

Advisory Board:

Michael Beck, President
Beck Evaluation & Testing
Associates, Inc.

Jennifer M. Conner, Assistant
Professor
Indiana University

Keith Cruse, Former Managing
Director
Texas Assessment Program

A Pre/Post Score Analysis for the
English 3 Course
Modules 1, 5, and 6

Florida Virtual School

Report 491, March 2014



Introduction

Florida Virtual School[®] (*FLVS*[®]) contracted with the *Educational Research Institute of America* (ERIA) to determine student learning in the FLVS English 3 course. As a student works through the course, he or she will take module assessments at the beginning and end of each of six English 3 course modules. The assessments for Modules 2, 3, and 4 consist of an essay that students work on over the course of the entire module – submitting their topics, theses, introductory paragraphs, body paragraphs, and conclusions all separately. Then students work through the revision process and submit a final completed product. There are no pretests for these three modules. Thus, pretest/posttest comparisons were not possible for these three modules. However, Modules 1, 5, and 6 include both pretests and posttests and consist primarily of student-selected response questions. ERIA analyzed the student test data for these three modules to determine the effectiveness of instruction as shown by the pretest and posttest student scores.

The English 3 course was designed by *Florida Virtual School*, an established leader in developing and providing virtual kindergarten through grade 12 education solutions to students worldwide. A nationally recognized e-learning model, FLVS, founded in 1997, was the country's first statewide Internet-based public high school. In 2000, the Florida Legislature established FLVS as an independent educational entity with a gubernatorial appointed board. FLVS funding is tied directly to student performance.

Each FLVS course has a real-time teacher who guides each student through the coursework, which is organized by modules and segments. As a student works through the modules of a course, he or she will connect with the teacher to take exams online and receive discussion-based assessments over the phone. Students do the work at their own pace and on their own time, but they interact with their teachers in multiple ways—including Live Lessons, phone calls, chat, texting, and email—throughout the course.

The data collection and analysis was designed to answer two questions:

1. Is the FLVS English 3 course effective in increasing the skills and strategies of students enrolled in the course?
2. Is the FLVS English 3 course equally effective in increasing the skills and strategies of all demographic students enrolled in the course?

STUDY DESIGN

The study was based on students' scores for Modules 1, 5, and 6. Modules 2, 3, and 4 have an essay assignment instead of a pretest/posttest assessment and gain scores were not able to be computed. Researchers at ERIA conducted Paired Comparison t tests to determine if the differences in the pretest and posttest scores were significantly different. The comparisons were conducted for each of the three modules independently as well as for the total for the three modules when scores were combined. The $\leq .05$ level of significance was used as the level at which differences would be considered statistically significant.

ERIA received data files from FLVS for each of the tests. All tests were scored by ERIA. Raw scores were converted to standards scores using a mean of 300 and a standard deviation of 50. This was done to assure a more normal distribution of test scores. The score transformation is linear and does not change performance levels in any way. In addition to the comparison of the combined module scores for the three modules, sub-group analyses were conducted for various demographic subgroups.

In addition, t tests effect size analyses were computed for each of the comparisons. Cohen's d statistic was used to determine the effect size. This statistic provides an indication of the strength of the effect of the treatment regardless of the statistical significance. Cohen's d statistic is interpreted as follows:

.20 to .49 = small effect

.50 to .79 = medium effect

.80+ = large effect

Description of the English 3 Course

The following course description was provided by FLVS:

In this course, students will acquire the language, reading, writing, and speaking/listening skills necessary for success in college, career, and beyond. Students will become critical readers and thinkers as they dive deeply into the texts presented throughout this course. Students will learn how to effectively research and integrate their findings, as well as cite their sources.

Description of the English 3 Assessments (Modules 1, 5, and 6)

The tests included pretests and posttests for each instructional module. The assessments were administered to each student when they began and completed each module. All tests were administered online. Table 1 provides the number of test item groups, the number of items in each group, and the average difficulty of the items at pretesting and posttesting. The average

difficulty is the average of the individual items across all the test items. Difficulty values can range from 0 to 1. Thus, if all the students get an item correct the difficulty would be 100 percent. Pretest and posttest items administered to each student were selected from the same item pool to assure comparability of pretests and posttests.

Table 1
Number of Item Groups and Items for
English 3 Modules 1, 5, and 6 Assessments

<i>Module Assessments</i>	<i>Test Item Groups</i>	<i>Number of Test Items</i>	<i>Item Average Difficulty</i>
Module 1 Multiple Choice Pretest	6	36	77%
Module 1 Multiple Choice Posttest	6	36	83%
Module 5 Multiple Choice Pretest	9	44	49%
Module 5 Multiple Choice Posttest	9	44	63%
Module 6 Multiple Choice Pretest	8	56	55%
Module 6 Multiple Choice Posttest	8	56	68%

Demographic Characteristics of the Student Population

The analyses of the demographic characteristics of the sample are included below. Only those students who were enrolled in the course and were administered the Module 1 pretest and posttest are included in Table 2. The table shows that the population was made up of mostly *Grade 11* students whose ethnic backgrounds were primarily *White, Black, or Hispanic*. The largest percentage of students were homeschooled (54 percent) and a sizable proportion (38 percent) enrolled in public school. *Females* outnumbered *males* by 63 percent to 37 percent. A significant percentage (27 percent) of the students were eligible for *Free or Reduced Lunch Programs*. In addition to the demographic groups included in Table 2, students who were classified as *Limited English Proficient (LEP)* or were receiving *Individual Educational Plans (IEP)* were also identified. However, the numbers of these students were less than one percent and the sample sizes were too small for comparative analyses.

Table 2
Demographic Characteristics of Students Comprising the Research Sample

<i>Grade Levels</i>						
	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>		
Number	2	44	496	82		
Percent	1%	7%	80%	13%		
<i>Ethnic Groups*</i>						
	<i>Hispanic</i>	<i>American Indian</i>	<i>Asian</i>	<i>Black</i>	<i>Hawaiian/Pacific Islander</i>	<i>White</i>
Number	196	39	25	65	5	525
Percent	31%	6%	4%	10%	1%	84%
*The total number of students across ethnic groups is larger than the total number of students in the study as a number of students selected more than one ethnic group. The percentage of students choosing only one ethnic group was 70 percent and the percentage choosing two or more ethnic groups was 30 percent.						
<i>Enrolled in School Type</i>						
	<i>Charter School</i>	<i>Homeschool</i>	<i>Private School</i>	<i>Public School</i>		
Number	5	334	45	234		
Percent	1%	54%	7%	38%		
<i>Gender and Free Lunch Eligibility</i>						
	<i>Gender</i>		<i>Eligible for Free Reduced Lunch Program</i>			
	<i>Male</i>	<i>Female</i>	<i>Yes</i>			
Number	234	390	168			
Percent	37%	63%	27%			

Results

Tables 3 to 6 provide the results of the Paired Comparison *t* tests. Table 3 shows that the increase from pretesting to posttesting for Module 1 was statistically significant ($\leq .0001$). The effect size for Module 1 was small perhaps because the pretest was moderately easy and thus the pretest scores were high. This prevented the posttest scores from showing a large increase.

Table 3
English 3 Module 1
Standard Score Paired Comparison *t*-test Results

<i>Test</i>	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>Standard Deviation</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretest	624	289.5	56.0	8.517	$\leq .0001$.43
Posttest	624	310.5	40.6			

Table 4 shows that the increase from pretesting to posttesting for Module 5 was statistically significant ($\leq .0001$) and the effect size was large.

Table 4
English 3 Module 5
Standard Score Paired Comparison *t*-test Results

<i>Test</i>	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>Standard Deviation</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretest	400	281.1	47.6	11.580	$\leq .0001$.82
Posttest	400	318.9	45.0			

Table 5 shows that the increase from pretesting to posttesting for Module 6 was statistically significant ($\leq .0001$) and the effect size was medium.

Table 5
English 3 Module 6
Standard Score Paired Comparison *t*-test Results

<i>Test</i>	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>Standard Deviation</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretest	221	282.5	52.1	8.935	$\leq .0001$.75
Posttest	221	317.5	41.0			

Table 6 shows that the increase from pretesting to posttesting for the combined scores for Modules 1, 5, and 6 was statistically significant ($\leq .0001$) and the effect size was large.

Table 6
English 3 Module 1, 5, and 6 Combined
Standard Score Paired Comparison *t*-test Results

<i>Test</i>	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>Standard Deviation</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretest	171	277.9	52.3	11.279	$\leq .0001$.99
Posttest	171	322.1	36.0			

Demographic Group Comparison on Combined Scores (Modules 1, 5, and 6)

Two analyses were conducted to determine the effect of demographic characteristics on the combined module scores. The first was a series of *Independent Sample t* tests to determine if the posttest mean scores of the various subgroups were significantly different from each other. The second was a series of *Paired Sample t* tests to determine if each of the subgroups made statistically different gain scores from pretesting to posttesting.

Independent Sample Comparisons

The results for the Independent Sample *t* tests showed there were no significant differences between demographic subgroups with the exception of the differences between males and females. The results showed no statistically significant posttest score differences for the following groups.

- Grade 10 students compared to Grade 11 students
- Grade 11 students compared to Grade 12 students
- Grade 10 students compared to Grade 12 students
- Free-reduced lunch eligible students compared to non-eligible students
- Minority students compared to non-minority students
- Homeschool students compared to public school students

There was a small statistically significant difference ($\leq .01$) between male and female students.

Paired Sample Comparisons

The paired sample comparisons were conducted to determine if each of the subgroups when compared independently from each other made statistically significant gains from pretesting to posttesting. There were a total of 171 students for whom pretest and posttest scores were available for all three modules. The demographic background data was available for 167 of

these students. Table 7 provides the comparison of scores for the total group of 167 students as well as for the following sub-groups:

- Grade Level
- Gender
- Eligibility for Free/Reduced Lunch Programs
- Minority and Non-Minority
- Type of School Attended

The total group of 167 students is provided for comparison to the sub-group results. Table 7 shows that the increases from pretesting to posttesting were all statistically significant.

The effect sizes were large for:

- all students
- grade 10 and 11 students
- males and females
- non-eligible for free/reduced lunch students
- non-minority students
- homeschool students and public school students

The effect sizes were medium for:

- grade 12 students
- eligible for free-reduced lunch students
- minority students

Table 7
Comparison of English 3 Demographic Groups

Group	Number	Mean	Standard	t-Test	Significance	Effect
<i>All Students</i>						
<i>Pretest</i>	167	277.7	52.4	11.081	≤.0001	.98
<i>Posttest</i>	167	321.7	36.0			
<i>Grade 10*</i>						
<i>Pretest</i>	20	268.2	72.3	4.619	≤.0001	.95
<i>Posttest</i>	20	325.3	45.4			
<i>Grade 11</i>						
<i>Pretest</i>	119	275.8	48.6	9.761	≤.0001	1.07
<i>Posttest</i>	119	321.1	34.8			
<i>Grade 12</i>						
<i>Pretest</i>	27	291.2	51.1	3.180	≤.0001	.69
<i>Posttest</i>	27	321.6	35.7			
<i>Males Only</i>						
<i>Pretest</i>	53	290.7	51.5	5.531	≤.0001	.91
<i>Posttest</i>	53	331.9	37.9			
<i>Females Only</i>						
<i>Pretest</i>	114	271.6	51.9	9.654	≤.0001	1.03
<i>Posttest</i>	114	316.9	34.2			
<i>Free/Reduced Lunch Only</i>						
<i>Pretest</i>	45	280.7	57.0	5.880	≤.0001	.79
<i>Posttest</i>	45	318.5	37.2			
<i>No Free/Reduced Lunch Only</i>						
<i>Pretest</i>	122	276.6	50.7	9.467	≤.0001	1.06
<i>Posttest</i>	122	322.9	35.6			
<i>Non-Minority Only</i>						
<i>Pretest</i>	141	275.8	52.0	10.793	≤.0001	1.03
<i>Posttest</i>	141	321.8	35.7			
<i>Minority Only</i>						
<i>Pretest</i>	26	287.9	54.0	3.084	≤.005	.70
<i>Posttest</i>	26	320.9	38.3			
<i>Homeschool Only**</i>						
<i>Pretest</i>	113	277.7	52.1	8.787	≤.0001	1.028
<i>Posttest</i>	113	323.4	35.1			
<i>Public School Only</i>						
<i>Pretest</i>	43	267.0	54.2	7.667	≤.0001	.99
<i>Posttest</i>	43	313.8	39.1			

*There was one grade 9 student for which no analysis was conducted.

**There were eight private school students and one charter school student. Those were too few to conduct an analysis.

Figures 1, 2, 3, 4, and 5 provide a graphic view of the increases shown in Table 7. In general, the increases were between 40 to 50 standard score points for each comparison group. An increase of 50 points would be one full standard deviation. These increases were just a bit less than a standard deviation, which is a reasonable increase for module tests that focus on a specific segment of instruction and for tests that had relatively high pretest scores and thus limited gains.

Figure 1
Combined Standard Score Increases by Grade Level

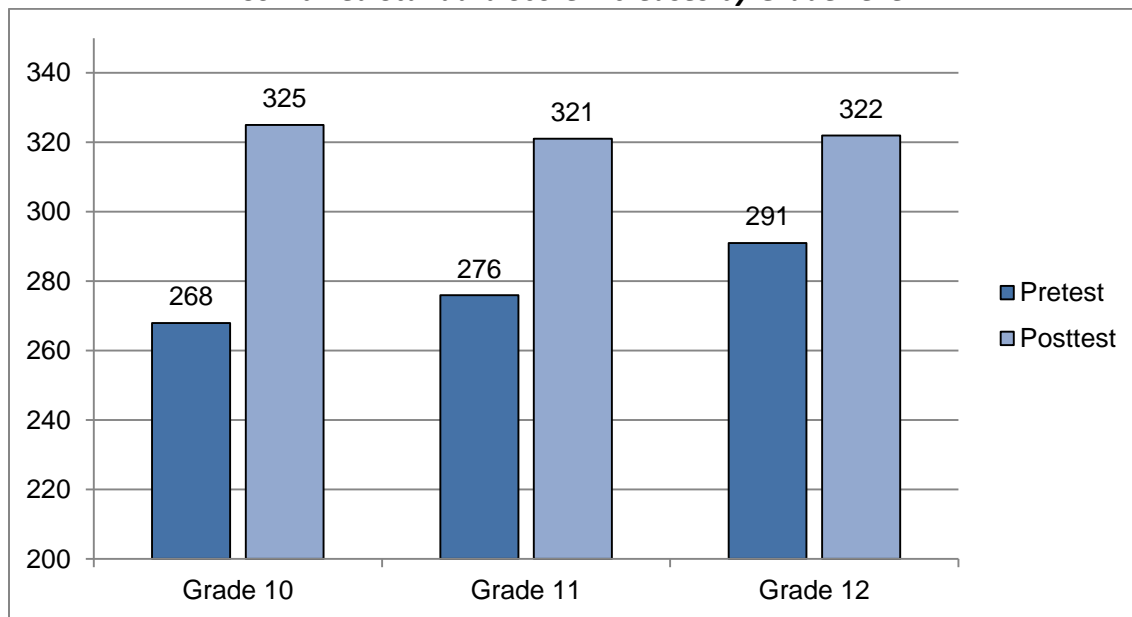


Figure 2
Combined Standard Score Increases by Gender

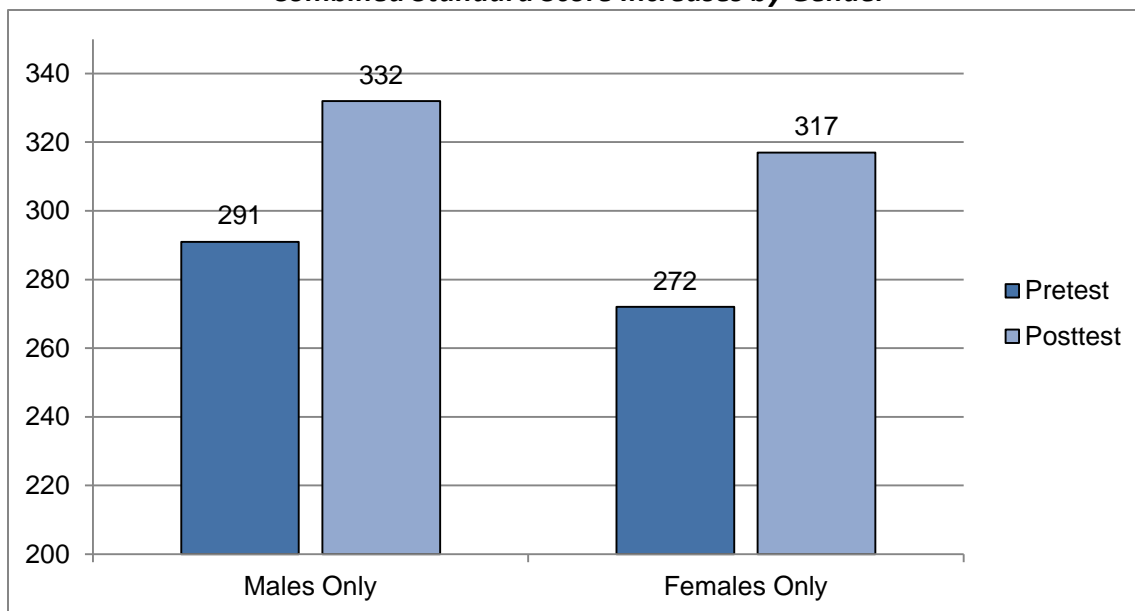


Figure 3
Combined Standard Score Increases by
Eligibility for Free/Reduced Price Lunch Program



Figure 4
Combined Standard Score Increases by Minority/Non-Minority Classification

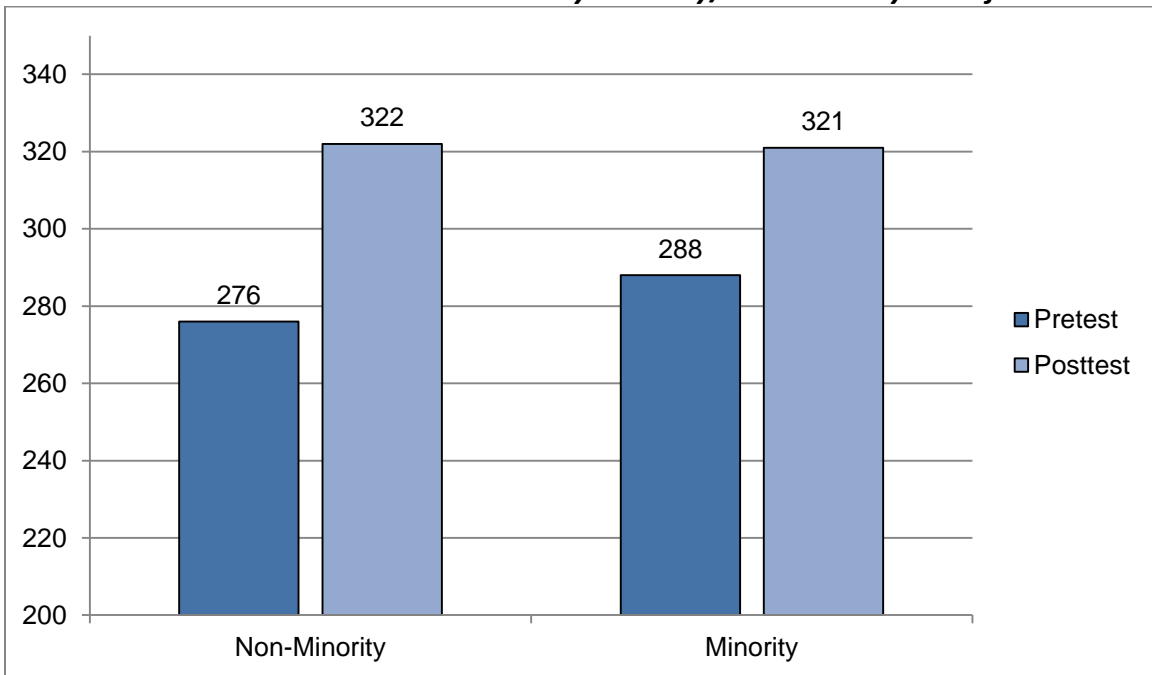
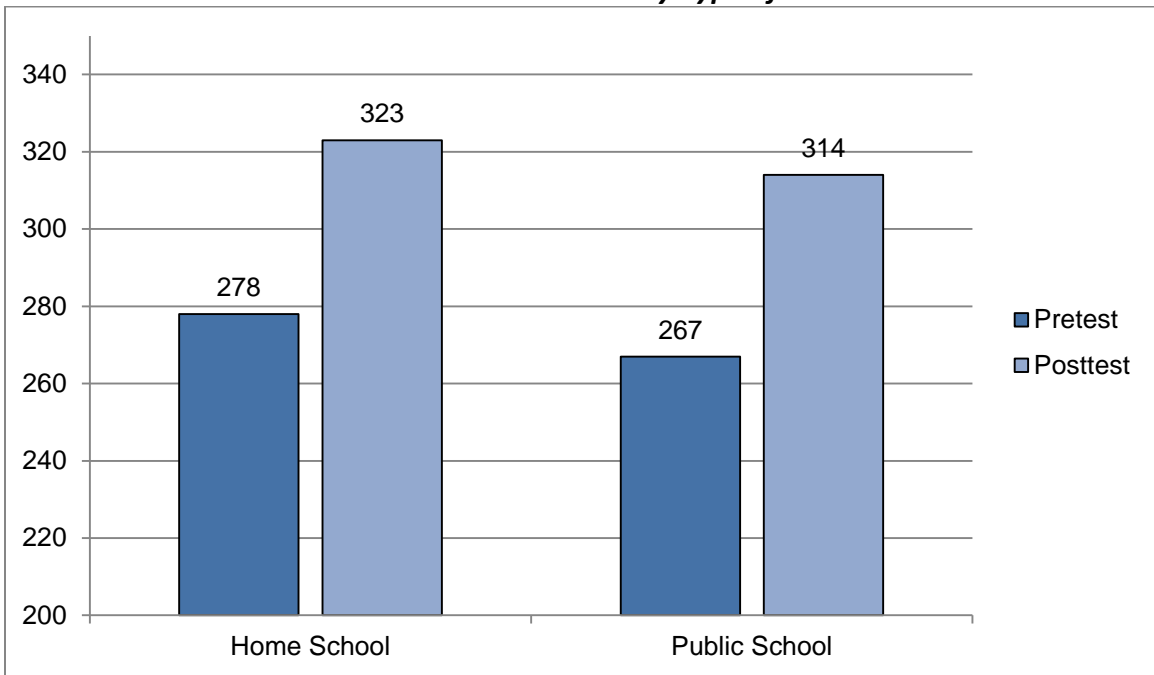


Figure 5
Combined Standard Score Increases by Type of School Attended



Conclusions

1. Is the FLVS English 3 course effective in increasing the skills and strategies of students enrolled in the course?

The study provides significant results that support the effectiveness of the instruction for FLVS English 3 modules instruction. The increase in scores was statistically significant from pretesting to posttesting for each of the three modules included in the study. In addition, the increase for all three modules combined was also statistically significant. The effect sizes were small for Module 1, large for Module 5, medium for Module 6, and large for all three modules combined.

The efficacy study supports the conclusion that the module instruction for English 3 is effective and gain scores are large for the combined module scores.

2. Is the FLVS English 3 course equally effective in increasing the skills and strategies of students in various demographic subgroups enrolled in the course?

The results show no statistically significant differences in posttest scores for the comparison of one demographic group to another with the exception of males compared to females with the males scoring a bit higher than the females ($\leq .01$).

When each group was analyzed comparing pretest scores to posttest scores, the results show that there were significant gain scores for all 11 of the demographic pretest/posttest group comparisons and the effect sizes for eight of the 11 comparisons were large. There were medium effect sizes for three groups, grade 12 students (.69), students eligible for free/reduced lunch programs (.79), and minority students (.70). These medium effect sizes were at the higher end of medium and not much different from the large effect sizes for the other eight groups. A summary of the comparisons between various demographic subgroups are shown in Table 8.

In summary, the results clearly show that there were no statistically significant differences between subgroups with the exception of a slight advantage of males compared to females. Further, the results show that each group showed statistically significant pretest/posttest gains with either large or medium effect sizes.

The efficacy study supports the conclusion that the instruction for English 3 is generally equally effective for all demographic groups.

Table 8
Summary of Comparison of English 3 Demographic Groups

Grade Level		
<i>Grade</i>	Statistically Significant	Effect Size
<i>10</i>	Yes	Large
<i>11</i>	Yes	Large
<i>12</i>	Yes	Medium
Gender		
<i>Gender</i>	Statistically Significant	Effect Size
<i>Male</i>	Yes	Large
<i>Female</i>	Yes	Large
Eligibility for Free/Reduced Lunch Program		
<i>Eligibility</i>	Statistically Significant	Effect Size
<i>Yes</i>	Yes	Medium
<i>No</i>	Yes	Large
Ethnic Category		
<i>Category</i>	Statistically Significant	Effect Size
<i>Minority</i>	Yes	Medium
<i>Non-Minority</i>	Yes	Large
School Type		
<i>School</i>	Statistically Significant	Effect Size
<i>Homeschool</i>	Yes	Large
<i>Public School</i>	Yes	Large