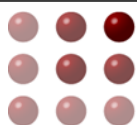# AN EVALUATION OF THE EFFECTIVENESS OF IMAGINE LEARNING FOR IMPROVING READING SKILLS

September 2018

SEGMEASUREMENT

BUILDING BETTER ASSESSMENTS / EVALUATING PRODUCT EFFECTIVENESS

Table of Contents

Executive Summary

Overview

During the 2017-2018 school year, SEG Measurement conducted a study of the effectiveness of Imagine Language & Literacy, an online system that provides individualized adaptive instruction and breaks down skills into component parts to help students become proficient readers. The study was conducted in three districts in Texas.

Context and Background

Research indicates that computer assisted instruction can positively impact students' performance in language and literacy development (Cassady & Smith, 2004; 2005; Cheung & Slavin, 2011; Macaruso & Rodman, 2011). Imagine Language & Literacy is instructional software designed to build language and literacy skills among students in kindergarten through sixth grade. To improve language and literacy achievement, Imagine Learning features instruction in phonemic awareness, phonics, vocabulary, fluency, comprehension, grammar, and language development (both academic and conversational).

Research indicates that Imagine Language & Literacy has a positive effect for literacy development among English language learners in grades K-5, for all students in grades K and 1, and for struggling readers in grades 2 and 3 (Cassidy, Smith, and Thomas, 2017; Elliot, S. 2014; Hobbs, 2016; Hobbs 2017). To date, effectiveness studies have not focused on the performance of students in fourth and fifth grade. The purpose for this research was to describe program impact for fourth and fifth grade students in Texas who used Imagine Language & Literacy as supplemental reading instruction.

Study Design

The study employed a quasi-experimental design with matched groups to compare the growth in reading skills between those students who used Imagine Language & Literacy as a supplemental part of their reading instruction (treatment group) and comparable students who did not use Imagine Language & Literacy as part of their reading instruction (control group). The growth in reading skills was assessed using the State of Texas Assessments of Academic Readiness (STAAR) Reading. Students' spring 2017 STAAR Reading scores served as the pretest and spring 2018 STAAR Reading scores served as the posttest.

Treatment and control group participants were statistically matched using propensity score matching. The students in each grade were matched based on prior reading skill, gender, and ethnicity. For each student who used Imagine Learning, a similar student who did not use Imagine Language & Literacy was determined. Only these matched students who took the posttest and met minimum requirements for using Imagine Language & Literacy were included in the analysis. This statistical matching provided increased rigor in the analyses and controlled for factors beyond product use that may have influenced students' performance. After creating matched groups of students who used Imagine Language & Literacy and students who did not use Imagine Learning, 1,282 fourth grade students and 1,064 fifth grade Texas students participated in the study.

The reading 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 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 STAAR Reading 2018 scores (dependent variable) between the treatment and control groups
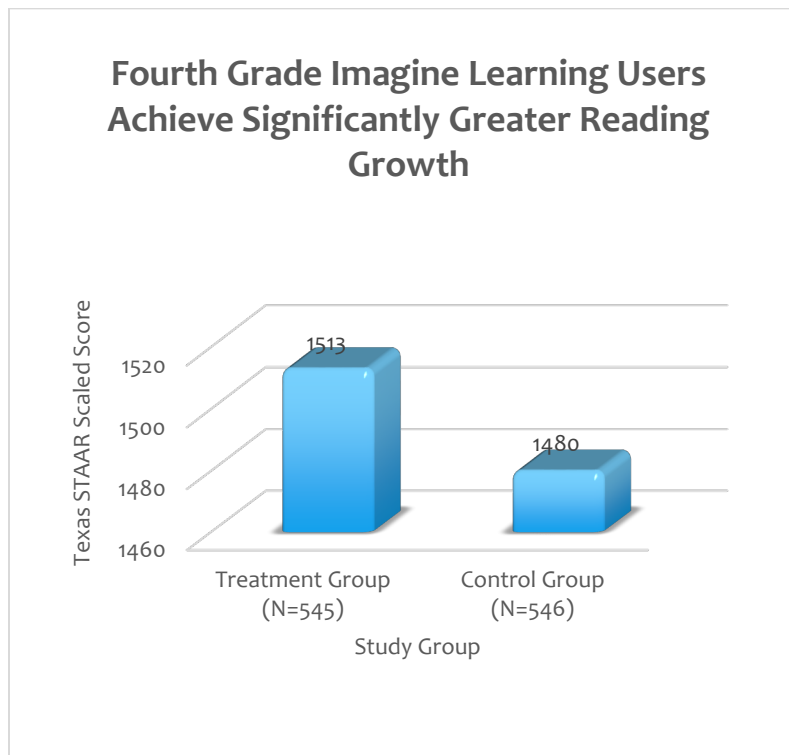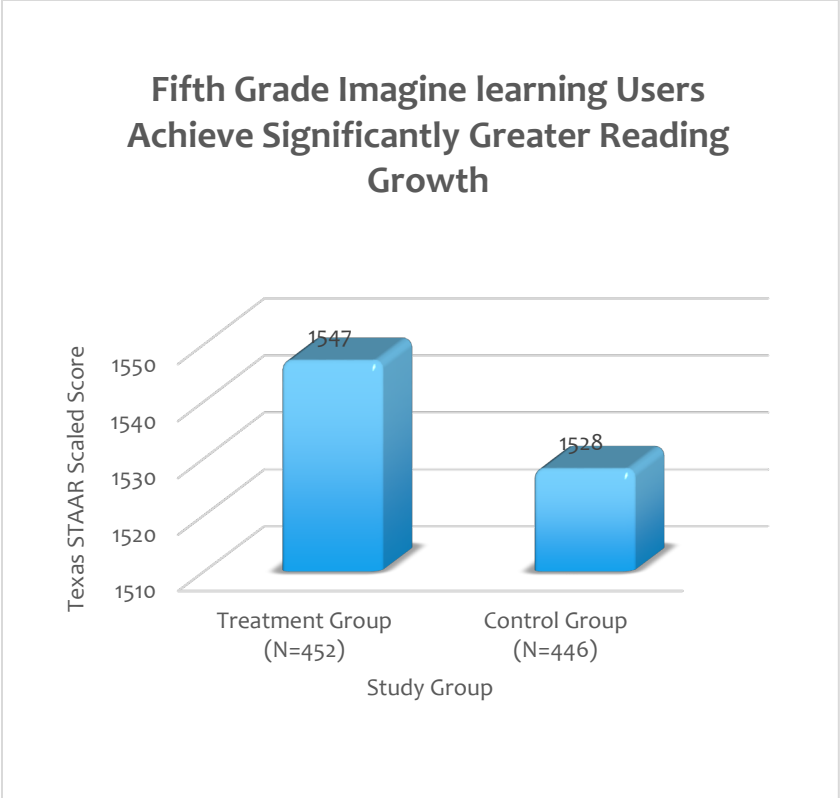
(independent variable) while controlling for the initial ability of the students from Spring 2017 STAAR Reading (covariate).

Study Results

Students who used Imagine Language & Literacy showed significantly more growth in reading skills than comparable students who did not use Imagine Learning. Fourth grade students using Imagine Language & Literacy showed about 33 points more growth on the assessment, or .24 standard deviations (ES=.24), than did fourth graders not using Imagine Learning. Fifth graders showed about 19 points more growth on the assessment, or .14 standard deviations (ES=.14), than did nonusers. For a student in fourth grade at the 50th percentile, this represents a gain to the 59th percentile. For the fifth grade, this represents a gain to the 56st percentile.

The average 2018 STAAR Reading test scores for the treatment and control group students in grades four and five are shown in Figures 1 and 2.

Fourth Grade Imagine Learning Users Achieve Significantly Greater Reading Growth

Fifth Grade Imagine learning Users Achieve Significantly Greater Reading Growth

The results indicate that Imagine Language & Literacy is effective for improving reading skills among fourth and fifth grade students who used the program.

Introduction

Overview

This study examines the effectiveness of Imagine Language & Literacy for improving the reading skills of fourth and fifth grade students. The year-long study (2017-2018 school year), conducted in three school districts in Texas, investigated the impact of Imagine Language & Literacy use among matched groups of fourth and fifth grade students using and not using the product. Reading skill growth among students using Imagine Language & Literacy (treatment group) was compared to reading skill growth among students who did not use Imagine Language & Literacy (control group). End-of-year STAAR Reading scores from the 2017-2018 school year were used to compare reading skill growth for the treatment and control group students, accounting for the initial reading level of students using the previous year STAAR Reading scores.

Research indicates that computer assisted instruction can positively impact students' performance in language and literacy development (Cassady & Smith, 2004; 2005; Cheung & Slavin, 2011; Macaruso & Rodman, 2011). Imagine Language & Literacy is instructional software designed to build language and literacy skills among students in kindergarten through sixth grade. To improve language and literacy achievement, Imagine Language & Literacy features instruction in phonemic awareness, phonics, vocabulary, fluency, comprehension, grammar, and language development (both academic and conversational). The program aligns with educational standards and addresses skills students need to become proficient in reading.

Imagine Language & Literacy is an adaptive supplemental program, used by more than 500,000 students nationwide. When students first use Imagine Language & Literacy, they complete an assessment that places them in content appropriate for their instructional needs. Struggling students may be placed in content that provides exposure to foundational skills necessary for becoming proficient readers, and advanced students may be placed in lessons that allow them to develop skills for comprehending complex literary and informational texts. Imagine Language & Literacy individualizes learning pathways for all students. When used in classrooms, Imagine Language & Literacy is a tool for differentiating instruction to meet students' instructional needs for literacy development.

Prior research has demonstrated that Imagine Language & Literacy has a positive effect for literacy development among English language learners in grades K-5, for all students in grades K and 1, and for struggling readers in grades 2 and 3 (Cassidy, Smith, and Thomas, 2017; Elliot, S. 2014; Hobbs, 2016; Hobbs 2017). To date, effectiveness studies have not focused on the performance of students in fourth and fifth grade. The purpose for this research is to describe program impact for fourth and fifth grade students in Texas who used Imagine Language & Literacy as supplemental reading instruction.

Methods and Procedures

Research Questions

The primary research question addressed by this study is: "Is Imagine Language & Literacy effective in improving students' reading skills?" The specific operational questions addressed to answer this are:
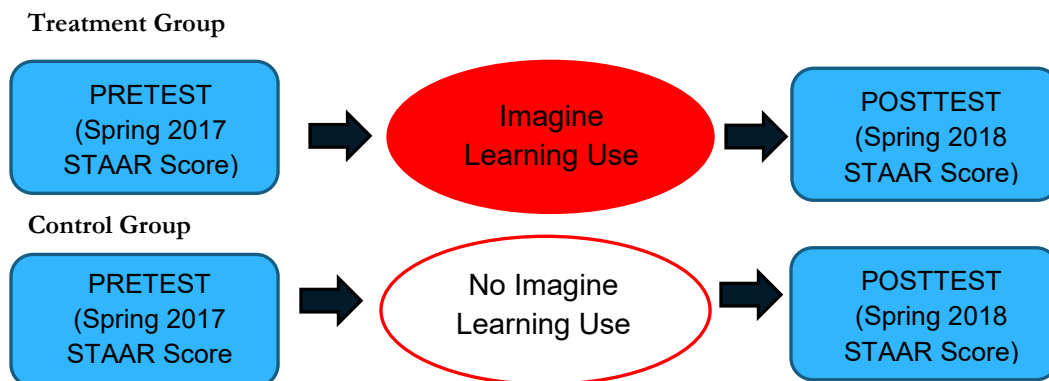
- Do students in fourth grade who receive supplemental instruction using Imagine Language & Literacy show larger gains in reading skills than comparable students who do not receive instruction using Imagine Learning?
- Do students in fifth grade who receive supplemental instruction using Imagine Language & Literacy show larger gains in reading skills than comparable students who do not receive instruction using Imagine Learning?

Study Design

The study employed a quasi-experimental design. A treatment group of students (students who used Imagine Learning) was compared to a control group of students (who did not use Imagine Learning) based on the end-of-year statewide STAAR Reading scores (posttest) adjusting for the initial reading ability of the students assessed using the prior year's STAAR Reading scores (pretest). The treatment group students received core literacy instruction and used Imagine Language & Literacy as supplemental instruction. The control group students received core literacy instruction and did not have access to Imagine Language & Literacy. The study design is depicted in Figure 1. Students were not randomly assigned to experimental groups; they were matched with respect to background and ability as described below.

Figure 1: Study Design

Treatment Group

| PRETEST (Spring 2017 STAAR Score) | → | Imagine Learning Use | → | POSTTEST (Spring 2018 STAAR Score) |

Control Group

| PRETEST (Spring 2017 STAAR Score) | → | No Imagine Learning Use | → | POSTTEST (Spring 2018 STAAR Score) |

Program Implementation

Students included in the study began using Imagine Language & Literacy by October 15th, 2017. To support local implementations, Imagine Learning offered five key supports for school personnel including: (1) access to Imagine University with online training materials; (2) training provided by Customer Success Managers; (3) access to Imagine Learning's Teacher Care call center; (4) onsite visits by Customer Success Managers; and (5) technical support as needed. These supports are typical supports offered to all Imagine Language & Literacy customers.

For this study, teachers received initial onsite training lasting 2-3 hours and follow-up training and support provided by local Customer Success Managers. All teachers were given access to Imagine University training videos, which are available on demand and accessible through the teacher portal. Teacher Care, which is a phone support system dedicated to answering teachers' questions about product features and functions, was available during the business hours for participating teachers.

School administrators and teachers determined models for implementing Imagine Language & Literacy at their sites. Implementation models varied across the study depending on local infrastructure and access to devices. Generally, students who used Imagine Language & Literacy used the program during lab rotations or during station rotations within their classrooms. The program was used as supplemental instruction and did not replace students' core literacy instruction. Imagine Learning users averaged 20 hours on the program for the durations of the school year. Students who did not receive access to Imagine Language & Literacy participated in literacy programs available at their schools.

Population

Participants in this study were drawn from three school districts. Students enrolled in grades four and five participated in the study. In each district, an equal number of treatment and control schools were recruited, with the exception of 1 district in which an additional control school was included. School districts selected schools to participate as treatment schools and identified schools with similar demographic characteristics as the treatment schools to act as control schools for the study. The largest school district participating in the study was a large urban district with a population of 32,682 students. The second school district participating was a suburban district with a population of 23,800 students, and the smallest district was a suburban district with a population of 9,800 students.

Seventeen schools participated in the study. Students in fifty-two fourth-grade classes participated (52 teachers), 24 contributing to the treatment group and 28 contributing to the control group. Students in fifty fifth-grade classrooms participated in the study (50 teachers), 19 contributing to the control group and 31 contributing to the treatment group.

The population was defined as those students in grades four and five who took the STAAR Reading in spring 2017, which served as the pretest. The distribution of students included in the population is described in Table 1.

Table 1:
Profile of Study Population

	Grade 4		Grade 5	
	Treatment	Control	Treatment	Control
Gender				
Female	399 (61%)	505 (69%)	399 (59%)	512 (67%)
Male	253 (39%)	230 (31%)	272 (41%)	254 (33%)
Ethnicity				
White or Caucasian	96 (15%	157 (21%)	108 (16%)	99 (13%)
Hispanic or Latino	336 (52%)	376 (51%)	343 (50%)	405 (54%)
Black or African American	182 (28%)	188 (25%)	201 (29%)	225 (30%)
Asian or Pacific Islander	24 (4%)	10 (2%)	25 (4%)	8 (1%)
Mixed Race or Other	11 (1%)	7 (1%)	8 (1%)	15 (2%)

Data Collection

At the outset of the study, data files were obtained from the three participating districts to acquire the necessary data for both the treatment and control groups. During August and September of 2017, SEG

Measurement provided specifications to the districts for providing a data export that included necessary data elements for establishing baseline equivalence and matching treatment and control groups. Each district then provided SEG Measurement with de-identified STAAR Reading performance data for spring 2017 and demographic information for each participating student. Each student was identified with a unique identifying number to both preserve confidentiality and to allow for later linking to the 2018 STAAR Reading scores (post test data).

The STAAR is the Texas statewide assessment program designed to measure the extent to which a student has learned and is able to apply the knowledge and skills defined in the Texas Essential Knowledge and Skills (TEKS), the Texas mandated curriculum. In grades 3–8, students are tested in mathematics and reading. In addition, students are tested in writing at grades 4 and 7, science at grades 5 and 8, and social studies at grade 8. The reading assessment scores were used as the primary measures in this study. An independent evaluation of the STAAR by Human Resources Research Organization (HUMRRO, 2016) found support for the validity and reliability of the 2016 STAAR Reading scores.

In the spring of 2018, SEG Measurement requested end-of-year data from each district. All participating districts provided SEG Measurement with the spring 2018 STAAR Reading scores for treatment and control students (data were received during June/July 2018). Imagine Learning provided usage data to SEG Measurement to ensure that only students who used the product were included in the analysis as treatment students. A minimum threshold of six hours of usage across the school year was established to ensure program exposure. The six-hour criterion reflected the completion of approximately 10 literacy lessons and is consistent with the criterion used in prior Imagine Learning studies.

Matched Sample

A multi-step process was used to select comparable groups for the study. Propensity score matching was used to help ensure comparability of the two study groups. Propensity score matching is widely recognized as effective in achieving group equivalence in the absence of randomization (Guo and Frazer, 1999). This technique identifies for each member of the treatment group, a corresponding member of the control group that is matched on ability and background. Propensity score matching was executed using logistic regression without replacement. To be eligible for matching the treatment control match needed to be within .05 (on a 0 to 1 Propensity score scale).

Treatment schools (using Imagine Learning) were identified by each district. Students from the schools identified as control schools served as the source for creating a comparable control group. For each student in the treatment group, a comparable student from the remaining students attending participating districts were selected to be included in the control group. Treatment students and comparable control students were matched such that each treatment student had a matching control student with similar characteristics including initial reading ability level (determined by spring 2017 STAAR Reading scores), gender, and ethnicity. Matching was done by grade.

While students were matched on initial ability, ANCOVA was also used to ensure that students were placed on a common baseline of initial starting reading skill. Using ANCOVA, we examined the difference in the posttest scores (dependent variables) between the treatment and control groups (independent variable) controlling for the initial skill level of the students (covariate). The spring 2017 STAAR Reading scores were 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.

Six hundred forty-one fourth grade treatment students were matched with 641 fourth grade control students. A suitable match could not be found for 11 treatment students in grade four. Five hundred and thirty-two fifth grade treatment students were matched with 532 fifth grade control group students. A suitable match could not be found for 140 treatment students in grade five.

As illustrated in tables 1 and 2 below, the two groups were well matched, nearly the same with respect to ability, gender and ethnicity. The treatment group for grade four had an average pretest score of 1381 while the control group had an average pretest score of 1403, a difference of 22 points (.16 SD) on the spring 2017 STAAR Reading. The treatment group for grade five had an average pretest score of 1479, while the control group had an average pretest score of 1486, a difference of 7 points (.05 SD) on the spring 2017 STAAR Reading.

Table 1:
Comparison of Initial Ability (Pretest Scores)

	Treatment	Control
Grade 4	1381	1403
Grade 5	1479	1486

Table 2:
Profile of Matched Samples

	Grade 4		Grade 5	
	Treatment	**Control**	**Treatment**	**Control**
Gender				
Female	391 (61%)	434 (68%)	329 (62%)	315 (59%)
Male	250 (39%)	207 (32%)	203 (38%)	217 (41%)
Ethnicity				
White or Caucasian	96 (15%)	119 (19%)	75 (14%)	79 (15%)
Hispanic or Latino	335 (52%)	333 (52%)	280 (53%)	273 (51%)
Black or African American	182 (28%)	172 (27%)	162 (30%)	165 (31%)
Asian or Pacific Islander	17 (3%)	10 (2%)	4 (1%)	7 (1%)
Mixed Race or Other	11 (2%)	7 (1%)	11 (2%)	8 (2%)

Analytic Sample

To be included in the final analytic sample, students from the matched samples were required to have posttest (STAAR 2018 Reading) results available and for the treatment group to have used Imagine Language & Literacy at least six hours across the school year. Based on these criteria, 545 fourth grade treatment students and 546 fourth grade control students were included in the final analyses. Based on these criteria, 454 fifth grade treatment students and 448 fifth grade control group students were included in the final analyses.

As illustrated in tables three and four below, the two groups were well matched, nearly the same with respect to ability, gender and ethnicity. The treatment group for grade four had an average pretest score of 1383, while the control group had an average pretest score of 1406, a difference of 23 points (.16 SD) on the spring

2017 STAAR Reading. The treatment group for grade five had an average pretest score of 1478, while the control group had an average pretest score of 1490, A difference of 11 points (.08 SD) on the spring 2017 STAAR Reading.

Table 3:
Comparison of Initial Ability (Pretest Scores)

	Treatment	Control
Grade 4	1383	1406
Grade 5	1478	1490

Table 4:
Profile of Students Included in the Analysis

	Grade 4		Grade 5	
	Treatment	Control	Treatment	Control
Gender				
Female	330 (61%)	363 (66%)	287 (63%)	264 (59%)
Male	215 (39%)	183 (34%)	167 (37%)	184 (41%)
Ethnicity				
White or Caucasian	83 (15%)	96 (18%)	64 (14%)	71 (16%)
Latino or Hispanic	292 (54%)	296 (54%)	242 (53%)	233 (52%)
Black or African American	144 (26%)	138 (25%)	135 (30%)	129 (29%)
Asian or Pacific Islander	16 (3%)	10 (2%)	5 (1%)	7 (1%)
Mixed Race or Other	10 (2%)	6 (1%)	8 (2%)	8 (2%)

Attrition

About 14%-15% of the students were not included in the final analysis either because they did not have a posttest score or failed to use the product at minimum specifications. The demographic profile for the fourth and fifth grade groups was comparable after attrition.

For the fourth-grade sample, the treatment group lost 96 students (15%) from the initial matched sample of 641 students. The fourth-grade control group lost 95 students (15%) from the initial matched sample of 641 students. In fifth-grade sample, the treatment group lost 78 students (14%) from the initial matched sample of 532 students. The fifth-grade control group lost 84 students (15%) from the initial matched sample of 532 students.

Results

Grade 4 Reading Skills Results

For fourth grade students, the results showed an effect size of .24 (Cohen's D) for the 2018 STAAR Reading assessment. Fourth grade students who used Imagine Language & Literacy achieved significantly higher
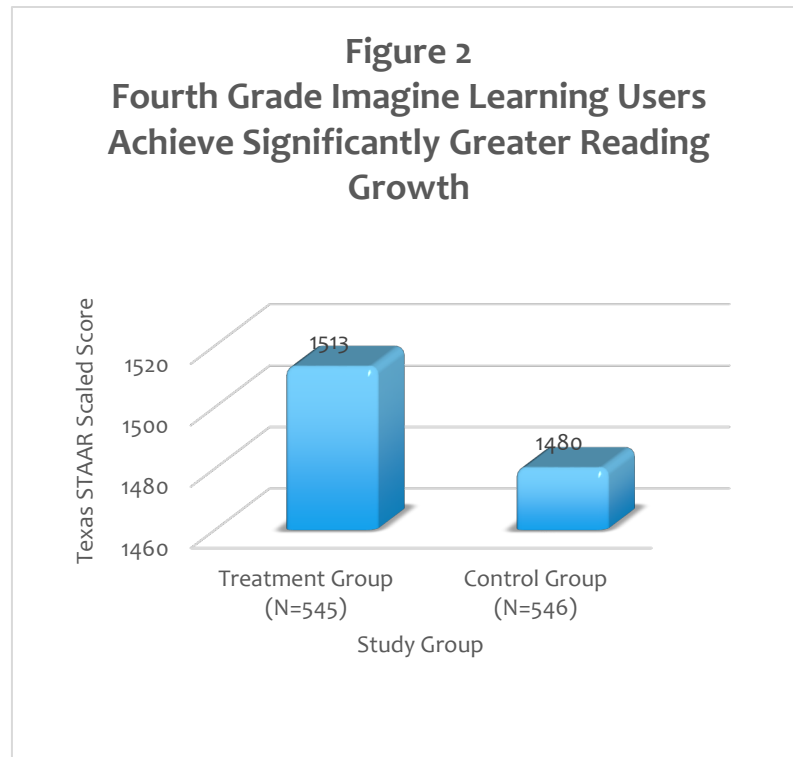
scores on the 2018 STAAR Reading assessment than students who did not use Imagine Language & Literacy ($F = 20.399$, df=2/1090; p=.001). For a student at the 50th percentile, an effect size of .24 would produce a gain to the 59th percentile. The results are illustrated in Figure 2 and summarized in Tables 5 and 6 below.

Table 5: ANCOVA of the
Treatment and Control Group 4th Grade Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	4499442.064	2	2249721.032	158.164	<.001
Intercept	8560457.117	1	8560457.117	601.833	<.001
Pretest	4360436.486	1	4360436.486	306.556	<.001
Study Group	290152.564	1	290152.564	20.399	<.001
Error	15475678.195	1088	14223.969		
Total	2464350106.000	1091			
Corrected Total	19975120.258	1090			

Table 6: Descriptive Comparison of the Treatment and Control Group
4th Grade Posttest Scores (Adjusted for Pretest Performance)

Group	Number of Students	Posttest Scores	
		Mean	Standard Deviation
Treatment	545	1513.20	135.042
Control	546	1480.48	134.884
Total	1092	1496.84	135.373

Figure 2
Fourth Grade Imagine Learning Users Achieve Significantly Greater Reading Growth
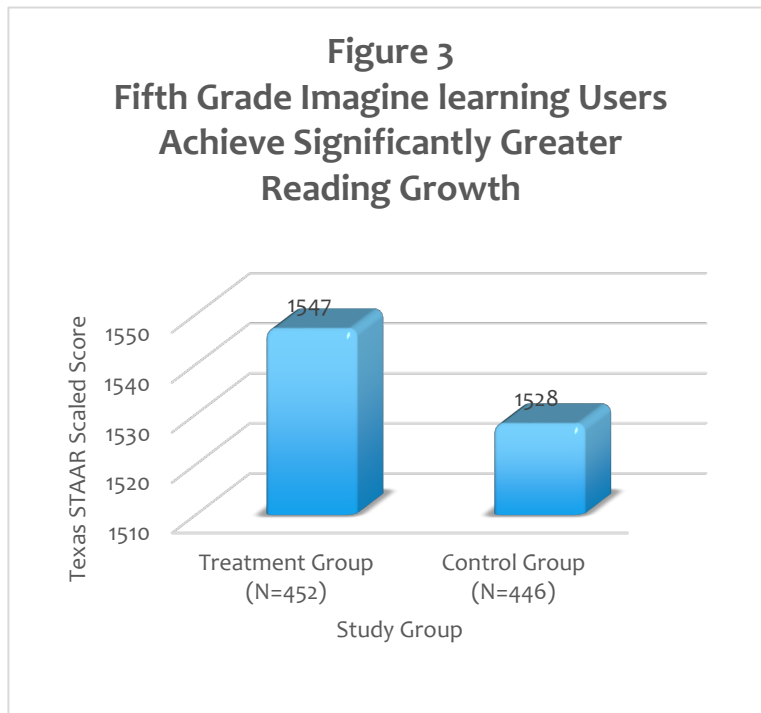
Grade 5 Reading Skills Results

For fifth grade, the results showed an effect size of .14 for the 2018 STAAR Reading assessment. Fifth grade students who used Imagine Language & Literacy achieved significantly higher scores on the 2018 STAAR Reading assessment than students who did not use Imagine Language & Literacy (F = 7.182, df=2/897; p =.008). For a student at the 50th percentile, an effect size of .14 would produce a gain to the 56th percentile. The results are illustrated in figure 3 and summarized in Tables 7 and 8 below.

Table 7: ANCOVA
of the Treatment and Control Group 5th Grade Posttest Scores

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	5807094.896	2	2903547.448	254.850	<.001
Intercept	2832201.536	1	2832201.536	248.588	<.001
Pretest	5773417.790	1	5773417.790	506.744	<.001
Study Group	81821.366	1	81821.366	7.182	<.008
Error	10196879.015	895	11393.161		
Total	2138721062.000	898			
Corrected Total	16003973.911	897			

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

Group	Number of Students	Posttest Scores	
		Mean	Standard Deviation
Treatment	452	1546.97	136.19
Control	446	1527.85	130.72
Total	898	1537.47	133.57

**Figure 3
Fifth Grade Imagine learning Users
Achieve Significantly Greater
Reading Growth**

Conclusion

The results observed in this study indicate that Imagine Language & Literacy is an effective tool for improving reading skills among students in fourth and fifth grade. The .24 effect size found in fourth grade and the .14 effect size observed in fifth compare favorably with research comparing the effects of educational technology applications and traditional methods (Cassidy, Smith, and Thomas, 2017; Elliot, S. 2014; Hobbs, 2016; Hobbs 2017). 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 reading achievement in K-12 classrooms. Similarly, IES reports an average effect size of .13 for similar intervention programs (Lipsey et. al., 2012). In this context, the effect sizes reported for this study can be interpreted as exceeding expectations for technology applications for 4th grade students, and meeting expected or typical effects for 5th grade students.

As with all research, this study is characterized by limitations and strengths that should be considered when interpreting the results of this study. For example, in quasi-experimental research designs, assignment to treatment and control conditions is not random. Through the use of propensity score matching and the controlling for initial ability using ANCOVA, we can be more certain Imagine Language & Literacy is responsible for the observed effects. The combination of propensity score matching and ANCOVA help ensure that the treatment and control groups are truly comparable based on baseline characteristics. Indeed, despite some minor attrition in both the treatment and control groups, baseline equivalence was maintained for the final analytic sample.

This study provides evidence that the Imagine Language & Literacy program, when used with fidelity, is effective in improving fourth and fifth grade students' reading skills.

References

Cassady, J., Smith, L., Thomas, C. (2017). Supporting emergent literacy for English language learners with computer-assisted instruction. Journal of Research in Reading 41(2): 350-369. DOI:10.1111/1467-9817.12110

Cheung, Alan C.K. and Robert Slavin (2013). The effectiveness of educational technology applications for enhancing reading achievement in K-12 classrooms: A meta-analysis. Educational Research Review, Volume 9, Pages 88-113Cassady, J. C., Smith, L. L., & Thomas, C. (2017). Supporting emergent literacy for English language learners with computer-assisted instruction. Journal of Research in Reading, DOI:10.1111/1467-9817.12110

Elliot, S. (2014). A study of the effectiveness of Imagine Learning on student reading achievement. SEG Measurement.

Guo, Shenyang Y. and Fraser, Mark W. (1999). Propensity Score Analysis: Statistical Methods and Applications New Jersey: Sage Publications.

Hobbs. J. (2017). Utah's Early Intervention Reading Software Program: 2016-2017 K-3 Program Evaluation Results. Evaluation and Training Institute.

Hobbs. J. (2016). Utah's Early Intervention Reading Software Program: 2015-2016 K-3 Program Evaluation Results. Evaluation and Training Institute.

Humrro (2016). Final Report: Independent Evaluation of the Validity and Reliability of STAAR Grades 3-8 Assessment Scores. Prepared under contract to the Texas Education Agency Contract 3436.

Lipsey, M., Puzio, K., Yun, C., Hebert, M, Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K., and Busick M. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. Washington DC: Institute of Education Sciences.

Macaruso, P., Rodman, A. (2011). Efficacy of computer-assisted instruction for the development of early literacy skills in young children. Reading Psychology 32(2): 172-196. DOI:10.1080/02702711003608071