Does Computer-Adaptive Curriculum Enhance Students' Achievement in Math? A Quasi-Experimental Study

March 2021

Chalie Patarapichayatham, PhD

Southern Methodist University

Victoria N. Locke, PhD

Istation

Sean Lewis, MA

Istation

Citation: Patarapichayatham, C., Locke, V. N., & Lewis, S. (2021, Apr 8-12). *Does Computer-Adaptive Curriculum Enhance Students' Achievement in Math? A Quasi-Experimental Study* [Paper Session]. AERA Virtual Annual Meeting.

Abstract

This study evaluates the effects of a computer-adaptive curriculum on students' math achievement in kindergarten through third grade. Istation's computer-adaptive curriculum was a treatment, and Istation's Indicators of Progress (ISIP™) Math assessments were used to measure students' math achievement. Data were collected in the 2019-2020 school year under a quasi-experimental research setting. Analysis of covariance (ANCOVA) results show that the use of computer-adaptive curriculum makes a measurable difference in students' achievement in math. Students who spent more time on computer-adaptive curriculum outperformed students who spent less time or no time on computer-adaptive curriculum across grades. Students in lower grades (kindergarten and first grade) had significantly higher gain scores than students in higher grades (second and third grades).

Keywords: computer-adaptive curriculum, students' math achievement, quasi-experimental research design

Introduction

Curriculum is important in an education system, constituting a shared goal between instructors and learners, and aids in planning the education procedure. Curriculum is essentially a series of activities and learning outcomes. Effective curriculum ensures that important concepts are taught through lessons, everyday life experiences, collaborative activities, and active instruction. Curriculum ranges from lessons developed by teachers to be used in their classrooms to professionally published textbooks. An evidence-based curriculum consists of practices that have been vetted through rigorous research studies with the ultimate goal of increasing the likelihood of positive student outcomes.

Computer-adaptive curriculum is an educational method which uses computer algorithms to orchestrate the interaction with a student and deliver customized lessons and learning activities to fit the unique needs of each student. It derives from various fields including computer science, AI, psychometrics, education, psychology, and brain science. Each student receives a unique experience based on their learning ability. The computer-adaptive curriculum allows students to drive their own learning pace that fits best with their ability.

Computer-adaptive curriculum often operates with computerized adaptive testing (CAT). CAT is a form of computer-based test that adapts to the examinee's ability level. It is a form of computer-administered test in which examinee responses determine the next item or set of items presented. CAT has many advantages over paper-pencil testing especially from administration, logistic, and scoring procedure standpoints. CAT has been very well known in education field in the past few decades.

Yilmaz (2017) investigated the effects of a particular adaptive math learning program (ALEKS) on math achievement and its impact on closing the achievement gap in math performance of middle school students. Student achievement data were collected using Northwest Educational Association (NWEA) MAP math scores. A quasi-experimental research design was conducted with a sample size of 1,110 students in fifth to ninth grades forming a comparison group and a treatment group of equal sizes in the 2014-2015 school year using ANCOVA. Results showed that mathematics instruction via ALEKS had a statistically significant positive effect on students' math achievement and growth levels measured by a normative end-of-year mathematics assessment when beginning-of-year scores are held constant.

Outhwaite et al. (2019) used a randomized control trial to investigate an effect of interactive math curriculum apps with 389 children ages 4 and 5 in the United Kingdom. After a 12-week intervention, results showed significantly greater math learning gains for students in the treatment group. They recommended that structured, content-rich, interactive apps can provide a vehicle for efficiently delivering high-quality math instruction for all students in a classroom context and can effectively raise achievement in early math.

McTiernam et al. (2015) used a randomized controlled trial to evaluate the impact of a frequency-building curriculum to increase the fluency of component mathematics skills in a sample of 28 males ages 9 to 11. Statistically significant differences between the treatment group and control group were found on measures of fluency, endurance, stability, and one subtest of the Wechsler Individual Achievement Test® for mathematical ability.

These studies have found a positive effect of computer-adaptive curriculum in math on students' achievement under the randomized controlled trial design as well as a quasi-experimental setting. These studies focused on a small sample size, so the generalizability of the findings may be limited. In order to improve the generalizability across the United States, we evaluate the impact of a computer-adaptive curriculum on students' math achievement in kindergarten through third grade across the United States. Our research question is whether the use of computer-adaptive curriculum makes a measurable difference in students' achievement in math, and if it varies by grade level.

Methodology

Measure

To evaluate the impact of a computer-adaptive curriculum on students' math achievement in a quasi-experimental research design, we used the computer-adaptive curriculum from Istation as a treatment variable and student's math performance score as an outcome variable. Istation has an evidence-based computer-adaptive curriculum and computer-adaptive assessment (ISIP Math).

Istation curriculum usage (measured in minutes) was used as a treatment effect to assign students into a treatment or comparison group, and the ISIP Math assessment was used to measure students' achievement in math. ISIP Math scores from September 2019 were used as a covariate to control for students' prior achievement in math. ISIP Math scores from October 2019, January 2020, and March 2020 were used in our analyses to answer our research question.

Sample

The data for this study came from the Istation database. We selected students across the United States that had data in the 2019-2020 school year from kindergarten through third grade. While the Istation database is extensive, we did not use all available students as this may introduce some selection bias into the results. Typically, students from higher socioeconomic status (SES) households have higher achievement than students from lower SES households (Lewis et al., 2019; Locke et al., 2021). Since the Istation database is slightly skewed toward schools that have higher percentages of students receiving free or reduced priced lunch (which is an indicator of lower SES) and this may impact the results, we selected a stratified sample according to SES at the school level. Sample stratification is a process of dividing members of the population into homogeneous subgroups before sampling. A stratified sample could thus claim to be more representative of the population than a simple random sampling or systematic sampling.

We used a school SES level variable to stratify samples in this study. We created four categories for SES, using the categories from the National Center for Education Statistics (NCES). SES 1 consists of schools that have 75% or more of their students enrolled in the free or reduced priced lunch (FRPL) program. SES 2 schools have 50% to 74.9% enrolled in FRPL, and SES 3 schools have 25% to 49.9% of students enrolled in FRPL. SES 4 schools have less than 25% of students enrolled in FRPL. Next, we calculated the percentage of students that were enrolled in each of the four SES levels according to enrollment data available from NCES for public and public charter schools and used this information to create sample targets. We selected 10,000 students per grade. Within each grade, 36% of these students were from SES 1, 16% from SES 2, 20% from SES 3, and 28% from SES 4.

In each school year, students take ISIP Math monthly from September to May. Some students take ISIP Math only at benchmarking assessment months three times a year, typically September, January, and May. Unfortunately, schools across the country closed around mid-March of 2020 due to the COVID-19 pandemic, and a majority of students stopped using the program. We decided to choose September, October, January, and March assessment scores for this study. Because October and March are not benchmarking assessment months, we encountered missing data for these months.

Missing data is not unusual in educational research, especially for longitudinal data. It is normal that some students do not assess every single month for a variety of reasons. There are many different ways to deal with missing data. Mean imputation is a popular method in educational research. However, once the missing values are replaced with the mean of the variable, the variability is reduced. The correlation between variables may be affected, and thus, results may be biased. A regression imputation method imputes a value that is predicted by a regression model using the expected values; however, the correlation will be overestimated, and the results may again be biased. Another popular method for a longitudinal data study is the last observation carries forward method where the missing data are replaced by the last observation. For example, if a student has a September score but the October score is missing, the October score is replaced by the September score. This method would definitely provide biased results because it would pull students' October mean scores down. It is not an appropriate method in a growth analysis. Keeping missing data as they are is another popular method in educational research. However, it is not an option in this study because this method does not allow us to compute the gain score between two data points at a student level.

Although we did not test whether the missingness in our data are missing completely at random (MCAR) or not, we assume that our data are MCAR by nature, and a simple list-wise deletion can be justified. With a list-wise deletion, students with any missing data on these variables in the analysis model are excluded. Keeping only students with complete data points may be somewhat biased. However, it allows us to compute the gain scores between two data points without implementing any imputation methods in the study such as a gain score from October to January or a gain score from October to March. For this reason, students with complete data points are kept. A final sample consisted of 25,574 students across grades. There were 4,402 in kindergarten, 4,755 in first grade, 4,691 in second grade, and 4,595 in third grade.

Because the sample size was reduced, we evaluated whether the final samples in each SES level still represent the population based on our stratification. Results show that our final samples represent our population and are shown in Table 1. Students in SES 1 increased from about 36% in the target sample to about 38% in the final sample, and students in SES 2 increased from 16% in the target sample to about 20% in the final sample. Students in SES 3 remain the same, whereas students in SES 4 decreased. Table 2 shows mean scores by grade.

Table 1 Final Sample

Grade	Target Sample				Final Sample			
	SES 1	SES 2	SES 3	SES 4	SES 1	SES 2	SES 3	SES 4
K	36%	16%	20%	28%	36%	19%	20%	25%
1	36%	16%	20%	28%	39%	19%	20%	22%
2	36%	16%	20%	28%	38%	21%	22%	19%
3	36%	16%	20%	28%	38%	21%	23%	18%

Table 2: Students' Mean Scores by Grade

Grade	Sample	September Mean Score (SD)	October Mean Score (SD)	January Mean Score (SD)	March Mean Score (SD)	Sept. to Oct. Gain Score	Sept. to Jan. Gain Score	Sept. to March Gain Score
K	4,402	303.70 (54.41)	318.84 (63.24)	361.86 (76.19)	377.44 (73.22)	15	58	74
1	4,755	393.47 (47.40)	410.23 (54.37)	445.51 (60.19)	461.20 (61.27)	17	52	68
2	4,691	477.75 (28.86)	483.42 (34.01)	500.27 (38.13)	510.49 (38.94)	6	23	33
3	4,595	505.30 (31.40)	511.88 (37.63)	525.05 (44.10)	532.35 (44.82)	7	20	27

In kindergarten, students grew 15 ISIP Math points from September to October. They grew 58 points from September to January and 74 points from September to March. In first grade, students grew 17, 52, and 68 points from September to October, September to January, and September to March, respectively. Second and third grade students had a similar growth pattern from September to October. Second grade students grew 6 points, and third grade students grew 7 points. The gain scores were 23 and 20 points for second and third grade students from September to January. Moreover, second grade students grew 33 points, and third grade students grew 27 points from September to March of the 2019-2020 school year.

Students' computer-adaptive curriculum usage was measured from zero minutes to 3,000 minutes in this study based on the total time students spent on Istation math curriculum. Zero-minute computer-adaptive curriculum usage means a student did not use the curriculum. The usage means were 548.28 minutes for kindergarten, 644.87 minutes for first grade, 636.70 minutes for second grade, and 568.47 minutes for third grade. We classified students into four different groups (quartiles) based on their computer-adaptive curriculum usage. They were in first, second, third, and fourth quartiles. Results of students' mean scores of September, October, January, and March by grade and by quartile of computer-adaptive curriculum usage are available in Table 3.

Table 3: Students' Mean Scores by Grade and by Computer-Adaptive Curriculum Usage Quartile

Grade	Usage Quartile	Sample	September Mean Score (SD)	October Mean Score (SD)	January Mean Score (SD)	March Mean Score (SD)	Sept. to Oct. Gain Score	Sept. to Jan. Gain Score	Sept. to March Gain Score
K	1	1,303	301.76 (53.08)	314.98 (62.15)	358.17 (73.09)	373.48 (71.69)	13	56	72
	2	1,121	302.63 (53.83)	315.37 (62.72)	354.65 (76.60)	369.49 (72.26)	13	52	67
	3	1,000	306.58 (55.09)	321.77 (63.66)	365.65 (77.59)	381.98 (73.24)	15	59	75
	4	978	304.60 (56.04)	324.96 (64.29)	371.15 (77.27)	387.20 (74.96)	20	67	83
1	1	1,079	391.11 (46.95)	407.88 (54.48)	443.02 (58.93)	457.65 (60.60)	17	52	67
	2	1,136	388.54 (45.25)	404.84 (51.42)	437.97 (58.03)	453.01 (59.04)	16	49	64
	3	1,180	393.82 (46.43)	411.76 (54.44)	446.68 (59.15)	462.79 (60.48)	18	53	69
	4	1,360	399.15 (49.72)	415.28 (56.10)	452.77 (62.98)	469.49 (63.23)	16	54	70

2	1	990	452.74 (31.31)	460.08 (35.48)	475.58 (42.71)	479.14 (40.87)	7	23	26
	2	1,117	453.95 (31.39)	461.68 (36.49)	479.33 (41.91)	483.20 (40.24)	8	25	29
	3	1,212	452.46 (31.00)	461.44 (36.09)	480.33 (41.50)	482.25 (39.21)	9	28	30
	4	1,372	454.29 (31.99)	463.32 (36.26)	483.91 (42.56)	486.12 (40.81)	9	30	32
3	1	1,107	476.80 (27.48)	483.07 (33.23)	499.44 (35.48)	508.55 (37.46)	6	23	32
	2	1,166	475.94 (29.03)	481.44 (34.15)	496.86 (37.39)	507.67 (37.58)	6	21	32
	3	1,253	478.79 (29.30)	483.78 (34.65)	501.45 (39.99)	512.35 (40.44)	5	23	34
	4	1,069	479.47 (29.44)	485.52 (33.83)	503.46 (39.04)	513.38 (39.83)	6	24	34

It is evident that the data were a little skewed at the computer-adaptive curriculum usage variable especially in quartile 1 in kindergarten and quartile 4 in first and second grades. It is evident that students improved their performance from September to October, January, and March across grades. Kindergarten and first grade students showed larger gain scores than second and third grade students over time. For example, in the first quartile of kindergarten, gain scores were 13, 56, and 72 for September to October, September to January, and September to March, respectively. On the other hand, in the first quartile of third grade, gain scores were 6, 23, and 32 for September to October, September to January, and September to March, respectively. It is also evident that students who spent more time on the adaptive curriculum scored higher in ISIP Math. For example, second grade students' gain scores from September to October were 7, 8, 9, and 9 for quartiles 1 to 4, respectively. Gain scores from September to January were 23, 25, 28, and 30, and gain scores from September to March were 26, 29, 30, and 32 for quartiles 1 to 4, respectively.

Model and Analysis

A true experimental design is powerful in educational and psychological studies. However, it can be difficult to implement in the classroom because it requires a fully random assignment between treatment and comparison groups, which many schools and teachers see as unfair. Like a true experimental design, a quasi-experimental design aims to establish a cause-and-effect relationship between an independent variable and a dependent variable. A quasi-experiment design, however, does not rely on random assignment. Subjects are assigned to groups based on non-random criteria. Quasi-experimental designs are a useful tool in situations where true experiments cannot be used for ethical or practical reasons.

A researcher's role is quite different in a true experimental design and a quasi-experimental design. A researcher usually designs the treatment and decides which subjects would receive it under a true experimental design. A researcher under a quasi-experimental design, on the other hand, often does not have control over the treatment. Instead, he or she studies subjects that received or did not receive treatments after the fact. Without randomization in a quasi-experimental design, it can be difficult to verify that confounding variables have been accounted for. However, a quasi-experimental design often provides higher external validity than a true experimental design because it often involves real-world interventions instead of artificial laboratory settings. It also provides higher internal validity than a non-experimental design study because it allows researchers to better control for confounding variables in the study.

Using the data from the Istation database, it is almost impossible to implement a true experimental design because we did not have a control over the treatment. A quasi-experimental design is an option for us because it allows us to investigate students who received or did not receive the treatment after the fact. For this reason, this study implements a quasi-experimental design to investigate whether the use of the computer-adaptive curriculum in math shows a measurable difference in students' achievement in math.

Next, we used a quartile of students' computer-adaptive curriculum usage to classify students into a treatment group or a comparison group. Students in the first and second quartiles of computer-adaptive curriculum usage were assigned to a comparison group. Students in the third and fourth quartiles of computer-adaptive curriculum usage were assigned to a treatment group. Students' mean scores from September, October, January, and March by grade in a treatment and a comparison group are available in Table 4.

We considered propensity score matching (PSM) to further control for equivalent groups between treatment and comparison. We considered two covariates, school SES and September score, because they were assumed to be related to the outcome measure (e.g., Brookhart et al., 2006; Rubin, 1997). Because school SES is already used in our stratified sample procedure and September score is used as a covariate for analyses, PSM is not needed.

Table 4: Students' Mean Scores by Grade and by Group

Grade	Group	Sample	September Mean Score (SD)	October Mean Score (SD)	January Mean Score (SD)	March Mean Score (SD)	Sept. to Oct. Gain Score	Sept. to Jan. Gain Score	Sept. to March Gain Score
K	Comparison	2,424	302.16 (53.42)	315.16 (62.40)	356.54 (74.74)	371.63 (71.97)	13	54	69
	Treatment	1,978	305.60 (55.56)	323.35 (63.97)	368.37 (77.46)	384.56 (74.12)	18	63	79
1	Comparison	2,215	389.79 (46.10)	406.32 (52.94)	440.43 (58.51)	455.27 (59.84)	17	51	65
	Treatment	2,540	396.67 (48.28)	413.64 (55.37)	449.94 (61.30)	466.38 (62.04)	17	53	70
2	Comparison	2,107	453.38 (31.35)	460.93 (36.02)	477.56 (42.32)	481.29 (40.58)	8	24	28
	Treatment	2,584	453.43 (31.54)	462.44 (36.19)	482.23 (42.10)	484.31 (40.11)	9	29	31
3	Comparison	2,273	476.36 (28.28)	482.23 (33.71)	498.12 (36.49)	508.10 (37.52)	6	22	32
	Treatment	2,322	479.10 (29.36)	484.58 (34.28)	502.37 (39.56)	512.82 (40.16)	5	23	34

Students in the treatment groups scored higher at each assessment month than students in comparison groups across grades. Students in the treatment groups had higher gain scores than students in the comparison groups across grades as well. In kindergarten, students in the treatment group scored 5 ISIP Math points higher than the comparison group (18 vs. 13) from September to October. They scored 9 points higher (63 vs. 54) from September to January and 10 points higher (79 vs. 69) from September to March. First grade students in treatment group scored 2 points higher than the comparison group (53 vs. 51) from September to January and 5 points higher (70 vs. 65) from September to March. In second grade, students in the treatment group scored 1, 5, and 3 points higher September to October (9 vs. 8), September to January (29 vs. 24), and September to March (31 vs. 28), respectively. Also, third grade students in the treatment group scored 1 and 2 points higher from September to January (23 vs. 23) and September to March (34 vs. 32), respectively.

Results so far show that the use of computer-adaptive curriculum in math makes a difference in students' achievement in math. Students who spent more time in computer-adaptive curriculum scored higher in ISIP Math across grades and across assessment months. Although results are encouraging, we did not control for students' prior achievement. To control for students' prior achievement in order to really investigate the effect of treatment, ISIP Math scores from September were used as a covariate in an ANCOVA analysis.

Analysis of covariance (ANCOVA) is a general linear model which blends ANOVA and regression model together. It is a mixture of continuous and categorical predictors. ANCOVA evaluates whether the means of a dependent variable (DV) or outcome variable are equal across levels of a categorical independent variable (IV) or predictor, while statistically controlling for the effects of other continuous variables that are not of primary interest, known as covariates (CV) or nuisance variables. ANCOVA assumes there is no interaction between the predictor and the covariate. If the covariate is significant and the predictor is significant, then we have enough evidence to say that there is a statistically significant difference between groups or levels when controlling for the covariate.

Nested data structures are very common in educational settings. Hierarchical models, such as hierarchical linear modeling (HLM), are suitable for a nested data structure. HLM is a regression-based analysis that takes the hierarchical structure of the data into account. Hierarchically structured data is nested data where groups of units are clustered together, such as students clustered within classrooms within schools. We took steps to control for the nested nature of the data. First, we used R statistical software to conduct a random selection of observations, independent of the school. We selected these cases within the four levels for the type of school described above. Since we stratified the sample based on a school SES variable and selected the cases in a random fashion, the school level variability is controlled, and a single level analysis such as ANCOVA is justified.

Two ANCOVA models were fit using R. Student group (i.e., treatment vs. comparison) was a predictor, and the September score was a covariate for both models.

The January score was the outcome variable for model 1, and the March score was the outcome variable for model 2. Results are available in Table 5.

Results

In model 1, the group of students (treatment vs. comparison) was a predictor, the September score was a covariate, and the January score was an outcome variable. ANCOVA results show that there is a statistically significant difference in the mean score for January and that students in the treatment group scored higher than students in the comparison group across all grades, after controlling for their prior achievement. In kindergarten, students in the treatment group scored 8 ISIP Math points higher (65 vs. 57) than students in the comparison group. First grade students in the treatment group scored 3 points higher (59 vs. 56) than students in the comparison group. Second grade students in the treatment group scored 5 points higher (29 vs. 24) than students in the comparison group, and third grade students in the treatment group scored 2 points higher (24 vs. 22) than students in the comparison group.

In model 2, the group of students (treatment vs. comparison) was a predictor, the September score was a covariate, and the March score was an outcome variable. ANCOVA results in model 2 show a similar pattern to model 1 results. Again, there is a statistically significant difference in the mean score of March. Students in the treatment group score significantly higher in March than students in the comparison group across all grades after controlling for their prior achievement. In kindergarten, students in the treatment group scored 10 points higher (98 vs. 88) than students in the comparison group. First grade students in the treatment group scored 4 points higher (76 vs. 72) than students in the comparison group. Second grade students in the treatment group scored 3 points higher (44 vs. 41) and third grade students in the treatment group scored 2 points higher (46 vs. 44). These values in models 1 and 2 could be interpreted as a gain score from the beginning-of-the-year assessment month to the middle-of-the-year assessment month as well as to March.

Table 5: Models 1 and 2 Results

Model	Grade		Estimate	SE	*t-value*	*p-value*
1	K	Intercept	57.12	4.61	12.38	<0.001
		Treatment	8.42	1.62	5.19	<0.001
		September score	0.99	0.01	66.77	<0.001
	1	Intercept	55.74	4.58	12.16	<0.001
		Treatment	2.72	1.11	2.47	<0.001
		September score	0.99	0.019	85.22	<0.001
	2	Intercept	24.01	5.95	4.03	<0.001
		Treatment	4.62	0.83	5.59	<0.001
		September score	1.00	0.013	76.62	<0.001
	3	Intercept	22.41	6.82	10.18	<0.001
		Treatment	1.79	0.82	2.17	0.030
		September score	0.89	0.01	63.11	<0.001
2	K	Intercept	88.02	4.49	19.60	<0.001
		Treatment	9.69	1.58	6.14	<0.001
		September score	0.94	0.01	64.99	<0.001
	1	Intercept	71.71	4.79	14.95	<0.001
		Treatment	4.33	1.15	3.77	<0.001
		September score	0.98	0.01	81.19	<0.001
	2	Intercept	41.13	5.57	7.38	<0.001
		Treatment	2.96	0.77	3.84	<0.001
		September score	0.97	0.01	79.41	<0.001
	3	Intercept	44.09	6.57	6.70	<0.001
		Treatment	2.05	0.79	2.58	<0.001
		September score	0.97	0.01	70.81	<0.001

Overall, the findings confirm the impact of a computer-adaptive curriculum in math on students' math achievement. The use of computer-adaptive curriculum makes a measurable difference in students' achievement in math. Students who spent more time on the computer-adaptive curriculum outperformed students who spent less time or no time on computer-adaptive curriculum across grades. Students in younger grades (kindergarten and first grade) had significantly higher gain scores from beginning of the year to January or March than students in higher grades (second and third grades).

Discussion

This study demonstrates that the computer-adaptive curriculum in math had positive effects on students' math achievement. Our results confirm findings from McTiernam et al. (2015), Yilmaz (2017), and Outhwaite et al. (2019). McTiernam and co-authors found statistically significant differences between the treatment group and control group on the impact of a frequency-building curriculum in math. Yilmaz found that mathematics instruction via ALEKS has a statistically significant positive effect on students' math achievement and growth levels measured by a normative end-of-year mathematics assessment when beginning-of-the-year scores are held constant. Outhwaite and co-authors found significantly greater math learning gains for the treatment group after 12 weeks of intervention with interactive math curriculum apps. The result also implies that Istation computer-adaptive curriculum in math helps students on their math achievement.

The result of this study implies that students who spend more time studying math lessons will have higher scores in math. This study also confirms that online assessment and computer-adaptive curriculum are practical since students could access lessons from anywhere at anytime with an internet connection. The online assessment and computer-adaptive curriculum could help students remain on track regardless of whether they physically attend instruction at school.

Limitations and Recommendations

While technology cannot replace a face-to-face classroom setting, it is important to understand that computer-adaptive curriculum can play an important role especially during the COVID-19 pandemic. Use of computer technology increased during the pandemic when schools closed or used remote or hybrid (in-person and remote) methods for teaching young students. This study confirms that computer-adaptive curriculum in math improves students' math achievement. Although this study investigated a computer-adaptive curriculum's effect on a student's achievement in math, the results may be generalized to other subjects as well. Other research on the use of technology in reading shows that it also helps students make greater gains and can help narrow achievement gaps (Sutter et al., 2019).

When using educational technology, it is important to note that most of these programs will provide usage recommendations. The cut-off point between the two groups closely resembles the Istation usage recommendation of 30 minutes per week. Schools that implemented close to the recommended usage levels or more saw greater gains than schools that did not, thus validating that implementing educational technology with fidelity will improve student outcomes.

There are a few limitations in this study. First, we investigated students in kindergarten through third grade. Results may be different in higher grades. Second, there are many other factors that could affect students' achievement in math from online assessment besides computer-adaptive math curriculum such as student engagement or mental health. Third, while ANCOVA analysis is quite well known in educational research, a longitudinal growth model may provide additional information which ANCOVA could not obtain. A multiple-group, piecewise growth model would be very suitable in this scenario because a piecewise growth model allows several segments of the slopes in a model. This model would also allow us to fit the model for both treatment and comparison groups simultaneously. It is a better way to control for measurement bias.

We also note that we used data from the 2019-2020 school year, which was disrupted by the COVID-19 pandemic school closures. Most schools in the United States closed in March 2020 and went to remote learning. While Istation made the assessment available for home progress monitoring, we did not use the April and May scores out of concern that they would bias the sample and vastly increase the amount of missing data. Evaluating the impact of the curriculum for an entire school year may change the results.

References

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*, 1149–1156.

Cook, T. D., Campbell, D. T. (1979). Quasi-experimentation: design & analysis issues for field settings. Houghton Mifflin company, Boston.

Lewis, S., Locke, V. N., & Patarapichayatham, C. (2019). Student engagement in online learning during COVID school closures predicts lower learning loss in Fall 2021. Dallas, TX: Istation.

Locke, V. N., Patarapichayatham, C., & Lewis, S. (2021). Learning Loss in Reading and Math in U.S. Schools Due to the COVID-19 Pandemic. Dallas, TX: Istation.

McTiernam, A., Holloway, J., Healy, O. & Hogan, M. (2015). A randomized controlled trial of the morningside math facts curriculum on fluency, stability, enhance and application outcomes. J Behav Educ (2016) 25:49–68.

Outhwaite, L. A., Faulder, M., Gulliford, A., & Pitchford, N. J. (2019). Raising early achievement in math with interactive apps: a randomized control trial. Journal of Educational Psychology, Vol. 111, No. 2, 284–298.

Patarapichayatham, C., Fahle, W., & Roden, T. R. (2013). ISIP Reading versus STAAR Reading: The Predictability Study. Dallas, TX: Istation.

Patarapichayatham, C., & Locke, V. N. (2019). An evaluation of Istation curriculum on student reading growth: quasi-experimental study using propensity score analysis. Dallas, TX: Istation.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, *127*(8_Part_2), 757–763.

Sutter, C. C., Campbell, L. O. & Lambie, G. W. (2019). Computer-adaptive reading to improve reading achievement among third-grade students at risk for reading failure. Journal of At-Risk Issues, 22, 2, 31-38.

Yilmaz, B. (2017). Effects of adaptive learning technologies on math achievement: a quantitative study of Aleks math software. Dissertation from University of Missouri, Kansas City.