Improving Student Learning Through Creative Collaboration:

A Study of the Impact of Tech4Learning's Pixie Software on Student Achievement

2011

SEG Measurement

This study was conducted by SEG Measurement, an independent educational research firm located in New Hope, Pennsylvania. This study was supported by a grant from Tech4Learning, Inc.

Improving Student Learning Through Creative Collaboration:
A Study of the Impact of Tech4Learning's Pixie Software on Student Achievement

Executive Summary

Introduction

Schools are under increasing pressure to develop students' academic skills. One way to achieve these goals is through the effective integration of technology in instruction. Tech4Learning designed Pixie, a collaborative creativity tool, to address this challenge. Pixie provides a rich environment for students to create projects and work collaboratively.

During the 2010-2011 school year, SEG Measurement conducted a national study with approximately 1,000 3rd, 4th, and 5th grade students, in 38 classrooms, in California, Georgia, Ohio, South Carolina, and Texas to evaluate the impact of using Pixie on student achievement.

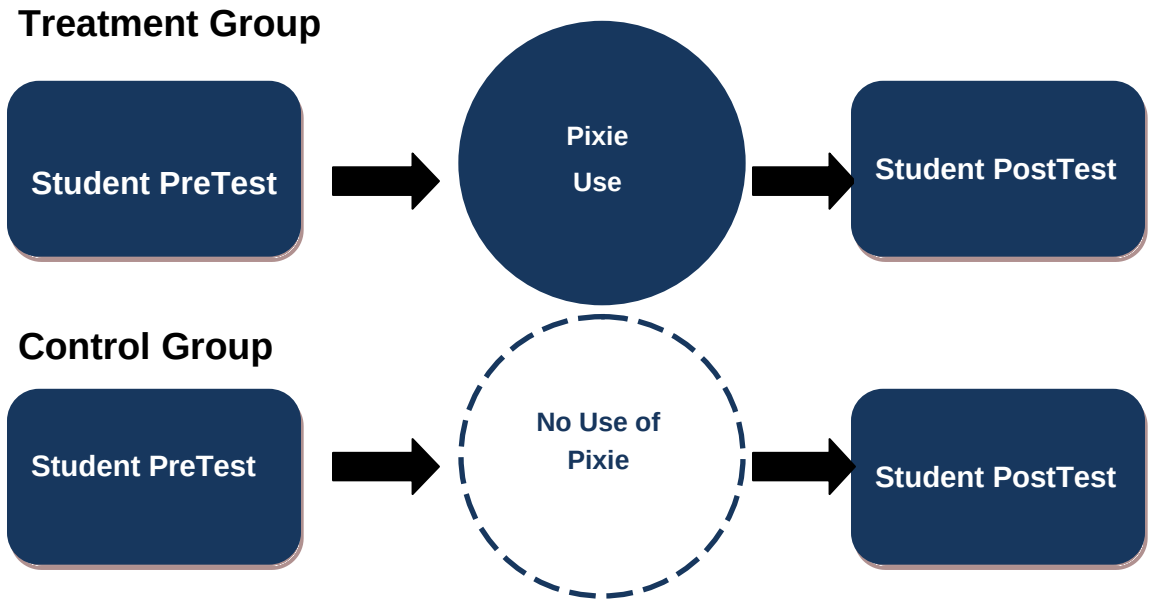
The goal of this study was to evaluate the impact of using Pixie on student learning. The results show that students who use Pixie learn significantly more than students who do not use Pixie. Students who used Pixie showed about one half year more of growth in Language Arts and in Mathematics than students who did not use Pixie.

Study Design

The primary question answered by this study is: Do students in grades 3, 4, and 5 show larger gains in Reading Comprehension and Mathematics skills if they use Pixie? The study also explored potential differences in growth between boys and girls and among students of different ethnic backgrounds.

The study compared two groups of students, matched in ability. The Treatment Group consisted of students who used Pixie; the Control Group consisted of students who did not use Pixie. The students in both groups were administered a pre-test in January 2011 and a post-test in May/June 2011 to evaluate the impact of Pixie use on their Reading Comprehension and Mathematics growth. This is illustrated below. (see Figure 1)

Figure 1: Study Design



The study compared the growth in Reading Comprehension and Mathematics Stanford 10 Achievement Test™ scores from the middle of the school year to the end of the school year. The results from the pretest and posttest were compared statistically to determine the level of growth in Reading Comprehension and Mathematics skills. In addition, a qualitative survey of teachers was conducted to obtain additional information and context for the study.

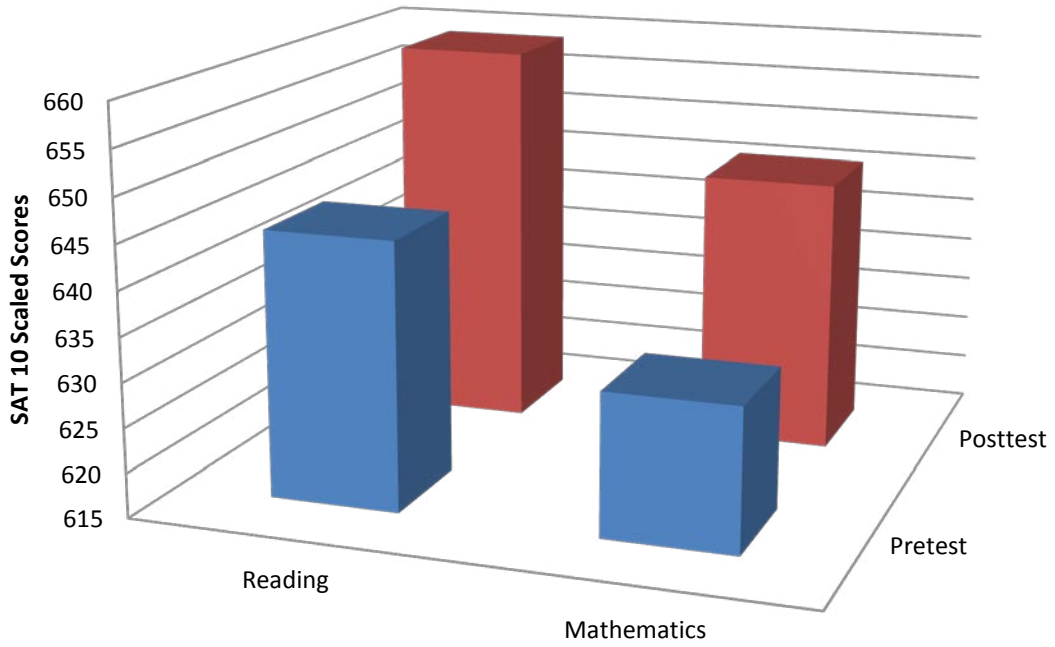
Students in the Treatment Group used Pixie about one to two hours weekly, while students in the Control Group did not use Pixie.

Results

The Treatment Group students who used Pixie showed substantial growth in Reading Comprehension and Mathematics during the course of the study (see Figure 2). During the course of the study, students in classes using Pixie increased their SAT 10 Reading Comprehension scale-scores by 14 points (Mean pretest=645; Mean posttest score=659) and their Mathematics scale-

scores by 15 points (Mean pretest=631; Mean posttest score=646). This means that the students in Pixie classes, on average, achieved about a full year of growth (for the typical student at the 50th percentile), during the second semester of the 2010-2011 school year in which the study was conducted.

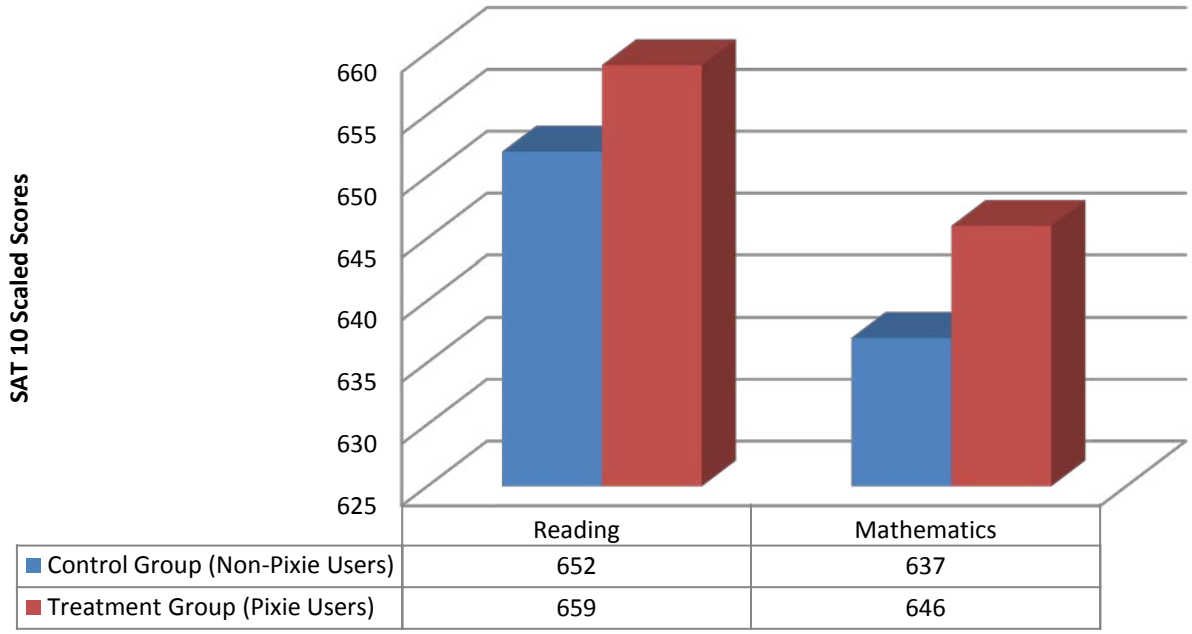
Figure 2
Comparison of Pretest and Posttest Scores for Students Using Pixie (Treatment Group)



	Reading	Mathematics
Pretest	645	631
Posttest	659	646

The greater academic growth observed for students in Pixie classes becomes even more visible when comparing these students against the Control Group, who did not use Pixie. The Treatment Group students showed statistically greater gains in Reading Comprehension (7 scale score points; Effect Size = .13) and Mathematics (9 scale score points; Effect Size = .16) than the Control Group classes (see Figure 3). This means that, on average, students who used Pixie showed about a half year's more growth than their peers who did not use Pixie.

Figure 3
Comparison of Posttest Scores for Treatment and Control Group Students



	Reading	Mathematics
Control Group (Non-Pixie Users)	652	637
Treatment Group (Pixie Users)	659	646

These effects indicate that the use of Pixie has a substantial impact on student Reading Comprehension and Mathematics skills growth. Pixie was found to be equally effective for boys and girls and for students of different ethnicities.

SEG Measurement

Summary

During the 2010-2011 school year (between January and June 2011), SEG Measurement conducted a national study with approximately 1,000 3rd, 4th, and 5th grade students, in 38 classrooms, in California, Georgia, Ohio, South Carolina, and Texas. Students who used Pixie showed meaningful growth in Reading Comprehension and Mathematics during the course of the study. Students in the Treatment Group classes increased their SAT 10 scores between 14-15 points, or about one year worth of growth. More significantly, Treatment Group students enrolled in classrooms using Pixie showed about one half year more of growth in Reading Comprehension and Mathematics than the Control Group students enrolled in classes not using Pixie. The Pixie users finished the year with scores that were 7 scale-score points higher in Reading Comprehension and 9 scale-score points higher in Mathematics on the SAT 10 assessments. The study also found that Pixie is equally effective for boys and girls and for students of different ethnic backgrounds.

The quantitative results were reinforced by the qualitative data provided by teachers in classes using Pixie. All of the teachers indicated that they were likely to use Pixie in the future, and nearly all of the teachers (92%) said they would recommend Pixie for use by others. Almost all (85%) of the teachers indicated that that Pixie was effective in improving student's attitudes toward school and learning. Nearly two thirds (61%) of the teachers indicated that that Pixie was effective in increasing their students' cognitive/intellectual growth.

The findings of this study provide substantial support for the effectiveness of Pixie in improving student Reading Comprehension and Mathematics skills.

SEG Measurement

Improving Student Learning Through Creative Collaboration:

A Study of the Effectiveness of Tech4Learning's Pixie Software on Student Achievement

Overview/Background of the Study

Schools are under increasing pressure to develop students' academic skills. One way to achieve these goals is through the effective integration of technology in instruction. Tech4Learning designed Pixie, a creative collaborative tool, to address this challenge. Pixie provides a rich environment for students to create projects and work collaboratively. This research study examined the impact of Pixie on student academic performance.

Effectiveness Study Goals and Overview

This report describes a nationwide study conducted during the 2010-2011 school year to evaluate the impact of using Pixie on student achievement. Specifically, the study compares the growth in academic skills of students in grades 3, 4 and 5 who used Pixie (Treatment Group) to those who did not use Pixie (Control Group). The study compared student academic growth in the Treatment and Control Groups. The study compared the growth in Reading Comprehension and Mathematics attained by students in the Treatment Group and Control Group between the middle and end of the 2010-2011 school year, as measured by the growth in Stanford 10 Achievement Test™ Abbreviated Battery (SAT 10) scores.

Research Questions

This study investigated the following questions:

1. Do students in grades 3, 4 and 5 using Pixie show larger gains in Reading Comprehension

and Mathematics skills than students who do not use Pixie?

2. Are there any differences in the Reading Comprehension and Mathematics skills between boys and girls in classes whose students use Pixie as compared to classes whose students do not use Pixie?
3. Are there any differences in the Reading Comprehension and Mathematics skills among ethnic groups in classes whose students use Pixie as compared to classes whose students do not use Pixie?

Student Sample

Between January and June 2011, approximately 1,000 students (N=826) in 38 classrooms in California, Georgia, Ohio, South Carolina, and Texas participated in a controlled study of Pixie effectiveness. Students in classes who used Pixie constituted the Treatment Group. Students in classes who did not use Pixie constituted the Control Group. There were approximately 408 students in the Treatment Group and approximately 415 students in the Control Group. Table 1 shows the number of students in each gender, ethnic, and grade category. (The total number of students listed for each background variable may be different since some schools were unable to provide complete background information.)

Table 1. Demographic Profile of Student Participants

Variable	Number (N) of Students	Percentage of Students
GENDER		
Male	**416**	**50%**
Female	**407**	**49%**
Total (All Gender)	**823**	
ETHNICITY		
Caucasian	**434**	**59%**
African American	**75**	**10%**
Hispanic	**113**	**15%**
Asian/Pacific Islander	**17**	**2%**
Mixed Race and Other	**102**	**14%**
Total (All Ethnicity)	**741**	

GRADE		
Grade 3	299	36%
Grade 4	310	38%
Grade 5	217	26%
Total (All Grades)	826	

In some cases, teachers did not provide complete background information for a student or a student did not take one of the tests included in the analyses. Where data was missing, the student's results were eliminated from those analyses.

Comparability of Study Groups

It is very important in a study comparing student academic growth to establish at the outset that the Treatment Group and Control Group are similar, particularly with respect to student academic ability, the outcome of interest. Demonstrating baseline equivalence of the sample (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 Pixie) or the result of some differences in the individuals that existed before we conducted the study.

Ideally, this matching is accomplished by sampling study participants of similar reading and math ability. However, any observed differences can be adjusted for statistically using analysis of covariance (ANCOVA). The Treatment Group and Control Group were compared with respect to initial Reading Comprehension and Mathematics ability, as well as their gender and ethnicity. The results indicate that the groups were similar in ability (see Table 2) and background (see Tables 3, 4 and 5).

Ability Comparison. The SAT 10 pretest scores were used to compare the initial Reading Comprehension and Mathematics levels for students in both the Treatment and Control Groups. The mean test scores for students in both Groups are presented in Table 2.

Table 2. Comparison of Initial Reading Comprehension and Mathematics levels (SAT 10 scores) for the Treatment Group and Control Group

STUDY GROUP	Reading Comprehension Mean	Reading Comprehension Standard Deviation	Mathematics Mean	Mathematics Standard Deviation
Treatment Group	**645.23**	44.18	**630.58**	48.84
	(N=397)		(N=397)	
Control Group	**643.57**	48.25	**628.33**	54.32
	(N=414)		(N=414)	

The Treatment and Control Groups were comparable in ability. There were no statistically significant differences in the Means between the Treatment and Control groups for Reading Comprehension (F=-.26, df=1/811, p<.61) or Mathematics (F=-.39, df=1/811, p<.53).

Gender and Ethnicity. The number of female and male students in both the Treatment and Control were computed and compared (see Table 3). A statistical comparison of the two study groups shows that the Treatment Group and Control Group were comparable with respect to gender and ethnicity. There were no statistical differences in the expected and observed frequencies for gender (chi square=2.02, df=1, p<.16) or ethnicity (chi square=5.25, df=1, p<.26).

Table 3. Comparison of the Gender Composition of the Treatment and Control Group

STUDY GROUP		Gender	
	Female	Male	Total
Treatment Group	224	199	423
Control Group	192	208	400
Total	416	407	823

Table 4. Comparison of the Ethnicity Composition of the Treatment and Control Group

STUDY GROUP	Ethnicity					
	Caucasian	African American	Hispanic	Asian or Pacific Islander	Mixed Race or Other	Total
Treatment Group	226	41	49	7	58	381
Control Group	208	34	64	10	44	360
Total	434	75	113	17	102	741

Description of the Pretest and Posttest

The academic growth of students was operationalized as the gains in Reading Comprehension and Mathematics ability between pre and posttest. The students participating in the study were measured using the Reading Comprehension and Mathematics Stanford Achievement Test™, Tenth Edition (SAT 10), Abbreviated Battery, Form A, 2002. The SAT 10 was used as both the pretest and posttest measure; students took the SAT 10 in January 2011 and then again at the end of May or in June 2011 at the end of the school year.

The Reading Comprehension and Mathematics subtests of the SAT 10 were used for this study. The Reading Comprehension subtest measures students' achievement within the framework of three types of materials or purposes for Reading: literary, informational, and functional text. Within each type of text, questions measure achievement in four modes of Comprehension: initial understanding, interpretation, critical analysis, and awareness and usage of Reading strategies. The Mathematics subtest measures the mathematics skills typically associated with the mathematic curriculum in US schools. Each subtest is 30 items in length (Stanford Achievement Test Series™, Tenth Edition, Technical manual; Harcourt, 2002).

The SAT 10 measures students' skill levels on a single vertical scale ranging from 200-900. The scale-scores represent equal units; differences between scores at any point in the scale represent the same amount of achievement variation. This allows for an accurate comparison of changes over

time. The scale is equivalent across forms and grade levels, to provide an accurate comparison across grade levels; a score at one grade level means that same thing at another grade level.

Reliability and Validity

The reliability of the SAT 10 ranges from .89 to .97 (KR-20 reliability coefficient; Harcourt, 2002). Several validity studies conducted for the SAT 10 have found strong evidence for the validity of SAT 10 scores; for example, content expert review found strong alignment with important Reading skills. Strong relationships were found between the SAT 10 and other measures of Reading ability. For a more complete discussion of the SAT 10 reliability and validity, readers are referred to the SAT 10 Technical Manual (Harcourt, 2002).

Description of the Treatment

The Treatment in this study was students' use of Pixie. Pixie is creativity software students can use to share ideas, imagination, and understanding through a combination of text, original artwork, voice narration, and images. Pixie is designed to motivate a wide range of learners and help them learn. Pixie is designed to help all students produce high-level work, achieve a sense of pride in their abilities, and foster the determination to achieve.

Creating with technology encourages thinking, creativity, and communication skills. Combining artwork with text and voice recording makes Pixie a rich tool for today's digital learners.
Pixie supports collaboration, team-building, and organization skills through real-time collaborative projects and allows students to instantly publish their work to share with family members.

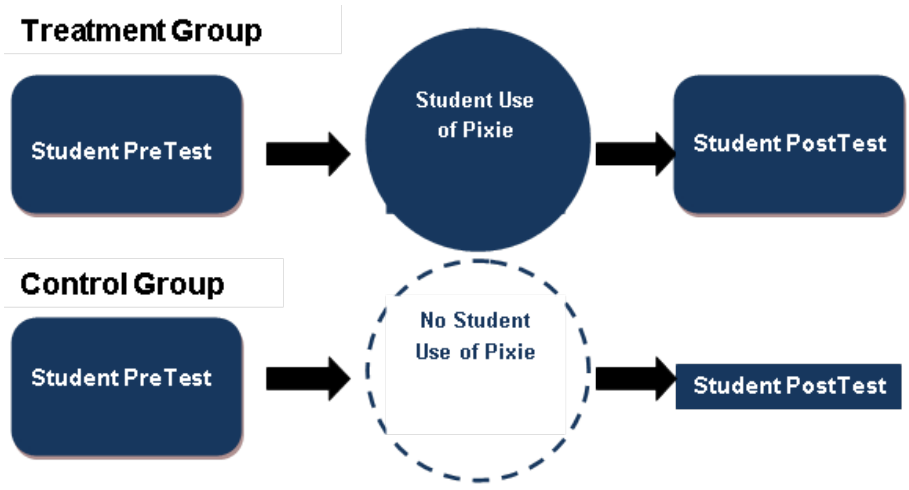
Students in the Treatment Group used Pixie between one and four hours per week, with most teachers reporting use one (46%) or two hours per week (31%). (Students in the Control Group did not use Pixie).

Study Design

The goal of this effectiveness study was to compare the academic growth of students in classes who used Pixie (Treatment) to students in classes who did not use Pixie (Control). Academic growth was measured using the Stanford 10 Reading Comprehension and Mathematics Tests. Students' growth in Reading Comprehension and Mathematics was measured by comparing their proficiency at the beginning of the second semester of the school year (January 2011) and again at the end of the school year (May and June 2011). Students in both the Treatment Group and the Control Group were administered the SAT 10 test as a pretest at the beginning and as a posttest at the conclusion of the school year. Students received approximately 12-20 weeks of instruction between the pretest and posttest. Students in the Treatment Group used Pixie, while those in the Control Group did not use Pixie. The results were then compared statistically.

The study employed a pre-post, Treatment-Control Group design. Since the students were not randomly assigned to the groups, this is considered a quasi-experimental design (see Figure 4 below).

Figure 4. Study Design



Data Collection

At the outset of the study, teachers were asked to provide background information about the participating students in order to characterize the sample, compare the differences between the study groups and facilitate the analysis of the Reading Comprehension and Mathematics gains between the study groups. This information included:

- Student grade level
- Student gender
- Student ethnicity
- Study group membership (Treatment or Control)

Teachers were also asked to provide some additional demographic and instructional information regarding Individual Education Plans (IEP) and disabilities. Due to the unavailability of information and/or privacy concerns, many teachers did not provide this additional information. Therefore, there was insufficient information to provide additional analyses examining these specific variables.

Teachers participating in the study were provided with SAT 10 test booklets and administration manuals for their grade level in January 2011. The teachers then administered the SAT 10 pretest (Reading Comprehension and Mathematics subtests) according to the administration instructions provided. The completed test booklets and answer sheets were then returned to SEG Measurement for processing. The answer sheets were scanned and entered into a database. Any questions that the students did not answer were scored as incorrect. Students answering fewer than four questions were removed from the analysis. All data was reviewed and checked for accuracy before scoring and analysis.

At the conclusion of the school year, in May or in June 2011, following approximately 12-20 weeks of instruction, teachers administered the SAT 10 posttest (Reading Comprehension and

Mathematics subtests). The SAT 10 pretest and posttest results were compared as a basis for evaluating the growth reported in this study.

Findings

Measuring Growth

The growth in Reading Comprehension and Mathematics skills for the Treatment Group and the Control Group was compared using a statistical procedure known as analysis of covariance (ANCOVA). This approach provides an accurate way to compare growth over time controlling for any potential differences in student skills between the two study groups that may have been present at the beginning of the study. Any differences in skill levels between Pixie Users Group and Control Group that may have existed at the beginning of the study were controlled to ensure that any differences in subsequent growth were the result of Pixie use and not merely the result of differences that existed at the start of the study. Using this method, we were able to compare differences as if the two groups were matched in initial Reading and Mathematics proficiency. While no procedure can completely eliminate differences that may exist at the outset of a study, ANCOVA is widely recognized as an effective way to control for differences.

Only students for whom matched pretest and posttest results were available were included in the analysis. The analysis looked only at those students who had taken the SAT 10 at the beginning of the second semester of the school year (pretest) and those who had taken the SAT 10 at the end of the school year (posttest). Students who left the class during this period or who joined the class during this period were not included in the growth comparisons.

Pre-Post Growth for Pixie Users

Students who were in classes that used Pixie showed substantial growth from pre- to posttest in Reading Comprehension and Mathematics. During the course of the study, students in classes using Pixie increased their SAT 10 Reading Comprehension scale-scores by 14 points (Mean

pretest=645; Mean posttest score=659) and their Mathematics scale-scores by 15 points (Mean pretest=631; Mean posttest score=646). While the growth achieved by students using Pixie is an important indicator of the effectiveness of Pixie, a more complete way to assess growth is to compare the growth achieved by students in classes using Pixie to students in classes that did not use Pixie. This allows us to see the unique contribution Pixie made to students' growth.

Comparison of Treatment Group Growth to Control Group Growth

The overall growth in Reading Comprehension and Mathematics skills as measured by the Reading Comprehension and Mathematics subtests of the SAT 10 for those students in the Treatment Group was compared to the growth in Reading Comprehension and Mathematics subtests of those students in the Control Group. Multivariate Analysis of Covariance (MANCOVA) was used to evaluate the difference in a composite Reading Comprehension and Mathematics skill score (dependent variable) between the Treatment and Control Groups (independent variable) controlling for the initial Reading and Mathematics levels of the students (covariate). The SAT 10 pretest scores were used as the covariate to place students in the Treatment Group and Control Group on the same baseline. The comparisons were based on 333 Treatment Group students and 334 Control Group students for whom both pretest measures and both posttest measures were available.

The results show a significant difference in a composite of the SAT 10 Reading Comprehension and Mathematics subtest posttest scores between the Treatment Group and the Control Group ($df=2/662$; $F=3.29$; $p<.04$) when initial Reading and Mathematics skills are controlled. The results, using Pillai's Trace, are summarized in Table 5 below.

Table 5. Multivariate Analysis of Covariance Comparison of Treatment and Control Group Reading Comprehension and Mathematics Posttest Scores

Effect		Value	F	Hypothesis df	Error df	Significance
Intercept	Pillai's Trace	.056	19.59	2.00	662.00	.01
Reading Pretest	Pillai's Trace	.288	133.99[a]	2.00	662.00	.01
Mathematics Pretest	Pillai's Trace	.181	73.2	2.00	662.00	.01
Study Group	Pillai's Trace	.010	3.29	2.00	662.00	.04

To provide a more complete understanding of these results for the separate Reading and Mathematics skill areas, the individual effects were examined separately using ANCOVA (see Table 6).

Table 6. Analysis of Covariance Comparison of the Treatment Group and Control Group Reading Comprehension and Mathematics Posttest Scores

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Signi-ficance
Corrected Model	Reading Posttest	826128.31	3	275376.10	177.08	.01
	Mathematics Posttest	939373.94[b]	3	313124.65	158.46	.01
Intercept	Reading Posttest	60014.20	1	60014.12	38.59	.01
	Mathematics Posttest	28358.35	1	28358.35	14.35	.01
Reading Pretest	Reading Posttest	414474.65	1	414474.65	266.52	.01
	Mathematics Posttest	95280.95	1	95280.95	48.22	.01
Mathematics Pretest	Reading Posttest	6064.14	1	6064.14	3.90	.05
	Mathematics Posttest	254255.53	1	254255.53	128.67	.01
Study Group	Reading Posttest	7020.14	1	7020.14	4.51	.03
	Mathematics Posttest	10507.81	1	10507.81	5.32	.02
Error	Reading Posttest	1031038.86	663	1555.11		
	Mathematics Posttest	1310108.33	663	1976.03		
Total	Reading Posttest	2.884E8	667			
	Mathematics Posttest	2.771E8	667			
Corrected Total	Reading Posttest	1857167.17	666			
	Mathematics Posttest	2249482.27	666			

Table 7. Descriptive Statistics Comparison of The Treatment Group and Control Group Reading Comprehension and Mathematics Posttest Scores (Adjusted for Pretest Covariate)

Dependent Variable	Group	N	Mean SAT 10	Standard Deviation SAT 10
Reading Posttest	Treatment	334	658.95	52.18
	Control	333	651.95	53.28
	Total	667	655.45	52.81
Mathematics Posttest	Treatment	334	646.45	57.68
	Control	333	637.34	58.28
	Total	667	641.90	58.12

Reading Comprehension Growth

The SAT 10 Reading Comprehension subtest scores, for those students in classes using Pixie (Treatment Group) were compared to the SAT 10 Reading Comprehension subtest scores of those students in classes who did not use Pixie (Control Group). ANCOVA was used to evaluate the difference in Reading subtest scores (dependent variable) between the Treatment and Control Groups (independent variable) controlling for the initial reading proficiency levels of the students (covariate). The SAT 10 pretest scores were used as the covariate to place students in the Treatment Group and the Control Group on the same baseline.

The results show a significant difference in Reading Comprehension between the Treatment Group and the Control Group ($df=1/667$; $F=4.51$; $p<.03$) when initial Reading proficiency is controlled. The average Reading Comprehension subtest score for students in the Treatment Group (Mean=658.95) was significantly greater than the average Reading Comprehension subtest score achieved by students in the Control Group (Mean=651.95). This represents an effect size of +.13 (Cohen's d). The results are summarized in Table 6 and 7 (see above).

Mathematics Growth

The SAT 10 Mathematics subtest scores, for those students in classes using Pixie (Treatment Group) were compared to the SAT 10 Mathematics subtest scores of those students in classes who did not use Pixie (Control Group). ANCOVA was used to evaluate the difference in Mathematics subtest scores (dependent variable) between the Treatment and Control Groups (independent variable) controlling for the initial mathematics proficiency levels of the students (covariate). The SAT 10 pretest scores were used as the covariate to place students in the Treatment Group and the Control Group on the same baseline.

The results show a significant difference in Mathematics between the Treatment Group and the Control Group ($df=1/667$; $F=5.32$; $p<.02$) when initial Mathematics proficiency is controlled. The average Mathematics subtest score for students in the Treatment Group (Mean= 646.45) was significantly greater than the average Mathematics subtest score achieved by students in the Control Group (Mean= 637.34). This represents an effect size of +.16 (Cohen's d). The results are summarized in Table 6 and 7 (see above).

Gender Results

We examined whether there were any differences in growth between male and female students between the Treatment and Control Group (main and interaction effects). To this end, the overall growth in Reading and Mathematics skills for the Treatment Group was compared to the overall growth in Reading and Mathematics skills within the Control Group as measured by the SAT 10. MANCOVA was used to evaluate the difference in a composite reading and mathematics score (dependent variable) between the Treatment and Control Groups (independent variable) of different genders (independent variable) controlling for the initial skill levels of the students (covariate). The SAT 10 pretest scores were used as the covariate to place students in the Treatment Group and the Control Group on the same baseline. The gender comparisons were based on 336 male students and 330 female students.

The main effect for study group memberships (Treatment and Control Group) was confirmed; there was a significant difference in a composite of the SAT 10 Reading Comprehension and Mathematics posttest scores between students in the Treatment and the Control Group when initial Reading and Mathematics proficiency levels are controlled (F=3.09; df=2/659 p<.05). There were no significant effects for the interaction between gender and study group membership (F=.35; df=2/659 p<.70). This indicates that Pixie was equally effective with boys and girls. The results, using Pillai's Trace, are summarized in Table 8 (see below).

Table 8. Multivariate Analysis of Covariance Comparison of Treatment and Control Group by Gender and Reading and Mathematics Posttest Scores

Effect		Value	F	Hypothesis df	Error df	Significance
Intercept	Pillai's Trace	.056	19.50	2	659	.01
Reading Pretest	Pillai's Trace	.280	127.85	2	659	.01
Mathematics Pretest	Pillai's Trace	.183	73.97	2	659	.01
Study Group	Pillai's Trace	.009	3.09	2	659	.05
Gender	Pillai's Trace	.001	.23	2	659	.80
Study Group by Gender	Pillai's Trace	.001	.35	2	659	.70

Table 9. Descriptive Statistics Comparison of The Treatment Group and Control Group by Gender Reading Comprehension and Mathematics Posttest Scores (Adjusted for Pretest Covariate)

Descriptive Statistics					
	Group	GENDER	Mean	Std. Deviation	N
Reading PostTest	Treatment	Male	651.95	54.21	160
		Female	665.38	49.52	174
		Total	658.95	52.18	334
	Control	Male	646.40	54.05	176
		Female	658.19	52.04	156
		Total	651.94	53.36	332
	Total	Male	649.04	54.12	336
		Female	661.98	50.77	330

		Total	655.45	52.85	666
Mathematics PostTest	Treatment	Male	642.44	60.14	160
		Female	650.14	55.25	174
		Total	646.45	57.68	334
	Control	Male	633.98	59.72	176
		Female	642.01	55.61	156
		Total	637.75	57.87	332
	Total	Male	638.01	59.98	336
		Female	646.29	55.48	330
		Total	642.11	57.90	666

Ethnicity Results

We examined whether there were any differences in growth between students in different ethnic groups between the Treatment and Control Groups (main and interaction effects). To this end, the overall growth in Reading and Mathematics skills for the Treatment Group was compared to the overall growth in Reading and Mathematics skills within the Control Group as measured by the SAT 10. MANCOVA was used to evaluate the difference in a composite reading and mathematics score (dependent variable) between the Treatment and Control Groups (independent variable) of different ethnicities (independent variable) controlling for the initial skill levels of the students (covariate). The SAT 10 pretest scores were used as the covariate to place students in the Treatment Group and the Control Group on the same baseline. The ethnic comparisons were based on 360 Caucasian students, 60 African American students, 91 Hispanic students, 14 Asian or Pacific Islander students and 79 students classified as mixed race or other.

There was no significant difference in a composite of the SAT 10 Reading Comprehension and Mathematics posttest scores between students in the Treatment and the Control Group when initial Reading and Mathematics proficiency levels are controlled ($F=.50$; $df=2/591$ $p<.61$). The failure to reconfirm the main effects was the result of the reduced sample size; many teachers were unable to provide the ethnicity of students, thereby reducing the N for this model. There were no significant effects for the interaction between ethnicity and study group membership ($F=.86$; $df=8/1184$;

p<.55). This indicates that teachers' use of Pixie was equally effective for students of different ethnic groups. The results, using Pillai's Trace, are summarized in Table 10 (see below).

Table 10. Multivariate Analysis of Covariance Comparison of Treatment and Control Group by Ethnicity and Reading and Mathematics Posttest Scores

Effect		Value	F	Hypothesis df	Error df	Significance
Intercept	Pillai's Trace	.055	17.31	2	591	.01
Reading Pretest	Pillai's Trace	.271	109.87	2	591	.01
Mathematics Pretest	Pillai's Trace	.186	67.60	2	591	.01
Study Group	Pillai's Trace	.002	.50	2	591	.61
Ethnicity	Pillai's Trace	.013	.97	8	1184	.45
Study Group by Gender	Pillai's Trace	.012	.86	8	1184	.55

Table 11. Descriptive Statistics Comparison of The Treatment Group and Control Group by Ethnicity Reading Comprehension and Mathematics Posttest Scores (Adjusted for Pretest Covariate)

Descriptive Statistics					
	GROUP	ETHNICITY	Mean	Std. Deviation	N
Reading Posttest	Treatment	Caucasian	667.24	53.307	173
		African American	638.97	39.052	32
		Hispanic	652.27	49.892	52
		Asian/Pacific Islander	670.75	57.183	8
		Mixed Race/Other	638.21	46.889	33
		Total	658.47	51.850	298
	Control	Caucasian	658.86	55.215	187
		African American	623.50	46.074	28
		Hispanic	649.97	41.189	39
		Asian/Pacific Islander	682.17	37.467	6
		Mixed Race/Other	643.37	50.466	46
		Total	652.62	52.820	306
	Total	Caucasian	662.89	54.393	360

		African American	631.75	42.818	60
		Hispanic	651.29	46.132	91
		Asian/Pacific Islander	675.64	48.322	14
		Mixed Race/Other	641.22	48.763	79
		Total	655.51	52.382	604
Math Posttest	Treatment	Caucasian	651.88	59.664	173
		African American	633.63	45.746	32
		Hispanic	645.42	58.675	52
		Asian/Pacific Islander	660.00	77.831	8
		Mixed Race/Other	636.91	53.876	33
		Total	647.35	58.103	298
	Control	Caucasian	647.49	58.829	187
		African American	617.46	45.828	28
		Hispanic	632.15	52.235	39
		Asian/Pacific Islander	685.00	23.749	6
		Mixed Race/Other	626.28	52.443	46
		Total	640.34	56.643	306
	Total	Caucasian	649.60	59.190	360
		African American	626.08	46.117	60
		Hispanic	639.74	56.094	91
		Asian/Pacific Islander	670.71	60.362	14
		Mixed Race/Other	630.72	52.965	79
		Total	643.80	57.428	604

Pixie Software Teacher Survey

To provide additional information and context for the quantitative study, we conducted a qualitative study of teachers in classes using Pixie. In May and June 2011, participating teachers were asked to complete an online survey. Teachers provided information about their background, school environment, use, and perceptions of Pixie.

Sample

Teacher Background characteristics. Approximately two thirds (68%; N=13) of the Treatment Group teachers completed the survey. Eighty five percent (85%) of the teachers were female and 15% were male. Nearly half (46%) of the teachers were between the ages of 31-40, another quarter (23%) were between the ages of 21-30 and the remaining 31% were between 41 and 60 years old. All of the participating teachers were Caucasian.

School and Class Size. Nearly half (46%) of the teachers reported teaching in schools with 401-600 students, a third (31%) taught in a school of 601-800 students, and the remaining quarter (23%) taught in schools with 801-1,000 students. Nearly two thirds (62%) of the teachers taught in classes with 21-30 students, another quarter (23%) had 11-20 students in their class and the remaining teachers (15%) reported 31 or more students in their classroom.

Pixie Use. More than half (54%) of the teachers using the Pixie software began using it with their students in August or September. And the remaining half (46%) started using Pixie in October or later in the Fall. Almost half (46%) reported that their students used Pixie software one hour per week, another third (31%) reported two hours of usage per week, and the remaining quarter (23%) reported that Pixie was used three or four hours per week.

Pixie Effectiveness. Nearly two thirds (62%) of the teachers indicated that that Pixie was

"excellent" in improving students' attitudes toward school and learning, and another quarter (23%) indicated that Pixie was "good" in improving students' attitudes. Nearly two thirds (61%) of the teachers indicated that that Pixie was "excellent" or "good" in increasing their students' cognitive/intellectual growth. The remaining third (38%) rated Pixie as "fair". Nearly three quarters of the teachers (70%) indicated that Pixie was "extremely well" or "very well" aligned with instructional goals and standards that guided their instruction. Another fifth (22%) viewed Pixie as moderately well aligned with instructional goals and standards.

Continued Use. When asked how likely they were to use Pixie in the future, all of the teachers indicated that they were "definitely" or "probably" likely to use Pixie. Nearly all of the teachers (92%) said they would "definitely" or "probably" recommend Pixie for use by others.

The teachers were asked, "In your opinion, what are the most effective aspects of Pixie for your students?"

- "I think that Pixie is so effective because it allows students to organize their thoughts and present them in a fun and coherent way."
- "Sometimes students have a hard time getting what they want on paper if doing something by hand. Pixie gives them guidance with the sticker function, to really bring to life what they already are designing in their head."
- "I think the engagement piece is the most influential part of the program's success in my room. Students are engaged in the lessons and want to explore ways to demonstrate understanding."
- "That the students can build presentations to present to the class."
- "Having the students create their own learning aids and projects in areas where they feel they need more repetition."
- "Writing about their drawings. Reading their own writing and other students' writing."
- "It is open ended and suitable for all grade levels. It allows for higher level thinking skills as students demonstrate what they have learned."

- "The ease of using the program is very important. I also like that it starts as a blank slate."
- "Ease of use when creating an original picture"
- "It's fun and easy for students to use."

The least effective aspects of Pixie for their students in the teachers' opinion were:

- "The time constraints that go into some of the projects. They can often spend more time illustrating a concept, rather than deepening their understanding of the concept being taught."
- "Time to use it."
- "I don't use the templates with my students."
- "I would need to know more about Pixie's other functions before determining what is least effective."
- "Pixie is a fun program so students can easily get off task or become too caught up in the design aspects of the project and less involved in the content portion."
- "The spell check was hard to find, and when you searched for stickers if it wasn't spelled exactly the sticker wouldn't come up."
- "When the students created something with paint and stickers on one slide then it is a great program."
- "It doesn't allow for practice of grade levels standards that are tested for the state."

Summary and Discussion

During the 2010-2011 school year (between January 2011 and June 2011), SEG Measurement conducted a national study with approximately 1,000 3rd, 4th and 5th grade students, in 38 classrooms, in California, Georgia, Ohio, South Carolina, and Texas, to evaluate the impact of Pixie use on student achievement. The goal of this study was to evaluate the impact of Pixie on student learning. The results show that students who use Pixie learn significantly more than students who do not use Pixie.

Students who used Pixie showed about one half year more of growth in Language Arts and Mathematics than students who did not use Pixie.

Study Design

The primary question answered by this study is: Do students in grades 3, 4, and 5 show larger gains in Reading Comprehension and Mathematics skills when using Pixie? The study also explored potential differences in growth between boys and girls and among students of different ethnic backgrounds.

The study compared two groups of students, matched in ability, using a quasi-experimental design. The Treatment Group consisted of students who used Pixie; the Control Group consisted of students who did not use Pixie. The students in both groups were administered a (Stanford 10™) pre-test at in January 2011 and a (Stanford 10™) post-test in May/June 2011 to evaluate the impact of Pixie use on their Reading Comprehension and Mathematics growth. The results from the pretest and posttest were compared statistically to determine the level of growth in Reading Comprehension and Mathematics skills. On average, students in the Treatment Group used Pixie about one to two hours weekly, while students in the Control Group did not use Pixie at all.

Results

The Treatment Group students who used Pixie showed substantial growth in Reading Comprehension and Mathematics during the course of the study. Students in classes using Pixie increased their SAT 10 Reading Comprehension scale-scores by 14 points and their Mathematics scale-scores by 15 points. This means that the students in Pixie classes, on average, achieved about one year of growth (for the typical student at the 50th percentile), during the course of the study from January 2011 to June 2011. These estimates are based on the average gains seen by students at the 50th percentile at grades 3, 4, and 5 provided by Harcourt (2002).

The greater academic growth observed for Pixie users becomes even more visible when comparing these students against the Control Group, who did not use Pixie. Students in the Treatment Group showed statistically greater gains in both Reading Comprehension and Mathematics than the Control Group. The Treatment Group students showed substantially greater gains in Reading Comprehension (7 scale score points; Effect Size= .13) and Mathematics (9 scale score points; Effect Size=.16) than the Control Group classes. This means that, on average, students in the Pixie classes showed about one half year's more growth than their peers in classes where Pixie was not used. Again, these estimates are based on the average gains seen by students at the 50th percentile at grades 3, 4, and 5 provided by Harcourt (2002).

These effects suggest that the use of Pixie has a substantial impact on student Reading Comprehension and Mathematics skills growth. The magnitude of the results is particularly meaningful in light of the timing of the study; students using Pixie showed these gains with only 15-22 weeks of instruction in a single school semester.

We also examined the impact of teachers' use of Pixie on both boys and girls and among students of different ethnic backgrounds to determine if the solution was differentially effective for major groups within the population. The solution was found to be equally effective for boys and girls and for students of different ethnicities. In short, the interaction between Pixie use and gender and ethnicity was not statistically significant.

Summary

During the 2010-2011 school year (between January and June 2011), SEG Measurement conducted a national study with approximately 1,000 3rd, 4th, and 5th grade students, in 38 classrooms, in California, Georgia, Ohio, South Carolina, and Texas. Students who used Pixie showed meaningful growth in Reading Comprehension and Mathematics during the course of the study. Students in Treatment Group classes increased their SAT 10 scores between 14-15 points, or about one years' worth of growth. More significantly, Treatment Group students enrolled in classrooms using Pixie showed about one half year's more of growth in Reading Comprehension and Mathematics than the Control Group students enrolled in classes where Pixie was not used. The Pixie users finished the year with scores that were 7 scale-score points higher in Reading Comprehension and 9 scale-score points higher in Mathematics on the SAT 10 assessments. The study also found that Pixie is equally effective for boys and girls and for students of different ethnic backgrounds.

The quantitative results were reinforced by the qualitative data provided by teachers in classes using Pixie. All of the teachers indicated that they were likely to use Pixie in the future, and nearly all of the teachers (92%) said they would recommend Pixie for use by others. Almost all (85%) of the teachers indicated that that Pixie was effective in improving students' attitudes toward school and learning. Nearly two thirds (61%) of the teachers indicated that that Pixie was effective in increasing their students' cognitive/intellectual growth.

The findings of this study provide substantial support for the effectiveness of Pixie in improving student Reading Comprehension and Mathematics skills.

SEG Measurement