RESEARCH REPORT

June 2023

Boosting Student Achievement with IXL's Diagnostic Snapshot

Christina Schonberg, Ph.D.

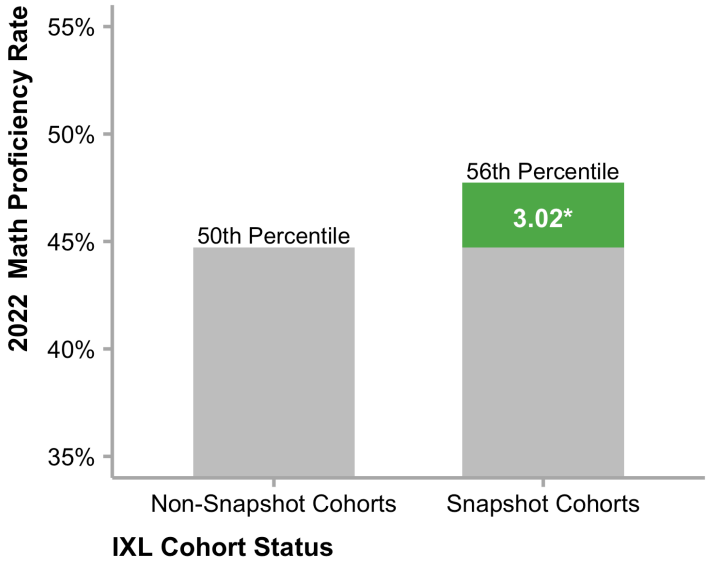
IXL LEARNING 777 Mariners Island Blvd., Suite 600, San Mateo, CA 94404
650-372-4040 | www.ixl.com

Executive Summary

IXL is an end-to-end teaching and learning solution that engages learners in grades Pre-K through 12 with a comprehensive curriculum and a first-of-its-kind assessment suite. A core component of IXL's assessment suite is the IXL Diagnostic, an interim assessment designed by a team of educators and mathematicians that uses Item Response Theory (Lord, 1980) to automatically adjust question difficulty based on students' response patterns. The diagnostic can be used in two ways. In *Real-Time* mode, IXL's Diagnostic provides students and educators with in-the-moment information about students' subject area knowledge, which is extremely helpful for daily instruction. In *Snapshot* mode, the diagnostic serves as a flexible, lightweight benchmark assessment that is designed to be used one or more times during the school year to easily assess and track students' grade-level proficiency over time. After completing the diagnostic—either in Real-Time or Snapshot mode—students receive personalized action plans that help them close gaps quickly.

The goal of the present study was to investigate the impact of completing IXL's Diagnostic Snapshot on student learning and achievement in grades 3 through 8. Using assessment data from four states (NY, TN, OK, and WI), we compared the math achievement of grade cohorts that completed at least one Snapshot during the 2021-22 school year to comparable grade cohorts that used IXL but did not complete a Snapshot during this time. Using 2021 assessment performance as pretest and 2022 assessment performance as posttest, we found[1]:

- **Using IXL's Diagnostic Snapshot improves student math achievement.** Grade cohorts that completed IXL's Diagnostic Snapshot performed better on their state math assessments than comparable IXL grade cohorts that did not complete IXL's Diagnostic Snapshot. Specifically, the proficiency rate[2] was about three percentage points higher for Snapshot cohorts, relative to comparable IXL cohorts that did not complete the Snapshot.



[1] *In all figures: * indicates significance at the .05 level*
[2] *Proficiency rate: percentage of students in a cohort classified as proficient or advanced on their state assessment*

Boosting Student Achievement with IXL's Diagnostic Snapshot

Background

IXL is an end-to-end teaching and learning solution that engages learners in grades Pre-K through 12 with a comprehensive curriculum, first-of-its-kind assessment suite, and personalized recommendations for meeting learning goals. It covers four main subject areas: mathematics, English language arts (ELA), science, and social studies. As of this writing, IXL is used by 1 in 4 students in the U.S. and by over 14.5 million students worldwide.

A core component of IXL's assessment suite is the IXL Diagnostic, an interim assessment developed by a collaborative team of educators and mathematicians that covers material aligned with the Common Core and other academic standards (see Bashkov et al., 2021). IXL's Diagnostic is especially valuable because it provides insights for students and educators about knowledge levels in key strands of math and ELA. The diagnostic uses Item Response Theory (Lord, 1980) to analyze student response patterns. Based on students' responses, the diagnostic then provides personalized action plans that lay out clear next steps for students to reach their learning goals.

IXL's Diagnostic is a reliable, valid tool that can be used in two ways to best meet students' and educators' needs. In *Real-Time* mode, students can use the diagnostic anytime, allowing for real-time assessment of their current knowledge. After completing the initial assessment, which takes only 45 minutes per subject, students can answer just a handful of diagnostic questions each week to keep their diagnostic results and personalized recommendations up to date, ensuring that they make meaningful progress on their learning goals. Teachers can use the immediate insights from the Real-Time Diagnostic to understand exactly what students need to improve on a day-to-day basis and to easily differentiate instruction. Many studies have examined the psychometric properties of the IXL Real-Time Diagnostic and have yielded desirable reliability and validity evidence, including coherent internal structure (IXL Learning, 2020a), multi-group measurement invariance (An et al., 2022), high reliability (IXL Learning, 2020a; Schonberg, 2021a), and strong predictive validity coefficients using multiple well-established assessments as criterion measures (An, 2021, 2022; IXL Learning, 2020b; Schonberg, 2021a, 2021b, 2022).

IXL's Diagnostic can also be used in *Snapshot* mode, which serves as a flexible, lightweight benchmark assessment. Snapshot mode allows administrators to capture student knowledge levels at a fixed point in time, across all students in a target grade level, school, or even an entire district. Unlike the Diagnostic's Real-Time mode, which requires brief but frequent (i.e., continuous) assessment to ensure that students' personalized recommendations stay up-to-date, Snapshot mode is designed to be used one or more times throughout the school year to provide a high-level overview of students' grade-level proficiency in key math and ELA strands. For example, a school administrator may administer the Diagnostic Snapshot for beginning-of-year benchmarking and then conduct a mid-year Snapshot to see how students have grown. Snapshot mode allows administrators to easily track student progress between Snapshots, which can help inform school- or district-level planning and decision-making. After completing the Diagnostic Snapshot, students receive a personalized action plan with specific suggestions for skill practice on IXL. Students can then keep these action plans up to date by using the diagnostic in Real-Time mode in between

Snapshots. Figure 1 illustrates one way that IXL's Diagnostic can be implemented over the course of a school year.

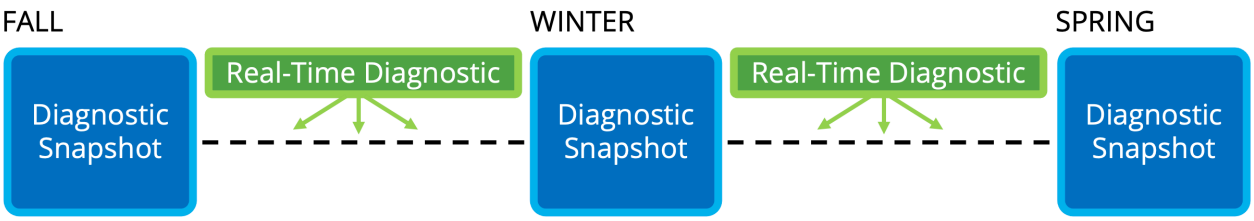


Figure 1. Example implementation of IXL's Diagnostic Snapshot and Real-Time Diagnostic. During Snapshot windows, students pinpoint their knowledge levels. Between Snapshot windows, students answer a few questions per week in the Real-Time Diagnostic, which keeps their levels and skill recommendations up to date.

Given its strong predictive validity, IXL's Diagnostic is a valuable tool in helping students make sure that they are on track to meet grade-level standards on end-of-year state assessments. The goal of the present study was to examine the impact of IXL's Diagnostic Snapshot on grade 3-8 students' math achievement in public schools across four geographically diverse states: New York, Oklahoma, Tennessee, and Wisconsin. These states were selected because they had sufficiently high numbers of IXL subscriptions both with and without Snapshot administrations, allowing us to compare Snapshot and non-Snapshot users.

The benefits of IXL usage are well-documented (see https://www.ixl.com/research); students who use IXL see significantly higher academic achievement relative to those who do not use IXL. The goal of this study was to investigate the added benefit of the Diagnostic Snapshot among IXL users. Among schools that use IXL, those that use benchmarking via the Diagnostic Snapshot are more likely to uncover and address students' knowledge gaps prior to end-of-year testing. Thus, we hypothesized that IXL users that systematically used the diagnostic as a benchmark—that is, those who completed a diagnostic assessment in Snapshot mode—would outperform IXL users who did not take the Diagnostic Snapshot. To test this hypothesis, we compared state assessment math proficiency rates among grade-level cohorts that completed IXL's Diagnostic Snapshot to those of comparable grade-level cohorts that used IXL but did not complete a Diagnostic Snapshot.

Study Design and Methodology

DATA SOURCES

Assessment and Demographic Data

Assessment and demographic data for students in grades 3 through 8 were obtained from the New York, Oklahoma, Tennessee, and Wisconsin Departments of Education. Math performance at pretest (2021) and posttest (2022) was measured using each state's end-of-year standardized assessments. The outcome measure was the percentage of students within a grade-level cohort reaching proficiency in math (i.e., the proficiency rate on the math assessment).

IXL Data

IXL Diagnostic Snapshot data were obtained from IXL's database. Specifically, we collected information about the number of Snapshots that each IXL school in the studied states completed during the 2021-22 school year. In addition, we also obtained IXL usage data (i.e., from the instructional component of our product informed by the diagnostic). Descriptives for IXL usage indicators—questions answered, skills proficient, and time spent—are presented in Table A1.

STUDY DESIGN

The goal of this study was to examine the effect of one of IXL's assessment components—the Diagnostic Snapshot—on student achievement. Therefore, all schools in the study were IXL schools. To assess the impact of the Snapshot, we used a quasi-experimental pretest-posttest control group design to compare the proficiency rates of grade-level cohorts that completed IXL's Diagnostic Snapshot during the 2021-22 school year to the proficiency rates of cohorts that used IXL but did not complete a Snapshot during this time (Figure 2). To control for baseline performance and demographic characteristics, we used one-to-one propensity score matching (described in more detail below) to match each Snapshot cohort to a similar cohort that did not use Snapshot.

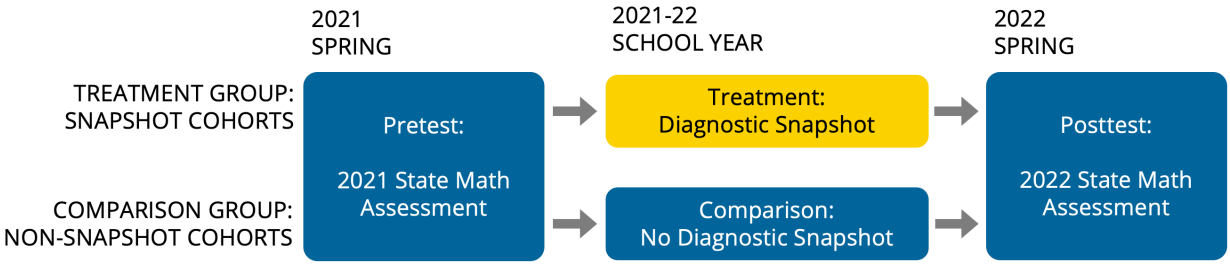


Figure 2. Study design

PARTICIPANTS

We defined treatment (Snapshot) and comparison (non-Snapshot) cohorts by identifying IXL accounts (i.e., schools or districts) that had completed at least one Snapshot during the study period. Treatment cohorts belonged to accounts that completed a Snapshot at least once, and comparison cohorts belonged to accounts that did not complete a Snapshot at all. We also included only cohorts that had non-missing pretest and posttest assessment data as well as non-missing demographic data. Using these criteria, we obtained a sample of 488 study cohorts for the analysis (treatment n = 244, comparison n = 244) after propensity score matching. Descriptive statistics for cohorts' IXL usage during the study period can be found in Table A1 (Appendix A).

PROPENSITY SCORE MATCHING

We conducted one-to-one propensity score matching without replacement using the *MatchIt* package in *R* (Ho et al., 2011; R Core Team, 2021) as a preprocessing step prior to analysis. A *propensity score* is the probability that a school cohort would be "assigned" to the treatment (i.e., Snapshot) group over the comparison group and is calculated using a combination of demographic

characteristics (i.e., covariates). In the absence of random assignment, propensity scores can be used to match comparison cohorts to treatment cohorts and create equivalent treatment and comparison groups. In a comparison of unmatched groups (e.g., Snapshot cohorts compared to all non-Snapshot cohorts), non-Snapshot cohorts could be very different from Snapshot cohorts on some dimensions. In contrast, using propensity score matching allows us to compare the performance of pairs of Snapshot and non-Snapshot cohorts that are very similar to each other. This comparison allows for a clearer attribution of the effect of the Diagnostic Snapshot and broader generalization of the findings to other IXL cohorts that are not yet using the Diagnostic Snapshot. Comparison cohorts were identified from 2,106 non-treatment cohorts that had non-missing assessment and demographic data. After matching, the resulting treatment and comparison groups had extremely similar demographic characteristics (see Table A2, Appendix A).

ANALYSIS

We specified and tested a multilevel model to account for clustering at the school and district levels (i.e., grade cohorts within a school tend to be more similar to each other than grade cohorts in other schools, and schools within a district tend to be more similar to each other than schools in other districts). In these models, we regressed 2022 state assessment grade-level proficiency rate on Snapshot cohort status (treatment or comparison) and covariates (baseline performance and demographic characteristics). We included covariates in the model because the absolute standardized mean differences (SMDs) for some covariates were greater than .05 after matching, indicating that these differences needed to be accounted for statistically. Specifically, we controlled for the following school-level demographic characteristics: percentage of male students; percentage of White students; percentage of economically disadvantaged students; percentage of students enrolled in special education; percentage of English-language learners; school size (number of students); student-teacher ratio; locale (city, suburb, town, or rural); and school Title I status. In addition, we controlled for grade level.

Following What Works Clearinghouse guidelines (WWC, 2020), each effect is accompanied by a test of statistical significance using a probability (*p*) value and a measure of effect size. The *p*-value is the probability of observing the current or more extreme data, assuming the effect is zero (Cohen, 1994). The smaller the *p*-value, the less likely it is that the result occurred at random; *p*-values less than .05 are considered statistically significant. Effect size is reported using Hedges' *g* and indicates the difference between treatment and control groups on an outcome measure in standard deviation units. For broad-scope educational assessments, moderate effect sizes range from about 0.10–0.20, and effect sizes of about 0.20 or higher are considered large (Kraft, 2020; Lipsey et al., 2012). We also report percentile gain, which is the expected change in IXL cohorts' percentile rank relative to non-IXL cohorts at the 50th percentile and is based on the effect size. Given that these analyses are at the grade cohort level, the effect sizes should be interpreted at the grade cohort level as well.

Results

We found that grade-level cohorts that used IXL's Diagnostic Snapshot outperformed comparable non-Snapshot cohorts on state math assessments in 2022. Specifically, the proficiency rate was about three percentage points higher for Snapshot cohorts relative to cohorts that used IXL but did not complete a Diagnostic Snapshot (Figure 3). The estimated treatment effect for the Diagnostic Snapshot was positive and statistically significant (b = 3.02, p = .028; see Table B in Appendix B for full model results). The effect size (Hedges' g) was 0.14, which corresponds to a percentile gain of six points.

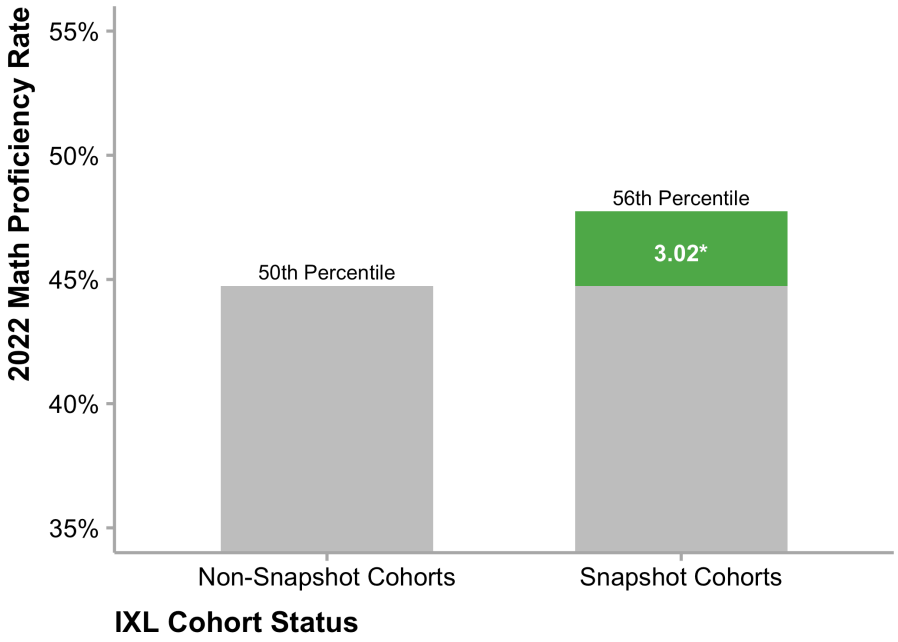


Figure 3. The impact of IXL's Diagnostic Snapshot

Discussion and Recommendations

A wide body of research has demonstrated that using IXL improves students' academic achievement. In this study, we investigated the added benefit of IXL's Diagnostic Snapshot by comparing IXL grade cohorts that completed at least one Snapshot to comparable IXL grade cohorts that did not. Controlling for baseline performance and demographic characteristics, we found that Snapshot cohorts outperformed non-Snapshot cohorts by about three percentage points on end-of-year state assessments.

These results show that using IXL's Diagnostic in Snapshot mode can play a key role in a successful IXL implementation. As a flexible, lightweight benchmark assessment, the Snapshot identifies students' knowledge levels at a specific time point and easily tracks growth over time. After students complete the Snapshot, they receive individualized action plans recommending skills that will help maximize their academic growth, whatever their starting point.

Over the course of the 2022-23 academic year, we found that Snapshot cohorts used IXL about twice as much as non-Snapshot cohorts (e.g., reaching proficiency in twice as many skills per week; see Table A1). It may be that using Snapshot not only led students to use IXL *more*, but also helped them use IXL *more effectively* by practicing the skills recommended in their personalized action plans. This more effective IXL usage, in turn, may be one of the reasons that the Snapshot cohorts outperformed the non-Snapshot cohorts on their end-of-year assessments. In addition, educators and school administrators may have used Snapshot results to inform their implementation of IXL, ensuring a timely, targeted intervention for students who needed it most.

The Diagnostic Snapshot is just one component of IXL's first-of-its-kind assessment suite. The assessment suite also includes a universal screener, which can identify students for intensive intervention in as little as 20 minutes. In addition to the benchmarking provided by the Diagnostic Snapshot, students and educators can also use the diagnostic in Real-Time mode, which provides immediate insights that support daily classroom instruction. The Real-Time Diagnostic and the Diagnostic Snapshot integrate seamlessly with IXL's comprehensive curriculum: using personalized action plans based on diagnostic performance, students can easily find exactly which skills they should practice next.

As an end-to-end teaching and learning solution, IXL can play a key role in helping students recover from the educational disruptions caused by the COVID-19 pandemic. IXL's assessment suite and personalized guidance identify content areas where students may be struggling and engage them with material at the appropriate level. By meeting students where they are, IXL can help students "catch up" by providing support for relearning missed or forgotten material. This combination of personalized learning and remediation has been suggested as a highly effective approach for both recovering from pandemic-related learning loss as well as boosting subsequent learning gains (Kaffenberger, 2021). The results of this study demonstrate that when IXL's Diagnostic and comprehensive curriculum are used in tandem, students experience significant gains. By taking advantage of the seamless integration between IXL's core components, students can maximize the impact of IXL and unlock their full academic potential.

References

An, X. (2021). *Assessing the predictive validity of the IXL Real-Time Diagnostic among marginalized students* (pp. 1–10). https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study-Marginalized-Students.pdf

An, X. (2022). *Examining the predictive validity of the IXL Real-Time Diagnostic using the Florida Standards Assessments as criterion* (pp. 1–8). https://www.ixl.com/materials/us/research/IXL_Real-Time_Diagnostic_Validation_Study_(FL_FSA).pdf

An, X., Xiong, Y., & Bashkov, B. M. (2022). *Multi-group measurement invariance of the IXL Real-Time Diagnostic math assessment* (pp. 1–10). https://www.ixl.com/materials/us/research/Multi-Group_Measurement_Invariance_of_the_IXL_Real-Time_Diagnostic_Math_Assessment.pdf

Bashkov, B. M., Mattison, K., & Hochstein, L. (2021). *IXL design principles: Core features grounded in learning science research* (pp. 1–18). https://www.ixl.com/research/IXL_Design_Principles.pdf

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*(12), 997-1003.

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. https://doi.org/10.18637/jss.v042.i08

IXL Learning. (2020a). *Assessing the construct validity and the predictive validity of the IXL Real-Time Diagnostic* (pp. 1–15). https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study-ILEARN.pdf

IXL Learning. (2020b). *Validation study of the IXL Real-Time Diagnostic using MAP Growth Assessments* (pp. 1–10). https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study.pdf

Kaffenberger, M. (2021). Modelling the long-run learning impact of the Covid-19 learning shock: Actions to (more than) mitigate loss. *International Journal of Educational Development, 81*, 102326. https://doi.org/10.1016/j.ijedudev.2020.102326

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms* (NCSER 2013-3000; pp. 1–54). National Center for Special Education Research, Institute of Education Sciences. http://ies.ed.gov/ncser

Lord, F. (1980). *Applications of Item Response Theory to practical testing problems*. Lawrence Erlbaum

Associates.

R Core Team. (2021). *R: A language and environment for statistical computing* (4.0.4). R Foundation for Statistical Computing. https://www.R-project.org/

Schonberg, C. (2021a). *Demonstrating grade-level predictive validity of the IXL Real-Time Diagnostic using the Virginia SOL as criterion* (pp. 1–12). https://www.ixl.com/materials/us/research/IXL_Real-Time_Diagnostic_Grade-Level_Validation_Study_(VA_SOL).pdf

Schonberg, C. (2021b). *The impact of IXL on math and ELA achievement in a Virginia school district* (pp. 1–13). https://www.ixl.com/research/The_Impact_of_IXL_in_a_Virginia_School_District.pdf

Schonberg, C. (2022). *Assessing the predictive validity of the IXL Real-Time Diagnostic using Star Assessments as criterion* (pp. 1–10). https://www.ixl.com/materials/us/research/Assessing_the_Predictive_Validity_of_the_IXL_Real-Time_Diagnostic_(Star).pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/Handbooks.

Appendix A: Descriptive Statistics

Table A1. IXL Math Usage Among IXL Diagnostic Snapshot and Non-Snapshot Cohorts During the 2021-22 School Year

Weekly IXL usage	Snapshot cohorts (*n* = 244)				Non-Snapshot cohorts (*n* = 244)			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Questions answered	38.22	33.00	0.00	158.01	17.54	20.84	0.00	136.26
Skills proficient	1.02	0.87	0.00	5.50	0.47	0.57	0.00	3.39
Time spent (in minutes)	13.87	12.46	0.00	83.35	7.46	9.35	0.00	52.15

Note. IXL uses a proprietary *SmartScore* to indicate a student's proficiency within a skill (i.e., a specific topic area within a subject). The SmartScore ranges from 0-100 and increases as students answer questions correctly; a SmartScore of 80 indicates skill proficiency.

Table A2. Sample Demographics

	Snapshot Cohorts	Non-Snapshot Cohorts
Pretest and posttest	*n* = 244	*n* = 244
2021 % proficient	39.10 (20.73)	36.79 (22.14)
2022 % proficient	43.90 (20.65)	39.77 (21.26)
Demographics		
Title I (*n* cohorts)	193	194
Locale (*n* cohorts)		
City	46	46
Suburb	87	80
Town	47	50
Rural	64	68
Grade level (*n* cohorts)		
3	53	52
4	52	38
5	46	50
6	40	41
7	27	32
8	26	31
School size (*n* students)	515 (278)	524 (291)
Student-teacher ratio	13.42 (2.74)	13.52 (3.31)
Gender: % male	51.69 (4.12)	51.96 (6.22)
Race: % White	60.06 (23.87)	63.28 (27.64)
% Economically disadvantaged	40.99 (23.96)	39.83 (21.97)
% English learners	6.88 (7.03)	6.60 (7.58)
% Special education	17.62 (6.49)	17.28 (6.41)

Note. Numbers in parentheses show standard deviations.

Appendix B: IXL Diagnostic Snapshot Analysis Results

Table B. Full Model Predicting 2022 State Assessment Proficiency Rate from Use of the IXL Diagnostic Snapshot and Covariates

Predictor	*b*	*SE*	95% CI	β	*t*	*p*
(Intercept)	44.73	2.20	40.51 – 48.88	0.14	20.326	<.001
Grade 4[1]	0.19	1.66	-3.08 – 3.35	0.01	0.113	.910
Grade 5[1]	0.10	1.66	-3.21 – 3.24	0.00	0.059	.953
Grade 6[1]	-0.97	1.78	-4.57 – 2.37	-0.05	-0.545	.586
Grade 7[1]	-2.42	2.04	-6.47 – 1.43	-0.12	-1.186	.236
Grade 8[1]	-8.25	2.18	-12.55 – -4.15	-0.39	-3.786	<.001
Race: % White[2]	-0.05	0.04	-0.11 – 0.02	-0.06	-1.312	.191
Gender: % male[2]	0.00	0.11	-0.20 – 0.21	0.00	-0.002	.998
% English learners[2]	-0.04	0.10	-0.23 – 0.15	-0.01	-0.435	.664
% Economically disadvantaged[2]	-0.26	0.04	-0.34 – -0.19	-0.29	-6.531	<.001
% Special education[2]	0.04	0.10	-0.16 – 0.23	0.01	0.342	.733
School size (*n* students)[2]	0.00	0.00	-0.01 – 0.00	-0.03	-0.902	.368
Student-teacher ratio[2]	0.31	0.23	-0.12 – 0.74	0.04	1.364	.174
Locale: rural[3]	-3.40	2.15	-7.47 – 0.72	-0.16	-1.581	.115
Locale: suburb[3]	-0.09	1.78	-3.45 – 3.36	0.00	-0.051	.959
Locale: town[3]	0.49	2.27	-3.71 – 4.98	0.02	0.216	.829
Title I school[4]	-2.68	1.68	-5.81 – 0.59	-0.13	-1.596	.112
2021 Math % proficient[2]	0.56	0.04	0.49 – 0.63	0.57	15.957	<.001
Completed IXL's Diagnostic Snapshot	**3.02**	**1.33**	**0.54 – 5.49**	**0.14**	**2.267**	**.028**

Note. Dependent variable: percentage of students reaching proficiency on the 2022 state math assessment. *b* = unstandardized regression coefficient, *SE* = standard error, CI = confidence interval, β = standardized regression coefficient.

[1] Dummy coded; grade 3 as reference group. [2] Grand-mean centered. [3] Dummy coded; city as reference group. [4] Dummy coded; non-Title I schools as reference group.