# AN EVALUATION OF THE EFFECTIVENESS OF LEARNBOP ON MATH SKILLS GROWTH IN FLORIDA CLASSROOMS

February 2017
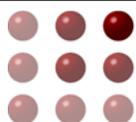
SEGMEASUREMENT
BUILDING BETTER ASSESSMENTS / EVALUATING PRODUCT EFFECTIVENESS

Contents

Executive Summary

Overview

During the 2015-2016 school year, SEG Measurement conducted a study of the effectiveness of LearnBop, an online system that provides individualized adaptive feedback and breaks down problems into smaller concepts to help students learn critical math skills.

After using a statistical technique to create matched groups of students who used LearnBop and students who did not use LearnBop, 438 students in sixth and eighth grade Florida classrooms participated in the study. Students in Florida using LearnBop showed significantly greater improvement in mathematics skills than students in Florida classrooms that did not use LearnBop. The effect size for sixth graders was .45 and the effect size for eighth graders was .28.

Study Design

The study employed a quasi-experimental design, comparing the growth in mathematics skills between two groups of students, students who used LearnBop as part of their mathematics instruction (treatment group) and comparable students who did not use LearnBop as part of their mathematical instruction (control group). The growth in mathematics skills was assessed using state assessments. The pretest was the Spring 2015 administration and the posttest was the Spring 2016 administration.

Treatment and control group participants were statistically matched using Propensity Score Matching. First, the schools were matched on attributes including average prior ability math level, enrollment size, poverty level, and type of school. Second, the students were matched within the matching schools on student characteristics including prior math performance level, gender, ethnicity, and eligibility for free or reduced lunch. For each student who used LearnBop, a matching student who did not use LearnBop was determined. Only these matched students were included in the analyses. This statistical matching provides increased rigor in the analyses and helps to control for factors beyond the product that may have an influence on performance.

The mathematics knowledge and skills growth in the treatment group and the control group was compared statistically using Analysis of Covariance (ANCOVA). ANCOVA provides a comparison between the treatment and

"Students in Florida classrooms using LearnBop showed significantly greater improvement in mathematics skills than matched students in Florida classrooms that did not use LearnBop."

"Sixth graders using LearnBop showed more than twice the mathematics growth that is typically seen and eighth grade LearnBop users showed 88% more growth than is typically seen on national assessments."

control group students, while adjusting for any potential differences in students' initial ability even though they were controlled for in the propensity score matching process. Specifically, we examined the difference in the Spring 2016 scores (dependent variable) between the treatment and control groups (independent variable) while controlling for the initial ability of the students from Spring 2015 (covariate).

Study Results

The analysis indicated that students that used LearnBop showed significantly more growth in math skills than comparable classrooms that did not use LearnBop. Sixth grade students using LearnBop showed about 10 points more growth on the assessment, or .45 standard deviations, than did sixth graders not using LearnBop (Effect Size=.45; F=65.895; p<.001). Eighth graders showed about 6 points more growth on the assessment, or .28 standard deviations, than did nonusers (Effect Size=.28; F=6.372; p=.013). For a student in sixth grade at the 50th percentile, this represents a gain to the 67th percentile. For a student in the eighth grade at the 50th percentile, this represents a gain to the 61st percentile.

The average (mean) mathematics test scores for the treatment and control group students are shown in Figures 1 and 2.

Figure 1: Comparison of Sixth Grade Posttest Scores for Treatment and Control Groups (Adjusted Means)
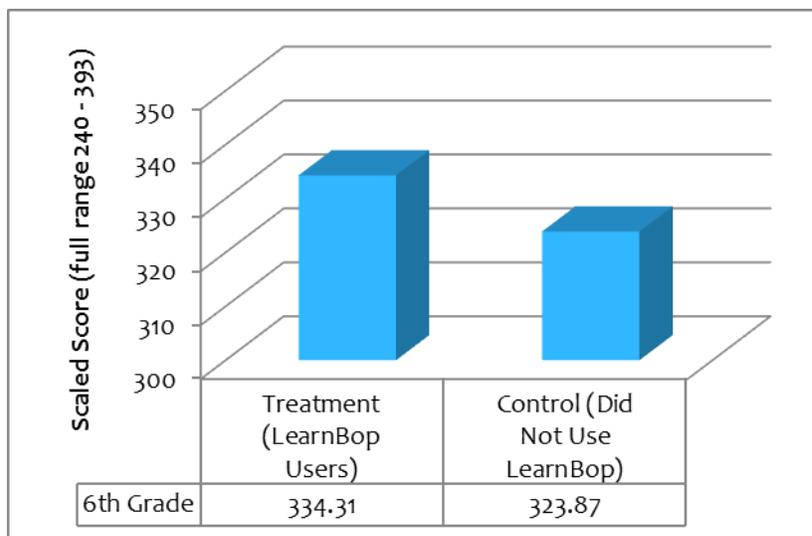
	Treatment (LearnBop Users)	Control (Did Not Use LearnBop)
6th Grade	334.31	323.87

Scaled Score (full range 240 - 393)

"Sixth graders using LearnBop showed about 10 points more growth on the assessment, or .45 standard deviations, than did sixth graders not using LearnBop . Eighth graders showed about 6 points more growth on the assessment, or .28 standard deviations, than did nonusers."

"The results support the effectiveness of LearnBop use in improving sixth and eighth grade students' math skills."

Figure 2: Comparison of Eighth Grade Posttest Scores for Treatment and Control Groups (Adjusted Means)
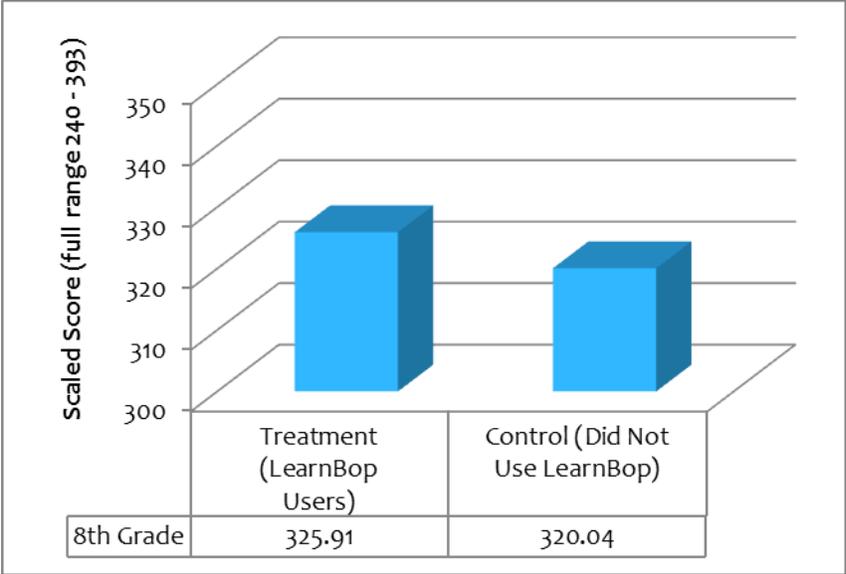
The results support the effectiveness of LearnBop use in improving sixth and eighth grade students' math skills.

Introduction

This study examines the impact of LearnBop on the math skills of sixth and eighth grade students in Florida. LearnBop is an online math product that provides one-on-one individualized and adaptive math tutoring on specific skills and concepts.

The study investigated the use of LearnBop during the 2015-2016 school year in 6th and 8th grade classes in a large district in Florida. Student learning in mathematics in classes using LearnBop (treatment group) was compared to student learning in mathematics in classes that did not use LearnBop (control group). State math assessments were used as the independent measures of students' mathematics skills.

About LearnBop

LearnBop uses dynamic math problems, called Bops, to help students learn standards aligned math concepts. Unlike most math problems, which start and end with one question, Bops are dynamic responding in real-time to student input. If a student asks for help or answers a question incorrectly, the initial problem is broken down into smaller, more manageable steps. Every step is connected to a standard aligned concept, allowing the teacher to see exactly which prerequisite concepts the student is missing.

LearnBop tracks every student's progress toward mastery as they learn. Mastery is the probability the student truly understands the concept. With the LearnBop dashboard, teachers have easy and quick access to critical data and powerful tools to inform instruction and guide learning. LearnBop enables teachers to see a holistic picture of their students' progress within and across grades while identifying mastery levels of standard aligned math concepts.

Study Findings

Research Questions

The primary research question addressed by this study focused on the effectiveness of LearnBop.

- Do students who use LearnBop show larger gains in mathematics skills than comparable students who do not use LearnBop?

Study Design

The study employed a pre/post treatment/control group comparison design. Since the students were not randomly assigned to the groups, this is considered a quasi-experimental design. The treatment group classes used LearnBop, while the control group classes used traditional mathematics instruction without the use of LearnBop. The study design is depicted in Figure 3.

Figure 3: Study Design
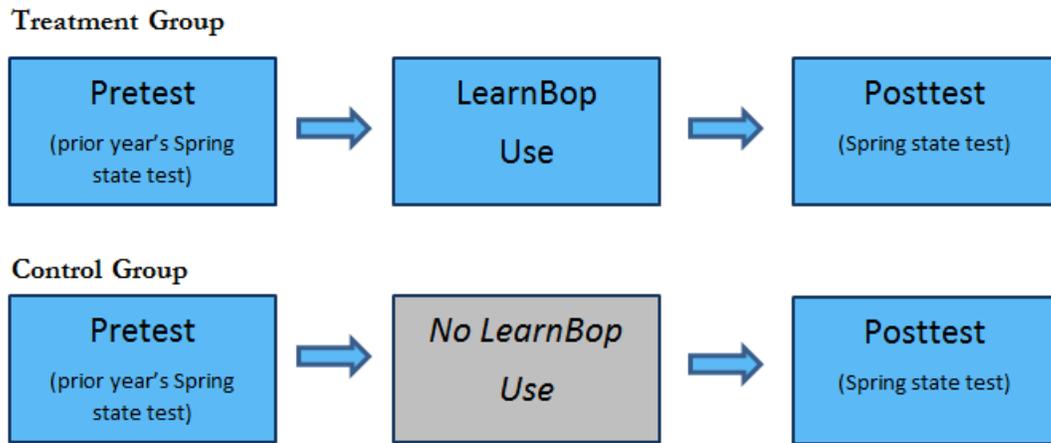
Data Collection

The district and SEG Measurement worked together to develop the specifications for a data export that would include all of the necessary data points to conduct the study while protecting the identity of the participants. The district provided the de-identified state test performance data for Spring 2015 and Spring 2016 for each student and also matched the performance data to LearnBop account information.

The LearnBop online system was used to track the number of logins for each class, the amount of time spent using LearnBop, the number of concepts mastered, and teacher logins. This usage information was used to determine which students were using LearnBop during the school year.

Participants

Students in classes in the district that did not use LearnBop composed the pool of possible Control Group students. Students in classes that used LearnBop a minimum of 120 minutes composed the Treatment Group. Students in classes that did not meet the minimum usage were excluded from the analyses.

The final set of participants were selected using a statistical matching technique called Propensity Score Matching. First, the schools were matched on attributes including average prior ability math level, enrollment size, poverty level, and type of school. Second, the students were matched within the matching schools on student characteristics including prior math performance level, gender, ethnicity, and eligibility for free or reduced lunch. For each student who used LearnBop, a matching student who did not use LearnBop was determined. Only these matched students were included in the analyses. The use of Propensity Score Matching provides increased rigor in the analyses and helps to control for factors beyond the product that may have an influence on performance. Teacher effectiveness data were not available for use in matching students.

There were twelve schools represented in these analyses. 438 students in were included in these analyses. The treatment group contained 219 students and the control group contained 219 students, with one control student matching each unique treatment student. The profile summaries of the students by grade are provided in Table 1. Some students were associated with multiple teachers and teacher demographics are not able to be reported.

Table 1: Profile of Participating Students

	Grade 6		Grade 8	
	Treatment	**Control**	**Treatment**	**Control**
Gender				
Female	54%	54%	48%	48%
Male	46	46	52	52
Ethnicity				
African American	9	9	47	47
Asian or Pacific Islander	3	3	0	0
Caucasian	54	54	12	12
Hispanic	32	32	41	41
Mixed Race or Other	2	2	0	0
Eligibility for Free or Reduced Lunch				
Yes	52	52	86	86
No	48	48	14	14

Description of the Outcome Measure

The Florida Standards Assessments were used to measure mathematics skills. All schools in Florida teach the Florida Standards which are measured through the Florida Standard Assessments. The grade 6 assessment includes items measuring ratio and proportional relationships, expressions and equations, geometry, statistics and probability, and the number system. The 8th grade assessment includes items measuring expressions and equations, functions, geometry, and statistics and probability and the number system. The assessments are administered via computer based testing.

Description of the Treatment

LearnBop uses dynamic math problems, called Bops, to help teach students standards aligned math concepts. Unlike most math problems, which start and end with one question, Bops are dynamic responding in real-time to student input. If a student asks for help or answers a question incorrectly, the initial problem is broken down into smaller, more manageable steps. Every step is connected to a standard aligned concept, allowing the teacher to see exactly which prerequisite concepts the student is missing.

The concepts covered within LearnBop for grade 6 include expressions and equations, geometry, ratios and proportional relationships, statistics and probability, and number system. The treatment group classes in this study used LearnBop at least 120 minutes and the average amount of use in sixth grade was 239 minutes. Among these classes, LearnBop was used an average of 60 minutes per week for up to 12 weeks.

The concepts covered within LearnBop for grade 8 include expressions and equations, functions, geometry, statistics and probability, and number system. The treatment group classes in this study used LearnBop at least 120 minutes and the average amount of use in eighth grade was 416 minutes. Among these classes, LearnBop was used an average of 112 minutes per week for up to 6 weeks.

Initial Comparability of the Groups

Propensity Score Matching was used to select comparable groups. A two step process was used. First, control schools that were similar to treatment schools were identified. School attributes used for matching included enrollment size, poverty level, prior ability level, and type of school. Second, students within the matched schools were matched such that each treatment student had a matching control student with similar characteristics including prior math ability level, gender, ethnicity, and eligibility for free or reduced lunch. All of the students included in the process were required to have Spring 2015 and Spring 2016 math test scores.

Even with the use of Propensity Score Matching, it is important to verify that the selected groups are indeed comparable. Demonstrating baseline equivalence of the treatment and control groups minimizes potential bias from selection in quasi-experimental designs that can alter effect size estimates. If the treatment group and the control group are not similar, we cannot be sure if the growth we see is due to the treatment (in this case, use of LearnBop) or the result of some differences in the individuals that existed before we conducted the study.

Pretest score (Spring 2015 state test) performance was used to compare the initial ability levels for students in the treatment and control groups. The treatment and control groups were found to have no statistical differences in initial ability (Grade 6: $F = 0.003$, $p = .957$, Grade 8: $F=0.641$, $p=.425$). Even though Propensity Score Matching was used and initial ability levels were comparable, analysis of covariance was used to allow for comparisons of Spring 2016 performance while accounting for Spring 2015 performance.

Analysis and Findings

The mathematics knowledge and skills of students using LearnBop (treatment group) was compared to the mathematics knowledge and skills of students who did not use LearnBop (control group).

Using Analysis of Covariance (ANCOVA), we examined the difference in the posttest scores (dependent variables) between the treatment and control groups (independent variable) controlling for the initial proficiency of the students (covariate). The Spring 2015 score was used as the covariate to place students in the treatment group and the control group on the same baseline. These analyses were run separately for each grade.

Grade 6 Math Skills Results

The results show an effect size of .45 for the state assessment. Sixth grade students who used LearnBop achieved significantly higher scores on the state assessment of math skills than students who did not use LearnBop (F = 65.895, p < .001). For a student at the 50th percentile, an effect size of .45 would produce a gain to the 67th percentile. The results are summarized in Tables 2 and 3 below.

Table 2: Analysis of Covariance of the Treatment and Control Group 6th Grade Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	120284.222	2	60142.111	498.404	<.001
Intercept	265.292	1	265.292	2.199	.139
Pretest	112141.195	1	112141.195	929.326	<.001
Study Group	7951.530	1	7951.530	65.895	<.001
Error	34873.463	289	120.669		
Total	31778640.000	292			
Corrected Total	155157.685	291			

**Table 3: Descriptive Statistics Comparison of the Treatment and Control Group
6th Grade Posttest Scores (Adjusted for Pretest Performance)**

Group	Number of Students	Posttest Scores	
		Mean	Standard Deviation
Treatment	146	334.31	20.38
Control	146	323.87	24.47

The passing score for the sixth grade state math assessment is 325. The average score of the treatment group exceeds this passing score and falls within the "Level 3 – Satisfactory" score range (325 – 338) while the average score of the control group does not meet the passing standard and falls within the "Level 2 – Below Satisfactory" score range (310-324). The average score of the sixth graders in this school district as reported by the state was 322.

Grade 8 Math Skills Results

The results show an effect size of .28 for the state assessment. Eighth grade students who used LearnBop achieved significantly higher scores on the state assessment of math skills than students who did not use LearnBop (F = 6.372, p = .013). For a student at the 50th percentile, an effect size of .28 would produce a gain to the 61st percentile. The results are summarized in Tables 4 and 5 below.

Table 4: Analysis of Covariance of the Treatment and Control Group 8th Grade Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	37688.095	2	18844.048	95.966	<.001
Intercept	174.742	1	174.742	.890	.347
Pretest	35383.986	1	35383.986	180.198	<.001
Study Group	1251.225	1	1251.225	6.372	.013
Error	28079.795	143	196.362		
Total	15295218.000	146			
Corrected Total	65767.890	145			

Table 5: Descriptive Statistics Comparison of the Treatment and Control Group 8th Grade Posttest Scores (Adjusted for Pretest Performance)

Group	Number of Students	Posttest Scores	
		Mean	Standard Deviation
Treatment	73	325.91	20.90
Control	73	320.04	21.08

The passing score for the eighth grade state math assessment is 337. Neither group's average score surpassed the passing standard. The average score for the treatment group was within the "Level 2 – Below Satisfactory" score range (322-336) while the average score for the control group was within the "Level 1 – Inadequate" score range (273-321). The average score of the eighth graders in this school district as reported by the state was 326.

Conclusion

A quasi-experimental, treatment/control, pre/post study using Propensity Score Matching to statistically match the students in the groups was conducted to evaluate the effectiveness of LearnBop use during the 2015-2016 school year in Florida classrooms.

Sixth and eighth grade students who used LearnBop showed significantly more growth in mathematics skills than students who did not use LearnBop (Effect size for 6th grade=.45, Effect size for 7th grade=.28).

The .45 effect size found in sixth grade and .28 effect size observed in eighth compare favorably with research comparing the effects of educational technology applications and traditional methods. Cheung and Slavin (2013) report an overall (mean) effect size of .15, based on a review of 84 studies examining the effects of educational technology applications on mathematics achievement in K-12 classrooms.

One important basis for evaluating these study results is to consider the additional contribution of LearnBop use to the typical mathematics growth expected in the grades levels studied. The average mathematics achievement gain (expressed as an effect size) for students in sixth grade on nationally normed tests is .40, and .32 for eighth graders (Lipsey, et al, 2012). While it is difficult to make definitive comparisons between this state assessment and national standardized assessments, we can use the results as a proxy for typical achievement at these grade levels. With the presumption that the effect sizes found in this study are improvements beyond the average effect from instruction, this means that Florida LearnBop users in sixth grade showed more than double the growth in mathematics than is typically expected (.45/.40). Eighth grade Florida LearnBop users showed 88% more growth in mathematics than is typically expected (.28/.30).

The results support the effectiveness of LearnBop use in improving sixth and eighth grade students' math skills in Florida.

References

Cheung, Alan C.K. and Robert Slavin (2013) The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. Educational Research Review, Volume 9, Pages 88-113.

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. ERIC