TECHNICAL APPENDIX

Efficacy analysis of Zearn Math in Tulsa

October 2022

Alisa Szatrowski, Ph.D.
Zearn

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 to the Eureka Math curriculum. The goal of this study was to isolate the impact of Zearn Math on student achievement, through quasi-experimental matching methods that facilitate causal inference.

This efficacy analysis was conducted in Tulsa, Oklahoma, a district with over 30,000 students, 14,000 of whom are in grades K-6. The student body is 79% economically disadvantaged, 38% multilingual learners, 14% students with disabilities, 11% gifted and talented and 60% Black and Latino students. (Tulsa Public Schools, 2022).

In grades K-6, there were 313 consistent Zearn Math users, those who completed three or more digital lessons per week, i.e., 90 or more digital lessons per year, who could be matched to assessment data from the 2021-2022 school year (Descriptive Table 2 contains a breakdown of the sample composition).[1]

Matching methodology

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 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 2021-2022 school year. The outcome under investigation was the average treatment effect 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 2021-2022 school year, operationalized as an average of three or more digital lessons per week; 90 or more digital lessons per year. The control group was selected from students in Tulsa with little to no Zearn Math usage, operationalized as an average of less than one digital lesson per week; fewer than 30 digital lessons per year.[2]

[1] The sample population of fidelity users differs from Tulsa's population, having a proportionally smaller population of students from disadvantaged backgrounds and fewer students starting at the lowest levels of achievement. The implications of this difference are discussed in the limitations section.

[2] 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.

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 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 was done to isolate the difference in outcomes to the intervention itself, and not to potential 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.

Using CEM, treatment students were put into matching strata with control students that were in the same grade and within five national percentile points on fall Math and English Language Arts (ELA) scores. 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, multilingual learner status, gifted and talented status and economic disadvantage.

This optimal matching method utilized Bertsekas' auction algorithm to produce combinatorial optimization such that treatment individuals were matched to others closest to them in the control pool and, 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; Rosenbaum, 2020).[3]

If a treatment student had no match within their grade 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.

[3] 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.

Test administration dates were provided for each student and the window between fall and spring assessments varied by up to five weeks between students. Usage rates were calculated by dividing the number of lessons a student completed between assessments by four less than the number of weeks between assessments to accommodate for holidays and school closures:

FIGURE 1

$$usage\ rate = \frac{Zearn\ Math\ lessons\ completed\ between\ fall\ \&\ spring\ assessments}{(Weeks\ between\ fall\ \&\ spring\ assessments - 4)}$$

The average window between assessments for students in the treatment and control groups were within .3 of a week of each other (See Descriptive Table 1 for more details).

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

Out of Tulsa's 313 consistent Zearn Math users, all but 18 were matched. Treatment and control groups differed by an average of 1.7 demographic factors and 2.1 percentile points in both math and ELA starting proficiencies. The 18 consistent Zearn Math users excluded from the study, due to lack of match, did not concentrate in any demographic category that would bias the sample (See Descriptive Table 1 for a breakdown of sample demographics).

The sample for this analysis is not entirely representative of the overall composition of Tulsa. In comparison to the district, a smaller proportion of the sample are economically disadvantaged, multilingual learners, students with disabilities or students of color. The sample also starts at higher levels of achievement. While two-thirds of K-6 students in the Tulsa are "Below Average" according to MAP proficiency categories, only 35% of the sample starts at this proficiency level (See Descriptive Table 2).

Matching methods attempt to compensate for this discrepancy by creating a comparison group of low- or non-using students that is demographically similar to the population of users. See the Conclusion and Limitations section for a discussion of the possible implications of this disparity.

Analysis

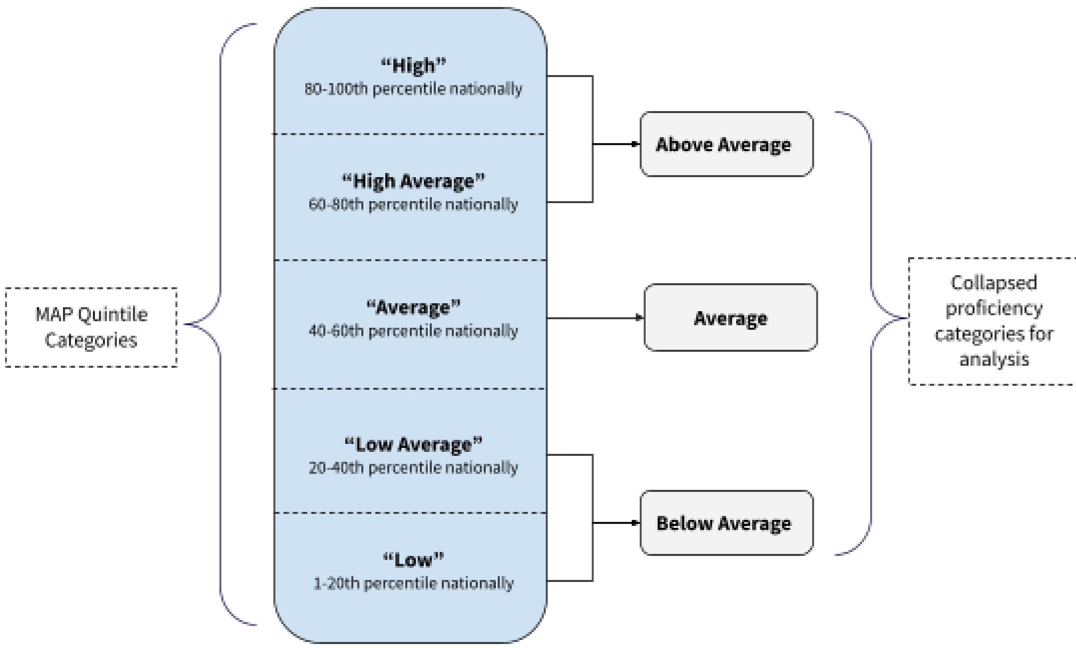
Once consistent Zearn Math users were matched to a similar group of low- or non-users, a difference of means analysis was conducted to quantify the impact of Zearn Math on student achievement. Means were calculated for treatment and control groups overall, as well as for groups disaggregated by starting math proficiency and demographic factors.

Academic growth was measured as the change in MAP scores between the fall 2021 and spring 2022 assessment administration. MAP is a mean-centered test with no absolute bar for proficiency. Tulsa defined proficiency as at or above the 50th percentile nationally and set annual district goals around students meeting this benchmark. Throughout this report, references to percent proficient will be based on the district definition of 50th percentile and above. Outcomes are reported in terms of change in national percentile points, change in percent proficient, change in MAP quintile (as measured nationally), and change in scale score.

The first analysis examined changes in national percentile points between fall 2021 and spring 2022, disaggregated by the average starting score for three collapsed proficiency levels. MAP provides quintiles labeled: “Low”, “Low Average”, “Average”, “High Average” and “High” representing 0-20th, 20-40th, 40-60th, 60-80th and 80-100th percentile respectively. In this analysis, “Below Average” includes all students in the 0-40th percentile, “Average” includes all students in the 40-60th percentile, and “Above Average” includes all students in the 60-100th percentile (See Figure 2).

FIGURE 2

Diagram of combined MAP quintiles



Difference in means t-tests were run on the average academic gains of consistent users versus the average academic gains 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:

FIGURE 3

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

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

FIGURE 4

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

On average, consistent Zearn Math users gained 11.8 percentile points whereas matched low- or non-users lost 2.0 national percentile points between fall and spring 2021-2022, a difference of 13.8 percentile points (effect size=.50). Gains were highest among consistent Zearn Math users who started the year in "Below Average". These students gained 22.3 percentile points while low- or non-users gained only 2.6, a difference of 19.6 percentile points (effect size=.62) (See Results Table 1). The outsized impact of Zearn Math use among students starting below proficiency has been previously reported by Zearn (2022a; 2022b) (See Results Table 4 for findings from the difference in means analysis).

RESULTS TABLE 1

Students with consistent Zearn Math usage grew significantly more than matched students with low or no Zearn Math usage

Growth in percentile points and years for Consistent Zearn users *(Treatment)* vs. Low- or Non-users *(Control)*, by starting proficiency

	Below average *(Quintiles 1 & 2)*	Average *(Quintile 3)*	Above average (Quintiles 4 & 5)
Treatment growth in percentile points	22.3	12.3	1.6
Treatment growth in years	1.5	1.6	1.4
Control growth in percentile points	2.6	-4.2	-4.9
Control growth in years	0.9	0.8	0.9
Growth difference in percentile points	19.640	16.511	6.511
Growth difference in years	0.6	0.8	0.5

MAP publishes expected growth ranges, for each grade band, between the fall and spring assessments, based on national norms (Thum & Kehfeld, 2020b). In this analysis, a year's growth was operationalized as MAP's expected change in score for each grade. The average of actual points earned by grade, for users and matched non- or low-users, was divided by the year's growth expectation to translate changes in scale score to years of growth. Then an average was taken of years growth by grade. Whereas non-users grew less than one-year, on average, consistent Zearn Math users grew 1.5 years during SY 2021-2022 (See Results Table 1).

In addition to absolute growth, mobility models compared the change in proficiency level for treatment and control students based on starting MAP quintiles. Across all quintiles, consistent Zearn Math users maintained or increased their proficiency levels at higher rates than non- or low-users.

Notably, for students starting at "Low", 2.6x as many students who used Zearn Math consistently, improved their achievement level, relative to students with little to no Zearn Math usage. Results Table 2 illustrates the mobility between achievement levels from the fall to spring, for each group.

RESULTS TABLE 2

Across all MAP achievement quintiles, students with consistent Zearn Math usage were more likely improve their achievement level

Change in proficiency level, between Fall 2021 and Spring 2022 among consistent Zearn Math users

		Spring 2022 Proficiency				
		Low	Low avg.	Average	High avg.	High
Fall 2021 Proficiency	Low	41%	28%	13%	10%	8%
	Low average	14%	20%	17%	31%	18%
	Average	1%	10%	33%	36%	20%
	High average	0%	0%	16%	38%	46%
	High	0%	0%	4%	16%	80%

Change in proficiency level, between Fall 2021 and Spring 2022 among low- or non- Zearn Math users

		Spring 2022 Proficiency				
		Low	Low avg.	Average	High avg.	High
Fall 2021 Proficiency	Low	78%	10%	8%	3%	3%
	Low average	40%	31%	15%	13%	1%
	Average	12%	27%	35%	18%	8%
	High average	2%	8%	23%	44%	24%
	High	0%	0%	4%	26%	70%

Subgroup Analysis

In addition to capturing changes in student achievement across all users, the analysis zoomed in to look at 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 3 demographic characteristics, subgroups did not always align on starting proficiency. Therefore differences in proficiency by demographic subgroup were reported as difference-in-difference rather than as actual percentile scores as depicted in Results Table 3 (See Descriptive Table 3 for a breakdown of starting and ending means, by subgroup).[4]

Across all subgroups, consistent Zearn Math users saw gains in percent proficient while non- or low-users saw drops in percent proficient, on average. Notably, Black and Latino students saw both

[4] The subgroup of students receiving Special Education services is not in any of the following tables because of the low-N of the sample.

the largest gains in students proficient in math (an additional 27%) among consistent Zearn Math users and the largest losses in students proficient in math (13% fewer) among non- or low-users.[5] For students of color, Zearn Math use made a dramatic impact on academic outcomes (See Results Tables 3 and 5 for more details).

RESULTS TABLE 3

Across all subgroups, there was an increase in students meeting proficiency amongst students consistently using Zearn Math

Change in percent proficient among students with consistent Zearn Math usage and those with little or no Zearn Math usage

	Consistent Users *(Treatment)*	Low- or Non-Users *(Control)*
All Students	20%	-4%
Female	18%	-5%
Male	21%	-3%
Black & Latino	27%	-13%
FRL	19%	-10%
ELL	13%	-9%
Gifted	9%	-7%

In addition to providing absolute scale and percentile scores, MAP provides an indicator of how much a student has grown compared to matched students who start at the same score or percentile for their grade level. Based on NWEA's MAP Norms Study, MAP assigns each student a projected growth expectation at the start of a growth period. At the end of a growth period, a student's "growth index" is calculated as the difference between their actual and projected growth. While this raw growth index cannot be compared across students and grades, a conditional growth index (CGI) is generated to compare a student's growth relative to the growth of a national sample of similarly matched peers. A value of 0 would indicate that a student has experienced mean growth while values above zero indicate growth exceeded projections and values below zero indicate growth fell below projections (Thum & Kuhfeld, 2020a).

The CGI is expressed in terms of standard deviations above and below the mean growth, calculated as:

FIGURE 5

$$CGI = \frac{Fall\ to\ spring\ observed\ growth - Fall\ to\ spring\ projected\ growth}{Standard\ deviation\ of\ growth}$$

[5] For each subgroup in treatment and control, percent proficient in the fall was subtracted from percent proficient in the spring. This change is depicted in Results Table 3. If the percent proficient within a subgroup was the same in the fall and spring, the change listed in Results Table 3 would be 0.

On average, consistent Zearn Math users showed growth that was .73 of a standard deviation above national norms. Non- and low-users showed growth that was .20 of a standard deviation below national norms. This was a difference of almost a full standard deviation, an effect size of .48.

This trend held across all subgroups, with consistent Zearn Math users significantly exceeding national growth norms and non-users falling below (or at) growth norms, on average (See Results Table 5).

Conclusion & Limitations

This analysis provides promising evidence of Zearn Math's positive impact on student achievement. In addition to positive changes in student performance overall, Black and Latino students who consistently used Zearn Math, as well as students who started below average in fall 2021, 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 at low levels of proficiency, further substantiates findings from efficacy analyses of Zearn Math's impact in other districts (Zearn, 2022a & 2022b).

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. 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 quasi-experimental methods allow researchers to control for observed confounders, a possibility exists that unobserved confounders 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.

This analysis examines 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 better aligned with the interests of this partner, for whom the results can help to encourage more universal fidelity usage of Zearn Math. However, utilizing fidelity usage 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 and those that started at the lowest levels of achievement. Therefore, it is possible that the impact of Zearn Math on a larger portion of students in Tulsa 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 mitigates some of this disparity by controlling for observable factors that impact academic growth such that they are as similar as possible between treatment and control groups.

Because findings from past studies (Zearn, 2022a; Zearn, 2022b) and this analysis show an outsized positive impact of Zearn Math on students starting at the lowest income levels and students from disadvantaged backgrounds, there is reason to believe that the sample underestimates the growth expected from more widespread adoption of Zearn Math. However, it is also possible that users in Tulsa 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 student population within Tulsa. It is possible that there are features specific to Tulsa 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.

Works Cited

Bertsekas, DP. (1981)." A new algorithm for the assignment problem." Math. Prog. 21:152–71

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *Stata Journal*, *9*(4), 524-546.

Iacus, S. M., G. King, and G. Porro. (2011). "Multivariate Matching Methods that are Monotonic Imbalance Bounding." Journal of the American Statistical Association, 106, Pp. 345–361.

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

Rosenbaum, P. R. (2020). Modern Algorithms for Matching in Observational Studies. *Annual Review of Statistics and Its Application*, *7*(1), 143-176. https://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, Pp. 279–306.

Thum, Y. M., & Kuhfeld, M. (2020a). NWEA 2020 MAP Growth Achievement Status and Growth Norms for Students and Schools. NWEA Research Report. Portland, OR: NWEA.

Thum, Y. M., & Kuhfeld, M. (2020b). NWEA 2020 MAP Growth Achievement Status and Growth Norms Tables for Students and Schools. NWEA Research Report. Portland, OR: NWEA.

Tulsa Public Schools. (2022). About Tulsa Public Schools. Retrieved September 12, 2022, from https://www.tulsaschools.org/about

Zearn. (2022a) *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.* https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf

Zearn. (2022b) *Consistent Zearn Usage Dramatically Reduces Learning Loss.* https://webassets.zearn.org/Implementation/Zearn_Impact_for_Students_Below_Grade_Level.pdf

Descriptive Table 1

Breakdown of sample matching characteristics		
	Treatment	**Control**
Total N's	295	295
Pre-scores (Fall assessment scores)		
Math scale score	171.3	171.1
Math percentile	51.5	51.2
ELA scale score	166.6	167.4
ELA percentile	49.6	49.8
Weeks between fall & spring assessments	36.2	35.9
Starting Performance Level		
Below Average	104	108
Average	81	74
Above Average	110	113
Grades		
Kindergarten	58	58
Grade 1	63	63
Grade 2	73	73
Grade 3	57	57
Grade 4	10	10
Grade 5	21	21
Grade 6	13	13
Demographic & academic subgroups		
Female	126	124
Male	169	171
Black & Latino	134	121
Students with disabilities	16	10
Economically disadvantaged	80	92
English language learners	72	68
Gifted	80	71

Descriptive Table 2

Sample composition			
	Treatment	**Control**	**Tulsa**
Demographic & academic subgroups			
Black & Latino	45%	41%	60%
Students with disabilities	5%	3%	14%
Economically disadvantaged	27%	31%	79%
Multilingual learners	24%	23%	38%
Starting Proficiency (Fall 2021)			
Below average	35%	37%	66%
Average	27%	25%	16%
Above average	37%	38%	18%

Descriptive Table 3

Fall and spring performance means by subgroup*

	Treatment fall 2021	Treatment spring 2022	Control fall 2021	Control spring 2022
All Students				
Math scale score (percentile)	171.3 (51.5)	192 (63.3)	171.1 (51.2)	185.4 (49.2)
Starting proficiency				
Below Average	152.5 (22.3)	176.8 (44.6)	153 (22.4)	168.7 (25.1)
Average	171.2 (50.4)	191.8 (62.7)	171.3 (50.6)	184.9 (46.4)
Above Average	189.3 (79.8)	206.6 (81.3)	188.3 (79.2)	201.7 (74.2)
Grades				
Kindergarten	135.6 (44.4)	160.7 (61.8)	135.8 (44.7)	154.8 (46.6)
Grade 1	154.8 (42.6)	178 (55.1)	154.7 (42.3)	172.2 (43.3)
Grade 2	175.6 (56.4)	195.9 (65.9)	175.4 (56.4)	189.8 (52.7)
Grade 3	189.6 (56.1)	207.9 (67.2)	188.7 (55.3)	200.9 (53.5)
Grade 4	211.5 (80.2)	222.4 (75.3)	211.8 (79.8)	221.8 (74.5)
Grade 5	208.4 (51.6)	223.2 (59.8)	208 (51.2)	213.6 (41.7)
Grade 6	215.9 (55.5)	234.3 (74.1)	215.7 (55.1)	219.5 (44.6)

*Top number is scale score, number in parentheses is percentile.

Descriptive Table 3 (cont.)

	Treatment fall 2021	Treatment spring 2022	Control fall 2021	Control spring 2022
Demographic & academic subgroups				
Female	168.5 (50.2)	189.1 (61.1)	168 (50.1)	181.3 (45.9)
Male	173.5 (52.4)	194.2 (64.9)	173.4 (52)	188.4 (51.6)
Black students	169.2 (40.3)	188.9 (53.4)	164.7 (42.6)	177.9 (36.8)
Latino students	173.1 (46.7)	193.8 (59.1)	169.8 (43)	183.8256 (40.6)
Students with disabilities	178 (39.8)	198.6 (56.3)	157.5 (32)	177.8 (41.3
Economically disadvantaged	164.7 (43.7)	183.6 (51.7)	167.2 (44.8)	180 (39.4)
English language learners	167.1 (44)	187.2 (53.8)	163.5 (38.8)	177.8 (36.9)
Gifted	192.9 (75.3)	211.3 (82)	198.9 (79.1)	211.6 (75.8)
*Top number is scale score, number in parentheses is percentile.				

Results Table 4

Comparison of changes in scores and proficiency between consistent Zearn users and low- or non-users					
	Treatment change in mean	**Control change in mean**	**Difference**	**Pooled SD**	**Cohen's d**
All Students					
Math scale score (SS)	20.678	14.325	6.353***	13.234	0.480
Math percentile	11.817	-1.983	13.800***	27.631	0.499
Math CGI	0.727	-0.200	0.928***	1.945	0.477
Math percent proficient	19.66%	-3.73%	23.39%***	56.20%	0.416
Starting proficiency					
Below average SS	24.260	15.741	8.519***	13.206	0.645
Below average percentile	22.269	2.630	19.640***	25.988	0.756
Below average CGI	0.929	-0.282	1.210***	1.822	0.664
Average SS	20.667	13.568	7.099***	10.099	0.703
Average percentile	12.309	-4.203	16.511***	22.668	0.728
Average CGI	0.771	-0.274	1.045***	1.407	0.743
Above average SS	17.300	13.469	3.831***	7.889	0.486
Above average percentile	1.573	-4.938	6.511***	13.949	0.467
Above average CGI	0.5068182	-0.0649558	0.572***	1.070	0.534
* p<.05 **p<.01 ***p<.001					

Results Table 4 (cont.)

Comparison of changes in scores and proficiency between consistent Zearn users and low- or non-users (cont.)

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
Grade					
Kindergarten scale score (SS)	25.086	19.052	6.034**	15.535	0.388
Kindergarten percentile	17.397	1.948	15.448***	33.120	0.466
Kindergarten CGI	0.707	0.448	0.259**	0.087	2.967
Kindergarten percent proficient	37.93%	6.90%	31.03%**	0.681	0.456
Grade 1 scale score (SS)	23.222	17.556	5.667**	13.714	0.413
Grade 1 percentile	12.524	0.984	11.540**	27.880	0.414
Grade 1 CGI	0.714	0.460	0.259**	0.646785	2.967
Grade 1 percent proficient	15.87%	-3.17%	19.05**	0.535	0.356
Grade 2 scale score (SS)	20.301	14.397	5.904***	13.069	0.452
Grade 2 percentile	9.493	-3.658	13.151***	28.305	0.465
Grade 2 CGI	0.671	0.493	0.178*	0.674	0.264
Grade 2 percent proficient	12.33%	-8.22%	20.55%**	0.492	0.417
Grade 3 scale score (SS)	18.246	12.211	6.035***	13.026	0.463
Grade 3 percentile	11.053	-1.807	12.860***	26.611	0.483
Grade 3 CGI	0.737	0.474	0.263**	0.695	0.379
Grade 3 percent proficient	22.81%	7.02%	15.79%*	0.492	0.321

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

Results Table 4 (cont.)

Comparison of changes in scores and proficiency between consistent Zearn users and low- or non-users (cont.)

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
Grade 4 scale score (SS)	10.900	10.000	0.900	9.422	0.096
Grade 4 percentile	-4.900	-5.300	0.400	17.245	0.023
Grade 4 CGI	1.000	0.900	0.100	0.316	0.316
Grade 4 percent proficient	0.00%	-10.00%	10.00%	0.316	0.316
Grade 5 scale score (SS)	14.857	5.619	9.238***	7.803	1.184
Grade 5 percentile	8.238	-9.571	17.810***	13.844	1.286
Grade 5 CGI	0.800	0.150	.650***	0.587	1.107
Grade 5 percent proficient	0.00%	-23.81%	23.80%	0.539	0.442
Grade 6 scale score (SS)	18.38462	3.846154	14.538***	8.442111	0.778
Grade 6 percentile	18.538	-10.462	29.000***	18.690	66.130
Grade 6 CGI	0.923	0.154	0.769***	0.439	1.754
Grade 6 percent proficient	30.77%	-38.46%	69.23%**	0.630	1.098

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

Results Table 5

Comparison of changes in scores and proficiency between consistent Zearn users and low- or non-users, by demographic and academic subgroup

	Treatment change in mean	Control change in mean	Difference	Pooled SD	Cohen's d
Subgroup					
Female SS	20.643	13.355	7.288***	10.488	0.695
Female percentile	10.905	-4.177	15.082***	21.498	0.702
Female percent proficient	18.25%	-4.84%	23.09%***	0.506	0.456
Female CGI	0.706	-0.376	1.082***	1.436	0.754
Male SS	20.704	15.029	5.675***	11.012	0.515
Male percentile	12.497	-0.392	12.889***	22.401	0.575
Male percent proficient	20.71%	-2.92%	23.63%***	0.502	0.471
Male CGI	0.744	-0.066	0.810***	1.479	0.547
Black student SS	19.750	13.200	6.550**	9.596	0.683
Black student percentile	13.125	-5.800	18.925***	20.281	0.933
Black student percent proficient	37.50%	-17.14%	54.64%***	0.539	1.014
Black student CGI	0.586	-0.447	1.032**	1.291	0.799

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

Results Table 5 (cont.)

	Treatment change in mean	Control change in mean	Difference*	Pooled SD	Cohen's d
Subgroup (cont.)					
Latino student SS	20.673	14.000	6.673***	10.032	0.665
Latino student percentile	12.391	-2.430	14.821***	20.260	0.732
Latino student percent proficient	16.36%	-9.30%	25.67%***	0.483	0.531
Latino student CGI	0.769	-0.272	1.041***	1.420	0.733
Economically disadvantaged SS	18.950	12.837	6.113***	10.577	0.578
Economically disadvantaged percentile	8.063	-5.348	13.410***	21.189	0.633
Economically disadvantaged percent proficient	18.75%	-9.78%	28.53%***	0.470	0.607
Economically disadvantaged CGI	0.363	-0.460	0.823***	1.459	0.564
English language learner SS	20.083	14.294	5.789***	9.700	0.597
English language learner percentile	9.819	-1.824	11.643***	19.918	0.585
English language learner percent proficient	12.50%	-8.82%	21.32%*	0.501	0.425
English language learner CGI	0.589	-0.316	0.905***	1.386	0.653
Gifted SS	18.388	12.718	5.669***	9.154	0.619
Gifted percentile	6.675	-3.296	9.971***	1.361	7.326
Gifted percent proficient	8.75%	-7.04%	15.79%**	0.365	0.433
Gifted CGI	0.744	-0.002	0.746***	1.179	0.633

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