Efficacy Analysis of Zearn Math in DC Public Schools

December 2022

Alisa Szatrowski, Ph.D.
Jessica Rickel, M.A.
Charlie Rosemond, M.S.
Zearn

Abstract

Analysis of consistent Zearn Math users and a comparable group of non-users in District of Columbia Public Schools (DCPS) shows that students who used Zearn Math consistently had higher levels of academic growth than similar students with low- or no-usage. This analysis used two quasi-experimental methods, Coarsened Exact Matching (CEM) and fixed effects, to examine the impact of Zearn Math across ~1,000 students who completed an average of 3+ Zearn Math lessons per week, during the 2018-2019, 2020-2021 and/or 2021-2022 school years. Findings showed that students who used Zearn Math consistently grew 36% more; the equivalent of 10.5 additional weeks of learning in one school year, as measured by the i-Ready Diagnostic. Students who started two or more years below grade-level had almost 2x the growth with consistent Zearn Math usage as similar students with low or no usage; the equivalent of an additional 23.7 weeks. Consistent Zearn Math usage was shown to result in a 32% reduction in summer slide, equivalent to five weeks of learning over the summer. Students with consistent Zearn Math usage in 2020-2021 maintained almost 75% of their additional growth a year after they stopped using Zearn Math.

Table of Contents

List of Figures

List of Tables

Executive Summary

A new study analyzing the impact of District of Columbia Public Schools' (DCPS) partnership with Zearn offers research-backed evidence that Zearn Math drives score growth for all students, resulting in long-term gains in recovered learning.

21,000+ students, grades 1-6
74% economically disadvantaged
78% Black and Latino
15% English learners

This report highlights findings from a multi-part study, examining:

- The impact of Zearn Math on within-year academic growth;
- Persistence of growth after students discontinue use; and
- The impact of Zearn Math on lessening summer learning loss.

The research, which uses two quasi-experimental methods, Coarsened Exact Matching (CEM) and fixed effects, to isolate the impact of Zearn Math on student achievement, compares students who consistently completed three or more Zearn Math lessons each week throughout the school year with similarly matched students who did not consistently use Zearn Math. It also looks at the same student in multiple years, between 2018-2022[1], to compare how they performed academically in years where they were consistent Zearn Math users vs. years where they had little to no usage. Unlike a standard correlational analysis, these methods facilitate causal inference and allow differences in outcomes to be more confidently attributed to Zearn Math and not to other variables.

Spanning three years of implementation, findings show:

Significant Within-Year Academic Growth

- **In one year, students who used Zearn Math consistently grew 36% more**; the equivalent of **10.5 additional weeks** of learning in one school year, as measured by the i-Ready assessment.

- **Students who started two or more years below grade-level,** a group i-Ready terms "striving learners" (Curriculum Associates, 2021), had **almost 2x the growth** with consistent Zearn Math usage as similar students with low or no usage; the equivalent of **an additional 23.7 weeks**.

[1] Methods that compare the same student in multiple years are adjusted to control for pandemic-related impacts to student learning.

- Across all i-Ready Placement Levels[2], consistent Zearn Math users maintained or increased their placement at higher rates than non- or low-users. For **students starting two or more grade levels below, 92% of consistent Zearn Math users moved up one or more levels** while only **62% of low- or non-users saw similar growth.**

- Consistent Zearn Math users had a **higher probability of achieving grade-level benchmarks** (+10.1%), **upward mobility in Placement Level** (+11.0%) and **meeting expected growth targets** (+14.9%) than similar students with low or no usage.
- Among matched **students in special education or multilingual learners** (MLLs) those who used Zearn Math consistently were **2x as likely to meet typical growth projections** as those with little or no usage.

- **Black and/or Latino Zearn Math users** were **1.3-1.4x more likely to move up one or more Placement Levels and meet expected growth projections** than Black and/or Latino students with low or no Zearn Math usage.

Zearn Math Stems Summer Learning Loss

- Students with consistent Zearn Math usage had a **32% reduction in summer slide[3]**, relative to low- or non-users. Students who consistently used Zearn Math during the prior school year **retained an additional five weeks of learning over the summer.**

Gains in Growth Persist a Year After Students Stop using Zearn Math

- Students with consistent Zearn Math usage in 2020-2021 maintained almost 75% of their additional growth, relative to low- or non-users, a year after they stopped using Zearn Math.

---

[2] i-Ready classifies students into five Placement Levels based on their performance: three or more grade levels below, two grade levels below, one grade level below, early on-grade level and on- or above-grade level.

[3] Summer slide is the learning loss that students experience over summer break, which shows up as a drop in BOY assessment scores from prior year EOY assessments taken just three months earlier. (Cooper, et. al., 1996)

Introduction

Zearn is the 501(c)(3) nonprofit educational organization behind Zearn Math, a top-rated math learning platform used by 1 in 4 elementary-school students and by more than 1 million middle-school students nationwide. This report summarizes findings from an efficacy analysis of the Zearn Math learning platform implemented as a digital complement in DC Public Schools (DCPS). The goal of this study was to isolate the impact of Zearn Math on student achievement, through quasi-experimental methods that facilitate causal inference.

DCPS is a district with 49,035 students, 21,621 of whom are in first to sixth grade. The student body is 74% economically disadvantaged, 15% students in special educations, 15% multilingual learners (MLLs), 45% at-risk[4] and 78% Black and/or Latino students. (DCPS, 2022)

Analyses assessed the impact of Zearn Math usage during the 2018-2019, 2020-2021 and 2021-2022 school years. In grades 1-6[5], 1,112 students consistently completed three or more digital lessons per week, approximately 90 or more digital lessons per year, and could be matched to assessment data from the 2018-2019, 2020-2021 and/or 2021-2022 school years.[6] (See Appendix A Tables A1 and A5 for a breakdown of the sample composition; see Appendix A Table A6 for student usage by year.)[7]

This study was designed to meet the What Works Clearinghouse (WWC) Meets WWC Group Design Standards with Reservations and to meet an Every Student Succeeds Act (ESSA) Tier 2 (Moderate) rating on the ESSA guidelines for evidence-based interventions. The study uses quasi-experimental matching methods to create baseline equivalency between treatment and control groups along major confounding factors. (See Appendix B for more information.)

---

[4] From DCPS At a Glance: "Students who are at-risk are those who qualify for Temporary Assistance for Needy Families (TANF), the Supplemental Nutrition Assistance Program (SNAP), have been identified as homeless during the academic year, who are under the care of the Child and Family Services Agency (CFSA or Foster Care), and who are high school students at least one year older than the expected age for their grade." (https://dcps.dc.gov/page/dcps-glance-enrollment)

[5] DCPS implements Zearn Math in grades K-6. However, Kindergarten lessons on Zearn differ from those in grades 1-6 in that they consist of fluency activities but do not contain a guided practice or formative assessment (Tower of Power). Because of these qualitative differences, this efficacy study focuses on implementation of Zearn Math in grades 1-6. Results including Kindergarten are footnoted throughout. Between school years 2018-2019, 2020-2021 and 2021-2022, DCPS had 295 fidelity users in Kindergarten.

[6] DCPS only used Zearn Math, with fidelity, in grade 6 during the 2021-2022 school year. School years 2018-2019 and 2020-2021 only included students in grades 1-5.

[7] The sample population of fidelity users differs from DCPS's population, having a proportionally smaller population of students from disadvantaged backgrounds. The implications of this difference are discussed in the limitations section.

Sample and Methodological Overview

In order to see maximum benefit from Zearn Math, students are advised to complete three or more digital lessons per week during the school year. Therefore, this analysis focuses on the impact of Zearn Math for students who completed an average of three or more digital lessons per week; approximately 90 or more digital lessons per year during the 2018-2019, 2020-2021 and/or 2021-2022 school years.[8]

To be included in the analysis, students also needed to have pre- and post- i-Ready Diagnostic scores, to facilitate comparison of academic growth between fidelity Zearn Math users and low- or non-users who completed fewer than 30 lessons in a year, i.e., fewer than one lesson per week.

Triangulated Findings through Multiple Quasi-experimental Methods

Data was analyzed using two quasi-experimental methods. The first was a quasi-experimental matching technique, Coarsened Exact Matching (CEM), that compared the academic growth of consistent Zearn Math users with a similar group of low- or non-users. This analysis examined Zearn Math's impact on the students in grades 1-6 who used Zearn Math consistently during the school years 2018-2019 (79 students), 2020-2021 (589 students) and 2021-2022 (322 students).[9] (See Appendix A Table A6 for a breakdown of student usage by year.) Outcomes of interest included Zearn Math's impact on student academic growth, stemming summer learning loss and long-term retention of learning.

The second was a fixed-effects analysis that compared a student's academic growth relative to grade-level expectations between a year in which a student was a low- or non-user of Zearn Math and a year in which they consistently used Zearn Math. This analysis examined Zearn Math's impact on the 1,091[10] students in grades 1-6 who were consistent users of Zearn Math for at least one year between 2018-2022 and low- or non-users another year during the study period.

The use of two quasi-experimental methods increases the robustness of findings. Both CEM and fixed effects models use techniques to approximate experimental conditions and facilitate causal inference in observational data, but they each have different strengths and limitations. By triangulating the primary analysis through two quasi-experimental methods, consistency in findings offers stronger

[8] This definition of treatment and control does not use an intention-to-treat (ITT) framework that would include in the treatment all students that had been offered Zearn Math (McCoy, 2017). While the ITT approach is the most efficacious for identifying the impact of a program under real-world implementation constraints, the goal for this study was to understand the impact of fidelity usage in the hopes of increasing fidelity usage of the platform across schools. This efficacy analysis examines the impact of Zearn Math, implemented with fidelity, vs. with little or no usage. The implications of Zearn's approach are discussed further in the limitations section.

[9] 43 students used Zearn Math, with fidelity for both years.

[10] The n for this study was lower than the n for the CEM within-year analysis because students needed to have two years of assessment data and specific patterns of usage for inclusion in this analysis. In contrast, CEM required only one year of assessment data.

support that methods have effectively controlled for confounders and isolated the impact of Zearn Math use on student achievement.

CEM matches a student with fidelity usage of Zearn Math (treatment) to a very similar student with little or no usage (control). This is done to approximate the treatment students' academic outcomes had they not used Zearn Math. In this analysis, students are compared within the same year. Thus, no adjustment needs to be made to compensate for pandemic-related impacts on assessment scores that confound the relationship between Zearn Math usage and academic outcomes. This adjustment would be necessary if student scores were being compared across different years.

Matched students are also in the same grade, so there is no need to adjust for grade-level changes in expected growth[11] as might be necessary if assessment scores were being compared across grade levels. The accuracy of the comparison, however, is based on a researcher's ability to match students on factors that could confound the relationship between Zearn Math use and academic outcomes, some of which may be unobservable. In other words, accuracy depends on the similarity of students being matched.[12]

In contrast, fixed effects models compare a student to themselves under different conditions to determine the impact of an intervention. These models facilitate causal inference because an individual serves as their own "control". In this sense, the models are more robust at controlling for between-student differences because students are "matched" to themselves.

However, fixed effects models do not control for factors that change, outside of the intervention, that might contribute to a difference in outcome (Allison, 2006). When an individual experiences changes that coincide with the intervention of interest, these must be controlled in the model or else estimates may be biased.

In this study, two major factors change for individuals between school years: 1) growth differences related to pandemic proximity (and whether a year is pre- or post-pandemic); and 2) a student's expected growth, depending on their grade. Fixed effects models require adjustments to compensate for these time-variant confounding factors and accuracy of estimates depend on the robustness of adjustments.

Because of the availability of multi-year data and large n's, Zearn Math efficacy could be analyzed using both quasi-experimental methods in one study. The complementary strengths and weaknesses of the two approaches provide greater confidence in the findings when they are consistent across

[11] Many assessments, including the i-Ready which is taken by DCPS students, project smaller changes in scale score growth as students go up in grade.

[12] WWC requires that studies using quasi-experimental methods illustrate baseline equivalence between treatment and control to qualify for the Meets Group Design standards with reservation designation. CEM calipers ensure that students have baseline equivalency, particularly on baseline math assessment scores. For WWC, fixed effects techniques comprise an acceptable statistical adjustment to create baseline equivalence (What Works Clearinghouse, 2022).

analyses, as seen in this study.

Long-term Retention of Learning and GIS mapping

While both the initial CEM and fixed effects analyses focused on within-year academic impacts of consistent Zearn Math usage, additional analyses were conducted to measure the lasting impact of Zearn Math after a student discontinued use. The first analysis examined “summer slide”, a phenomenon of learning loss that students experience in the two- to three-month school break (Cooper, et. al., 1996). Students were matched on spring end-of-year (EOY) assessments and the outcome of interest was fall beginning-of-year (BOY) assessment scores to see if consistent Zearn Math users better retained math knowledge over the summer months. In essence, does Zearn Math lessen the “summer slide?”

The second analysis looked at the extent to which consistent Zearn Math users from 2020-2021 retained additional learnings through 2021-2022, relative to similarly matched peers, despite being low- or non-users during the 2021-2022 school year.

Finally, GIS mapping was used to capture patterns of school-level implementation across the DCPS district. Maps also projected impact by school, based on student performance, to highlight opportunities to maximize Zearn Math implementation in future years. This analysis captures the potential districtwide impact with an implementation strategy that targets the highest impact schools.

i-Ready Growth

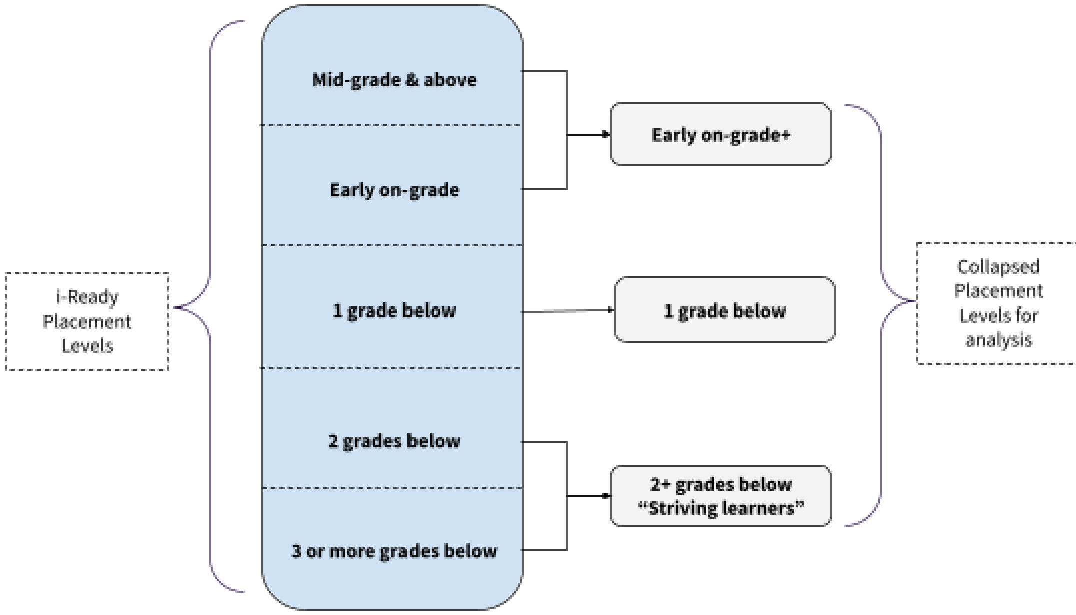
Academic growth was measured as the change in i-Ready aggregate math scores between the fall and spring assessment administration during each year of the study. i-Ready classifies students into five Placement Levels based on their scale score and grade: three or more grade levels below, two grade levels below, one grade level below, early on-grade level and on- or above-grade level.

Because i-Ready provides expected growth projections for students based on their grade and starting Placement Level (Curriculum Associates, 2022b), results are reported in terms of absolute growth (change in scale scores) and actual growth as a percentage of projected growth (change in scale score divided by projected growth).

Throughout the report, growth and mobility are disaggregated by average baseline score for three collapsed Placement Levels. In this analysis, students scoring three or more levels below and two levels below are combined into a two or more grades below category that i-Ready terms “striving learners” (Curriculum Associates, 2021). Students one grade level below are kept in their own below grade-level category. Students in Placement Levels early on-grade, and mid-grade and above are combined into a grade-level category early on-grade+ (See Figure 1).

FIGURE 1

Diagram of Combined i-Ready Levels



In addition to growth in scale score points, growth among consistent Zearn Math users and low- or non-users was translated into weeks of learning. i-Ready guidelines instruct that expected growth measures represent 30 weeks of learning, therefore weekly growth expectations can be calculated as 1/30th of a student's annual expected growth (Curriculum Associates, 2022a).

Because growth expectations vary by grade and starting Placement Level, weekly growth expectations were calculated as a mean of the growth expectations of students in a given sub-sample, divided by 30. This measure was then used to translate growth for each subgroup into weeks.

To translate scale score point changes to weeks of growth, a group's expected scale score point growth, per week, was calculated by taking the average growth expectation for students in a given sub-sample (whether by Placement Level or demographic factor) and dividing it by 30.

$$Weekly\ growth\ expectation = \frac{mean\ growth\ expectation}{30}$$

Scale score point changes were divided by this weekly growth expectation to calculate the weeks of growth represented by the change in scale score:

$$Weeks\ of\ growth = \frac{scale\ score\ change}{weekly\ growth\ expectation}$$

Findings also focus on the impact of Zearn Math on students' Placement Level growth; achievement of early on-grade Placement Level and above;[13] probability of moving up one or more Placement Levels; and probability of meeting growth expectations.

CEM: Impact by Matched Group Comparison

Quasi-experimental matching techniques were used to isolate the impact of Zearn Math on student achievement. Consistent Zearn Math users were matched with low- or non-users, in the same grade, on starting math and English Language Arts (ELA) achievement scores, along with seven student characteristics. The goal of matching was to create 1:1 pairings between similar students, differing primarily on Zearn Math usage during the 2018-2022[14] school years. The outcome under investigation was the average treatment effect on the treated, as controls were selected to match individuals in the treatment group.

In order to see maximum benefit from Zearn Math, students are advised to complete three or more digital lessons per week during the school year. Therefore, the treatment group was composed of students who consistently used Zearn Math during the 2018-2019, 2020-2021 and/or 2021-2022 school years, operationalized as an average of three or more digital lessons per week; approximately 90 or more digital lessons per year. The control group was selected from other students in the district with little to no Zearn Math usage, operationalized as an average of fewer than one digital lesson per week; fewer than 30 digital lessons per year, during the same school year.[15]

Drawing causal inference from observational data is challenging because factors that impact a person's likelihood to receive an intervention may also impact their outcomes. Therefore the differences in outcomes observed between individuals may not be caused by the intervention itself, but by other confounding factors that imbalance the treatment and control groups (Stuart, 2008; Iacus et al., 2011).

Matching methods were used to balance the composition of confounding factors between individuals

---

[13] i-Ready considers early on-grade to be at standard at the beginning of a school year, but below standard at the end of a school year. Therefore this measure is used as one of several to benchmark students' changes in Placement Level.

[14] DCPS assessed students in fall 2019, but did not assess students in spring 2020, due to disruptions from COVID. Therefore 2019-2020 student performance, as it relates to Zearn Math usage, is excluded from this analysis.

[15] That students who reach fidelity and those with little to no usage may have unmeasurable differences is not considered a confounding factor by WWC but what WWC terms, "imperfect overlap in the characteristic between the conditions'. WWC provides the example of a program based on voluntary enrollment in which students who volunteer could differ from those who did not in hard to measure qualities like introversion vs. extroversion. They clarify that, "The WWC does not consider this to be a confounding factor, but the selection mechanism and potential difference in unmeasured characteristics are reasons that QEDs are limited to a rating of Meets WWC Group Design Standards with Reservations, if the baseline equivalence requirement is satisfied' (2020, p. 82).

who consistently used Zearn Math (the treatment group) and a comparison group of individuals who had little to no Zearn Math usage (the control group). This is done to isolate the difference in outcomes from the intervention itself, separate from any impact due to potentially confounding factors.

This efficacy analysis used a two-step Coarsened Exact Matching (CEM) method with optimal matching to create a control group that was as similar as possible to the treatment group of consistent Zearn Math users. CEM is a technique that simulates block sampling by matching students on covariates, demographic and academic factors that may be related both to a student's likelihood of using Zearn Math consistently and their academic performance (Blackwell et al., 2009; Iacus et al., 2011). The effectiveness of matching is conditional on the ability of observable factors to capture the selection process that sorted individuals into treatment and control. Models that do not capture major factors may produce biased estimates[16].

DCPS students take the i-Ready assessment for math and either the Reading Inventory (RI) or the Dynamic Inventory of Basic Early Literacy Skills (DIBELS), largely based on grade level[17]. Around 80% of the sample had fall ELA assessment scores, but 20% did not. Using CEM, treatment students were put into matching strata with control students that were in the same academic year; same grade; took the same fall ELA assessment; were within three scale score points on the fall i-Ready Diagnostic; and within either 30 points on the (RI) or 10 points on the DIBELS[18], depending on which assessment a student took. Then, within strata, treatment students were matched to control students with whom they shared at least four of seven other student characteristics: school, gender, race, ethnicity, special education status, MLLs and at-risk status. The 20% of the sample without fall ELA scores were matched with other students with missing scores who met all other matching criteria.

This optimal matching method utilized Bertsekas' auction algorithm. This algorithm was used to produce combinatorial optimization such that treatment individuals were matched to others closest to them in the control pool. When controls were the best-fit match for more than one treatment individual, the pairing went to the individual for whom the next best pairing was the farthest (1981;

---

[16] This potential for bias does not exclude a study from meeting WWC's Group Design Standards with Reservations as long as baseline equivalency can be established. According to WWC: "In QED studies, confounding is almost always a potential issue due to the selection of a sample, because some unobserved factors may have contributed to the outcome. The WWC accounts for this issue by not allowing a QED study to receive the highest rating" (What Works Clearinghouse, 2020).

[17] The majority of DCPS students in grades K-2 took the DIBELS assessment while most students in grades 3-6 took the RI assessment.

[18] Different thresholds for proximity were selected based on the average size of the proficiency bands and projected yearly growth for each assessment. Closeness of fit on math scores was prioritized because of greater relevance to the outcomes of interest. WWC guidelines require that baseline equivalence be established using "a pretest in the same domain as the outcome" (What Works Clearinghouse, 2022). Therefore close mean baseline math scores are required for baseline equivalence.

Rosenbaum, 2020).[19]

If a treatment student had no match within their academic year, grade, ELA assessment type and score strata with whom they shared at least four characteristics, they were excluded from the treatment group. The caliper that limited match difference to no more than three characteristics was selected to maximize inclusion in the sample, prevent biasing through uneven patterns of exclusion and still ensure similarity between groups.

For more information on Zearn's methodological approach, see ***Efficacy Analysis Methodology: Zearn's approach to Coarsened Exact Matching***.

Out of the district's 1,157 incidents of year-long fidelity usage,[20] all but 167 were matched. Treatment and control pairs differed by an average of 1.57 demographic factors, 1.22 scale score points on i-Ready Diagnostic math, 4.65 points on DIBELS and 14.98 points on RI. Mean pretest math scores between treatment and control students differed by .02 scale score points on the i-Ready Diagnostic. This is less than .05 of a standard deviation of the combined means. According to WWC, "Baseline differences less than or equal to 0.05 standard deviations in absolute value automatically satisfy the baseline equivalence standard and do not require statistical adjustment" (WWC, 2022, p. 53).

The 167 consistently using students excluded from the study due to lack of match did not concentrate in any demographic category that would bias the sample (see Appendix A Tables A1 through A4 for a breakdown of sample demographics).

Difference of Means

Once consistent Zearn Math users were matched to a similar group of low- or non-users, a difference of means analysis was conducted to quantify the impact of Zearn Math on student achievement during 2018-2019, 2020-2021 and 2021-2022. Means were calculated for treatment and control groups overall as well as for groups disaggregated by starting math Placement Level, demographic factors and year.[21]

---

[19] In other words, if Control Student A was the best match for Treatment Student 1 and Treatment Student 2, sharing 6 out of 7 characteristics with each, Control Student A could still only be matched with either Treatment Student 1 or Treatment Student 2. If the next best match for Treatment Student 1, Control Student B, shared 4 characteristics, and the next best match for Treatment Student 2, Control Student C, shared 5 characteristics, then Treatment Student 1 would be matched with Control A and Treatment Student 2 would be matched with Control C. In this way, the algorithm of optimal matching balances the closeness of any individual match with its impact on the closeness of the overall group match.

[20] 43 students used Zearn Math, with fidelity, for two years. Because the matched sample was pulled separately for 2018-2019, 2020-2021 and 2021-2022, these individuals could be in the matched treatment sample for two separate years based on their usage. Students could also be in the treatment group one year and control group another. The unit of analysis was a year of Zearn Math fidelity usage.

[21] Appendix A Tables A9, A10, A13, and A14 contain difference in means t-tests disaggregated for 2020-2021 and 2021-2022. Because of the small sample size for 2018-2019, 82 fidelity Zearn Math users, differences in means for treatment and control were not statistically significant for that year alone. Therefore, Results Tables for

Difference in means t-tests were run on the average academic gains of consistent Zearn Math users vs. those of low- or non-users to determine if the impact of treatment was statistically significant. Given *SD*=standard deviations and *n*=number of observations per group, t-tests were conducted as:

$$t = \frac{mean_{treatment} - mean_{control}}{\sqrt{\frac{SD^2_{treatment}}{n_{treatment}} + \frac{SD^2_{control}}{n_{control}}}}$$

Effect size was calculated with *Cohen's d* which divides the difference in means between treatment and control by the pooled standard deviations:

$$Cohen's\ d = \frac{mean_{treatment} - mean_{control}}{pooled\ SD}$$

On average, consistent Zearn Math users gained 28.6 scale score points whereas matched low- or non-users gained 21.1 scale score points between fall and spring i-Ready assessments, a difference of 7.6 scale score points (effect size=.27). Students in the sample were expected to grow an average of .72 scale score points per week. Therefore, this translated to an additional 10.5 weeks of growth, or learning equivalent to one-third of a year relative to low- or non-users[22]. (See Appendix A Table A8 for findings from the difference in means analysis.)

Gains were highest among consistent Zearn Math users who started the year as "striving learners", i-Ready's term for students who are two or more grade levels below (Curriculum Associates, 2021). These students gained 38.9 scale score points while low- or non-users gained 20.8, a difference of 18.1 scale score points (effect size=.73) (see Results Table 1). Translated into weeks of growth,[23] students who started as striving learners grew an additional 23.7 weeks with consistent Zearn Math usage, relative to those with low or no usage. The outsized impact of Zearn Math use among students starting below Placement Level has been reported by Zearn in previous efficacy analyses (2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c).

2018-2019 alone are omitted, but 2018-2019 Zearn Math fidelity users and low- or non-users are included in aggregate models.

[22] i-Ready year-long growth is defined as 30 weeks. (i-Ready, 2022)

[23] A student's expected scale score point growth, per week, was calculated by taking the average growth expectation for students in the sample that were at that Placement Level and dividing it by 30 as i-Ready calculates expected growth based on the typical student's growth over 30 weeks of instruction. (i-Ready, 2022)

RESULTS TABLE 1

Growth Across i-Ready Placement Levels

Growth in scale scores and weeks for Consistent Zearn Math users *(Treatment)* vs. Low- or Non-users *(Control)*, by starting Placement Level, all years*

	All students	2+ grade levels below	1 grade level below	Early on-grade and above
Treatment growth in scale score points	28.6	38.9	31.8	23.8
Treatment growth in weeks	39.9	50.9	40.3	37.1
Control growth in scale score points	21.1	20.8	24.0	18.4
Control growth in weeks	29.4	27.2	30.4	28.7
Growth difference in scale score points	7.6	18.1	7.8	5.4
Growth difference in weeks	10.5	23.7	9.8	8.4

*On average, students in the sample were expected to grow .72 points per week. Disaggregated by starting Placement Level, those who started 2+ grade levels below were expected to grow .77 scale score points per week, those who started 1 grade below were expected to grow .79 points per week, and those who started early on-grade were expected to grow .64 points per week.

In addition to capturing changes in student achievement across all users, the analysis zoomed in on how Zearn Math use impacted the performance of student subgroups. Because pairs of consistent Zearn Math users and low- or non-users were allowed to mismatch on up to three demographic characteristics, subgroups did not always align on starting scale scores. Therefore, differences in academic growth by demographic subgroup were reported as difference-in-difference[24] rather than as absolute scale scores (see Appendix A Table A7 for a breakdown of starting and ending means, by subgroup).

[24] All groups with the exception of students in special education and MLLs had baseline differences <.05 of a standard deviation which satisfies baseline equivalence without adjustment, according to WWC. MLL and students in special education had differences <.25 of a standard deviation, satisfying baseline equivalency with a difference-in-difference adjustment (2022). (See Appendix A Table A7 for full details on baseline equivalence.)

RESULTS TABLE 2

Growth Across Student Subgroups

Growth in scale scores and weeks for Consistent Zearn Math users (Treatment) vs. Low- or Non-users (Control), by student subgroup, all years*

	Consistent Zearn Math users		Low- or non-users		Difference	
	Growth in scale score points	Growth in weeks**	Growth in scale score points	Growth in weeks	Scale score	Weeks
All Students	28.6	39.9	21.1	29.4	7.6	10.5
Female SS	28.2	39.7	20.5	28.9	7.7	10.8
Male SS	29.1	40.2	21.6	29.8	7.5	10.3
Black and/or Latino SS	29.5	38.8	18.3	24.0	11.2	14.8
Special education SS	33.6	44.1	15.5	20.4	18.1	23.7
MLL SS	26.7	33.3	10.3	12.8	16.5	20.5
At-risk SS	27.3	34.8	19.1	24.4	8.2	10.4

*On average, students in the sample were expected to grow .72 points per week. Disaggregated by subgroup, weekly growth expectations were .71 for female students, .72 for male students, .76 for Black or Latino students, .76 for students in special education, .80 for MLL students, and .78 for at-risk students.

**Typical growth projections are based on a 30 week academic year.

Across subgroups, consistent Zearn Math users saw larger gains in scale score points than matched low- or non-users. MLLs and students in special education saw more than double the gains with consistent Zearn Math usage as similar MLL and students in special education that had little to no Zearn math usage. These additional scale score points translated to over 10 additional weeks of learning. MLL and students in special education grew 20 weeks more with consistent Zearn Math usage than matched MLLs and students in special education with low or no Zearn Math usage.

Median growth as a percentage of a student's typical growth projection was 104% across all students with low or no use.[25] The median growth for students from historically marginalized subgroups with low or no Zearn Math usage was less than 100% of typical growth projections. In contrast, students with consistent Zearn Math usage exceeded typical growth by 36% and all subgroups exceeded typical growth projections (see Results Table 3).

[25] Note that 104% of typical growth is the median growth value for students with low or no usage while 98% is the mean.

RESULTS TABLE 3

Growth as a Percentage of i-Ready Typical Growth Projections

Median % of typical growth reached by consistent Zearn Math users and low- or non-users, by subgroup, 2018-2019, 2020-2021 & 2021-2022

	Consistent Users	Low- or Non-Users	Difference
All students	136%	104%	32%
Female	133%	100%	33%
Male	138%	104%	34%
Black and/or Latino	131%	89%	42%
Special education	152%	65%	87%
MLL	128%	57%	71%
At-risk	**	**	**

**Excluded due to lack of statistical significance. Full results available in Appendix A Tables A11 and A15.

Similarly, consistent Zearn Math users had a higher percentage of students reaching early on-grade level or above than non- or low-users. Notably, twice as many MLLs who consistently used Zearn Math went from below the early on-grade Placement Level to early on-grade or above than similar MLLs who were low- or non-users (see Results Table 4 & Appendix A Table A12).[26]

RESULTS TABLE 4

Change in Students Reaching i-Ready Early On-Grade or Above

Change in percent early-on grade and above among students with consistent Zearn Math usage and those with little or no Zearn Math usage

	Consistent Users	Low- or Non-Users
All Students	+38%	+27%
Female	+38%	+27%
Male	+38%	+28%
Black or Latino	46%	+30%
Special education	**	**
MLL	+46%	+16%
At-risk	+46%	+33%

**Excluded due to lack of statistical significance. Full results available in Appendix A Tables A8 and A12.

[26] For each subgroup in treatment and control, percent early on-grade or above in the fall was subtracted from percent early on-grade or above in the spring. This change is depicted in Appendix A Tables A8 and A12. If the percent early on-grade or above within a subgroup was the same in the fall and spring, the change listed in Appendix A Tables A8 and A12 would be 0.

Consistent Zearn Math users also met individual growth projections at higher rates than low- or non-users. Black and/or Latino students who consistently used Zearn Math met growth projections at 1.4x the rate of those who were low- or non-users. Rates of meeting growth projections were particularly amplified among students in special education and MLLs who used Zearn Math consistently. Those students met their expected growth at almost 2x the rate of non- or low-users.

RESULTS TABLE 5

Percent of Students Meeting or Exceeding i-Ready Growth Projections

Percent meeting growth expectation among students with consistent Zearn Math usage and those with little or no Zearn Math usage

	Consistent Users	Low- or Non-Users
All Students	68.69%	53.74%
Female	68.41%	53.05%
Male	68.97%	54.42%
Black or Latino	64.36%	46.29%
Special education	70.69%	38.24%
MLL	61.17%	29.87%
At-risk	**	**

**Excluded due to lack of statistical significance. Full results available in Appendix A Tables A11 and A15.

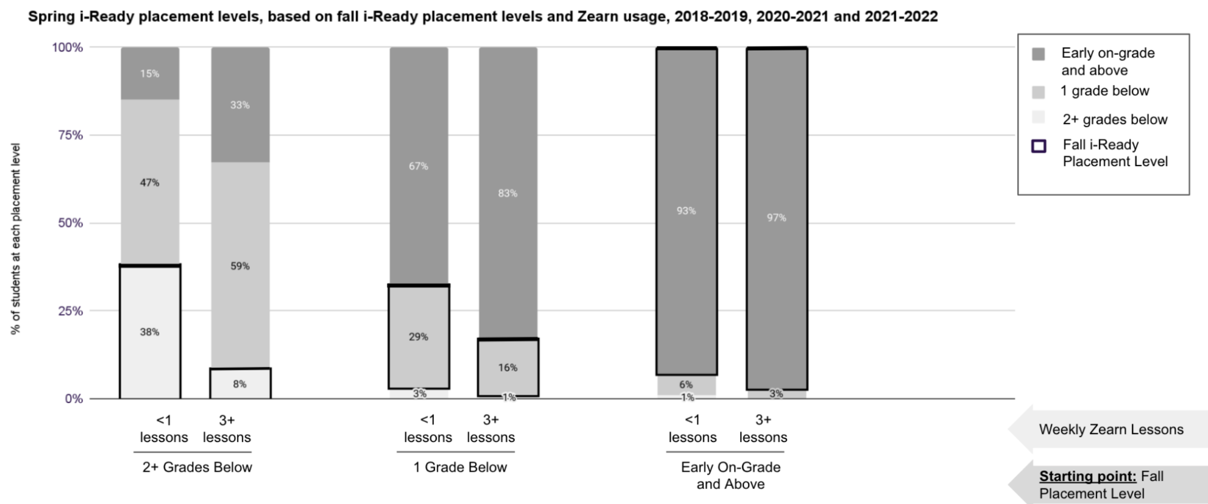
Placement Level Mobility

Based on starting i-Ready Placement Level, mobility models compared the change in Placement Level for treatment and control students. i-Ready classifies students into five Placement Levels based on their performance: three or more grade levels below, two grade levels below, one grade level below, early on-grade level and on- or above-grade level. In this analysis and throughout the report, students scoring three or more levels below and two levels below are combined into a two or more grades below category that i-Ready research terms “striving learners” (Curriculum Associates, 2021). Additionally, students one grade level below are kept in their own below grade-level category. Finally,students in Placement Levels early on-grade and mid-grade and above are combined into a grade-level category of early on-grade+ (See Figure 1).

Across all Placement Levels, consistent Zearn Math users maintained or increased their placement at higher rates than non- or low-users. Notably, for students starting two or more grade levels below, 92% of consistent Zearn Math users moved up one or more levels while only 62% of low- or non-users saw similar growth. Figure 2 illustrates the mobility between Placement Levels from the fall to spring i-Ready assessments.

FIGURE 2

Change in i-Ready Placement Level



Probability Models and Placement Level Mobility

Logistic regression models calculated the difference in probability of achieving Placement Level mobility among students with consistent Zearn Math usage and similarly matched students with low or no usage. Outcomes included: the probability of moving up at least one Placement Level, probability of scoring early on-grade or higher and the probability of meeting expected growth projections.

In addition to probabilities, outcomes are expressed as relative likelihoods, i.e., the probability of the outcome among consistent Zearn Math users vs. those with little to no Zearn Math usage.[27] Consistent Zearn Math users had a higher probability of achieving grade-level benchmarks (+10.1%), upward mobility in Placement Level (+11.0%) and meeting expected growth targets (+14.9%). This means, for example, that a 15% increase exists in the probability of meeting projected growth targets associated with consistent Zearn Math usage. (See Results Table 6 and Appendix A Table A16 for more details.)

[27] Because the groups have been matched on students' starting achievement scores, grade, year and seven academic and demographic factors, these variables are already controlled in the models.

RESULTS TABLE 6

Probability of Achieving Academic Growth and Early On-Grade Placement Level

Probabilities of growth among students with consistent users and low- or non-users				
	Zearn Math users	**Low- or non-users**	**Difference**	**Relative likelihood**
Probability of moving up 1+ levels	69.0%	58.0%	+11.0%	1.19
Probability of scoring early on-grade or higher	85.4%	75.3%	+10.1%	1.13
Probability of meeting expected growth	68.7%	53.7%	+14.9%	1.28

The relative probabilities of Placement Level mobility between consistent Zearn Math users and low- or non-users were amplified among traditionally disadvantaged students. Black and/or Latino Zearn Math users were 1.3-1.4x more likely to move up one or more Placement Levels and meet typical growth projections than Black and/or Latino students with low or no Zearn Math usage.

Students in special education and MLLs had the largest gains in probability of meeting growth expectations from Zearn Math. Students in special education who used Zearn Math had a 70.7% probability of meeting growth projections while students in special education and little or no Zearn Math usage had a 38.2% probability. MLLs who used Zearn Math consistently had a 61.2% probability of meeting growth projections while those with little to no use had a 29.9% probability. For students in each subgroup, consistent use of Zearn Math was associated with almost 2x the likelihood of meeting growth projections. (See Appendix A Table A16 for complete results.)

RESULTS TABLE 7

Probability of Placement Level Growth and Meeting i-Ready Projected Growth

Increase in probability of Placement Level growth and meeting growth projections among students with consistent Zearn Math usage		
	Probability increase among Zearn Math users	**Probability relative to non-users**
Black or Latino probability of moving up 1+ levels	+18.52%	1.33x
Black or Latino probability of meeting expected growth	+18.07%	1.39x
Special education probability of moving up 1+ levels	+28.80%	1.61x
Special education probability of meeting expected growth	+32.45%	1.85x
MLL probability of moving up 1+ levels	+27.68%	1.58x
MLL probability of meeting expected growth	+31.29%	2.05x
At-risk probability of moving up 1+ levels	+11.05%	1.19x
At-risk probability of meeting expected growth	**	**

**Excluded due to lack of statistical significance. Full results available in Appendix A Table A16.

Fixed Effects: Within-Student Comparison

As part of the sample of 1,112 consistent Zearn Math users in grades 1-6, DCPS had a sample of 1,091 students who used Zearn Math consistently for at least one year and had low or no usage for at least one year, between 2018-2021.[28] (See Appendix A Tables A17, A18, and A19 for a breakdown of the fixed effects sample by grade, year and usage.) Using fixed effects models, students were compared to themselves on the outcome of academic growth between the year in which they consistently used Zearn Math and the year in which they had little to no usage.

Fixed effects methods are considered quasi-experimental because they compare a student to themselves under different conditions.[29] Essentially, students become their own control group, thereby matching perfectly on all time-invariant endogenous characteristics. In contrast to CEM, which pairs students to others that are similar across observable factors, fixed effects are more robust at controlling for differences in student characteristics, associated with the outcome of interest, because students are “matched” to themselves.

In this analysis, students’ academic growth was compared between the year in which they were a low- or non-user of Zearn Math and the year in which they used Zearn Math with fidelity. While fixed effects models create more robust matching on time-invariant characteristics of students than CEM, they do not control for time-variant characteristics that change for the individual in conjunction with changes from the treatment to control condition.

CEM is able to control for these time-variant factors by matching students within the same grade and year, something fixed effects models cannot do because students have the experience of using Zearn Math with fidelity and having low or no Zearn Math usage in different grades and years.[30] In order for fixed effects models to be robust in the analysis of Zearn Math’s impact on students’ academic growth, adjustments must be made to control for two confounding factors: 1) the change in students’ expected growth, which lessens as students go up in grade level; and 2) growth differences related to pandemic proximity (and whether a year is pre- or post-pandemic).

These factors have the potential to confound the relationship between Zearn Math usage and academic growth because declining growth expectations mean that with or without Zearn Math usage, students would grow less in a year when they were in a higher grade. For example, first graders who start the year on grade level are expected to grow 21 points in one year. In contrast, sixth graders who

---

[28] The n for this analysis was lower than the within year CEM analysis because students needed to have two years of assessment data and specific patterns of usage for inclusion. In contrast CEM required only one year of assessment data.

[29] WWC identifies fixed effects as an acceptable statistical technique to satisfy baseline equivalence. (2022)

[30] In every year of the study, there were some students in the sample using Zearn Math with fidelity while others had low or no usage. (See Appendix A Table A18 for more details.)

start the year on grade level are only expected to grow 13 points in one year. Therefore, the fact that a student grew more or less between years might be caused by the change in their grade growth expectation, not the intervention of Zearn Math.

In addition, most students saw a dramatic decline in academic growth from 2018-2019 to 2020-2021 because of the pandemic's disruption to learning. If a student grew more in 2018-2019 than 2020-2021 or 2021-2022, the cause might be pandemic disruptions, not Zearn Math usage. This analysis adjusts for those two factors that confound the relationship between Zearn Math use and student academic growth, even within the same student.

Scale score growth expectations from fall to spring are based on each student's grade and initial Placement Level. For every student, however, regardless of grade or initial Placement Level, i-Ready projections represent 30 weeks of learning (i-Ready, 2022). In order to compare a student's growth between different grades, growth was translated into weeks of learning by dividing a student's actual growth by 1/30 of their expected growth for a given year:

$$weeks\ of\ growth\ =\ 30\ *\ \left(\frac{actual\ growth}{expected\ growth}\right)$$

This means that students were compared across a measure that is comparable between grades, i.e., how many weeks of learning they achieved. Regardless of grade level, every student is expected to achieve 30 weeks of learning as defined by the growth projection. Therefore the percentage of their typical growth projection that they achieve can be compared between lower and higher grades.

The second potentially confounding factor for school years 2018-2019, 2020-2021 and 2021-2022, was the large year-to-year changes in average growth related to pandemic proximity and missed learning. In order to control for this factor, a pandemic adjustor was calculated as the average weeks of growth achieved by all DCPS students, in grades 1-6, in a given year:

$$pandemic\ adjustment\ =\ DCPS\ mean\ weeks\ of\ growth\ for\ year$$

The table below displays the average weeks of growth across all DCPS students, in grades 1-6, in each of the three study years. Unsurprisingly, during the 2020-2021 school year, students achieved less than half of the growth, relative to expectation, as students in 2018-2019 and 2021-2022.[31]

[31] Note that i-Ready adjusts growth projections based on starting proficiency. Therefore similar levels of growth, relative to projection in years 2018-2019 and 2021-2022 do not necessarily mean that students have recovered missed learning and are performing at pre-pandemic levels of academic achievement.

DESCRIPTIVE TABLE 1

Average DCPS student growth, relative to i-Ready growth projections, by year, for students in grades 1-6	
	Average weeks of growth
2018-2019	41.8
2020-2021	17.3
2021-2022	40.2

Each student's growth, as a percentage of BOY projection based on their fall Placement Level, was mean-centered around the pandemic adjustment for a given year:

$$mean\ centered\ growth = \%\ projected\ growth\ achieved - pandemic\ adjustment$$

Therefore, the growth measure in the fixed effects models reflects the percentage of projected growth an individual achieved relative to that year's average growth among DCPS students. By mean-centering around pandemic learning norms and comparing the relative growth measure within a student, rather than between students, the fixed effects models attempt to isolate the differences in growth to those attributable to consistent Zearn Math usage.

On average, students' scale score growth during years in which they used Zearn Math with fidelity was the equivalent of 34% of their expected growth more than they achieved in years when they had low or no Zearn Math usage. This translates into 10.2 additional weeks of growth with consistent Zearn Math usage vs. without. (Fixed effects results can be found in Appendix A Table A20.)

The implication of this finding is that, for any given year between 2018-2022, on average, a student using Zearn Math consistently would have a score that represented 10.2 weeks of learning more than how much they would have learned if they had low or no Zearn Math usage. As students work to catch up and move forward from pandemic-related learning loss, this is evidence that Zearn Math provides the within-school-year pace necessary to recover missed learning and keep students progressing on grade-level. This finding is similar to the results from the CEM analysis above.

RESULTS TABLE 8

Average Growth on the i-Ready Exam with and without Zearn Math Usage

Fixed effects analysis: average growth on the i-Ready exam, by grade, between fall and spring assessment administration in DCPS, for students in years when they consistently used Zearn Math vs. years when they had low or no Zearn Math usage*

	Consistent Zearn usage (weeks of growth, relative to average DCPS student)	**Low or no Zearn usage** (weeks of growth, relative to average DCPS student)	**Difference** (additional weeks of growth with Zearn Math usage)
All students	+13.23	+3.04	+10.18***
Female	+13.73	+2.22	+11.5***
Male	+12.57	+3.90	+8.67***
Black or Latino	+10.76	+0.50	+10.27***
Special education	+16.09	+3.32	+12.77*
MLL	**	**	**
At-risk	**	**	**

*Results based on FE model comparing growth relative to grade level expectations for students with at least one year of fidelity Zearn Math usage and one year of low or no Zearn Math usage over school years 2018-2019, 2020-2021, and 2021-2022.

**Excluded due to lack of statistical significance. Full results available in Appendix A Table A20.

Additional models were run to look at how the impacts of Zearn Math use differed based on demographic subgroup characteristics. Because time-invariant characteristics drop out of fixed effects models, the sample was segmented by demographics prior to the within-student fixed effects analysis. Analysis was conducted for students in the following subgroups: Black and/or Latino students, economically disadvantaged ("at-risk") students, female students, male students, students in special education and MLLs.

Black and/or Latino and female students with consistent Zearn Math usage showed similar growth to students overall, relative to low- or non-usage. Students in special education saw even larger gains of 12.8 additional weeks of growth when they used Zearn Math consistently. (See Appendix A Table A20 for full results.)

Summer Slide

The next analysis looked at whether similar scoring students who used Zearn Math with fidelity had lower levels of math learning loss during summer break, a time when many students lose academic gains with no in-person or digital instruction (Cooper et. al., 1996). In other words, does Zearn Math lessen "summer slide"?

Summer slide learning loss shows up as a drop in BOY assessment scores in comparison to prior EOY assessments taken just three months earlier. On average, summer losses are larger in math than reading and more common for high-poverty students. Procedure-based rather than conceptual knowledge is also more susceptible to loss (Cooper et. al., 1996). As schools seek to catch students up and move them forward in their math learning, reducing summer learning loss is particularly important.

This analysis looks at whether students who consistently used Zearn Math were more likely to retain their prior-year knowledge into the next school year than similar students who learned without the support of Zearn Math digital lessons.

For this analysis, students with consistent Zearn Math usage during the 2018-2019 or 2020-2021 school years were matched with low- or non-users, in the same grade, based on their spring[32] i-Ready and RI or DIBELS EOY results and seven demographic and academic factors, using the same calipers as all other CEM analyses.[33] Because the focus of this analysis was whether Zearn Math lessened "summer slide," fall 2020 and fall 2021 results were compared as the outcome of interest for each cohort. If Zearn Math better supports retention of math knowledge, then consistent Zearn Math users would have higher fall i-Ready assessment scores.

In grades 1-5, a total of 964 students used Zearn Math during either the 2018-2019 or the 2020-2021 school year. Students also had spring i-Ready assessment scores from 2019 or 2021 and fall i-Ready assessment scores from 2019 or 2021, respectively.[34] Of those students, all but 219 were matched. Treatment and control pairs differed by an average of 1.8 demographic factors, 1.16 scale score points on i-Ready Diagnostic in math, 4.56 points on DIBELS and 14.91 points on RI.[35] The 219 consistently using students excluded from the study due to lack of match did not concentrate in any demographic

---

[32] Note that, unlike prior CEM analyses, students were matched on spring scores, not fall.

[33] Treatment students were put into matching strata with control students that were in the same grade, took the same spring ELA assessment, were within three scale score points on the spring i-Ready Diagnostic in math and within either 30 points on the (RI) or 10 points on the DIBELS spring assessment, depending on which assessment a student took for ELA. Then, within strata, treatment students were matched to control students with whom they shared at least four of seven other student characteristics: school, gender, race, ethnicity, special education status, MLLs, and at risk status.

[34] Because the summer slide analysis required a different set of assessment scores than the within year CEM analysis, a higher number of fidelity users from 2020-2021 could be matched to the necessary scores to be included in the sample. While the within year efficacy analysis had 696 students, eligible for match, from 2020-2021, the summer slide analysis had 888 from that year.

[35] Students overall and all groups with the exception of at-risk, Black and/or Latino, MLLs and students in special education have baseline differences <.05 of a standard deviation which satisfies baseline equivalence without adjustment, according to WWC. Groups with baseline mean differences >.05 had differences <.25 of a standard deviation, satisfying baseline equivalency with a difference-in-difference adjustment (2022). (See Appendix A Table A22 for full details on baseline equivalence.)

category that would bias the sample (See Appendix A Table A21 for a breakdown of sample demographics and Appendix A Table A22 for spring score averages by treatment and control).

RESULTS TABLE 9

Summer Learning Loss Across Subgroups

Additional growth in scale score points and weeks, retained by students who consistently used Zearn Math during the 2020-2021 school year, relative to matched peers with little or no usage

	Consistent users	Low or non-Users	Difference
All students summer slide in scale score points	-8.3	-12.2	3.9
All students summer slide in weeks	-10.7	-15.7	5.0
Female summer slide in scale score points	-9.07	-12.13	3.06
Female summer slide in weeks	-11.9	-15.9	4.0
Male growth summer slide in scale score points	-7.58	-12.32	4.7
Male summer slide in weeks	-9.9	-16.1	6.2
Black or Latino summer slide in scale score points	-11.23	-16.13	4.9
Black or Latino summer slide in weeks	-14.1	-20.2	6.2
Special education summer slide in scale score points	**	**	**
Special education summer slide in weeks	**	**	**
MLL summer slide in scale score points	**	**	**
MLL summer slide in weeks	**	**	**
At-risk summer slide in scale score points	-10.66	-20.46	9.8
At-risk summer slide in weeks	-13.0	-25.0	12.0

**Excluded due to lack of statistical significance. Full results available in Appendix A Table A23. For spring and fall scale scores see Appendix A Table A22.

On average, students with consistent Zearn Math usage in 2018-2019 or 2020-2021 had a 32% reduction in summer slide, relative to low- or non-users. Over the summer, students who consistently used Zearn Math during the prior school year lost 3.9 fewer scale score points (effect size=.17) relative to matched low- or non-users. In other words, they lost 5.0 fewer weeks of learning over the summer (see Appendix Table A23 for findings from the difference in means analysis).

At-risk students saw even larger differences in learning loss with consistent Zearn Math users losing 48% less than matched at-risk students with low or no usage. (See Appendix A Table A23 for findings from the difference in means analysis.)

The difference in summer slide means that even among students with the same EOY test scores, Zearn Math users have a 12% increase in their probability of starting the next year in the early on-grade or above Placement Level and a 33% decrease in their probability of starting the next year 2+ grade levels below. In other words, students with low or no Zearn Math usage the prior year are significantly more likely to fall back a Placement Level or more over the summer.

RESULTS TABLE 10

BOY Placement Level Distribution for Matched Treatment and Control Students

Probability of scoring in each fall placement level, based on prior year Zearn Math usage, for students matched on EOY scores and additional demographic and academic factors

	2+ grade levels below	1 grade level below	Early on-grade+
Treatment	14.4%	45.8%	39.9%
Control	21.3%	43.0%	35.7%
Difference	-7.0%	2.8%	4.2%
Increase/decrease in probability	33% decrease in probability	7% increase in probability	12% increase in probability

Note: All results statistically significant at p<.01

Two-Year Impact

Results from the within-year CEM and fixed effects analyses show promising evidence that fidelity usage of Zearn Math has a robust impact on student growth, resulting in approximately 35% more learning, on average, for DCPS students. Summer slide analyses showed that students who learned with Zearn Math lost five fewer weeks of learning over the summer, relative to demographically similar students with the same spring scores who learned without Zearn Math the prior year. In this section, models were also run to examine whether students who used Zearn Math with fidelity during the 2020-2021[36] school year retained additional learning, relative to their matched peers, a year later, at the end of the 2021-2022 school year.

As states consider investments in education interventions to catch students up and move them forward in math, it is important that students retain their gains in learning beyond a year. Therefore, longitudinal success of Zearn Math in increasing student learning, even beyond a student's time on the learning platform, is an important measure of efficacy.

This analysis focused on the two year outcomes of a matched sample of students with fidelity usage during the 2020-2021 school year and a similar group of students with low or no usage in 2020-2021. In

[36] 2020-2021 was selected as the base year for long term impact as the 2018-2019 cohort of fidelity users was a substantially smaller sample, made smaller still by the necessity of students having 2020-2021 and/or 2021-2022 post-assessment scores and low or no Zearn Math usage in subsequent years. Similar to many districts and states across the country, DCPS did not assess students at the end of the 2019-2020 school year. As additional years of post-pandemic Zearn Math usage and assessment scores become available in districts across the country, Zearn intends to continue analyzing the long-term impacts of consistent usage.

order to eliminate confounding factors from additional Zearn Math usage outside of 2020-2021 on the impacts of interest, users were removed from the matching pool if they had usage in 2021-2022 that exceeded an average of one lesson/week, approximately 30 lessons/year, regardless of whether they were fidelity or low- or non-users during 2020-2021. Additionally, students had to have spring 2022 i-Ready assessment scores to be included in the matched sample.

For this analysis, the matched sample was selected using the same matching process, calipers and criteria as the initial CEM one-year efficacy analysis. In total, 515 students with fidelity Zearn Math usage in 2020-2021, met the criteria for inclusion in the matching pool, having low or no usage in school year 2021-2022, i-Ready fall assessment scores for 2020 and i-Ready spring assessment scores for 2022. Of those 515 students, all but 113 were matched.

Treatment and control pairs differed by an average of 1.77 demographic factors, 1.28 scale score points on the i-Ready Diagnostic in math, 4.82 points on DIBELS and 14.32 points on RI.[37] The 113 consistently using students excluded from the study, due to lack of match, did not concentrate in any demographic category that would bias the sample (See Appendix A Table A24 for a breakdown of sample demographics and Appendix A Table A25 for a breakdown of starting and ending scale scores by subgroup.)

On average, students with consistent Zearn Math usage in 2020-2021 maintained 74% of their growth, relative to low- or non-users, a year after they stopped using Zearn Math. In this two-year impact sample, students who consistently used Zearn Math during the 2020-2021 school year, ended the 2020-2021 school year 12.9 points higher[38] than matched low- or non-users. At the end of the 2021-2022 school year, students who consistently used Zearn Math during the 2020-2021 school year, ended the 2021-2022 school year 9.6 points higher than matched low- or non-users (effect size=.31); the equivalent of 13.3 additional weeks of learning. (See Appendix A Table A27 for findings from the difference in means analysis.)

[37] All groups with the exception of Black and/or Latino students, MLLs, and students in special education had baseline differences <.05 of a standard deviation which satisfies baseline equivalence without adjustment, according to WWC. Black and/or Latino and MLL students had differences <.25 of a standard deviation, satisfying baseline equivalency with a difference-in-difference adjustment. Students in special education did not meet the standard for baseline equivalence with adjustment and their results are not reported in findings (2022). (See Appendix A Table A25 for full details on baseline equivalence.)

[38] Because the two-year impact results are based on a subsample from the sample used in the within-year CEM analysis, year one growth is not identical to the growth measured in the within-year analysis.

RESULTS TABLE 11

One- and Two-Year Growth Among Fidelity Users Relative to Low- or Non-Users

Additional growth in scale score points and weeks, retained by students who consistently used Zearn Math during the 2020-2021 school year, relative to matched peers with little or no usage

	EOY 2021	EOY2022
All students growth difference in scale score points	12.9	9.6
All students growth difference in weeks	16.8	13.3
Female growth difference in scale score points	14	7.8
Female growth difference in weeks	18.2	10.8
Male growth difference in scale score points	11.7	11.0
Male growth difference in weeks	15.2	15.2
Black or Latino growth difference in scale score points	15.8	8.7
Black or Latino growth difference in weeks	20.2	11.8

ⱡⱡStudents in special education excluded as the subgroup does not satisfy baseline equivalence even with statistical adjustment for the two-year impact analysis.
**MLLs and at-risk students excluded due to lack of statistical significance. Starting and ending scale score means for each year are available in Appendix A Table A26

Results from logistic regression models showed that a year after discontinuing use of Zearn Math, students who used consistently during the 2020-2021 school year were still more likely than matched peers to have increased their Placement Level and achieved early on-grade or above. (See Appendix A Table A28 for full logistic regression results.)

RESULTS TABLE 12

Probability of Achieving Growth Benchmarks Among Consistent Zearn Math Users

Additional probability of level growth achieving early on-grade or above among students who consistently used Zearn Math during the 2020-2021 school year, relative to matched peers with little or no usage

	EOY2022
Increase in probability of moving up 1+ levels w/consistent Zearn Math usage	+12.9%
Relative likelihood of moving up 1+ levels w/consistent Zearn Math usage	1.72x
Increase in probability of scoring early on-grade+ w/ consistent Zearn Math usage	+11.9%
Relative likelihood of scoring early on-grade+ w/ consistent Zearn Math usage	1.79x

*All results are statistically significant. See Appendix A Table A28 for more detail.

**Probability and relative likelihood of meeting expected growth is omitted as growth projections change annually based on a student's starting placement level, such that expectations would be dissimilar in fall 2021 if students performed differently in 2020-2021, despite initial matching.

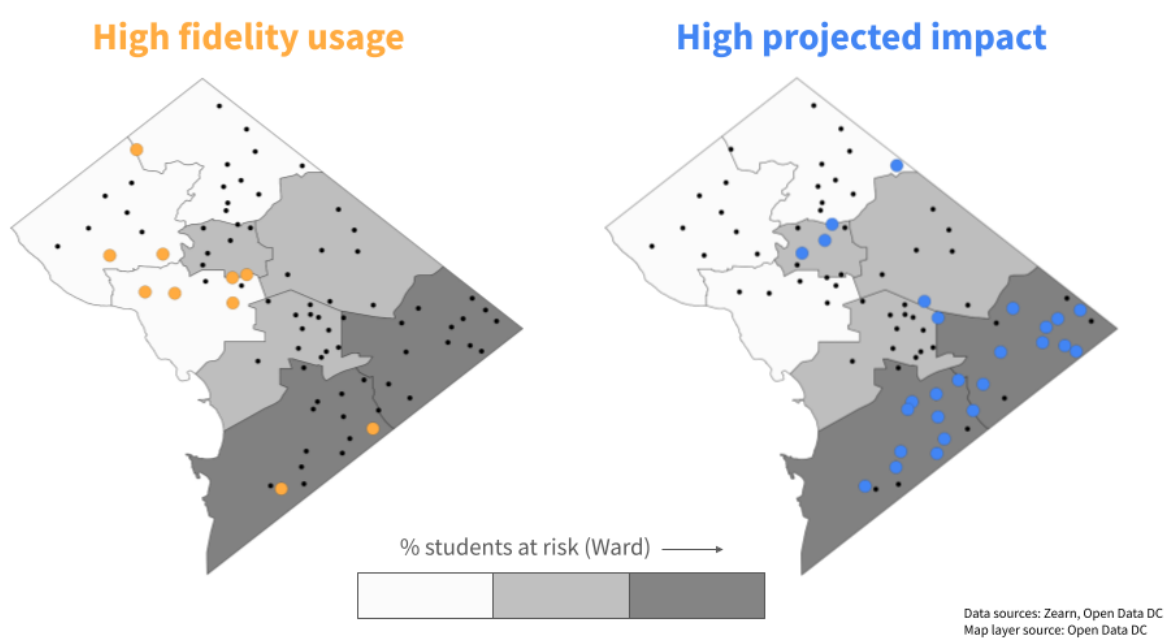
Trends in Usage

Given the strong evidence of impact in DCPS schools, the final analysis looked at where in the district Zearn Math is being implemented with fidelity. Then impact projections of medium and high were calculated based on the percentage of students in Placement Level 2+ grade levels below on the i-Ready exam, in each school. This criteria was selected as Zearn Math has consistently shown an outsized impact among students performing below standard, both in this analysis and in previous efficacy analyses (2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c).

The first map in Figure 3 shows the geographic distribution of non-charter DCPS schools with the highest fidelity usage of Zearn Math. The high-fidelity usage schools, highlighted in orange, represent the schools that rank in the top third of non-charter DCPS schools[39] in terms of percent of students using Zearn Math with fidelity in grades K-6, during the school year 2021-2022.[40]

Figure 3

Schools with the Highest Fidelity Usage vs. Schools with the Highest Projected Impact, SY 2021-2022



[39] Only schools with at least one user who reached fidelity were included in the ranking before the thirds were defined.

[40] In 2021-2022, the schools with the highest rates of fidelity usage had between 3.0-11.2% of students, in grades K-6, completing 3+ lessons per week.

These dots are overlaid on a map of DC wards, color-coded to represent the percent of the student body attending schools in that ward from high poverty backgrounds. The light-colored areas represent the wards with the lowest levels of poverty while the dark gray areas represent the wards with the highest levels of poverty. Low, medium and high poverty labels were assigned to wards with <40%, 40-75% and 75%+ percent of students defined as at-risk, respectively. The cutoff of 40% was selected to differentiate between high- and mid-income schools because schools are eligible for Title I funding if their student populations are at least 40% economically disadvantaged (U.S. Department of Education, 2018). The 75%+ cutoff between mid- and low-income schools was selected because the National Center for Education Statistics defines "high poverty" schools as those in which 75% of students are free- or reduced-lunch eligible (2022).

The second map in Figure 3 highlights, in blue, the non-charter DCPS schools with the highest projected impact of fidelity implementation on student math achievement scores, based on the percent of the student body scoring at the early on-grade Placement Level and above. The schools in blue have fewer than 8.5% of students scoring at the early on-grade Placement Level or above on the i-Ready assessment. Because Zearn Math has an outsized impact on students starting below standard, these schools would likely see the most growth from fidelity implementation.

Findings show tremendous opportunities for impact in DCPS schools with strategic implementation of Zearn Math. While 2020-2021 represented a peak in Zearn Math usage across DCPS, with some schools reaching 43% fidelity implementation among students, fidelity implementation declined in 2021-2022, with the highest using school reaching 11% fidelity implementation. Given evidence of the efficacy of Zearn Math in DCPS, these declines represent a missed opportunity in student learning for both the short- and long-term.

Maps show how current usage patterns compare to projected impacts. In DCPS, Zearn Math is implemented with the highest fidelity in the city's wealthiest wards and highest achieving schools. Based on achievement-related projections, DCPS would see the largest boost in learning if schools in the southeastern wards, where the highest concentration of high-poverty students attend school, implemented Zearn Math with fidelity. Zearn Math hopes to partner with DCPS to support high levels of implementation in the most challenged schools and realize maximum growth for DCPS students.

Conclusion and Limitations

This analysis provides promising evidence of Zearn Math's positive impact on student achievement. In addition to positive changes in student performance overall, MLLs, students in special education and Black and/or Latino students who consistently used Zearn Math, as well as students who started below the early on-grade Placement Level, saw even larger gains than the average student. The finding that Zearn Math impacts all students positively, but is associated with even more growth among those starting below the standard, further substantiates findings from efficacy analyses of Zearn Math's impact in other districts (2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c).

In addition to greater growth in raw scale scores, consistent Zearn Math users were more likely than closely matched low- and non-users to move up a Placement Level, end the year early on-grade level or above, and meet expected growth projections. There is also promising evidence that academic gains from fidelity usage of Zearn Math are retained by students even after they discontinue use.

Results from the within-year impact of consistent Zearn Math usage were similar across analyses conducted with two quasi-experimental methods: CEM and fixed effects. Findings from both showed an average of 35% more growth with the consistent use of Zearn Math as a digital complement to in-class instruction. The consistency of findings across methods provides more confidence in the accuracy of results.

Two additional analyses were conducted to examine whether or not students who consistently used Zearn Math better retained math knowledge after periods of no use. An analysis of Zearn Math's impact on summer learning loss found that students who used Zearn Math consistently lost less to summer slide than those with the same spring assessment scores (and other matching factors) who learned through in-person instruction alone.

A separate analysis showed that students who used Zearn Math in 2020-2021, then largely discontinued use for 2021-2022, maintained 74% of their additional growth relative to matched low- or non-users at the end of 2021-2022. This means that the benefits from Zearn Math held even beyond a student's time on the platform.

Findings from both studies support that students who learn on the Zearn Math platform better retain their learnings as measured by the i-Ready Diagnostic.

By matching students closely on starting scores in both math and ELA, grade and seven demographic and academic factors, treatment and control groups were similar along major confounding characteristics. This technique better isolated the impact of Zearn Math usage as an explanatory factor for differences in academic growth and performance than less rigorous correlational analyses. Confirmation of findings through an additional quasi-experimental method added robustness and confidence to the analytic results. For both students overall and disadvantaged subgroups, Zearn Math usage appears to drive higher levels of academic growth.

Despite the strong findings from this analysis, some limitations are present. While CEM allows researchers to control for observed confounders, a possibility exists that unobserved factors mediate the relationship between Zearn Math usage and academic performance. Even with robust quasi-experimental methods, accuracy of estimates is limited by the ability to model all variables relevant to selection into treatment and control.

Fixed effects analysis better controls for time-invariant endogenous confounders within students, but does not control for time-variant factors that confound the relationship between Zearn Math usage and academic performance. In this analysis, accuracy of estimates from the fixed effects model depends on the accuracy of adjustment for pandemic-related disruptions and changes in grade-level, both of which have the potential to impact students' academic performance in a given year. The use of both CEM and fixed effects analyses, and the consistency of results, provides greater confidence in the accuracy of findings.

This analysis examined the impact of fidelity usage of Zearn Math rather than utilizing an intention-to-treat analytic framework that would define the treatment group as all students to whom Zearn Math was available (McCoy, 2017). The focus on fidelity usage is better aligned with the interests of DCPS, for whom the results may help to encourage more universal fidelity usage of Zearn Math. Utilizing fidelity as the benchmark for treatment means that estimates may be biased as this usage represents the best version of implementation which may exceed "typical use".

In addition, the sample is dissimilar from the larger student body, having a proportionally smaller population of students from disadvantaged backgrounds. Therefore, it is possible that the impact of Zearn Math on a larger portion of students in DCPS might show a different effect size. The impact of this difference between sample composition and the overall student population is difficult to definitively disentangle with quasi-experimental methods. Matching and fixed effects models mitigate some of this disparity by controlling for factors that impact academic growth such that they are as similar as possible between treatment and control groups.

Findings from past studies (2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c) and this analysis show an outsized positive impact of Zearn Math on students from disadvantaged backgrounds. Thus, it is possible that the sample underestimates the growth expected from more widespread adoption of Zearn Math. It is also possible, however, that DCPS users are systematically different from non-users in a way that is not observable in the data.

GIS mapping indicates that an implementation strategy targeting the lowest performing schools would have even larger academic impacts across the district. However, it is also possible that users in DCPS are systematically different from non-users in a way that is not observable in the data, but makes them more likely to benefit from Zearn Math or more likely to have greater academic growth irrespective of Zearn Math use.

Finally, this study was conducted on a small population of DCPS students. It is possible that there are features specific to DCPS that facilitate large gains with Zearn Math usage that may not be present in other districts. The geographic specificity of this study may limit the generalizability to a more nationally representative population.

With robust methods and the expansion of efficacy studies to multiple districts across the country, continued replication of trends and findings will provide even stronger evidence of Zearn Math's efficacy moving forward. Zearn plans to continue this work over the coming months and years.

References

Allison, P. D. (2009). *Fixed effects regression models*. SAGE.

Bertsekas, D.P. (1981). A new algorithm for the assignment problem. *Math Programming*, *21*, 152–71. https://doi.org/10.1007/BF01584237

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524-546. https://doi.org/10.1177/1536867X0900900402

Cooper, H., Nye, B., Charlton, K., Lindsay, J., and Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227–268. https://doi.org/10.3102/00346543066003227

Curriculum Associates. (2021). *i-Ready improves outcomes for striving learners, Evidence to support addressing unfinished learning*. Curriculum Associates. https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-striving-learners-research-brief.pdf

Curriculum Associates. (2022a). *The impact of i-Ready Personalized Instruction during the 2020-2021 school year, evidence to support historically marginalized student groups*. Curriculum Associates. https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-efficacy-during-covid-2022.pdf

Curriculum Associates. (2022b) *Using i-Ready Diagnostic scores to meet the Learning Assistance Program (LAP) data reporting guidelines*. Washington Office of Superintendent of Public Instruction. https://www.k12.wa.us/sites/default/files/public/lap/i-Ready%20LAP%20Reporting%20Guidelines%20-2019.pdf

DCPS. (2022). *DCPS at a glance: enrollment* [Data table]. DCPS. https://dcps.dc.gov/page/dcps-glance-enrollment

DCPS. (2022). *DC public schools* [Data vector shapefile]. Open Data DC in the Office of the Chief Technology Officer, Government of the District of Columbia. https://opendata.dc.gov/datasets/4ac321b2d409438ebd76a6569ad94034/explore?location=38.894944%2C-77.015000%2C12.52

Empower K-12. (2022). *Positive indicators for academic recovery in DC*. Empower K-12. https://static1.squarespace.com/static/5f9857f027d55d2170cd92ac/t/630498dd85a2717611d8124d/1661245664787/EK12+Spring+2022+Unfinished+Learning+Update+Brief.pdf

Iacus, S. M., King, G., & Porro, G. (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association*, *106*(493), 345–361. https://doi.org/10.1198/jasa.2011.tm09599

McCoy, C.E. (2017). Understanding the intention-to-treat principle in randomized controlled trials. *The Western Journal of Emergency Medicine*, *18*(6), 1075–1078. https://doi.org/10.5811/westjem.2017.8.35985

National Center for Education Statistics. (2022). Concentration of public school students eligible for free or reduced-price lunch. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/clb

REL Midwest. (2019). *ESSA tiers of evidence: What you need to know*. Regional Educational Laboratory at American Institutes for Research. https://ies.ed.gov/ncee/edlabs/regions/midwest/pdf/blogs/RELMW-ESSA-Tiers-Video-Handout-508.pdf

Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, *7*(1), 143-176. https://dx.doi.org/10.1146/annurev-statistics-031219-041058

Stuart, E. A., & Rubin, D.B. (2008). Matching with multiple control groups with adjustment for group differences. *Journal of Educational and Behavioral Statistics*, *33*(3), 279–306. https://doi.org/10.3102/1076998607306078

Szatrowski, A. (2022a) *Efficacy analysis of Zearn Math: Findings from a large Midwestern urban district.* Zearn. https://webassets.zearn.org/Implementation/MidwesternUrbanDistrictTechnicalPaper.pdf

Szatrowski, A. (2022b) *Efficacy analysis of Zearn Math in Nebraska*. Zearn. https://webassets.zearn.org/Implementation/NebraskaTechnicalAppendix.pdf

Szatrowski, A. (2022c) *Efficacy analysis of Zearn Math in Tulsa*. Zearn. https://webassets.zearn.org/Implementation/TulsaTechnicalPaper.pdf

U.S. Census Bureau. (2022). *Wards from 2022* [Data vector shapefile]. Open Data DC in the Office of the Chief Technology Officer, Government of the District of Columbia. https://opendata.dc.gov/datasets/wards-from-2022/explore?location=38.893755%2C-77.014450%2C12.69

U.S. Department of Education. (2018). *Title I, Part A program*. Title I, Part A Program. https://www2.ed.gov/programs/titleiparta/index.html

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

Zearn. (2022a). *Consistent Zearn usage dramatically reduces learning loss.* Zearn. https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf

Zearn. (2022b). *Economically disadvantaged students, multilingual learners, and Black and Latino students experience double-digit increases in proficiency on 2022 end-of-year assessments with consistent Zearn usage*. Zearn. https://webassets.zearn.org/Implementation/EfficacyResearch_Zearn_Impact_for_Student_Subgroups.pdf

Zearn. (2022c). *For students who scored below grade level in math, consistent Zearn usage tied to 2 grade levels of growth in 2 years of pandemic learning—nearly double the gains of curriculum alone.* Zearn. https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf

Appendix A

Table A1

CEM within year impact analysis: breakdown of sample matching characteristics, 2018-2019, 2020-2021 and 2021-2022		
	Treatment	**Control**
Total N's	990	990
Pre-scores (Fall assessment scores)*		
Math scale score	453.6	453.6
ELA scale score (RI)	887.3	888.0
ELA scale score (DIBELS)	347.6	346.6
Starting Placement Level (N's)		
2+ grade levels below	83	86
1 grade levels below	439	429
Early-on grade and above	468	475
Grades (N's)		
Grade 1	105	105
Grade 2	112	112
Grade 3	233	233
Grade 4	196	196
Grade 5	241	241
Grade 6	103	103
Demographic & academic subgroups (N's)		
Male	493	498
Female	497	492
Black and/or Latino	463	512
Special education	58	34
MLL	103	77
At-risk	157	171

* 5 G2 students had an RI instead of DIBELS; 241 were missing ELA scores (1 in G1, 91 in G3,77 in G4, 69 in G5, and 3 in G6)

Table A2

CEM within year impact analysis: breakdown of sample matching characteristics, SY 2018-2019		
	Treatment	**Control**
Total N's	79	79
Pre-scores (Fall 2018 assessment scores)*		
Math scale score	420.0	420.1
ELA scale score (RI)	584.4	581.5
ELA scale score (DIBELS)	196.2	196.2
Starting Placement Level (N's)		
2+ grade levels below	4	4
1 grade levels below	53	53
Early on-grade and above	22	22
Grades (N's)		
Grade 1	21	21
Grade 2	23	23
Grade 3	18	18
Grade 4	16	16
Grade 5	1	1
Grade 6	0	0
Demographic & academic subgroups (N's)		
Male	36	34
Female	43	45
Black and/or Latino	74	68
Special education	1	0
MLL	8	6
At-risk	39	37
* 1 G1 student was missing an ELA score.		

Table A3

CEM within year impact analysis: breakdown of sample matching characteristics, SY 2020-2021		
	Treatment	**Control**
Total N's	589	589
Pre-scores (Fall 2020 assessment scores)*		
Math scale score	446.5	446.5
ELA scale score (RI)	754.1	754.1
ELA scale score (DIBELS)	346.0	345.1
Starting Placement Level (N's)		
2+ grade levels below	51	53
1 grade levels below	271	264
Early on-grade and above	267	272
Grades (N's)		
Grade 1	74	74
Grade 2	75	75
Grade 3	177	177
Grade 4	143	143
Grade 5	120	120
Grade 6	0	0
Demographic & academic subgroups (N's)		
Male	304	308
Female	285	281
Black and/or Latino	263	320
Special education	40	27
MLL	84	62
At-risk	84	97

*5 G2 students had an RI instead of DIBELS; 236 were missing ELA scores (90 in G3, 77in G4, 69 in G5).

Table A4

CEM within year impact analysis: breakdown of sample matching characteristics, SY 2021-2022

	Treatment	Control
Total N's	322	322
Pre-scores (Fall 2021 assessment scores)*		
Math scale score	474.8	474.7
ELA scale score (RI)	981.9	983.2
ELA scale score (DIBELS)	357.2	355.9
Starting Placement Level (N's)		
2+ grade levels below	28	29
1 grade levels below	115	112
Early on-grade and above	179	181
Grades (N's)		
Grade 1	10	10
Grade 2	14	14
Grade 3	38	38
Grade 4	37	37
Grade 5	120	120
Grade 6	103	103
Demographic & academic subgroups (N's)		
Male	153	156
Female	169	166
Black and/or Latino	126	124
Special education	17	7
MLL	11	9
At-risk	34	37

*4 were missing ELA scores (1 in G3, 3 in G6).

Table A5

Sample composition			
	Treatment	**Control**	**DCPS**
Demographic & academic subgroups			
Black or Latino	47%	52%	78%
Special education	6%	3%	15%
MLL	10%	8%	15%
At-risk	16%	17%	45%

Table A6

Students who used Zearn Math w/fidelity (i.e., 3+ lessons/week, approximately 90+ lessons/year), by grade and year*			
	2018-2019	**2020-2021**	**2021-2022**
Kindergarten**	45	230	20
Grade 1	21	78	11
Grade 2	25	85	15
Grade 3	19	211	44
Grade 4	16	172	41
Grade 5	1	150	138
Grade 6	0	0	130

*53 students used Zearn Math w/fidelity for two years.

**Because Zearn Math lessons for Kindergarten differ substantially from Zearn Math for grades 1-6, Kindergarten is excluded from the main analysis. See Footnote 2 for more detail.

Table A7

CEM within year impact analysis: fall and spring scale score means, by subgroup, 2018-2019, 2020-2021 and 2021-2022							
	Treatment fall	**Treatment spring**	**Control fall**	**Control spring**	**Fall mean difference**	**Pooled SD**	**Difference in SDs***
All Students							
Math scale score	453.59	482.20	453.61	474.70	-0.02	38.69299	0.00
Starting Placement Level							
2+ grade levels below	415.39	454.31	416.16	436.95	-0.78	30.71634	-0.03
1 grade levels below	434.76	466.57	434.64	458.68	0.12	33.0033	0.00
Early on-grade and above	478.03	501.86	477.52	495.96	0.52	28.3528	0.02
Grade Level							
Grade 1	391.35	423.22	391.62	416.43	-0.27	25.61347	-0.01
Grade 2	419.62	449.07	419.86	444.05	-0.24	26.37871	-0.01
Grade 3	442.92	469.39	442.89	463.68	0.03	21.64181	0.00
Grade 4	460.13	492.08	460.12	477.60	0.01	21.60793	0.00
Grade 5	483.32	511.37	483.21	502.42	0.12	21.71436	0.01
Grade 6	496.11	520.54	496.08	521.77	0.03	23.18928	0.00

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment, according to WWC. Differences <.25 of a standard deviation satisfy baseline equivalent with adjustment of difference-in-difference (2022).

Table A7 (cont.)

CEM within year impact analysis: fall and spring scale score means, by subgroup, 2018-2019, 2020-2021 and 2021-2022 (cont.)

	Treatment fall	Treatment spring	Control fall	Control spring	Fall mean difference	Pooled SD	Difference in SDs*
Demographic & academic subgroups							
Male	451.80	480.88	453.55	475.15	-1.75	38.37007	-0.05
Female	455.36	483.56	453.66	474.20	1.70	39.0104	0.04
Black or Latino	438.68	468.22	440.15	458.46	-1.47	37.53774	-0.04
Special education	435.26	468.86	430.15	445.68	5.11	39.80328	0.13
MLL	430.91	457.65	428.42	438.69	2.50	33.21806	0.08
At-risk	432.23	459.51	432.25	451.36	-0.02	37.005	0.00

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment, according to WWC. Differences <.25 of a standard deviation satisfy baseline equivalent with adjustment of difference-in-difference (2022).

Table A8

CEM within year impact analysis: comparison of changes in scores and Placement Level between consistent Zearn users and low- or non-users, 2018-2019, 2020-2021 & 2021-2022

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
All Students					
Math scale score (SS)	28.64	21.07	7.56***	27.65	0.274
Math percent early on-grade or above	38.08%	27.27%	10.81%***	0.484	0.223
Starting Placement Level					
2+ grade levels below SS	38.93	20.79	18.14***	24.63	0.737
1 grade below SS	31.82	24.04	7.78***	21.16	0.368
Early on-grade and above SS	23.83	18.44	5.38***	21.01	0.256
Grade Level					
Grade 1 SS	31.87	24.81	7.056*	24.00	0.294
Grade 1 percent early on-grade or above	46.67%	38.10%	8.57%	0.550	0.156
Grade 2 SS	29.46	24.20	5.26	26.77	0.196
Grade 2 percent early on-grade or above	42.86%	39.29%	3.57%	0.553	0.065
Grade 3 SS	26.46	20.79	5.68**	21.80	0.261
Grade 3 percent early on-grade or above	45.06%	32.19%	12.88%**	0.538	0.239
Grade 4 SS	31.949	17.470	14.48***	23.01	0.629
Grade 4 percent early on-grade or above	39.80%	20.41%	19.39%***	0.527	0.368
Grade 5 SS	28.046	19.22	8.83***	18.55	0.476
Grade 5 percent early on-grade or above	28.63%	19.09%	9.54%*	0.460	0.208
Grade 6 SS	24.44	25.69	-1.25	16.06	-0.078
Grade 6 percent early on-grade or above	27.18%	24.27%	2.91%	0.460	0.063

* $p<.05$ ** $p<.01$ *** $p<.001$

Table A9

CEM within year impact analysis: comparison of changes in scores and Placement Level between consistent Zearn Math users and low- or non-users, 2020-2021					
	Treatment change in mean	**Control change in mean**	**Difference**	**Pooled SD**	**Cohen's d**
All Students					
Math scale score (SS)	25.998	15.876	10.12***	30.01	0.337
Math percent early on-grade or above	36.67%	22.58%	14.09%***	0.53	0.267
Starting Placement Level					
2+ grade levels below SS	37.98	18.00	19.98***	25.87	0.772
1 grade below SS	29.97	19.25	10.72***	22.46	0.477
early on-grade and above SS	19.68	12.18	7.49***	23.20	0.323
Grade Level					
Grade 1 SS	29.35	18.84	10.51*	25.12	0.418
Grade 1 percent early on-grade or above	37.84%	27.03%	10.81%	0.55	0.197
Grade 2 SS	25.27	18.93	6.33	28.37	0.223
Grade 2 percent early on-grade or above	0.32	0.31	1.33%	0.56	0.024
Grade 3 SS	23.28	18.27	5.02*	22.74	0.221
Grade 3 percent early on-grade or above	0.41	0.28	12.43%*	0.54	0.229
Grade 4 SS	29.55	12.31	17.245***	23.05	0.748
Grade 4 percent early on-grade or above	0.38	0.14	24.48%***	0.53	0.458
Grade 5 SS	24.16	12.87	11.29***	20.89	0.541
Grade 5 percent early on-grade or above	0.31	0.17	14.17%*	0.49	0.289
* p<.05 **p<.01 ***p<.001					

Table A10

CEM within year impact analysis: comparison of changes in scores and Placement Level between consistent Zearn Math users and low- or non-users, 2021-2022					
	Treatment change in mean	**Control change in mean**	**Difference**	**Pooled SD**	**Cohen's d**
All Students					
Math scale score (SS)	32.01	26.80	5.22***	22.42	0.233
Math percent early on-grade or above	36.34%	28.57%	7.76%***	0.391	0.199
Starting Placement Level					
2+ grade levels below SS	38.07	24.93	13.14*	20.79	0.632
1 grade below SS	34.70	27.35	7.36**	16.94	0.434
early on-grade and above SS	29.34	26.75	2.58	14.84	0.174
Grade Level					
Grade 1 SS**	47.60	33.10	14.50**	14.58	0.995
Grade 1 percent early on-grade or above	90.00%	60.00%	30.00%	0.44	0.675
Grade 2 SS**	31.79	27.14	4.64	18.329	0.253
Grade 2 percent early on-grade or above	57.14%	50.00%	7.14%	0.508	0.141
Grade 3 SS	38.58	30.47	8.11*	14.41	0.562
Grade 3 percent early on-grade or above	55.26%	39.47%	15.79%	0.503	0.314
Grade 4 SS	42.14	28.84	13.30**	19.77	0.673
Grade 4 percent early on-grade or above	51.35%	35.14%	16.22%	0.499	0.325
Grade 5 SS	32.04	25.38	6.66***	14.24	0.468
Grade 5 percent early on-grade or above	26.67%	21.67%	5.00%	0.429	0.117
Grade 6 SS	24.44	25.69	-1.25	16.06	-0.078
Grade 6 percent early on-grade or above	27.18%	24.27%	2.91%	0.460	0.063

* p<.05 **p<.01 ***p<.001
**n<20

Table A11

CEM within year impact analysis: comparison of growth in scale score related to expectation and % meeting growth expectation between consistent Zearn Math users and low- or non-users, 2018-2019, 2020-2021 & 2021-2022

	Treatment mean	Control mean	Difference	Pooled SD	Cohen's d
All Students					
Math scale score growth relative to expectation	7.10	-0.45	7.55***	27.67	0.273
Math median % of typical growth achieved	135.86%	104.35%	31.51%***	--	--
Math percent meeting growth expectation	68.69%	53.74%	14.95%***	0.65	0.230
Starting Placement Level					
2+ grades below growth relative to expectation	15.94	-2.13	18.07***	24.19	0.747
2+ grades below median % of typical growth achieved	166.67%	87.23%	79.44%***	--	--
2+ grades below percent meeting expectation	79.52%	45.35%	34.17%***	0.49	0.702
1 grade below growth relative to expectation	8.13	0.35	7.78***	21.17	0.367
1 grade below median % of typical growth achieved	134.62%	103.85%	30.77%***	--	--
1 grade below percent meeting expectation	68.79%	54.55%	14.25%***	0.49	0.293
Early on-grade+ growth relative to expectation	4.58	-0.87	5.44***	21.39	0.255
Early on-grade+ median % of typical growth achieved	132.67%	105.56%	27.11%**	--	--
Early on-grade+ percent meeting expectation	66.67%	54.53%	12.14%***	0.49	0.248

* p<.05 **p<.01 ***p<.001

Table A11 (cont.)

CEM within year impact analysis: comparison of growth in scale score related to expectation and % meeting growth expectation between consistent Zearn Math users and low- or non-users, 2018-2019, 2020-2021 & 2021-2022 (cont.)					
	Treatment mean	**Control mean**	**Difference**	**Pooled SD**	**Cohen's d**
Grade Level					
Grade 1 growth relative to expectation	4.64	-2.39	7.03*	22.77	0.309
Grade 1 median % of typical growth achieved	117.24%	100.00%	17.24%	--	--
Grade 1 percent meeting expectation	60.95%	53.33%	7.62%	0.50	0.154
Grade 2 growth relative to expectation	5.66	0.44	5.22	25.66	0.204
Grade 2 median % of typical growth achieved	126.92%	100.00%	26.92%	--	--
Grade 2 percent meeting expectation	59.82%	50.89%	8.93%	0.50	0.179
Grade 3 growth relative to expectation	1.30	-4.37	5.67**	21.52	0.264
Grade 3 median % of typical growth achieved	111.54%	90.00%	21.54%*	--	--
Grade 3 percent meeting expectation	57.94%	47.21%	10.73%*	0.50	0.215
Grade 4 growth relative to expectation	9.39	-5.05	14.43***	22.90	0.630
Grade 4 median % of typical growth achieved	1.391304	0.8722826	51.90%***	--	--
Grade 4 percent meeting expectation	70.41%	46.94%	23.47%***	0.49	0.476
Grade 5 growth relative to expectation	10.94	2.10	8.85***	18.39	0.481
Grade 5 median % of typical growth achieved	161.11%	114.29%	46.83%***	--	--
Grade 5 percent meeting expectation	79.25%	56.85%	22.41%***	0.47	0.480
Grade 6 growth relative to expectation	10.99	12.23	-1.24	16.07	-0.077
Grade 6 median % of typical growth achieved	192.31%	192.86%	-0.55%	--	--
Grade 6 percent meeting expectation	82.52%	77.67%	4.85%	0.40	0.121

* p<.05 **p<.01 ***p<.001

Table A12

CEM within year impact analysis: comparison of changes in scores and Placement Level between consistent Zearn Math users and low- or non-users, by subgroup, 2018-2019, 2020-2021 & 2021-2022					
	Treatment change in mean	**Control change in mean**	**Difference**	**Pooled SD**	**Cohen's d**
Subgroup					
Female SS	28.20	20.54	7.66***	21.72	0.352
Female percent early on-grade or above	37.83%	26.63%	11.20%***	0.527	0.221
Male SS	29.08	21.59	7.49***	21.67	0.346
Male percent early on-grade or above	38.34%	27.91%	10.43%**	0.508	0.205
Black or Latino SS	29.54	18.30	11.24***	24.84	0.452
Black or Latino percent early on-grade or above	46.22%	29.69%	16.53%***	0.542	0.305
Special education SS	33.60	15.53	18.07**	26.34	0.686
Special education percent early on-grade or above	43.10%	26.47%	16.35%	0.507	0.328
MLL SS	26.74	10.27	16.47***	28.06	0.587
MLL percent early on-grade or above	45.63%	15.58%	30.05%***	0.596	0.504
At-risk SS	27.28	19.11	8.17**	27.17	0.301
At-risk percent early on-grade or above	46.50%	32.75%	13.75%*	0.554	0.248

* p<.05 **p<.01 ***p<.001

Table A13

CEM within year impact analysis: comparison of changes in scores and Placement Level between consistent Zearn Math users and low- or non-users, by subgroup, 2020-2021

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
Subgroup					
Female SS	25.48	14.57	10.91***	24.21	0.45
Female percent early on-grade or above	36.14%	18.86%	17.28%***	0.54	0.32
Male SS	26.48	17.06	9.42***	22.96	0.41
Male percent early on-grade or above	37.17%	25.97%	11.20%**	0.52	0.21
Black or Latino SS	25.34	11.80	13.54***	26.47	0.51
Black or Latino percent early on-grade or above	41.83%	19.06%	22.76%***	0.55	0.42
Special education SS	30.95	10.56	20.39**	27.83	0.73
Special education percent early on-grade or above	45.00%	18.52%	26.48%*	0.51	0.52
MLL SS	24.54	5.97	18.57***	29.23	0.64
MLL percent early on-grade or above	40.48%	4.84%	35.64%***	0.60	0.60
At-risk SS	22.64	10.36	12.28**	30.21	0.41
At-risk percent early on-grade or above	39.29%	17.53%	21.76%**	0.56	0.39

* p<.05 **p<.01 ***p<.001

Table A14

CEM within year impact analysis: comparison of changes in scores and Placement Level between consistent Zearn Math users and low- or non-users, by subgroup, 2021-2022

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
Subgroup					
Female SS	31.11	27.09	4.02*	15.14	0.27
Female percent early on-grade or above	37.25%	25.00%	12.25%*	0.46	0.26
Male SS	33.01	26.48	6.53***	17.38	0.38
Male percent early on-grade or above	35.50%	31.93%	3.58%	0.49	0.07
Black and/or Latino SS	35.46	24.70	10.76***	18.93	0.57
Black and/or Latino percent early on-grade or above	48.41%	38.71%	9.70%	0.51	0.19
Special education SS**	41.12	34.71	6.400	17.37	0.37
Special education percent early on-grade or above	41.18%	57.14%	-15.97%	0.51	-0.31
MLL SS**	40.64	25.00	15.64*	16.83	0.93
MLL percent early on-grade or above	81.82%	44.44%	37.37%	0.49	0.76
At-risk SS	34.44	23.81	10.63*	19.91	0.53
At-risk percent early on-grade or above	50.00%	37.84%	12.16%	0.53	0.23

* p<.05 **p<.01 ***p<.001
**n<20

Table A15

CEM within year impact analysis: comparison of growth in scale score related to expectation and % meeting growth expectation between consistent Zearn Math users and low- or non-users, by subgroup, 2018-2019, 2020-2021 & 2021-2022

	Treatment mean	Control mean	Difference	Pooled SD	Cohen's d
Subgroup					
Female growth relative to expectation SS	6.94	-0.86	7.80***	21.79	0.358
Female median % of typical growth achieved	133.33%	100.00%	33.33%***	--	--
Female percent meeting growth expectation	68.41%	53.05%	15.36%***	0.49	0.314
Male growth relative to expectation SS	7.27	-0.04	7.32***	21.40	0.342
Male median % of typical growth achieved	138.46%	104.45%	34.01%***	--	--
Male percent meeting growth expectation	68.97%	54.42%	14.55%***	0.49	0.299
Black and/or Latino growth relative to expectation SS	6.52	-4.36	10.79***	24.42	0.446
Black and/or Latino median % of typical growth achieved	131.03%	88.89%	42.15%***	--	--
Black and/or Latino percent meeting growth expectation	64.36%	46.29%	18.07%***	0.50	0.363
Special education growth relative to expectation SS	11.07	-7.85	18.92***	26.17	0.723
Special education median % of typical growth achieved	151.64%	64.64%	87.00%*	--	--
Special education percent meeting growth expectation	70.69%	38.24%	32.45%**	0.50	0.656
MLL growth relative to expectation SS	2.65	-13.75	16.40***	27.43	0.598
MLL median % of typical growth achieved	128.00%	57.14%	70.86%***	--	--
MLL percent meeting growth expectation	61.17%	29.87%	31.29%***	0.50	0.625
At-risk growth relative to expectation SS	3.53	-4.20	7.73**	26.66	0.290
At-risk median % of typical growth achieved	116.67%	88.89%	27.78%	--	--
At-risk percent meeting growth expectation	56.05%	46.20%	9.85%	0.50	0.197

* p<.05 **p<.01 ***p<.001

Table A16

CEM within year impact analysis: logistic regression probabilities of early on-grade Placement Level and Placement Level growth among students with consistent Zearn usage and those with little or no Zearn usage, 2018-2019, 2020-2021 & 2021-2022					
	Zearn Math users	Low- or non-users	Difference	Relative likelihood	OR
All students					
Probability of moving up 1+ levels	68.99%	57.98%	11.01%	1.19	1.61***
Probability of scoring early on-grade or higher	85.35%	75.25%	10.10%	1.13	1.92***
Probability of meeting expected growth	68.69%	53.74%	14.95%	1.28	1.89***
Demographic & academic subgroups					
Male probability of moving up 1+ levels	67.14%	57.43%	9.71%	1.17	1.51**
Male probability of scoring early on-grade or higher	84.79%	76.91%	7.88%	1.10	1.67**
Male probability of meeting expected growth	68.97%	54.42%	14.55%	1.27	1.86***
Female probability of moving up 1+ levels	70.82%	58.54%	12.29%	1.21	1.72***
Female probability of scoring early on-grade or higher	85.92%	73.58%	12.34%	1.17	2.19***
Female probability of meeting expected growth	68.41%	53.05%	15.36%	1.29	1.92***
* p<.05 **p<.01 ***p<.001					

Table A16 (cont.)

CEM within year impact analysis: logistic regression probabilities of early on-grade Placement Level and Placement Level growth among students with consistent Zearn Math usage and those with little or no Zearn Math usage, 2018-2019, 2020-2021 and 2021-2022 (cont.)

	Zearn Math users	Low- or non-users	Difference	Relative likelihood	OR
Black or Latino probability of moving up 1+ levels	75.16%	56.64%	18.52%	1.33	2.32***
Black or Latino probability of scoring early on-grade or higher	77.97%	59.18%	18.79%	1.32	2.44***
Black or Latino probability of meeting expected growth	64.36%	46.29%	18.07%	1.39	2.10***
Special education probability of moving up 1+ levels	75.86%	47.06%	28.80%	1.61	3.54**
Special education probability of scoring early on-grade or higher	65.52%	44.12%	21.40%	1.49	2.41*
Special education probability of meeting expected growth	70.69%	38.24%	32.45%	1.85	3.90**
MLL probability of moving up 1+ levels	75.73%	48.05%	27.68%	1.58	3.37***
MLL probability of scoring early on-grade or higher	68.93%	36.36%	32.57%	1.90	3.88***
MLL probability of meeting expected growth	61.17%	29.87%	31.29%	2.05	3.70***
At-risk probability of moving up 1+ levels	70.70%	59.65%	11.05%	1.19	1.63*
At-risk probability of scoring early on-grade or higher	71.97%	54.97%	17.00%	1.31	2.10**
At-risk probability of meeting expected growth	56.05%	46.20%	9.85%	1.21	1.49
* p<.05 **p<.01 ***p<.001					

Table A17

Breakdown of the fixed effects sample by years of usage		
# of students	Years of low to no usage	Years of fidelity usage
681	1	1
395	2	1
15	1	2

Table A18

Breakdown of the fixed effects sample by usage and year		
School year	Fidelity users	Low- or non-users
2018-2019	69	517
2020-2021	571	154
2021-2022	241	784

Table A19

Breakdown of the fixed effects sample by usage and grade		
Grade	Fidelity users	Low- or non-users
Grade 1	84	280
Grade 2	78	357
Grade 3	191	283
Grade 4	211	189
Grade 5	237	201
Grade 6	80	145

Table A20

Fixed effects analysis: average growth on the i-Ready exam, by grade, between fall and spring assessment administration in DCPS, for students in years when they consistently used Zearn Math vs. years when they had low or no Zearn Math usage*			
	Consistent Zearn Math (weeks of growth, relative to average DCPS student)	**Low or no Zearn Math** (weeks of growth, relative to average DCPS student)	**Difference** (additional weeks of growth with consistent Zearn Math)
All students	+13.23	+3.04	+10.18***
Female	+13.73	+2.22	+11.5***
Male	+12.57	+3.90	+8.67***
Black or Latino	+10.76	+0.50	+10.27***
Special education	+16.09	+3.32	+12.77*
MLL	+6.93	+3.66	+3.26
At Risk	+4.91	-+1.59	+6.50

*Results based on a fixed effects model comparing growth relative to grade level expectations for students with at least one year of fidelity Zearn Math usage and one year of low or no Zearn Math usage over school years 2018-2019, 2020-2021 and 2021-2022.

Table A21

Summer slide analysis: breakdown of sample matching characteristics, 2018-2019 & 2020-2021		
	Treatment	**Control**
Total N's	745	745
Pre-scores (Spring assessment scores- spring 2019 & spring 2021)*		
Math scale score	459.4	459.3
ELA scale score (RI)	895.6	895.9
ELA scale score (DIBELS)	441.9	442.4
Starting Placement Level (N's)		
2+ grade levels below	12	14
1 grade levels below	145	147
Early on-grade and above	588	584
Grades (N's)		
Grade 1	196	196
Grade 2	114	114
Grade 3	171	171
Grade 4	150	150
Grade 5	114	114
Demographic & academic subgroups (N's)		
Male	370	341
Female	375	403
Black and/or Latino	364	405
Special education	48	27
MLL	95	72
At-risk	121	137
* 11 G2 students had an RI instead of DIBELS; 113 were missing ELA scores (24 in G3, 50 in G4, 39 in G5)		

Table A22

Summer slide analysis: spring and fall scale score means, by subgroup, summer 2019 & summer 2021							
	Treatment Spring	**Treatment Fall**	**Control Spring**	**Control Fall**	**Spring mean difference**	**Pooled SD**	**Difference in SDs***
All Students							
Math scale score	459.37	451.04	459.26	447.05	0.11	37.92	0.00
Starting Placement Level (N's)							
2+ grade levels below	413.83	425.33	413.64	418.64	0.19	25.14	0.01
1 grade levels below	422.19	420.96	421.82	415.67	0.37	30.91	0.01
Early on-grade and above	469.47	458.99	469.78	455.63	-0.31	32.83	-0.01
Grade Level							
Grade 1	420.43	406.54	420.29	406.15	0.14	23.81	0.01
Grade 2	443.54	437.66	443.47	432.23	0.07	27.98	0.00
Grade 3	465.15	458.42	465.09	456.27	0.06	24.05	0.00
Grade 4	485.37	476.72	485.25	470.41	0.13	24.42	0.01
Grade 5	499.29	496.10	499.09	487.62	0.20	24.62	0.01

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment, according to WWC. Differences <.25 of a standard deviation satisfy baseline equivalent with adjustment of difference-in-difference (2022).

Table A22 (cont.)

Summer slide analysis: spring and fall scale score means, by subgroup, summer 2019 & summer 2021							
	Treatment Spring	**Treatment Fall**	**Control Spring**	**Control Fall**	**Spring mean difference**	**Pooled SD**	**Difference in SDs***
Demographic & academic subgroups							
Male	460.86	453.27	459.64	447.32	1.22	38.40	0.03
Female	457.91	448.84	458.94	446.80	-1.03	37.50	-0.03
Black or Latino	448.57	437.34	452.00	435.87	-3.43	38.48	-0.09
Special education	443.56	437.31	448.30	434.93	-4.73	43.43	-0.11
MLL	439.81	430.78	434.42	420.50	5.39	37.56	0.14
At-risk	439.69	429.02	443.80	423.34	-4.11	38.14	-0.11

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment, according to WWC. Differences <.25 of a standard deviation satisfy baseline equivalent with adjustment of difference-in-difference (2022).

Table A23

Summer learning loss in scale score points for prior year Consistent Zearn users *(Treatment)* vs. Low- or Non-users *(Control)*, by subgroup					
	Treatment change in mean	**Control change in mean**	**Difference**	**Pooled SD**	**Cohen's d**
All Students					
Math scale score (SS)	-8.33	-12.21	3.88***	23.26	0.167
Starting Placement Level					
2+ grade levels below SS	11.40	5.00	6.40	18.86	0.339
1 grade levels below SS	-1.23	-6.15	4.92*	17.41	0.282
Early on-grade and above SS	-10.48	-14.15	3.66**	20.36	0.180
Grade Level					
Grade 1 SS	-13.89	-14.13	0.24	22.61	0.01
Grade 2 SS	-5.89	-11.27	5.39	22.12	0.24
Grade 3 SS	-6.74	-8.82	2.08	18.00	0.12
Grade 4 SS	-8.65	-14.84	6.19**	18.68	0.33
Grade 5 SS	-3.18	-11.46	8.29***	17.81	0.47
Subgroup					
Female SS	-9.07	-12.13	3.06*	19.8866	0.15
Male SS	-7.58	-12.32	4.74**	20.5970	0.23
Black or Latino SS	-11.23	-16.13	4.90**	22.9246	0.21
Special education SS	-6.25	-13.37	7.12	23.5795	0.30
MLL SS	-9.03	-13.92	4.89	21.0927	0.23
At-Risk SS	-10.66	-20.46	9.80**	24.8936	0.39
* p<.05 **p<.01 ***p<.001					

Table A24

Two year impact analysis: breakdown of sample matching characteristics		
	Treatment	**Control**
Total N's	402	402
Pre-scores (SY 20-21 Fall assessment scores)*		
Math scale score	442.9	442.9
ELA scale score (RI)	731.3	730.4
ELA scale score (DIBELS)	341.6	340.9
Starting Placement Level (N's)		
2+ grade levels below	41	40
1 grade levels below	210	208
Early on-grade and above	151	154
20-21 Grade (N's)		
Grade 1	61	61
Grade 2	42	42
Grade 3	101	101
Grade 4	118	118
Grade 5	80	80
Demographic & academic subgroups (N's)		
Male	213	195
Female	189	207
Black and/or Latino	194	261
Special education	31	20
MLL	52	49
At-risk	67	87

*3 G2 students had an RI instead of DIBELS; 163 were missing ELA scores (41 in G3, 70 in G4, and 52 in G5)

Table A25

Two-year impact analysis: matched sample fall 2020 and spring 2022 scale score means, by subgroup							
	Treatment Fall 2020	Treatment Spring 2022	Control Fall 2020	Control Spring 2022	Fall mean difference	Pooled SD	Difference in SDs*
All Students							
Math scale score	442.85	486.34	442.86	476.77	0.00	35.00	0.00
Starting Placement Level (N's)							
2+ grade levels below	407.37	456.44	406.85	443.38	0.52	30.62	0.02
1 grade levels below	433.28	478.17	432.95	468.80	0.33	30.91	0.01
Early on-grade and above	465.79	505.81	465.58	496.21	0.21	26.15	0.01
Grade Level							
Grade 1	391.92	434.33	392.16	429.49	-0.25	26.13	-0.01
Grade 2	422.55	463.86	422.55	460.55	0.00	31.63	0.00
Grade 3	439.56	480.54	439.49	480.83	0.08	21.64	0.00
Grade 4	457.75	504.30	457.65	489.81	0.09	19.66	0.00
Grade 5	474.53	518.61	474.60	496.99	-0.08	20.70	0.00
Demographic & academic subgroups							
Male	442.74	488.25	444.13	478.64	-1.39	36.86	-0.04
Female	442.98	484.17	441.66	475.02	1.32	33.00	0.04
Black or Latino	432.53	470.47	438.92	468.19	-6.39	35.65	-0.18
Special education	428.16	474.55	417.30	432.85	10.86	39.14	0.278+
MLL	425.37	471.56	428.69	467.69	-3.33	34.81	-0.10
At-risk	430.39	460.85	431.38	457.30	-0.99	36.87	-0.03

*According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment, according to WWC. Differences <.25 of a standard deviation satisfy baseline equivalent with adjustment of difference-in-difference (2022).
+Does not satisfy baseline equivalence even with adjustment.

Table A26

Two-year impact analysis: matched sample fall 2020 and spring 2022 scale score means, by subgroup										
	Treatment				Control				Difference	
	Fall 2020	Spring 2021	Fall 2021	Spring 2022	Fall 2020	Spring 2021	Fall 2021	Spring 2022	EOY 2021	EOY 2022
All Students										
Math scale score	442.9	469.7	460.0	486.3	442.9	456.8	448.5	476.8	12.9	9.6***
Starting Placement Level										
2+ grade levels below	407.4	444.4	435.9	456.4	406.9	419.3	415.1	443.4	24.6	12.5*
1 grade levels below	433.3	462.5	452.0	478.2	433.0	449.8	440.9	468.8	12.4	9.0***
Early on-grade and above	465.8	486.5	477.6	505.8	465.6	476.0	467.3	496.2	10.3	9.4**
Grade Level										
Grade 1	391.9	421.0	403.6	434.3	392.2	411.8	397.9	429.5	9.5	5.1
Grade 2	422.5	444.1	433.7	463.9	422.5	432.8	422.7	460.5	11.4	3.3
Grade 3	439.6	463.9	453.9	480.5	439.5	460.0	450.9	480.8	3.8	-0.4
Grade 4	457.7	487.6	478.7	504.3	457.7	467.9	461.7	489.8	19.7	14.4***
Grade 5	474.5	501.1	496.8	518.6	474.6	483.4	478.0	497.0	17.8	21.7***
Subgroups										
Male	442.7	469.8	460.3	488.3	444.1	459.5	448.9	478.6	11.7	11.0***
Female	443.0	469.6	459.6	484.2	441.7	454.3	448.1	475.0	14.0	7.8**
Black or Latino	432.5	458.7	445.5	470.5	438.9	449.3	440.4	468.2	15.8	8.7**
Special education	428.2	460.3	448.3	474.5	417.3	408.5	404.9	432.9	40.9	30.8*
MLL	425.4	453.9	442.7	471.6	428.7	443.3	435.5	467.7	13.9	7.2
At-risk	430.4	451.8	437.3	460.9	431.4	440.7	437.3	457.3	12.1	4.5

* p<.05 **p<.01 ***p<.001

Table A27

Two-year impact analysis: comparison of changes in scores and Placement Level, fall 2020 to spring 2022, between students who consistently used Zearn Math in 2020-2021, relative to matched peers with little or no usage					
	Treatment change in mean	**Control change in mean**	**Difference**	**Pooled SD**	**Cohen's d**
All Students					
Math scale score (SS)	43.49	33.92	9.567***	30.827	0.310
Math percent early on-grade or above	39.30%	26.62%	12.69%***	0.530	0.239
Starting Placement Level					
2+ grade levels below SS	49.07	36.53	12.55*	26.155	0.480
1 grade levels below SS	44.89	35.85	9.03***	22.832	0.396
Early on-grade SS	40.02	30.63	9.39**	30.163	0.311
Grade Level					
Grade 1 SS	42.41	37.33	5.08	28.367	0.179
Grade 1 percent early on-grade or above	31.15%	16.39%	14.75%	0.590	0.250
Grade 2 SS	41.31	38.00	3.31	35.076	0.094
Grade 2 percent early on-grade or above	30.95%	26.19%	4.76%	0.632	0.075
Grade 3 SS	40.98	41.35	-0.366	22.362	-0.016
Grade 3 percent early on-grade or above	39.60%	39.60%	0.00%	0.539	0.000
Grade 4 SS	46.55	32.161	14.390***	22.053	0.653
Grade 4 percent early on-grade or above	45.76%	27.12%	18.64%**	0.525	0.355
Grade 5 SS	44.09	22.39	21.70***	29.056	0.747
Grade 5 percent early on-grade or above	40.00%	17.50%	22.5%**	0.494	0.456
* p<.05 **p<.01 ***p<.001					

Table A27 (cont.)

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
Subgroup					
Female SS	41.20	33.36	7.83**	24.670	0.318
Female percent early on-grade or above	37.04%	25.60%	11.43%*	0.553	0.207
Male SS	45.52	34.51	11.009***	27.708	0.397
Male percent early on-grade or above	41.31%	27.69%	13.62%*	0.540	0.252
Black or Latino SS	37.94	29.28	8.66**	28.506	0.304
Black or Latino percent early on-grade or above	39.69%	24.14%	15.55%**	0.577	0.270
Special education SS	++	++	++	++	++
Special education percent early on-grade or above	++	++	++	++	++
MLL SS	46.19	39.00	7.19	26.60	0.270
MLL percent early on-grade or above	57.69%	36.73%	20.96%	0.558	0.375
At-risk SS	30.46	25.92	4.54	30.62	0.148
At-risk percent early on-grade or above	31.34%	16.09%	15.25%	0.611	0.250

* p<.05 **p<.01 ***p<.001
++Subgroup does not satisfy baseline equivalence even with statistical adjustment for the two-year impact analysis.

Table A28

Two-year impact analysis: logistic regression probabilities of early on-grade Placement Level and Placement Level growth, fall 2020 to spring 2022, between students who consistently used Zearn Math during the 2020-2021 school year and matched peers with little or no usage

	Probability for consistent Zearn Math users	Probability for low- or non-Zearn Math users	Difference	OR
All Students				
Probability of moving up 1+ levels	66.42%	53.48%	12.94%	1.72***
Probability of scoring early on-grade or higher	76.87%	64.93%	11.94%	1.79***
Demographic & academic subgroups				
Male probability of moving up 1+ levels	68.54%	56.41%	12.13%	1.68*
Male probability of scoring early on-grade or higher	77.93%	67.69%	10.24%	1.69*
Female probability of moving up 1+ levels	64.02%	50.72%	13.30%	1.73**
Female probability of scoring early on-grade or higher	75.66%	62.32%	13.34%	1.875**
Black or Latino probability of moving up 1+ levels	63.40%	49.43%	13.98%	1.77**
Black or Latino probability of scoring early on-grade or higher	64.43%	50.96%	13.48%	1.74**
Special education probability of moving up 1+ levels	70.97%	30.00%	40.97%	5.7**
Special education probability of scoring early on-grade or higher	58.06%	15.00%	43.06%	7.85**
MLL probability of moving up 1+ levels	71.15%	67.35%	3.81%	1.20
MLL probability of scoring early on-grade or higher	73.08%	55.10%	17.97%	2.12
At-risk probability of moving up 1+ levels	59.70%	42.53%	17.17%	2.00*
At-risk probability of scoring early on-grade or higher	55.22%	36.78%	18.44%	2.12*

* p<.05 **p<.01 ***p<.001

Appendix B

This study was designed to meet the What Works Clearinghouse (WWC) “Meets WWC Group Design Standards with Reservations'' rating and to meet an ESSA Tier 2 (Moderate) rating on the ESSA guidelines for evidence-based interventions. This Appendix provides more detail about the criteria for these designations and how this impact study meets those criteria.

What Works Clearinghouse provides ratings of randomized control trials (RCTs) and quasi-experimental designs (QEDs) against their Group Design standards. There are three possible ratings: Meets WWC Standards without Reservations, Meets WWC Standards with Reservations, or Does Not Meet WWC Standards. Because QED studies that establish baseline equivalence or use acceptable statistical adjustments “reduce, but likely do not eliminate, the potential bias associated with the group assignment procedures”, Meets WWC Standards with Reservations is the highest possible rating for QEDs (What Works Clearinghouse, 2022).

This study uses quasi-experimental matching methods to create baseline equivalency between treatment and control groups along major confounding factors. Consistent Zearn Math users were matched with low- or non-users, in the same grade, on starting math and English Language Arts (ELA) standardized test scores, along with seven student characteristics using a two-step Coarsened Exact Matching (CEM) method with optimal matching. CEM is a technique that simulates block sampling by matching students on covariates related both to a student’s likelihood of using Zearn Math consistently and their academic performance (Blackwell et al., 2009; Iacus et al., 2011).

A QED study must satisfy several criteria to meet the WWC standard of “Meets WWC Standards with Reservations”. The first is that the outcome measure “meets four standards: (1) face validity, (2) reliability, (3) not over aligned with the intervention, and (4) consistent data collection procedures” (What Works Clearinghouse, 2022). In this study, the primary outcome of growth in math is the i-Ready Diagnostic. WWC considers standardized tests that are routinely administered in educational settings, like i-Ready Diagnostic, to meet these standards.

The next criteria is the elimination of confounding factors (What Works Clearinghouse, 2022). By matching fidelity users to low- or non-users within three scale score points on their pre-score i-Ready Diagnostic and within either 30 points on the Reading Inventory (RI) or 10 points on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), as well as at least four of seven other student characteristics: school, gender, race, ethnicity, special education status, multilingual learners (MLLs), and at-risk status, the design of this study creates two groups that are academically and demographically similar on the most relevant and measurable confounding factors that would impact academic growth.

While CEM allows researchers to control for observed confounders, a possibility exists that there are unmeasured factors that differentiate the comparison groups of students who reach fidelity and those with little to no usage. For example, it is possible that an unmeasured characteristic allows fidelity users to reach higher usage than would be possible for low- or non-users. However, this type of unmeasurable attribute is what WWC refers to as, “imperfect overlap in the characteristic between the conditions” which they term a selection mechanism, not a confounding factor (2020, p. 82).

This possibility of an unmeasured characteristic that could bias estimates is similar to an example provided by WWC of a program based on voluntary enrollment in which students who volunteer could differ from those who did not in hard-to-measure qualities like introversion vs. extroversion. It clarifies that, “the WWC does not consider this to be a confounding factor, but the selection mechanism and potential difference in unmeasured characteristics are reasons that QEDs are limited to a rating of Meets WWC Group Design Standards with Reservations, if the baseline equivalence requirement is satisfied” (2020, p. 82).

The final criteria for a quasi-experimental study to meet WWC Standards with Reservations is illustrating baseline equivalence between treatment and control groups. This can be done with a pre-intervention measure that is the same as the outcome measure (2022). In this case, i-Ready Diagnostic math scores are used as a pre-intervention measure of baseline equivalence and as the outcome measure of the study.

According to WWC, baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment. Differences <.25 of a standard deviation satisfy baseline equivalent with statistical adjustment. Difference-in-difference and fixed effects are both acceptable statistical adjustments (2022). All groups in this study meet the criteria for baseline equivalence either without or with adjustment, with the exception of students in special education in the two year impact (see Appendix B Table B1). Results for that subgroup are not reported in the two-year impact results as they do not qualify as baseline equivalent even with statistical adjustment.

Table B1

Study qualification for WWC baseline equivalence standards, by analysis and subgroup*

	Within-year CEM	Summer slide	Two-year impact
All students	Meets	Meets	Meets
Grades	Meets	Meets	Meets
Placement levels	Meets	Meets	Meets
Female	Meets	Meets	Meets
Male	Meets	Meets	Meets
Black or Latino	Meets	Meets w/adjustment	Meets w/adjustment
MLL	Meets w/adjustment	Meets w/adjustment	Meets w/adjustment
Special education	Meets w/adjustment	Meets w/adjustment	Does not meet
At-risk	Meets	Meets w/adjustment	Meets

*Baseline differences <.05 of a standard deviation satisfy baseline equivalence without adjustment. Differences <.25 of a standard deviation satisfy baseline equivalent with statistical adjustment.
**See Appendix Tables A7, A22, and A25 for baseline equivalence data.

WWC Essa Tier 2 designation requires a strong quasi-experimental research design that would qualify for Meets WWC Standards with Reservations. In addition, an ESSA Tier 2 rating requires a minimum of 350 students. All studies in this analysis have sample sizes exceeding 350 (see Appendix B, Table B2 for sample sizes for each analysis.) In addition, the study must have been conducted in more than one school. This study spans 54 schools.

Table B2

Sample size of all DCPS analyses

	Treatment sample	Control sample	Total sample
Within-year CEM	990	990	1980
Within-year fixed effects	--	--	1091
Summer slide	745	745	1490
Two-year impact	402	402	804

*In order to qualify for ESSA Tier 2, a study must include at least 350 participants.

Finally, findings must be statistically significant and there can be “no strong negative findings from experimental or quasi-experimental studies” (Regional Educational Laboratory at American Institutes for Research, 2019, p. 2). Results from this study show statistically significant positive impacts from the implementation of Zearn Math. There have been no strong negative findings from other experimental or quasi-experimental studies, while there have been statistically significant positive findings from other QED Zearn studies (see 2022a, 2022b, 2022c; Szatrowski, 2022a, 2022b, 2022c).