ESSA RESEARCH REPORT

August 2021

Assessing the Impact of IXL Math over Three Years: A Quasi-Experimental Study

Bozhidar M. Bashkov, Ph.D.

IXL LEARNING 777 Mariners Island Blvd., Suite 600, San Mateo, CA 94404

650-372-4040 | www.ixl.com

Third-Party Validated ESSA Evidence for IXL Math

This study performed a rigorous empirical evaluation of the efficacy of IXL Math, a popular PreK-12 educational technology platform designed to supplement a conventional standards-aligned curriculum. In both propensity score matching (pre-processing) and multiple regression analysis, the study controlled for baseline performance, grade level, and key school and student characteristics.

Results revealed statistically significant larger gains among IXL schools relative to comparable non-IXL schools over the course of a three-year intervention in grades 3-8. Thus, the study not only built upon the large body of extant evidence supporting the efficacy of IXL Math, but it also did so using methodology that meets the requirements for ESSA Tier 2 (moderate evidence) as well as WWC standards with reservations. Last but not least, this study provided a formal evaluation of the benefits of a three-year intervention.

An independent review determined that the study meets the requirements for ESSA Tier 2 and WWC standards with reservations (see Mislevy, Seftor, & Wei, 2021).

Assessing the Impact of IXL Math over Three Years: A Quasi-Experimental Study

Background

Educational technology plays an increasingly important role in K-12 education in the U.S. and across the globe. Many educational applications and services have been developed in the past two decades to help meet students' learning needs and alleviate teachers' workload. In the U.S., this momentum was at least in part spurred by the No Child Left Behind (NCLB) Act of 2001, which placed mandates on end-of-year testing in grades 3-8 for all public schools in order to propel more elementary and middle-school students into proficiency on state academic achievement standards (NCLB, 2001). More recently, the Every Student Succeeds Act (ESSA) of 2015, which replaced NCLB, gave states more freedom in establishing their own standards, but also placed more emphasis on evidence-based interventions. Specifically, the new legislation outlined four tiers of evidence: strong, moderate, promising, and "demonstrates a rationale," each associated with a research study design ranging from most to least rigorous. To help educators select products shown to be effective, the U.S. Department of Education issued *Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments* for their interpretation of the ESSA tiers of evidence that draw on What Works Clearinghouse (WWC) standards (U.S. Department of Education, 2016). Both ESSA and WWC standards call for rigorous research designs and statistical controls in order to obtain accurate estimates of intervention effectiveness and minimize or eliminate altogether the effects of prior achievement, demographic variables, or self-selection bias.

The focus of the present study is IXL, a personalized learning platform used by more than 12 million students worldwide, spanning Pre-K through 12th grade. IXL is offered in the four subject areas (i.e., English language arts, mathematics, science, and social studies), as well as Spanish; however, the subject of interest in this study is IXL Math. Rooted deeply in learning sciences research, IXL Math was built on well-established learning progression theories and math education principles (e.g., cardinality, concreteness fading). Other noteworthy features include active learning (meaning students play an active role in directing their own learning), an adaptive engine that recommends and presents tasks at an appropriate difficulty level and sequence in each student's unique learning progression, immediate feedback, and mastery goal orientation (see Bashkov, Mattison, & Hochstein, 2021).

Review of the Literature

Numerous studies have examined the efficacy of IXL Math and have generally reported promising results. First and foremost, IXL Learning researchers have examined the efficacy of IXL Math in over 60,000 schools across more than 30 U.S. states at the school or grade-cohort level, primarily in grades 3-8 (e.g., An, 2021a, 2021b; IXL Learning, 2019). Controlling for baseline performance and key demographic variables, these studies have found statistically significant differences in end-of-year state assessment performance in favor of schools that have implemented IXL Math compared to non-IXL schools, with an average effect size of about 0.12 standard deviation (*SD*) units and an expected percentile gain of about five points. Similarly, a third-party research firm found an effect size of 0.12 *SD* and a 5-percentile gain on the Oregon Assessment of Knowledge and Skills (OAKS) math test among fourth- and fifth-graders following a one-year IXL Math implementation (Empirical Education, 2013). Finally, a quasi-experimental study using one-to-one propensity score matching of treatment and control K-12 students in a large virtual public charter school found that IXL students who mastered 25 or more skills over the course of a semester outperformed comparable non-IXL students by more than 2 points on NWEA MAP math and reading tests (IXL Learning, 2019).

The purpose of the current study was two-fold. First, this study aimed to perform a rigorous empirical evaluation of the efficacy of IXL Math using sound methodology that meets WWC standards with reservations as well as the requirements for ESSA Tier 2 (moderate evidence). Second, the present study examined the effectiveness of IXL Math over a three-year period. As such, this study not only built upon the large body of evidence summarized above, but it also provided a formal evaluation of the benefits of a three-year intervention.

Methodology

STUDY DESIGN

The study used a quasi-experimental pretest-posttest control group design (see Figure 1). Both the assignment and the analysis were at the school level. Specifically, each treatment (i.e., IXL Math) school was matched to a similar comparison (i.e., non-IXL) school using one-to-one propensity score matching in order to establish baseline equivalence of pretest performance and grade level[1] and to control for key student demographics and school characteristics known to impact achievement. This preliminary step is explained in more detail further below.

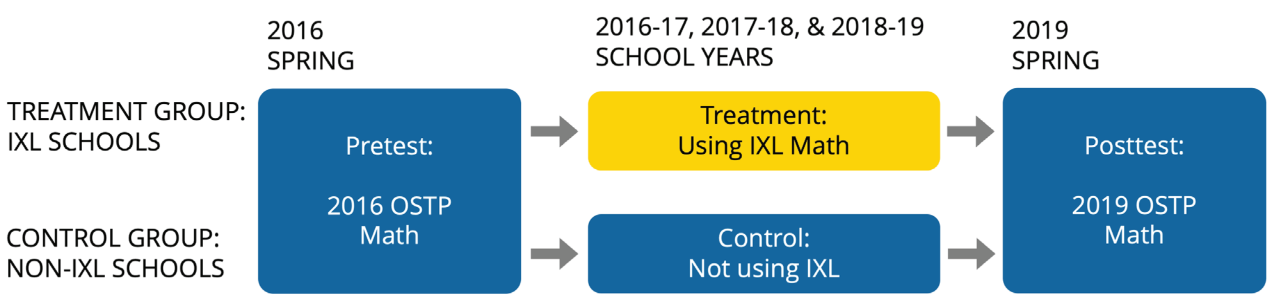

Figure 1. Study design

[1] *Given the assignment and analysis were at the school level, we established grade-level equivalence by classifying and matching on school type: Elementary (serving grades 3-5; 34.37%), Elementary/Middle (serving grades 3-8; 42.25%), or Middle (serving grades 6-8; 23.38%). All but two intervention schools were matched exactly on school type; one middle school was matched to an elementary school, and another middle school was matched to an elementary/middle school. See Table 3 for the full breakdown of school type for treatment and comparison schools.*

DATA SOURCES

The study examined public schools in the state of Oklahoma and included data from three sources: IXL usage data, school performance data, and school demographic data. IXL usage data were retrieved from IXL's internal database and provided information on whether a school had adopted IXL Math before and during the intervention period and the extent to which students used the product (Table 1). Data on school performance and some demographic characteristics were obtained from the Oklahoma State Department of Education. Other school characteristics, such as school type, location, and Title I status were retrieved from the National Center for Education Statistics.

INTERVENTION

IXL Math offers thousands of hyperspecific "micro-skills" aligned to K-12 state standards, ranging from foundational skills focused on teaching young learners how to count to advanced skills where students solve complex problems (see Bashkov et al., 2021). Skill practice is designed to be an independent activity, so students can work on IXL as part of their assigned homework or live in the classroom along with their peers. Using information from in-skill practice and its diagnostic assessment, IXL provides each student with personalized next steps at the right level of rigor. Thus, the IXL Math intervention is both student- and teacher-driven and can take many forms; however, all IXL usage is captured and summarized in three main usage indicators (see Table 1).

Table 1. IXL Math Usage

IXL Usage (per student per week)	*M*	*SD*	Min	Max
Time spent (in minutes)	10.61	9.15	0.51	64.87
Questions answered	25.87	17.62	2.16	99.13
Skills proficient	0.58	0.42	0.05	2.06

Note. N = 179 (treatment group only). Skills proficient = number of skills in which students reached proficiency (i.e., a SmartScore of 80 or higher, out of 100).

The intervention duration was up to three years, lasting from the fall of 2016 to the spring of 2019. However, given the assignment and analysis were at the school level, there was no strict intervention intensity cutoff at the individual student level; a student would be considered an IXL user if they answered at least one question on IXL. Instead, a school was considered a treatment school if at least 10% of its students used IXL Math during any of the three school years on average. For example, if 50% of students at a given school used IXL Math during two of the three school years, then the school was still considered a treatment school, as the average user percentage over the three-year period was greater than the 10% threshold. This specific cutoff was chosen in order to exclude from the analysis schools whose students had access to IXL Math but did not use the product, as well as to accommodate a wide range of product usage across schools. Altogether, 45 schools used IXL Math in any one year in the 2016-2019 period, 41 schools used it in any two years, and 93 schools used the product in all three years. One- and two-year treatment schools started using IXL later than three-year schools, but had similar levels of weekly IXL usage in the latest intervention year (2018-19). Descriptive statistics on IXL Math usage for the treatment group averaged across students and schools are presented in Table 1. The comparison group comprised schools that did not adopt IXL Math at all during any of the three school years.

PROPENSITY SCORE MATCHING

One-to-one propensity score matching without replacement (Ho et al., 2011) was used to create equivalent treatment and comparison groups in the absence of random assignment. Using a set of covariates, each treatment school was matched to a comparison school with an identical or very similar propensity score value (see Figure A1 in Appendix A). A propensity score is the probability that a school with certain characteristics would be assigned to the treatment group (over the comparison group). Comparison schools were identified from the 822 non-treatment schools in the state. The resulting sample comprised 179 treatment and 179 comparison schools with nearly identical characteristics. These schools served students in grades 3-8 and were classified as treatment or comparison schools based on their implementation and use of IXL Math during the three-year period they were under study, as described above. Absolute standardized mean differences after matching for all covariates were below WWC's cutoff of 0.25, with more than half being below 0.05 (see Figure A2 in Appendix A). Nevertheless, all covariates used in matching were included in the outcome analysis to adjust for these small differences.

Table 2. Descriptive Statistics for the Matched Sample (Continuous Variables)

	Intervention schools			Comparison schools			Estimated effect		
	n	*M*	*SD*	*n*	*M*	*SD*	*b*	*p*	*g*
Outcome measure	179	38.07	16.58	179	34.17	15.10	3.613	<.001	0.228
Pretest measure	179	68.48	14.69	179	67.91	13.66			
School size	179	416.64	217.42	179	412.91	259.98			
Gender (% male)	179	51.81	3.08	179	52.08	3.41			
Race (% white)	179	50.14	18.95	179	52.40	17.40			
% SPED	179	17.84	5.86	179	17.68	5.68			
% Low SES	179	64.56	19.42	179	64.19	15.33			

Note. Absolute standardized mean differences for baseline characteristics are presented visually in Appendix A.

Table 3. Descriptive Statistics for the Matched Sample (Categorical Variables)

	Intervention schools		Comparison schools	
	n	%	*n*	%
School type				
Elementary	75	41.90	76	42.46
Elementary/Middle	68	37.99	69	38.55
Middle	36	20.11	34	18.99
School location				
City	23	12.85	17	9.50
Suburb	31	17.32	27	15.08
Town	46	25.70	50	27.93
Rural	79	44.13	85	47.49
Title I	161	89.94	164	91.62

Note. Absolute standardized mean differences for baseline characteristics are presented visually in Appendix A. *N* = 179 schools in each group.

COVARIATE AND OUTCOME MEASURES

Achievement Measure. Math achievement at baseline (pretest) and after the intervention (posttest) was measured as the school proficiency rate on the Oklahoma School Testing Program (OSTP) Math assessment. OSTP tests measure students' progress in learning on the Oklahoma Academic Standards for English language arts, mathematics, and science in grades 3-8. More information about these assessments can be found on the Oklahoma State Department of Education website (https://sde.ok.gov/state-testing-resources).

Covariates. Per WWC protocols (WWC, 2018), covariates for both matching and analysis included baseline achievement, school characteristics, and student demographics as follows:

- 2016 OSTP Math proficiency rate (pretest; % students reaching "proficient" or "advanced")
- School type (elementary, elementary/middle, or middle)
- School location (city, suburb, town, or rural)
- School size (number of students)
- Title I status (yes/no)
- Gender (% male students)
- Race (% white students)
- Special education (% students with disabilities)
- Socioeconomic status (% economically disadvantaged students)

ANALYSIS

Once the data were pre-processed via matching, a multiple regression model was used to estimate the average treatment effect of IXL Math following IXL implementation of up to three years. Specifically, this model regressed the outcome (2019 OSTP Math proficiency rate) on the binary IXL Math predictor distinguishing treatment from comparison schools as well as all of the covariates used in matching. Inclusion of the covariates served two purposes. First, it accounted for any absolute standardized mean differences at baseline greater than .05. Second, it allowed us to estimate the treatment effect more precisely by controlling for key school-level characteristics and student demographics suspected to be related to the outcome (math achievement). Per WWC guidelines, Hedges' *g* was computed as a measure of effect size using the model-adjusted (i.e., estimated) means and the pooled unadjusted (i.e., observed) standard deviation (WWC, 2020). It is important to note that given the outcome (i.e., proficiency rate) is computed at the cluster level, and the study analysis is also at the cluster (i.e., school) level, the effect size should be interpreted at the school, and not the student level.

Results

Descriptive statistics on observed (i.e., unadjusted) OSTP Math proficiency rates for treatment and comparison schools in 2016 (pretest) and 2019 (posttest) are presented in Table 2. Note that the sharp decline in proficiency rates from pretest to posttest is likely due, at least in part, to the more rigorous proficiency standards adopted by the state during the intervention period (Oklahoma State Department of Education, 2018).

The estimated average treatment effect for IXL Math—controlling for baseline performance, grade level, and key demographic variables—was positive and statistically significant ($b = 3.613$, $SE = 0.979$, $p < .001$). The effect size indicative of the practical significance of this effect was moderately large (Hedges' $g = 0.228$). Recall that this effect size pertains to the school-level proficiency rate difference between treatment and comparison schools and should be interpreted at the school level. Model-adjusted means for the outcome are presented in Table 4. Full model results are presented in Table B1 in Appendix B.

Table 4. **Model-adjusted Means for Outcome Proficiency Rates**

Group	Estimated Mean	*SE*	*df*	Lower 95% CI	Upper 95% CI
Treatment	36.06	1.04	344	34.02	38.11
Comparison	32.45	1.08	344	30.33	34.57

Note. SE = standard error; *df* = degrees of freedom; CI = confidence interval.

Conclusion

The present study sought to conduct an empirical evaluation of the effectiveness of a popular educational software for PreK-12 mathematics, namely IXL Math. Following strict ESSA and WWC study design guidelines, this study found sizable gains for IXL Math schools relative to comparable non-IXL schools, controlling for prior achievement and key demographic variables. More specifically, the study examined the efficacy of IXL Math adoption of up to three years, indicating IXL Math can have significant benefits for students in schools adopting the product over a similar time period.

Results from this study complement prior research showing the effectiveness of IXL Math in grades 3-8. More studies are needed to examine its efficacy outside this grade range. However, a major obstacle in this effort may be the lack of standardized achievement data in lower elementary and high school grades. Nonetheless, IXL Math passed a rigorous quasi-experimental test of its efficacy, demonstrating its ability to boost student achievement and propel more students to proficiency in math.

References

An, X. (2021a). *Measuring the Impact of IXL Math and IXL Language Arts in Kentucky Schools*. https://www.ixl.com/research/Impact-of-IXL-in-Kentucky.pdf

An, X. (2021b). *Associations between IXL Usage and IXL Impact in Math and ELA: Analysis of Data from Seven States*. https://www.ixl.com/research/Impact_of_IXL_in_7_states.pdf

Bashkov, B. M., Mattison, K., & Hochstein, L. (2021). *IXL Design Principles: Core Features Grounded in Learning Science Research*. https://www.ixl.com/research/IXL_Design_Principles.pdf

Empirical Education. (2013). *A Study of Student Achievement, Teacher Perceptions, and IXL Math*. https://www.empiricaleducation.com/ixl/

Ho, D., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1-28.

IXL Learning. (2018). *Measuring the Impact of IXL Math and IXL Language Arts in Smarter Balanced States*. https://www.ixl.com/research/The-IXL-Effect-Smarter-Balanced-States.pdf

IXL Learning. (2019). *The Impact of IXL Math and IXL ELA on Student Achievement in Grades Pre-K to 12*. https://www.ixl.com/research/ESSA-Research-Report.pdf

Mislevy, J., Seftor, N., & Wei, X. (2021). *IXL Math: Nonregulatory ESSA Standards Evidence Review & What Works Clearinghouse Standards Review*. SRI. https://www.sri.com/wp-content/uploads/2021/10/QED-IXL-Math-11122021.pdf

No Child Left Behind Act of 2001, 20 U.S.C. 70 § 6301 *et seq*.

Oklahoma State Department of Education. (2018). *Oklahoma ESSA Consolidated State Plan*. https://sde.ok.gov/ok-essa-state-plan

U.S. Department of Education. (2016). *Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments*. https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf

What Works Clearinghouse. (2018). *Review protocol for primary mathematics Version 4.0 (December 2018)*. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_pm_protocol_v4.0_508.pdf

What Works Clearinghouse. (2020). *What Works Clearinghouse procedures and standards handbooks*. https://ies.ed.gov/ncee/wwc/Handbooks

Appendix A

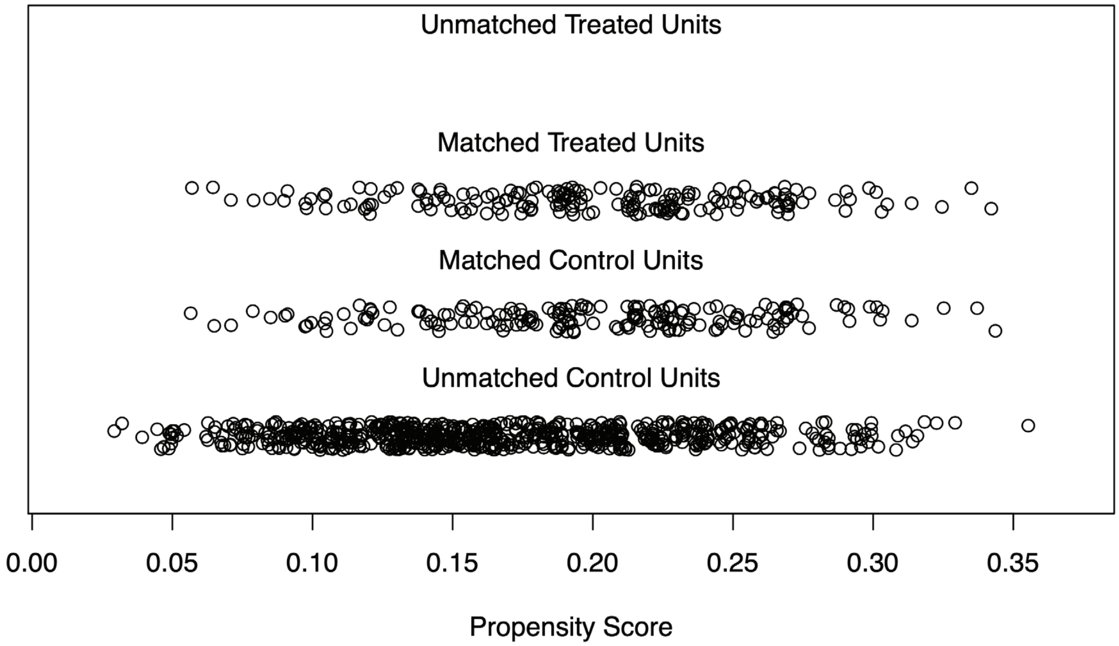

Figure A1. Distribution of Propensity Scores

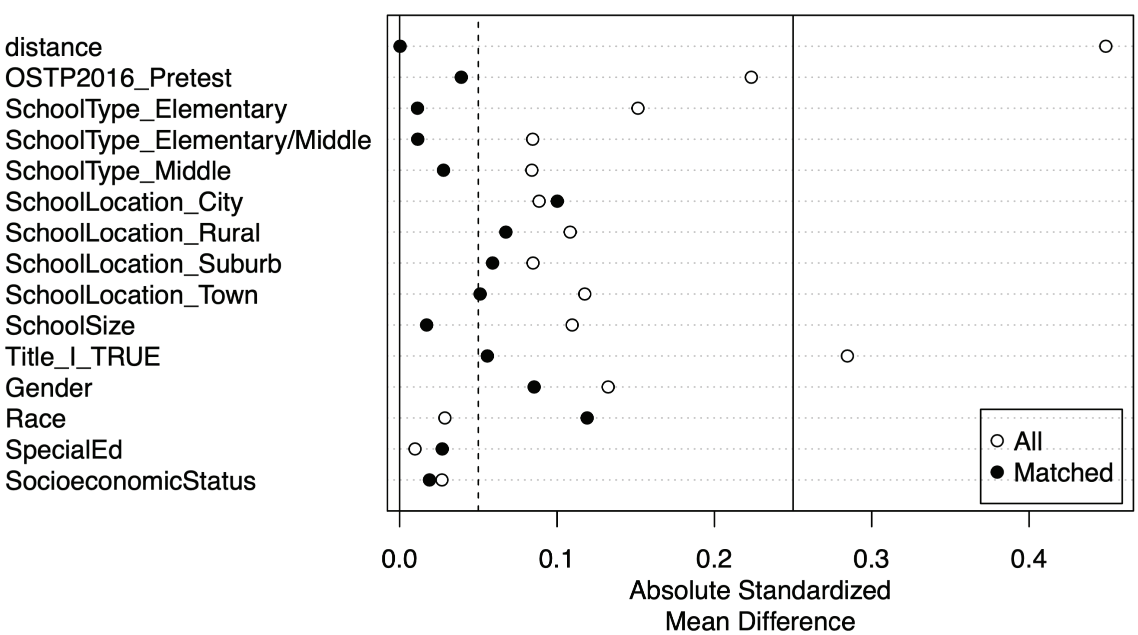

Figure A2. Absolute Standardized Mean Differences for Covariates Before and After Matching

Appendix B

Table B1. Full Regression Results

Predictor	*b*	*SE*	*t*	*p*
(Intercept)	34.607	2.376	14.564	<.001
Gender: % male[1]	0.493	0.154	3.214	.001
Race: % white[1]	0.108	0.036	2.986	.003
% Economically disadvantaged[1]	-0.230	0.040	-5.705	<.001
% Students with disabilities[1]	0.075	0.090	0.835	.405
School type: elementary/middle[2]	-7.739	1.185	-6.533	<.001
School type: middle[2]	-10.499	1.418	-7.405	<.001
School size (*N* students)[1]	0.007	0.003	2.517	.012
School location: rural[3]	1.069	2.005	0.533	.594
School location: suburb[3]	-0.830	2.016	-0.412	.681
School location: town[3]	1.297	1.970	0.658	.511
Title I[4]	1.878	1.910	0.983	.326
2016 math proficiency[1]	0.553	0.043	13.003	<.001
IXL Math school	**3.613**	**0.979**	**3.692**	**<.001**

Note. Dependent variable: Percent of students reaching proficiency on 2019 OSTP Math. *b* = unstandardized regression coefficient, *SE* = standard error.
[1] Grand mean-centered.
[2] Dummy coded; Elementary school as reference group.
[3] Dummy coded; location “City” as reference group.
[4] Dummy coded; Non-Title I schools as reference group.