Measuring the Effectiveness of Zearn Math in Louisiana[1]

July 2021

Shirin A. Hashim
Harvard Graduate School of Education

[1] This summary report has not been formally reviewed or edited. The findings and conclusions expressed are solely those of the author who is a former employee of Zearn, with no financial stake in the organization, and a former research fellow with the Louisiana Department of Education. Funding for this project was provided by Overdeck Family Foundation, which provides general operating support to Zearn.

Overview

This report provides an overview of initial findings on the effectiveness of Zearn Math in raising math test scores on Louisiana's statewide test in recent years. Using administrative data from the Louisiana Department of Education (LDOE) and Zearn Math usage metrics, I find that programmatic usage of Zearn Math increased math achievement on LEAP 2025 by about 0.03 standard deviation units, on average, across 3rd, 4th, and 5th grade[2].

Zearn is a non-profit organization that developed Zearn Math, a math curriculum that combines in-person instruction with digital lessons. Over one-tenth of educators indicated that they regularly used Zearn Math in their elementary school classrooms in the 2019-2020 school year according to a nationally representative survey (Doan et al., 2020). Louisiana has had a particularly high concentration of schools that programmatically use Zearn Math following LDOE deeming it as a Tier 1 curriculum in 2016 (LDOE, 2016). In the 2018-19 school year, nearly 80% of Louisiana's public and charter elementary schools had some degree of activity on the online platform, of which of over one-fourth used consistently throughout the school year.

While a randomized control trial (RCT) would have been the ideal research design to demonstrate the causal impact of the program, this non-experimental approach yields positive evidence suggesting that using Zearn Math increased students' math scores. I plan to further probe these findings by exploring impact heterogeneity, performing placebo tests with unaffected outcomes, and contextualizing alternative curricula in future research.

Data

Data were obtained from LDOE and Zearn. LDOE provided average 3rd, 4th, and 5th grade LEAP 2025 math scaled scores for each grade and school for five years spanning 2015 through 2019[3]. I mean-standardize these scaled scores by grade and year using student-level means and standard deviations from the LEAP 2025 technical reports in order to account for potential changes to the

[2] This would be considered a small- to medium-sized effect, comparable to the median effect size of 0.03 observed across large-scale causal studies of preK-12 education interventions with standardized achievement outcomes (Kraft, 2020). Alternatively, we can contextualize this in terms of typical growth on most measures of student achievement, which is about 0.25 standard deviation units from one grade to the next (Hanushek et al., 2012). In this case, programmatic usage of Zearn Math has an estimated effect size comparable to the academic growth that occurs over approximately 1.5 months of the school year, on average.

[3] One limitation of this analysis is that it is being conducted at the school-grade-level rather than at the classroom-level due to the nature of the Zearn Math data collection. However, curricular adoption often occurs at the school- or district-level, which supports defining programmatic usage by school-grade. This is further evidenced by the fact that nearly all the schools that programmatically used Zearn Math in at least one grade-level also used it programmatically in another grade-level (see Table 1). Still, it is possible that there was variation in implementation at the classroom-level that is not reflected here.

assessment over time and differences in the distribution of scores by grade-level[4]. I obtained school-level demographic data and grade-level enrollment data from the state's publicly available sources.

Zearn Math provided data on the average number of students who were active on the online platform and the average number of days students logged in. In my primary specification, I define programmatic usage as at least 50% of enrolled students using the digital component of Zearn Math for at least 50% of the number of state-required school days (i.e., 84 days)[5]. Zearn Math data are available from 2017-2019 for grades 1-5. These are the years in which digital usage were captured, as well the years in which Zearn Math was an approved curriculum in Louisiana.

Table 1 describes the characteristics of the main analysis sample, which is comprised of 3rd-5th grades in Louisiana public and charter schools from 2015-2019 with scores on the math LEAP 2025 assessment. Panel A includes all the LDOE schools in my dataset. Panel B includes the subset of school-grades that used Zearn Math programmatically in each year. It is worth noting that despite the lack of random assignment to the program, the schools that used Zearn Math in the 2018-19 school year are similar to the state overall across the available characteristics.

Empirical Approach

The main challenge in estimating the effect of the program on student achievement is that schools that adopted Zearn Math may differ in both observable and unobservable ways from those that did not. For example, the schools that were first to adopt Zearn Math may have had teachers who were more focused on curriculum quality or more motivated to try a new curriculum. A simple comparison of outcomes between grades that used Zearn Math and those that did not could yield biased estimates of the impact because of unobserved factors, such as teacher motivation, that may be associated with both curricular adoption and achievement. As a first attempt at eliminating some of these sources of omitted variable bias, I fit the following models, which include controls for school-by-grade and grade-by-year:

(1) $Score_{sgt} = \beta_0 + \beta_1 Zearn_{sgt} + \gamma X_{st} + \mu_{sg} + \lambda_{gt} + \epsilon_{sgt}$
(2) $Score_{sgt} = \beta_0 + \beta_1 ZearnYears_{sgt} + \gamma X_{st} + \mu_{sg} + \lambda_{gt} + \epsilon_{sgt}$
(3) $Score_{sgt} = \beta_0 + \beta_1 Years1_{sgt} + \beta_2 Years2_{sgt} + \gamma X_{st} + \mu_{sg} + \lambda_{gt} + \epsilon_{sgt}$

The outcome of interest, *Score*, represents the mean standardized math score for school *s* and grade *g* in year *t*. In Model 1, *Zearn* is a dichotomous variable indicating whether grade *g* in school *s* in year *t* used Zearn Math. In Models 2 and 3, I account for the fact that students in some grades may have used Zearn Math in previous years and grades. *ZearnYears* represents the number of years in which the cohort of students in grade *g* in school *s* would have used Zearn Math as of year *t*, corresponding to Column 4 in Table 2. For example, if students in school *s* first used Zearn Math as 4th graders in 2018 and then as 5th graders in 2019, the total number of years the 2019 5th grade cohort would have

[4] In some grades and years, the state only reports the means and standard deviations separately for the computer-based administration and the paper-based administration of the test. In these cases, I use the statistics for the paper-based assessment, which reflect over 95% of test takers.
[5] Alternative definitions of programmatic usage, for example, 30% of school days or 40% of school days, or 30% of students active or 40% of students active, do not materially change the findings presented here.

used Zearn Math is two years as of 2019. In Model 3, I allow for each additional year of Zearn Math to have a non-linear association with math scores by creating indicator variables for the number of years grade *g* in school *s* would have used Zearn as of year *t*.

These models include school-by-grade fixed effects, μ_{sg}, so identification of the estimate of the impact of Zearn Math is being driven by within-school-grade differences in usage. In other words, relative student achievement at a given school and grade is measured against the relative student achievement in the same school and grade in years when it did or did not use Zearn Math (Model 1) or used Zearn Math for a different number of years (Models 2 and 3). These fixed effects eliminate sources of omitted variable bias originating from differences across school-grades that are constant over time and may be associated with curricular decisions. For example, they would account for a given school having a particularly effective set of 3rd grade teachers, or if the neighborhoods assigned to a school were consistently more or less advantaged than the school demographics would imply.

I also include grade-by-year fixed effects, λ_{gt}, to account for potential secular trends experienced by students statewide, while allowing these to differ by grade-level. This fixed effect would take into account, for instance, if achievement were rising statewide in a particular grade-level due to changes in the test. Lastly, I include a vector of controls, X_{st}, for school-year gender and racial composition, enrollment, the percent of students who are classified as economically disadvantaged (% ED) and the percent of students classified as having limited English proficiency (% LEP) to account for the fact that the composition of students is not constant over time in each school.

Initial Findings

Table 3 shows the main results from fitting these models. The estimate on *Zearn* in Column 1 implies that programmatic use of Zearn Math is associated with a 0.034 standard deviation increase in math scores, on average, relative to years in which that school-grade did not use Zearn Math programmatically or at all. This estimate, however, does not account for the fact that some students would have used Zearn Math in a previous grade-level in a previous year at their school. If the effect of using Zearn Math is positive and compounding or persistent in later years, then this may be an under- or over-estimate of using Zearn Math for one year. For example, the estimate may be too high if the students who used Zearn Math in year *t* also used Zearn Math in a previous grade-level in a previous year such that the one-year estimate is in fact a multiple-year effect of using the program. The estimate could also be too low if the students who used Zearn Math in year *t* are being compared to students who did not use the program in year *t* but did use it in a previous grade-level in a previous year, such that the comparison cohort scores are inflated by historical use of the program.

Columns 2 and 3 aim to adjust for this potential cohort exposure by measuring the impact of each year that students in a particular school-grade used Zearn Math[6]. In Column 2, the estimate on

[6] One limitation of this cohort-level analysis is that it does not consider student mobility. For context, previous research has shown that the student non-structural mobility rate across elementary/middle grades in Louisiana has previously been around 15% (Maroulis et al., 2016). In other words, about 15% of students switch schools despite having the option to remain at their current school. The assumption being made here that all students attended the same school in the previous two years may be biasing the estimated impact of Zearn Math towards zero.

ZearnYears thus implies that each year that students would have used Zearn Math programmatically is associated with a 0.027 standard deviation increase in math scores, on average. For example, a 5th grade cohort that used Zearn Math for the first time as 4th graders and then again as 5th graders, is estimated to have math scores that are an average of about 0.06 standard deviations higher relative to 5th grade cohorts in other years at the school who had never used Zearn Math programmatically.

Column 3 disaggregates the effect of having used Zearn Math for one versus two years. The estimate on *Years1* indicates that having used Zearn Math for one of the past three years is associated with an approximately 0.029 standard deviation increase in math scores, on average. Having used the program for two of the past three years is associated with an average of an approximately 0.048 standard deviation increase in math scores as seen by the estimate on *Years2,* though this is not statistically significant at conventional levels.

Taken altogether, there is evidence to suggest that using Zearn Math increased mean LEAP 2025 math scores in the year in which it was used, and that there may be some additional gains from multiple years of usage, as well, though these are more ambiguous, in part due to the small sample of school-grades that used Zearn Math for multiple years relative to the effect size.

Discussion

This work comes at a critical time when pandemic-related school closures have likely hastened the reliance on technology-based materials, expanding Zearn Math's presence across Louisiana and the nation in the last year. As our educational system continues to depend on remote or hybrid learning environments, it is imperative that we understand what resources effectively support student learning and growth.

To that end, there is more work to be done. While I provide preliminary evidence that consistent usage of Zearn Math led to higher student achievement, the potential for heterogenous effects warrants additional research. For example, this initial report does not investigate differences by grades, school characteristics, teacher effectiveness, or fidelity of implementation. Further, robustness checks could be conducted with additional data. For example, science, social studies, English Language Arts (ELA), or middle school math LEAP 2025 scores could be used in placebo tests since these grades and subjects would be arguably unaffected by changes to an elementary math curriculum. Alternative methods, such as propensity score matching, can also be used to supplement these findings.

It is also important to probe the potential mechanisms underlying any measured impacts, especially given the mixed evidence on the impact of curricular choice on student learning (see, for example: Blazar et al. 2019; Kane et al., 2016). It may be that benefits of specific curricula emerge when they are implemented with more classroom support, more intensive usage, or more targeted teacher training. Future research should examine whether Zearn Math has been advantageous over existing curricula across these various implementation factors and better understand the "business as usual" practices that Zearn Math is being measured against.

Lastly, a non-experimental method, while providing an informative and feasible approach to evaluate the impact of Zearn Math, will be limited in terms of causal inference relative to an RCT. Any future opportunity to measure the impact of Zearn Math in an experimental setting would provide more conclusive evidence to support these initial findings.

Table 1: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	# Schools	# School-Grades	% Female	% White	% Black	% Asian	% Hispanic	% ED	% LEP	Average School Enrollment	Standardized LEAP 2025 Math Score
(A) All LDOE											
2015	826	2,232	0.486	0.443	0.452	0.015	0.061	0.721	0.033	596	0.007
2016	825	2,227	0.486	0.442	0.445	0.015	0.067	0.755	0.037	597	0.012
2017	827	2,227	0.486	0.436	0.443	0.016	0.072	0.747	0.041	602	0.009
2018	833	2,238	0.487	0.432	0.440	0.016	0.077	0.722	0.044	605	-0.002
2019	818	2,199	0.487	0.429	0.436	0.016	0.083	0.729	0.045	626	0.006
(B) Programmatically Used Zearn Math											
2015	--	--		--	--	--	--	--	--	--	--
2016	--	--		--	--	--	--	--	--	--	--
2017	8	9	0.519	0.478	0.412	0.005	0.047	0.711	0.009	1211	-0.204
2018	72	119	0.489	0.407	0.459	0.017	0.071	0.693	0.041	590	0.052
2019	150	277	0.482	0.421	0.470	0.012	0.063	0.726	0.030	600	0.019

Notes: Panel A includes all LDOE schools with at least one valid 3rd-5th grade mean scaled score and demographic data. Panel B is the subset of schools that used Zearn Math programmatically (at least 50% of students using for 50% of required school days) in the given year and does not account for students in those grades who may have historically used Zearn Math. School demographic characteristics are enrollment weighted averages of school-year data. Standardized test scores are weighted by the number of students tested. They are not equal to 0 in Panel A because they are standardized using the student-level mean and standard deviations by year and grade, and grade-levels do not necessarily have an equal number of students testing.

Table 2: Cohort Usage of Zearn Math

	(1) Year *t*	(2) Year *t*-1	(3) Year *t*-2	(4) # Zearn Years (in year *t*)	(5) N (school-grade-years)
(A)	X			1	339
(B)		X		1	41
(C)			X	1	7
(D)	X	X		2	66

Notes: Column 5 is the number of school-grade-years in which students would have used Zearn Math in the years indicated by columns (1)-(3), and the number of years indicated by column (4). School-grade-years in each row are not mutually exclusive cohorts of students. For example, a cohort of students who used Zearn Math in 2017 for the first time as 3rd graders, and then did not use in 2018 and 2019 as 4th and 5th graders, would be included in rows A, B, and C. There was 1 school-grade-year that used Zearn Math for 3 years that is dropped from the analysis to avoid drawing inferences from a single observation. Findings are not sensitive to the inclusion of this observation.

Table 3: Fixed Effects Estimates of the Impact of Zearn Math on LEAP 2025 Math Scores

	(1)	(2)	(3)
Zearn	0.034**		
	(0.016)		
ZearnYears		0.027**	
		(0.013)	
Years1			0.029*
			(0.016)
Years2			0.048
			(0.031)
R^2	0.861	0.861	0.861
N School-Grade-Years	11,008	11,008	11,008
N School-Grades	2,499	2,499	2,499

Notes: School clustered standard errors in parentheses (* p<.10 ** p<.05 *** p<0.01). All estimates are calculated with school-grade and grade-year fixed effects, are weighted by the number of students tested, and include controls for average school-year student composition by gender, race/ethnicity, economic disadvantage, limited English proficiency, and enrollment. There are 1,113 school-grade-year observations with fewer than 5 years of data. Limiting the sample to a balanced panel does not change the findings.

References

Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S.,...& Kurlaender, M. (2019). *Learning by the Book: Comparing math achievement growth by textbook in six Common Core states*. Cambridge, MA: Center for Education Policy Research, Harvard University. Retrieved from: https://cepr.harvard.edu/files/cepr/files/cepr-curriculum-report_learning-by-the-book.pdf

Doan, S., Grant, D., Henry, D., Kaufman, J.H., Lawrence, R.A., Tuma, A.P., Setodji, C.M., Stelitano, L., Woo, A., and Young, C.J. (2020) *American Instructional Resources Surveys: 2020 Technical Documentation and Survey Results*. Creative Commons Attribution 4.0 International Public License. Retrieved from: https://www.rand.org/pubs/research_reports/RRA134-4.html

Hanushek, Eric & Peterson, P.E. & Woessmann, L. (2012). "Is the U.S. catching up?: International and state trends in student achievement." *Education Next*, 12, pp. 24-33.

Kane, T., Owens, A., Marinell, W., Thal, D., & Staiger, D. (2016). *Teaching Higher: Educators' Perspectives on Common Core Implementation.* Cambridge, MA: Center for Education Policy Research, Harvard University. Retrieved from http://cepr.harvard.edu/files/cepr/files/teaching-higher-report.pdf?m=1454988762

Kraft, M.A. (May 2020). "Interpreting Effect Sizes of Education Interventions." *Educational Researcher,* vol. 49, no. 4, pp. 241–253, doi:10.3102/0013189X20912798.

Louisiana Department of Education (LDOE). (2016). *Instructional Materials Evaluation – Student Standard Review.* Retrieved from: https://www.louisianabelieves.com/docs/default-source/curricular-resources/zearn-inc---zearn-math-grades-1-4.pdf?sfvrsn=4.

Maroulis, S., Santillano, R., Harris, D.N., and Jabbar, H. (May 2016). "What Happened to Student Mobility After the New Orleans' Market-Based School Reforms?" Education Research Alliance for New Orleans: An Initiative of Tulane University. Retrieved from: https://educationresearchalliancenola.org/files/publications/ERA-Policy-Brief-Mobility-Student-Performance-160512.pdf