
		Did Not Attend Summer Program	6.7	34.5	57.5	84.1
	Does not have IEP	Attended Summer Program	10.6	36.5	58.6	82.3
		Did Not Attend Summer Program	10.9	36.4	58.4	83.1
Math	Has IEP	Attended Summer Program	6.1	33.8	61.4	84.2
		Did Not Attend Summer Program	8.0	34.3	58.0	86.7
	Does not have IEP	Attended Summer Program	11.2	36.7	59.3	83.0
		Did Not Attend Summer Program	11.6	37.1	59.6	83.4

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C8. Average NCE scores on the fall Star Reading and Math assessments by whether students attended summer programming, English Learner status, and by their respective NCE performance groups

Star Assessment	Special Education Status	Group	Below Average	Low Average	High Average	Above Average
Reading	ELs	Attended Summer Program	6.3	35.1	57.6	80.9
		Did Not Attend Summer Program	7.9	35.4	57.5	83.6
	Non-ELs	Attended Summer Program	8.0	36.3	58.5	82.3
		Did Not Attend Summer Program	9.1	36.1	58.4	83.2
Math	ELs	Attended Summer Program	8.6	36.8	61.0	88.3
		Did Not Attend Summer Program	10.2	36.2	59.0	82.1
	Non-ELs	Attended Summer Program	8.5	36.1	59.2	82.3
		Did Not Attend Summer Program	9.8	36.5	59.5	84.0

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

Table C9. Average NCE scores on the fall Star Reading and Math assessments by percentage of summer program days attended and by NCE performance groups

Star Assessment	Summer Program Days Attended	Below Average	Low Average	High Average	Above Average
Reading	**1%-24% days**	7.9	36.0	58.4	82.9
	25%-49% days	6.8	34.8	59.0	83.2
	50%-74% days	7.4	35.9	58.7	84.7
	75%-99% days	8.0	36.7	58.4	81.8
Math	**1%-24% days**	9.0	35.9	57.8	85.4
	25%-49% days	7.3	35.8	60.1	82.4
	50%-74% days	8.3	35.7	59.1	83.5
	75%-99% days	8.7	36.5	59.8	82.8

Source: Data from Qlik Summer Program Enrollment and Attendance, data accessed on August 16, 2021; Qlik Total Student Enrollment Yearly, data accessed on October 6, 2021; Qlik Academic Screeners, data accessed on October 27, 2021.

References

Arpino, B., & Cannas, M. (2017). Propensity score matching with clustered data. An application to the estimation of impact of caesarean section on the Apgar score. *Statistics in Medicine, 35,* 2074-2091. https://doi.org/10.1002/sim.6880

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424. https://doi.org/10.1080/00273171.2011.568786

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15,* 199-236. https://doi.org/10.1093/pan/mpl013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42,* 1-28. https://www.jstatsoft.org/v42/i08/

Jacovidis, J. N., Foelber, K. J., & Horst, S. J. (2017). The effect of propensity score matching method on the quantity and quality of matches. *The Journal of Experimental Education, 85,* 535-558. https://doi.org/10.1080/00220973.2016.1250209

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. UQlik https://www.R-project.org/.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25,* 1-21. https://doi.org/10.1214/09-STS313

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of Translational Medicine, 7*(1), 1-11. https://doi.org/10.21037%2Fatm.2018.12.10