The Impact of Zearn Math on Student Achievement in Virginia

A Matched-Comparison Study

December 2025

Kevin Huang, Noman Khanani, and Kirk Walters

This study examined the effects of consistent usage of Zearn Math, a digital learning platform, on math performance and growth among Virginia's grades 4–6 students during the 2024–25 school year. Using quasi-experimental matching methods and statewide administrative data, WestEd analyzed outcomes for approximately 7,500 consistent Zearn users and matched peers with low or no Zearn usage.

- Across all analyses, in total and by subgroup, **consistent Zearn usage was associated with statistically significant**, **positive gains on state math assessment scores**.
- In particular, **students using Zearn with fidelity (3+ lessons per week) outperformed their matched peers** by 0.34 standard deviations—equivalent to a 13-percentile-point increase.
- **Fidelity users were also 18 percentage points more likely to demonstrate growth over time** compared with nonusers (an effect size of 0.49).

Overview

High-quality instructional materials and targeted digital learning supports play a central role in accelerating grade-level math learning, especially as students continue to recover from pandemic-era learning disruptions. Virginia has increasingly invested in such resources to strengthen core instruction and provide additional, personalized practice for students. This report summarizes a study by WestEd examining the impact of one widely used tool, Zearn Math, on grades 4–6 students' academic achievement and growth.

Zearn is a nonprofit educational organization that created Zearn Math, an online K–8 math learning platform. Teachers and tutors using Zearn Math assign students lessons to complete on the platform. Upon logging in, students engage in guided lesson activities designed to enhance their understanding of mathematical concepts and procedures. All Zearn Math lessons include fluency activities, videos led by onscreen teachers with pause points to solve math problems, and a closing mastery-based quiz on which students must score 100 percent in order to advance to the next lesson. Lessons cover every concept of K–8 math and take approximately 30 minutes to complete.

To accelerate math learning for Virginia students, the Virginia Math High Impact Statewide Tutoring Program offers school divisions in the Commonwealth access to Zearn Math.

What the Study Examined

This study explored the impact of using Zearn consistently—completing one or more lessons per week, on average—during the 2024–25 school year for students across Virginia public schools in grades 4 through 6. Specifically, the study focused on two key research questions:

1. What is the effect of consistent Zearn use on achievement and growth for students in grades 4 to 6 compared to nonusers? To what extent does the effect differ by usage dosage?
2. What is the effect of Zearn use *with fidelity* (3+ lessons per week) on achievement and growth, by subgroup, for students in grades 4 to 6 compared to nonusers?

Data Sources

WestEd received deidentified data from the Virginia Department of Education that included results from the state's annual end-of-year Standards of Learning (SOL) math and reading assessments for grades 3–6 from both the 2023–24 and 2024–25 school years, along with SOL Growth Assessments in math and reading for the same years. The data set also included progress indicators derived from the Growth Assessment, student demographic and attendance records, and Zearn usage data reported as the total number of on-or-above grade-level lessons completed in 2024–25.

Because most Zearn users with data from prior state assessment scores in 2024–25 were enrolled in grades 4–6, the impact analysis focused on students in these three grade levels. To examine patterns of usage, researchers grouped students according to the total number of grade-level lessons they completed over the course of the year:

- **Level 1 users:** students who completed between 30 and 59 lessons annually—an average of about 1 to 1.9 lessons per week

- **Level 2 users:** students who completed between 60 and 89 lessons annually, averaging roughly 2 to 2.9 lessons per week
- **Level 3 users:** students who completed 90 or more lessons over the year, corresponding to an average of at least 3 lessons per week, which reflects Zearn usage at fidelity (the study also refers to this group as *fidelity users*)
- **Nonusers:** students who completed fewer than 30 lessons during the year (fewer than 1 lesson per week, on average), as well those who did not complete any Zearn lessons

Study Sample

After cleaning the data; removing duplicate records; and restricting the sample to those who were not missing prior SOL scores, outcome, and demographic data, the final analytic sample comprised 7,514 consistent Zearn users (i.e., those completing an average of one or more lessons per week) and a matched group of non–Zearn users drawn from a pool of 148,256 students. Among consistent users, 4,912 were classified as Level 1 users, 1,585 as Level 2 users, and 1,015 as Level 3 users (fidelity users). To establish baseline equivalence, consistent users were matched to nonusers based on baseline assessment scores and demographic characteristics. Tables A1–A4 in the Appendix provide additional details on the sample.

Methodology

To assess whether Zearn usage was associated with stronger math outcomes, the analysis used a treatment-on-the-treated, quasi-experimental design to compare grades 4–6 students who regularly used Zearn with a similar group of students who did not use the program or used it inconsistently (referred to as *nonusers*).

Research Question 1

Given preexisting differences in prior achievement and demographics between the groups, for Research Question 1, each Zearn user was matched to a nonuser using their baseline test scores, grade level, and background characteristics (i.e., gender, race/ethnicity, English Learner status, economic disadvantage, disability status, and attendance patterns). This matching process helps ensure that any differences between the two groups are more likely due to Zearn usage rather than differences in the characteristics of the students being compared.

Research Question 2

For Research Question 2, which focused on subgroups of students using Zearn with fidelity, each fidelity user was matched to three nonusers instead of one. This approach increases statistical power because sample sizes within subgroups were often small.

Measuring Student Achievement and Growth

Zearn's impact was evaluated in two primary ways. First, to measure impact on student achievement, end-of-year math SOL assessment scores were converted to a common scale using *z*-scores to support comparisons across grades. Effect sizes (Hedges' *g*) summarized the magnitude of these differences (Hedges, 1982). Second, to measure impact on student growth, the state's progress indicator was used to determine whether students moved up, stayed the same, or moved down a performance level from one year to the next. A simplified growth measure classified students as having grown if they increased at least one performance level. After matched groups were established, statistical models estimated differences in outcomes between Zearn users and their matched peers on both measures.

Subgroup results were reported only when at least 100 students within a particular group used Zearn with fidelity. As a result, analyses were not

conducted for the following subgroups: prior math achievement (few students scored at the Basic or Below Basic levels), American Indian or Alaska Native, Native Hawaiian or Pacific Islander, multiple races, students with disabilities, and students who were chronically absent. For grade-by-grade analyses, original test scores (rather than *z*-scores) were used to enhance interpretability.

Additional methodological details are described in the Appendix.

Study Findings

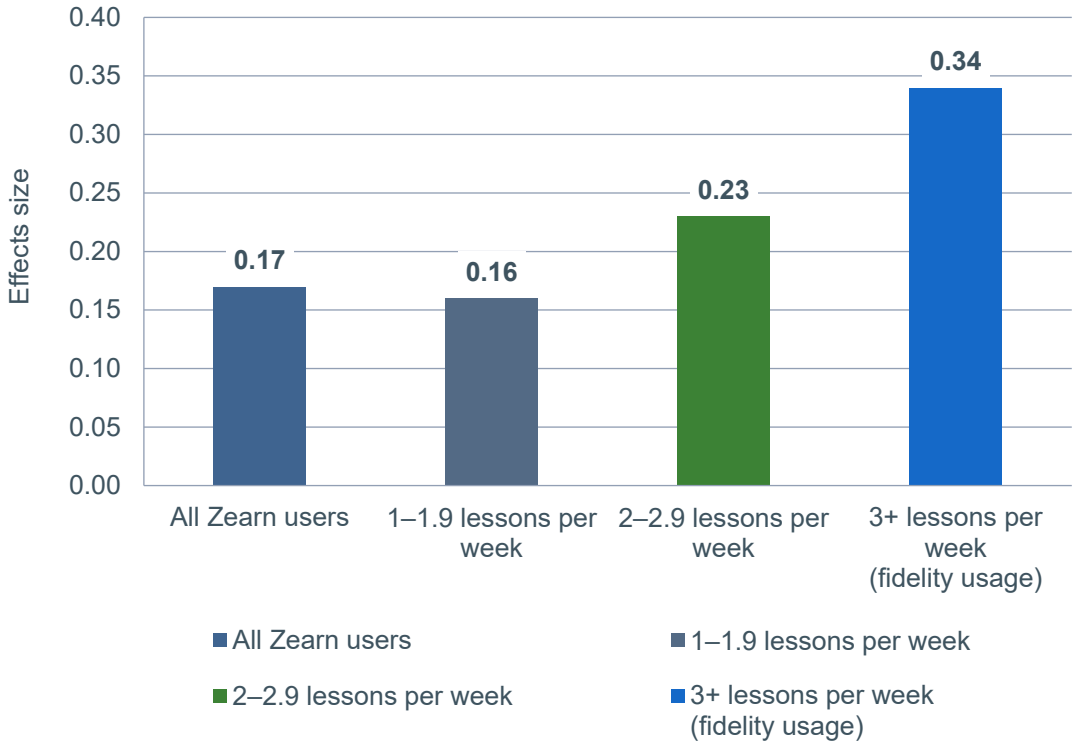
This section describes findings related to the overall impacts of Zearn Math on students' standardized scores (*z*-scores) and growth (based on the state's progress indicator), followed by the impacts on various subgroups of fidelity users.

Overall Impacts of Zearn Math

Consistent Zearn usage across all usage levels was associated with positive impacts on math achievement in grades 4–6. Impacts were largest for fidelity users.

The overall impact of Zearn usage, comparing all consistent Zearn users with nonusers, was 0.17 standard deviations (Figure 1). Impacts were greater with larger dosage, and Zearn had the most notable positive impact on fidelity users. The effect size of 0.34 standard deviations for fidelity users is equivalent to a student moving from the 50th to the 63rd percentile, a 13-percentile-point increase.

Figure 1. Impact of Zearn Usage on Student Achievement on State Standardized Math Assessment, in Total and by Usage Level

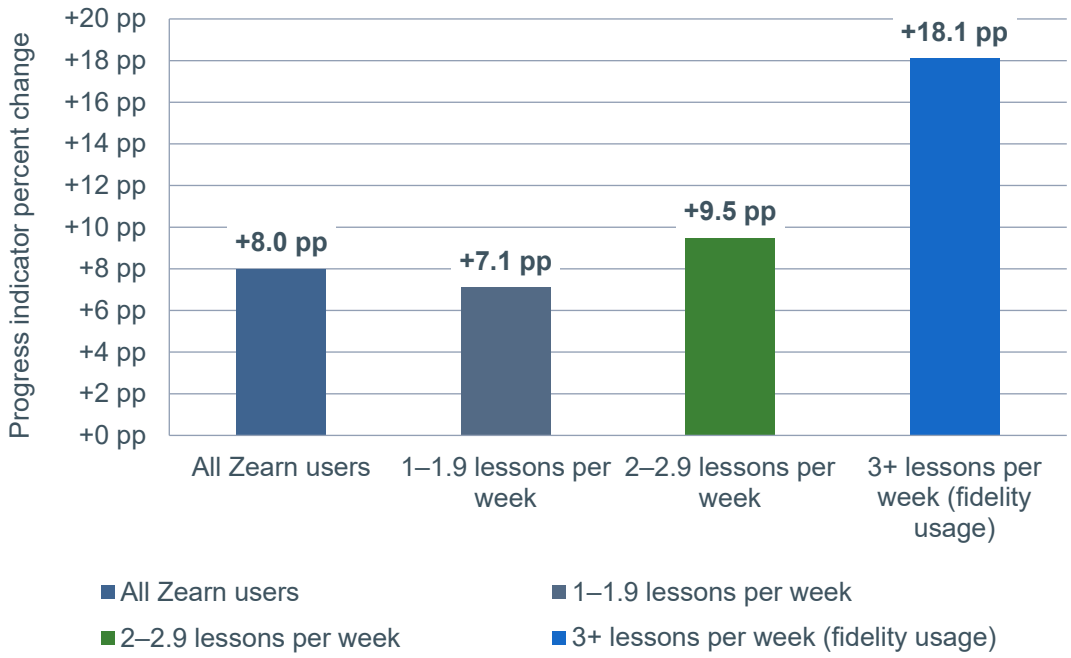


Note. Effect sizes are Hedges' *g*. The matched groups based on the propensity score matching approach were used to estimate the impact. All results are statistically significant at the .001 level.

Consistent Zearn usage across all usage levels was associated with positive growth on the state's progress indicator.

On average, consistent Zearn users were 8 percentage points more likely to improve at least one progress indicator level compared to their matched peers (Figure 2). Impacts on growth were highest for fidelity users, who were 18 percentage points more likely to improve at least one level than their matched peers.

Figure 2. Impact of Zearn Usage on Student Growth, in Total and by Usage Level



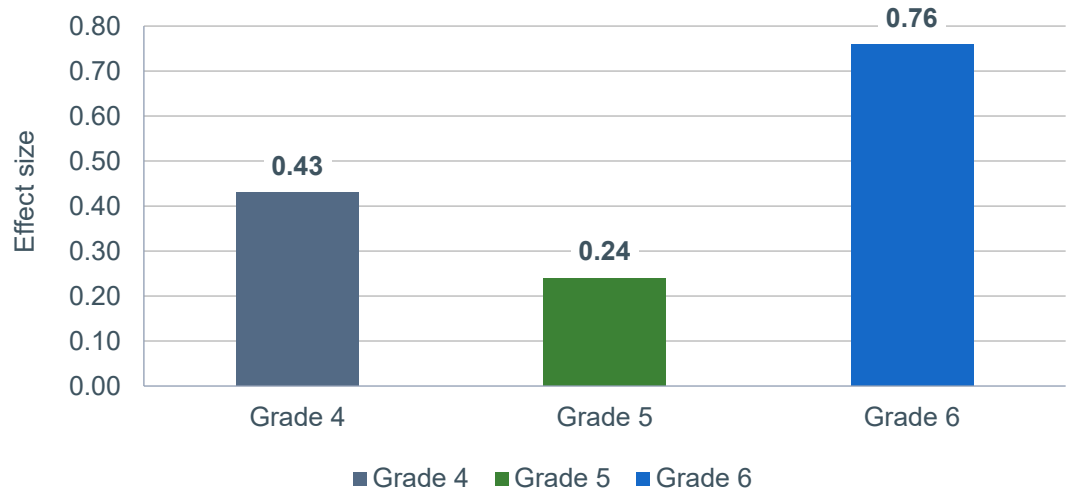
Note. Effect sizes (Cox index) are 0.23, 0.21, 0.28, and 0.49, respectively. All results are statistically significant at the .001 level.

Impacts of Zearn Use With Fidelity on Subgroups of Students

When used with fidelity, Zearn had strong positive impacts on achievement for all subgroups.

For students who completed three or more lessons per week, statistically significant positive impacts were found across all subgroups on student achievement. Hedges' *g* ranged from 0.24to 0.56 standard deviations (see Figures 3–5 below). As shown in Figure 3, impacts were particularly large in grades 4 and 6 (effect size of 0.43 and 0.76, respectively).

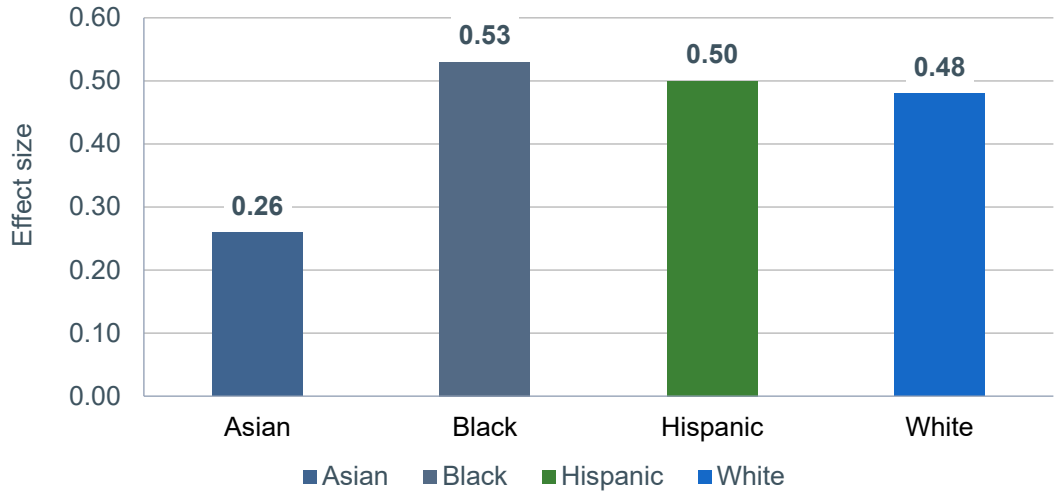
Figure 3. Impact of Zearn Fidelity Usage on Student Achievement, by Grade Level



Note. Effect sizes are Hedges' *g*. All results are statistically significant at the .001 level. The impact estimate for grade 6 students may not be reliable due to a small sample size. Effects in terms of scale scores are 21.5, 14.3, and 35.3 for grades 4, 5, and 6, respectively.

For Black, Hispanic, and White students, impacts were similar at approximately 0.50 standard deviations (Figure 4). Impacts were smaller, but still meaningfully large, for Asian students.

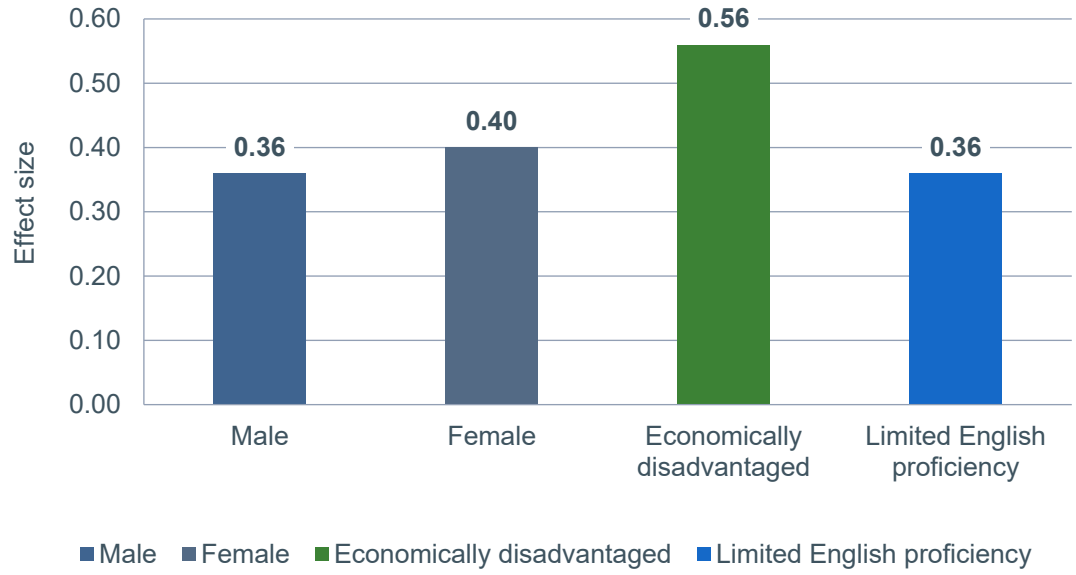
Figure 4. Impact of Zearn Fidelity Usage on Student Achievement, by Race/Ethnicity



Note. Effect sizes are Hedges' *g*. All results are statistically significant at the .001 level.

Across all demographic groups explored, impacts were also consistently large. For economically disadvantaged students, in particular, impacts were largest (0.56 standard deviations), as shown in Figure 5.

Figure 5. Impact of Zearn Fidelity Usage on Student Achievement, by Demographic Group

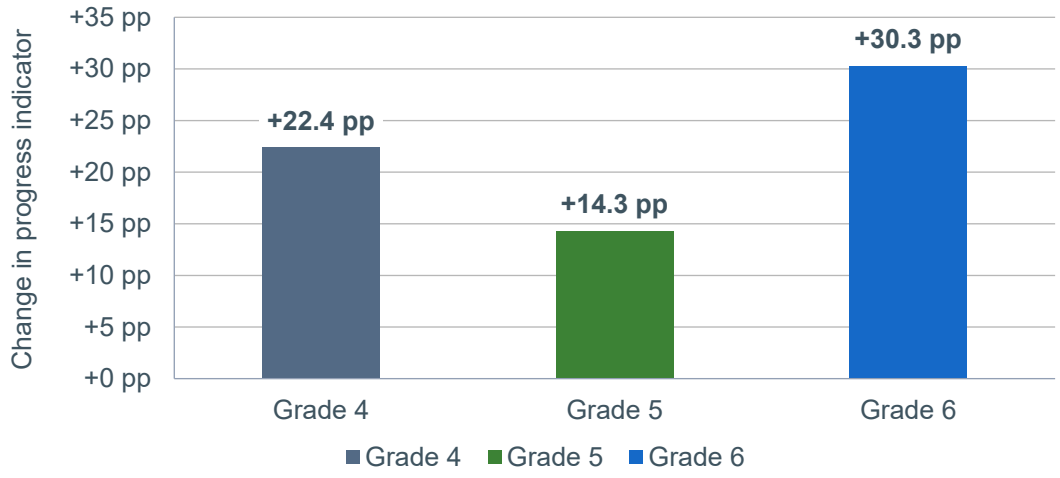


Note. Effect sizes are Hedges' *g*. All results are statistically significant at the .001 level.

When used with fidelity, Zearn had strong positive impacts on growth for all subgroups.

For students who completed three or more lessons per week, statistically significant positive impacts were found across all subgroups. The effect sizes across all subgroups, based on the Cox index (Sánchez-Meca et al., 2003), ranged from 0.39 to 0.86 standard deviations (Figures 6–8). Like the results for achievement, impacts on growth were particularly large in grades 4 and 6 (effect size of 0.59 and 0.86, respectively) (Figure 6).

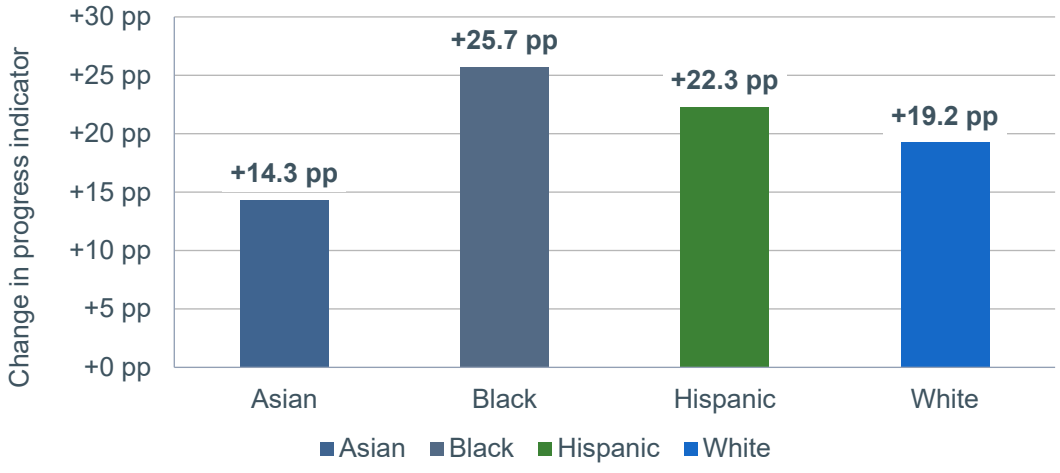
Figure 6. Impact of Zearn Fidelity Usage on Student Growth, by Grade Level



Note. Effect size (Cox index) is 0.59, 0.40, and 0.86, respectively. All results are statistically significant at the .001 level. The impact estimate for grade 6 students may not be reliable due to a small sample size.

Black, Hispanic, and White fidelity users were approximately 20 percentage points more likely than nonusers of the same racial groups to experience growth in proficiency level (Figure 7).

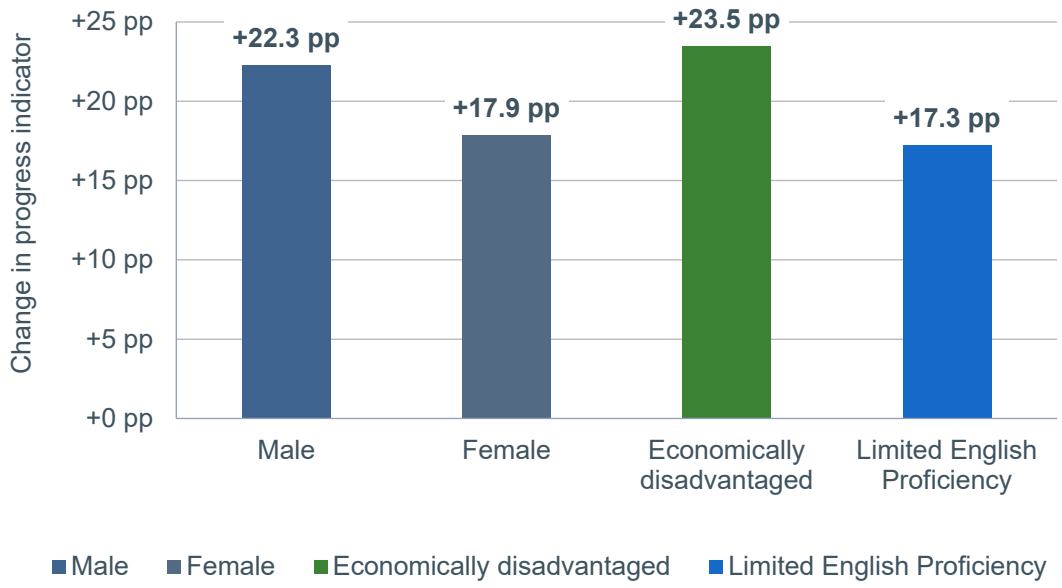
Figure 7. Impact of Zearn Fidelity Usage on Student Growth, by Race/Ethnicity



Note. Effect size (Cox index) is 0.39, 0.71, 0.58, and 0.53, respectively. All results are statistically significant at the .001 level.

Across all demographic groups explored, impacts on growth were consistently large, ranging from 17 to 24 percentage points (Figure 8).

Figure 8. Impact of Zearn Fidelity Usage on Student Growth, by Demographic Group



Note. Effect size (Cox index) is 0.62, 0.48, 0.64, and 0.45, respectively. All results are statistically significant at the .001 level.

Implications

The findings suggest that consistent implementation of Zearn Math in grades 4–6—particularly at the fidelity level (90+ lessons annually)—can meaningfully improve math achievement and growth. All effect sizes reported in this study can be interpreted as large relative to the typical impacts of field-based education interventions of a similar cost (Kraft, 2020). Districts implementing Zearn should prioritize sustained usage, ensure alignment with core instruction, and support teachers in integrating digital lessons into their classroom practice.

These findings are consistent with other recent studies of the Zearn Math platform using similar methodologies to explore effects on end-of-year math

scores (Szatrowski, 2022; Szatrowski et al., 2025). For example, one recent evaluation in Louisiana found an effect size of 0.20 standard deviations on end-of-year standardized math scores for students in grades 4 through 8 (Storey & Neitzel, 2025).

In addition, the study's results underscore the importance of monitoring usage patterns throughout the year to identify classrooms or schools that may require additional support to reach fidelity. Because fidelity users demonstrated the strongest gains, districts may benefit from establishing clear expectations for weekly lesson completion and providing routines that help teachers and students meet those targets. Furthermore, districts and schools should consider whether and how Zearn usage differs by how it is implemented in core and small-group settings. Finally, given that usage varied across student groups, districts should examine participation data to ensure equitable access and to address barriers that may prevent certain students from engaging consistently with the program.

Limitations

Although matching methods help control for baseline differences between Zearn users and their matched peers, unmeasured or unobservable factors may still influence outcomes. Additionally, in this study, only approximately 7,500 students out of the total grades 4–6 student population were consistent Zearn users, making it unclear how well the study sample reflects the broader student population in these grades. Consistent users also differed from nonusers in multiple ways: They were more likely to be Asian and higher achieving and less likely to be economically disadvantaged and have a disability. These differences may limit the extent to which findings generalize to all students. Relatedly, subgroup analyses of fidelity users with smaller sample sizes (for example, n = 113 for grade 6) should be interpreted cautiously. Finally, the study reflects only 1 year of implementation (2024–25). Results may shift as districts gain experience and as usage patterns evolve over time.

References

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*(2), 490.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Education Researcher*, *49*(4), 241–253.

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*(4), 448.

Storey, N. S., & Neitzel, A. J. (2025). *Evaluation of Zearn Supplemental with dedicated implementation support*. Center for Research and Reform in Education, The Johns Hopkins University.

Szatrowski, A. (2022). *Efficacy analysis of Zearn Math: Findings from a large Midwestern urban district*.

Szatrowski, A., Riolo, M., & Rickel, J. (2025). *Impact study methodology: Quasi-experimental treatment-on-the-treated impact research*.

Appendix. Study Sample and Methodology

This appendix provides a more detailed account of the study's sample and methodology.

Study Sample

The study sample was pooled from 155,770 grades 4–6 students in Virginia with nonmissing enrollment, baseline assessment, and demographic information. Of these, 7,514 were consistent Zearn users (i.e., those completing an average of one or more lessons per week). A breakdown of Zearn usage for the entire sample is provided below in Table A1.

Table A1. Sample of Eligible Students for Matching

Zearn usage level	Frequency	Percent
Non–Zearn users	148,256	95.2
Zearn usage Level 1	4,914	3.2
Zearn usage Level 2	1,585	1.0
Zearn usage Level 3 (fidelity usage)	1,015	0.7
Total	**155,770**	**100.0**

Table A2 summarizes the characteristics of each subsample. Consistent Zearn users had higher prior-year Standards of Learning (SOL) math proficiency rates than nonusers, with fidelity users (Level 3 users) showing the highest rates. Fidelity users were also more often Asian, male, and students in grade 4 or 5 and were less likely to be economically disadvantaged or identified with a disability.

Table A2. Characteristics of Pre-Matched Study Sample, by Zearn Usage

Characteristic	Nonusers	Level 1 users	Level 2 users	Level 3 users
Proficient on prior-year SOL math proficiency	69%	82%***	88%***	92%***
Grade 4	35%	46%***	41%***	42%***
Grade 5	36%	31%***	40%***	47%***
Grade 6	29%	28%***	19%***	11%***
Asian	6%	8%***	11%***	20%***
Black	23%	22%	21%**	16%***
Hispanic	18%	16%***	15%***	14%***
White	45%	47%***	44%	41%**
Male	50%	55%***	57%***	57%***
Female	50%	45%***	42%***	43%***
Economically disadvantaged	49%	47%**	42%***	32%***
Limited English proficiency	16%	15%**	16%	20%***
Students with a disability	15%	8%***	7%***	6%***

Note. Asterisks indicate whether the value in a cell is statistically different from the "Nonuser" cell in the same row. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$

As indicated earlier, the subgroup analysis for RQ2 was conducted solely based on the fidelity users (90+ lessons annually) and their matched comparison group of nonusers. In addition, to obtain more robust estimates, subgroup impacts were estimated only for groups with at least 100 fidelity

users (see Table A3 for sample size by subgroup for fidelity users). Because of that, the analysis was not conducted for the following subgroups: prior math achievement (a small number of students fell at the Basic and Below Basic levels); American Indian or Alaska Native, Hawaiian, and multiple races; students with a disability; and chronically absent students.

Table A3. Subgroup Sample Size of Fidelity Users

Characteristic	Sample size
Pass/proficient on prior year SOL math	933
Fail/Basic/Below Basic on prior-year SOL math	82
Grade 4	426
Grade 5	476
Grade 6	113
Asian	203
Black	165
Hispanic	147
White	419
Multirace	79
Male	578
Female	437
Economically disadvantaged	323
Limited English proficiency	204
Students with a disability	56
Chronically absent	39

Note. Subgroup analyses were not run on groups in which the sample size fell under 100 students.

Methodology

The analysis was based on a treatment-on-the-treated quasi-experimental design (QED) in which our WestEd research team compared students in grades 4–6 who consistently completed Zearn lessons to a matched group of nonusers. Zearn users, in total and by usage level, were matched to nonusers using propensity score matching (PSM) and coarsened exact matching (CEM). The CEM was intended to be used to assess the robustness of results based on PSM. For Research Question 1 (RQ1), comparing consistent Zearn users with the matched group, we report findings based on the PSM one-to-one match. For Research Question 2 (RQ2), comparing fidelity users with their matched peers by subgroup, the findings are based on the PSM one-to-three match to increase the statistical power and the precision of the estimates because the sample size for subgroups tended to be smaller than the full sample.

Matching variables included prior-year math and reading scores, student demographics, and grade level. Student demographics included gender, ethnicity, economically disadvantaged status, limited English proficiency indicator, disability indicator, and chronic absenteeism indicator. Table A4 reflects that the matched samples achieved strong balance, with all differences between the two groups being under 0.05 standard deviations.

Table A4. Effect Size Differences Between Matched Samples

Characteristic	Standardized effect size difference
Prior-year SOL math proficiency	0.00
Prior-year SOL reading proficiency	0.01
Grade 4	0.00
Grade 5	0.00

Characteristic	Standardized effect size difference
Asian	0.00
Black	0.02
Hispanic	0.02
White	-0.03
Male	0.02
Economically disadvantaged	0.02
Limited English proficiency	0.02
Students with a disability	0.02

Note. Effect size measured using Cohen's *d*. Grade 6 is missing because one grade had to serve as the reference group in the matching model.

Because traditional SOL test scores are not vertically scaled, we converted the scale scores to *z*-scores using the mean and standard deviation for each grade in reading and math. The resulting *z*-scores enabled us to conduct the analysis across grade levels.

In addition to the scale scores, the impact of Zearn Math was estimated based on the progress indicator, which is determined by a vertical scale resulting from the Growth Assessments. The vertical scale has eight levels, with Level I the lowest and Level VIII the highest. The following scenario uses the math assessment as an example for how the vertical scale works:

- A 3rd grade student with a score in the range of 1367–1400 will be placed in Level IV.
- When this student moves up to grade 4 the next year, if their score is in the 1431–1510 range, they will be placed into Level V.

- On the other hand, if their score is in the 1397–1430 range, they will be placed into Level IV, the same level they were in the previous year.

In our analysis, we created a "growth" indicator and used it as one of the outcome measures. For students moving up to the next level, the growth indicator equals 1; for students remaining at the same level or moving down to a lower level, it equals 0.

After identifying the matched groups, a series of regression models were used to estimate treatment impacts based on scale scores (or *z*-scores), with the student as the unit of analysis. Logistic regression models were used because the outcome variable was a binary indicator of growth. The student demographics used for matching were also included as covariates in the impact models. Effect size (Hedges' *g* for the impacts based on scale scores or the Cox index for the growth indicator) was computed for each impact estimate.

Finally, for the by-grade analysis, to help with interpretation, the scale scores (not the converted *z*-scores) were also used to examine the impacts.

Suggested citation: Huang, K., Khanani, N., & Walters, K. (2025). *The impact of Zearn Math on student achievement in Virginia: A matched comparison study*. WestEd.

WestEd is a nonpartisan, nonprofit organization that aims to improve the lives of children and adults at all ages of learning and development. We do this by addressing challenges in education and human development, increasing opportunity, and helping build communities where all can thrive. WestEd staff conduct and apply research, provide technical assistance, and support professional learning. We work with early learning educators, classroom teachers, local and state leaders, and policymakers at all levels of government.

For more information, visit WestEd.org. For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the E-Bulletin, our semimonthly e-newsletter, at WestEd.org/subscribe.