

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 of the
English 4: Florida College Prep Course
Modules 1 to 4

Florida Virtual School

Report 496, September, 2014

Contents

| | |
|---|----|
| Introduction | 2 |
| Study Design..... | 3 |
| Description of the English 4: Florida College Prep Course | 4 |
| Description of the English 4: Florida College Prep Course Assessments..... | 4 |
| (Modules 1 to 4)..... | 4 |
| Demographic Characteristics of the Student Population | 4 |
| Results..... | 6 |
| Demographic Group Comparison on Combined Scores (Modules 1, 2, 3, and 4)..... | 7 |
| Paired Sample Comparisons | 8 |
| Comparisons of Gains by Demographic Subgroups..... | 12 |
| Conclusions | 16 |

Introduction

Florida Virtual School® (*FLVS*®) contracted with the *Educational Research Institute of America* (ERIA) to determine student learning in the FLVS English 4: Florida College Prep (FCP) course. As a student works through the course, he or she will take module assessments at the beginning and end of each of four English 4: FCP course modules. Modules 1 to 4 include both pretests and posttests and consist of student-selected response questions. ERIA analyzed the student test data for these four modules to determine the effectiveness of instruction as shown by the pretest and posttest student scores.

The English 4: FCP course was designed by *FLVS*, 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 state-wide 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 4: Florida College Prep course effective in increasing the skills and strategies of students enrolled in the course?
2. Is the *FLVS* English 4: Florida College Prep course equally effective in increasing the skills and strategies of students with different demographic characteristics?

Study Design

The study was based on students' scores for Modules 1, 2, 3, and 4. Researchers at ERIA conducted *Paired Comparison t-tests* to determine if the differences in the pretest and posttest scores were significantly different. ERIA received data files from FLVS for each of the tests. All tests were scored by ERIA. Raw scores were converted to standard 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.

Comparisons were conducted for each of the four modules independently as well as for the total for the four modules when scores were combined. The $\leq .05$ level of significance was used as the level at which differences would be considered statistically significant.

In addition to the comparison of the combined module scores for the four modules, analyses were conducted for the sub-groups of various demographic groups. *Paired Comparison t-tests* were used to determine if each of the sub-groups of the five demographic groups made statistically significant gains. In addition, *Repeated Measure Analysis of Variance (ANOVA)* analyses were computed to determine if the gains score of one sub-group was significantly larger than the gain score of another sub-group within the same demographic group.

In addition *t-tests* effect size analyses were computed for each of the comparisons using *Cohen's d* statistic. For the *Paired Comparison ANOVA* the *Partial Eta Squared* statistic was used to determine effect sizes. Both the *Cohen d* and the *Partial Eta Squared* statistics provide an indication of the strength of the effect size 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

Partial Eta Squared is interpreted as follows:

- .10 to .24 = small effect
- .25 to .39 = medium effect
- .40+ = large effect

Description of the English 4: Florida College Prep 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 4: Florida College Prep Course Assessments (Modules 1 to 4)

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 4: Florida College Prep Module Assessments 1 to 4

| <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 | 32 | 96 | 68% |
| Module 1 Multiple Choice Posttest | 32 | 96 | 78% |
| Module 2 Multiple Choice Pretest | 20 | 60 | 63% |
| Module 2 Multiple Choice Posttest | 20 | 60 | 79% |
| Module 3 Multiple Choice Pretest | 20 | 60 | 52% |
| Module 3 Multiple Choice Posttest | 20 | 60 | 70% |
| Module 4 Multiple Choice Pretest | 20 | 60 | 52% |
| Module 4 Multiple Choice Posttest | 20 | 60 | 70% |

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 12* students whose ethnic backgrounds were primarily White, Black, or Hispanic. The

largest percentage of students were public school students (82 percent) and a sizable proportion (15 percent) were homeschooled students. Males and females were represented in almost equal numbers. A significant percentage (42 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 Characteristic of Students Comprising the Research Sample

| <i>Grade Levels</i> | | | | | | |
|---|-----------------------|------------------------|--|----------------------|----------------------------------|--------------|
| | <i>10</i> | <i>11</i> | <i>12</i> | | | |
| Number | 4 | 114 | 222 | | | |
| Percent | 1% | 33% | 65% | | | |
| <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 | 54 | 14 | 8 | 66 | 4 | 273 |
| Percent | 16% | 4% | 3% | 19% | 1% | 80% |
| *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 80 percent and the percentage choosing two or more ethnic groups was 20 percent. | | | | | | |
| <i>Enrolled in School Type</i> | | | | | | |
| | <i>Charter School</i> | <i>Homeschool</i> | <i>Private School</i> | <i>Public School</i> | | |
| Number | 7 | 50 | 6 | 278 | | |
| Percent | 2% | 15% | 2% | 82% | | |
| <i>Gender, Individual Education Plan, Free Lunch Eligibility for Free/Reduced Lunch Program, and Limited English Proficiency</i> | | | | | | |
| | <i>Gender</i> | | <i>Eligible for Free Reduced Lunch Program</i> | | | |
| | <i>Male</i> | <i>Female</i> | <i>Yes</i> | | | |
| Number | 175 | 167 | 143 | | | |
| Percent | 51% | 49% | 42% | | | |

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 medium.

Table 3
English 4 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 | 384 | 299 | 17.6 | 12.394 | $\leq .0001$ | .67 |
| Posttest | 384 | 311 | 18.1 | | | |

Table 4 shows that the increase from pretesting to posttesting for Module 2 was statistically significant ($\leq .0001$) and the effect size was large.

Table 4
English 4 Module 2
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 | 383 | 299 | 10.2 | 18.199 | $\leq .0001$ | .95 |
| Posttest | 383 | 309 | 10.8 | | | |

Table 5 shows that the increase from pretesting to posttesting for Module 3 was statistically significant ($\leq .0001$) and the effect size was large.

Table 5
English 4 Module 3
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 | 288 | 299 | 8.9 | 15.343 | $\leq .0001$ | 1.01 |
| Posttest | 288 | 309 | 10.7 | | | |

Table 6 shows that the increase from pretesting to posttesting for Module 4 was statistically significant ($\leq .0001$) and the effect size was large.

Table 6
English 4 Module 4
Standard Score Paired Comparison *t*-test Results

| Test | Number of Students | Mean Standard Score | Standard Deviation | <i>t</i> -test | Significance | Effect Size |
|----------|--------------------|---------------------|--------------------|----------------|--------------|-------------|
| Pretest | 281 | 299 | 9.2 | 14.166 | $\leq .0001$ | .96 |
| Posttest | 281 | 308 | 9.6 | | | |

Table 7 shows that the increase from pretesting to posttesting for the combined scores for Modules 1 to 4 combined was statistically significant ($\leq .0001$) and the effect size was large.

Table 7
English 4 Module 1, 2, 3, and 4 Combined
Standard Score Paired Comparison *t*-test Results

| Test | Number of Students | Mean Standard Score | Standard Deviation | <i>t</i> -test | Significance | Effect Size |
|----------|--------------------|---------------------|--------------------|----------------|--------------|-------------|
| Pretest | 189 | 285 | 45.0 | 18.142 | $\leq .0001$ | 1.28 |
| Posttest | 189 | 343 | 45.2 | | | |

Demographic Group Comparison on Combined Scores (Modules 1, 2, 3, and 4)

Two analyses were conducted to determine the effect of demographic characteristics on the combined module scores. The first was a series of *Paired Sample Comparisons (Dependent Sample t tests)* to determine if each of the subgroups made statistically different gain scores from pretesting to posttesting.

A second analysis was conducted to determine if any of subgroups of a particular demographic group made larger gains than another subgroup. For example the first analysis showed that both male and female students made statistically significant gains. However, that does not answer the question as to whether males made larger gains than females or whether the reverse was true. To test this concern, a *Repeated Measures Analysis of Variance (ANOVA)* was carried out for each of the five demographic groups.

Paired Sample Comparisons

In the first analysis, the paired sample comparisons (Dependent Sample t-tests) 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 four modules. The demographic background data was available for 143 of these students. Table 8 provides the comparison of scores for the total group of 143 students as well as for the following sub-groups:

- Grade Level
- Minority and Non-Minority
- Public and Homeschool Students
- Male and Female Students
- Qualified for Free/Reduced Price Lunch Programs and Non-Qualified Students

The total group of 143 students is provided for comparison to the sub-group results. Table 8 shows that the increases from pretesting to posttesting were all statistically significant. The effect sizes were also large for all of the sub-groups.

Table 8
Comparison of English 4 Demographic Groups

| Group | Number | Mean | Standard | t-Test | Significance | Effect |
|--------------------------------|--------|------|----------|--------|--------------|--------|
| <i>All Students</i> | | | | | | |
| <i>Pretest</i> | 143 | 286 | 45.6 | 15.958 | ≤.0001 | 1.29 |
| <i>Posttest</i> | 143 | 344 | 44.6 | | | |
| <i>Grade 11</i> | | | | | | |
| <i>Pretest</i> | 50 | 281 | 44.9 | 9.941 | ≤.0001 | 1.48 |
| <i>Posttest</i> | 50 | 346 | 43.1 | | | |
| <i>Grade 12</i> | | | | | | |
| <i>Pretest</i> | 92 | 290 | 46.1 | 12.442 | ≤.0001 | 1.17 |
| <i>Posttest</i> | 92 | 343 | 45.8 | | | |
| <i>Minority</i> | | | | | | |
| <i>Pretest</i> | 31 | 283 | 49.7 | 6.296 | ≤.0001 | 1.11 |
| <i>Posttest</i> | 31 | 335 | 43.5 | | | |
| <i>Non-Minority</i> | | | | | | |
| <i>Pretest</i> | 109 | 288 | 44.8 | 14.465 | ≤.0001 | 1.31 |
| <i>Posttest</i> | 109 | 347 | 45.4 | | | |
| <i>Public School Students*</i> | | | | | | |
| <i>Pretest</i> | 110 | 286 | 43.4 | 13.445 | ≤.0001 | 1.28 |
| <i>Posttest</i> | 110 | 341 | 42.7 | | | |

| <i>Home School Students</i> | | | | | | |
|---|----|-----|------|--------|--------|------|
| <i>Pretest</i> | 25 | 293 | 58.1 | 7.468 | ≤.0001 | 1.29 |
| <i>Posttest</i> | 25 | 363 | 49.8 | | | |
| <i>Male Students</i> | | | | | | |
| <i>Pretest</i> | 60 | 275 | 45.0 | 13.748 | ≤.0001 | 1.40 |
| <i>Posttest</i> | 60 | 341 | 49.5 | | | |
| <i>Female Students</i> | | | | | | |
| <i>Pretest</i> | 83 | 295 | 44.5 | 10.156 | ≤.0001 | 1.27 |
| <i>Posttest</i> | 83 | 347 | 40.9 | | | |
| <i>Non-Qualified for Free/Reduced Lunch Program</i> | | | | | | |
| <i>Pretest</i> | 94 | 292 | 45.2 | 12.831 | ≤.0001 | 1.28 |
| <i>Posttest</i> | 94 | 350 | 45.1 | | | |
| <i>Qualified for Free/Reduced Lunch Program</i> | | | | | | |
| <i>Pretest</i> | 49 | 276 | 45.0 | 9.399 | ≤.0001 | 1.33 |
| <i>Posttest</i> | 49 | 334 | 42.1 | | | |

*There were two private school students and six charter school students. 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 8. In general, the increases were between 50 to 70 standard score points for each comparison group. An increase of 50 points would be one full standard deviation so these increases were greater than a standard deviation which is a reasonable increase for module tests which focus on a specific segment of instruction and for tests which had relatively high pretest scores and thus perhaps limited gains because of those high pretest scores.

Figure 1
Combined Standard Score Increases by Grade Level

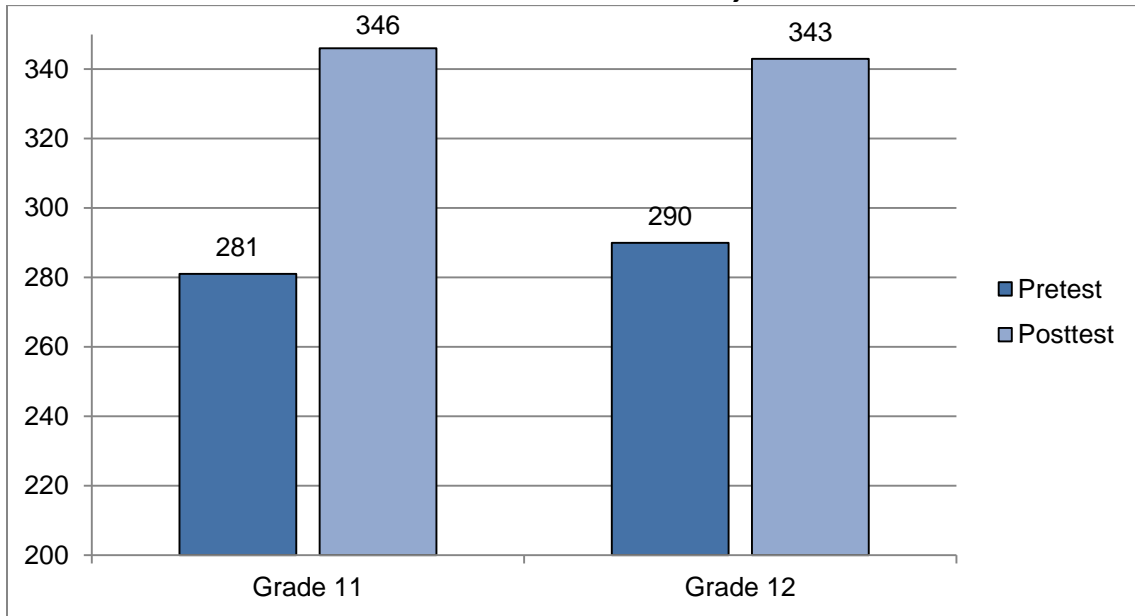


Figure 2
Combined Standard Score Increases by Minority/Non-Minority Classification

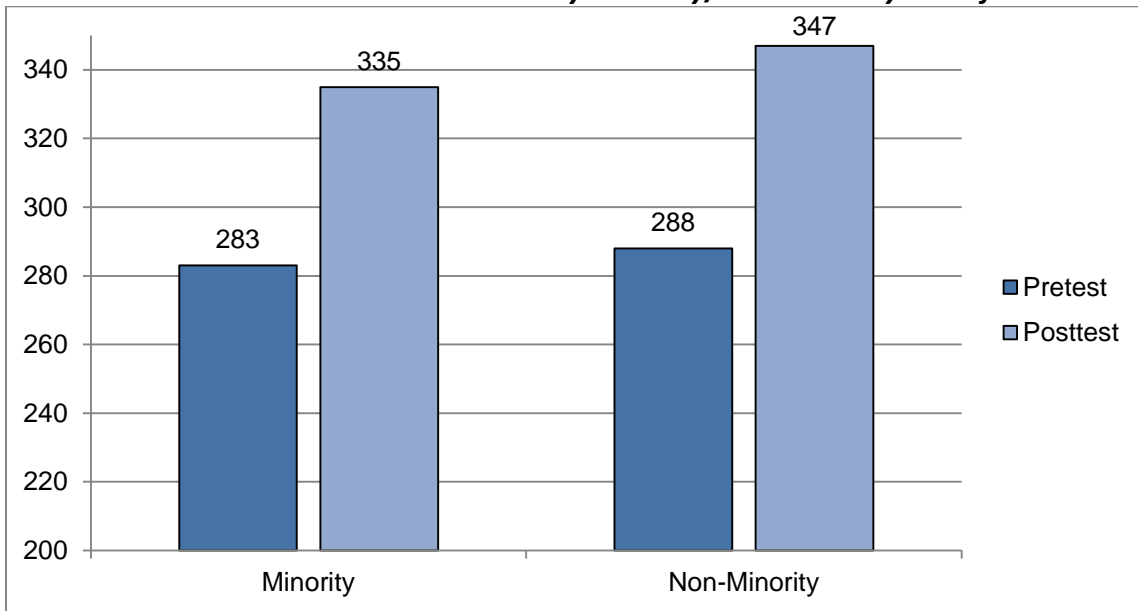


Figure 3
Combined Standard Score Increases by
Type of School Attended

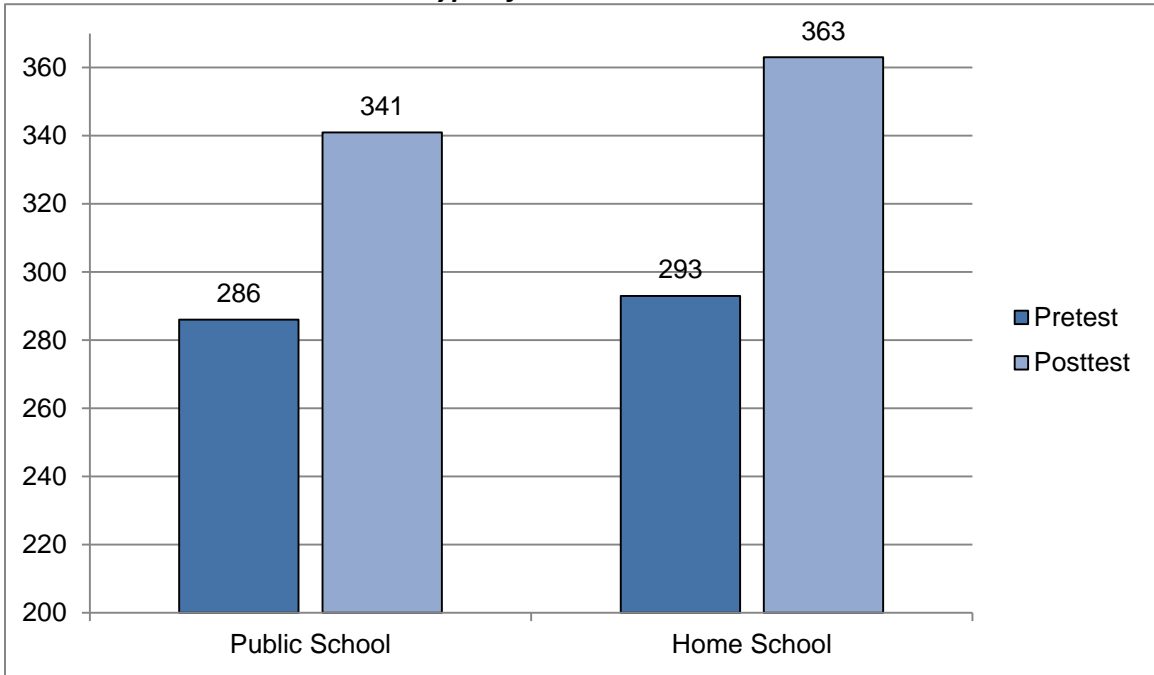


Figure 4
Combined Standard Score Increases by Gender

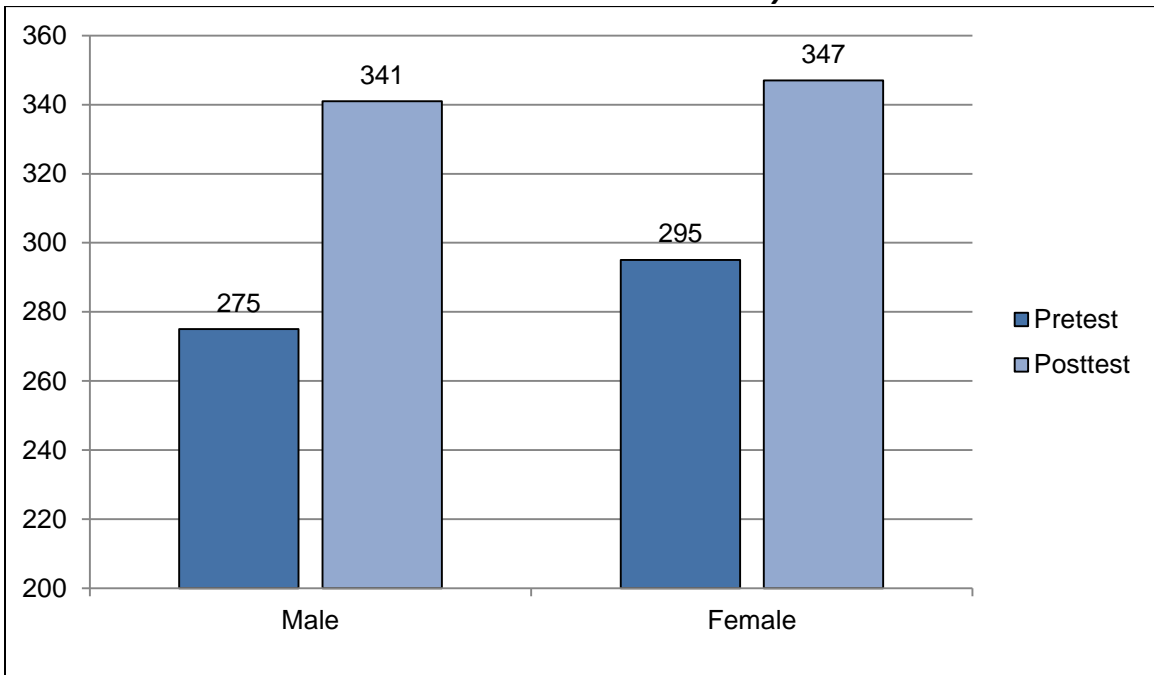
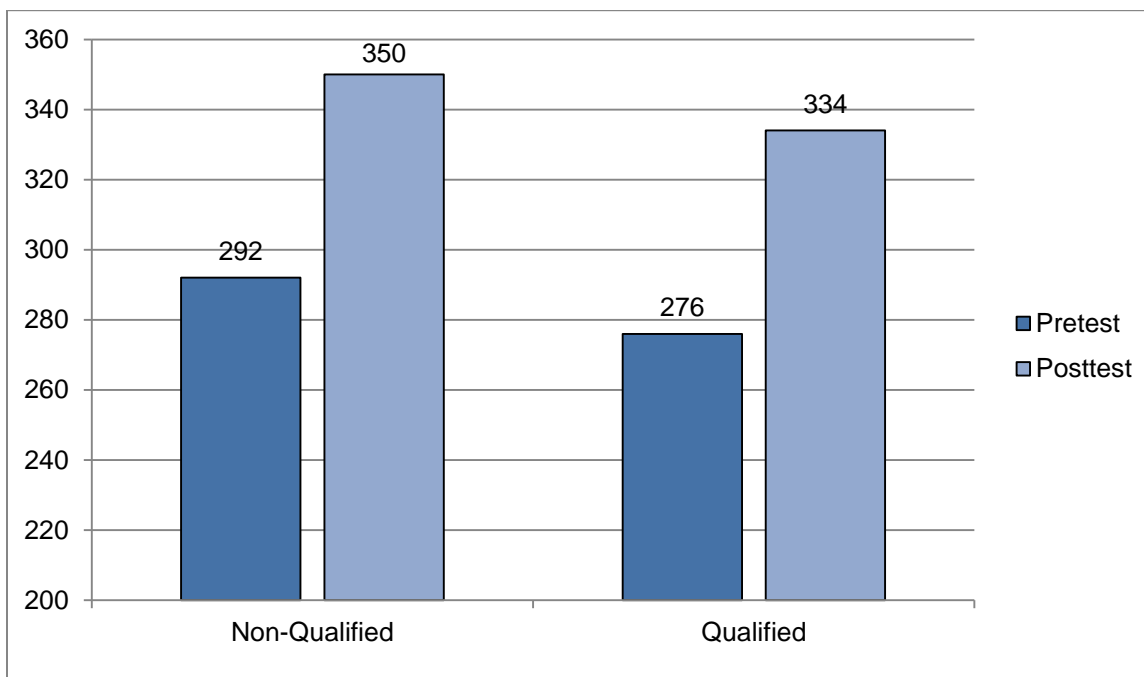


Figure 5

Combined Standard Score Increases by Qualification for Free/Reduced Lunch Programs



Comparisons of Gains by Demographic Subgroups

In the second analysis, the results of the five *Repeated Measures ANOVA* tests show that none of the comparisons were statistically significant. The strength of effect sizes are all very small.

All of the effect sizes were very small which would be expected with non-statistically significant differences.

1. There was no statistically significant effect **when grade 11 students were compared to grade 12 students**, $F(1, 141) = 2.434$, $p = \leq .121$, *Partial Eta Squared* 0.017.
2. There was no statistically significant effect when **minority students were compared to non-minority students**, $F(1, 141) = .426$, $p = \leq .515$, *Partial Eta Squared* 0.003.
3. There was no statistically significant effect when **home school students were compared to public school students**, $F(1, 133) = 2.292$, $p = \leq .132$, *Partial Eta Squared* 0.017.
4. There was no statistically significant effect when **male students were compared to female students**, $F(1, 141) = 3.433$, $p = \leq .066$, *Partial Eta Squared* 0.024.
5. There was no statistically significant effect when **free/reduced lunch eligible were compared to free/reduced lunch ineligible**, $F(1, 141) = .006$, $p = \leq .938$, *Partial Eta Squared* 0.0001.

A graphic view of the effect sizes are shown in figures 6 to 10. The charts show that the subgroups of the five demographic groups made very similar gains. Any differences as shown by the *Repeated Measures ANOVA* tests were statistically non-significant.

Figure 6
Gain Scores Comparing
Grade 11 and Grade 12 Students

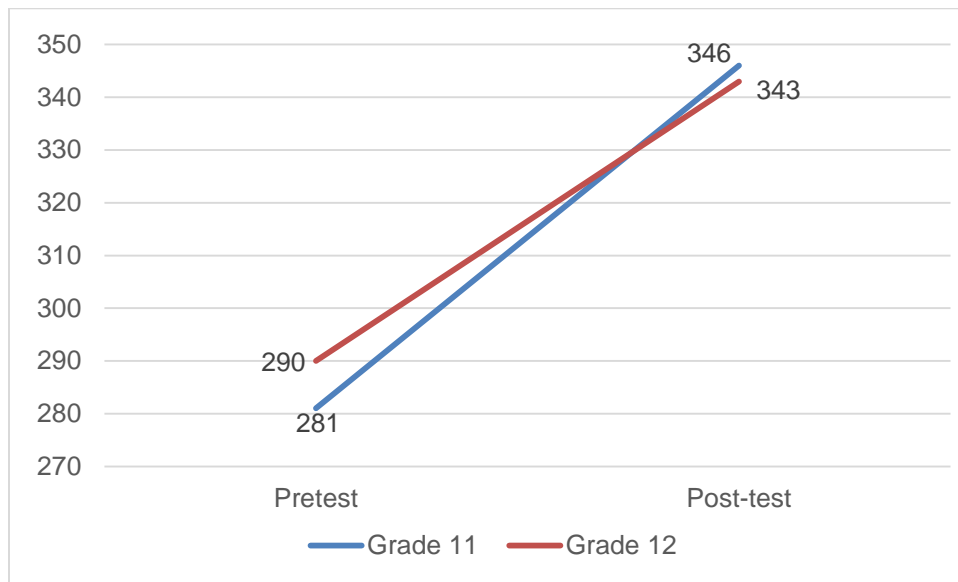


Figure 7
Gain Scores Comparing
Minority Students and Non-Minority Students

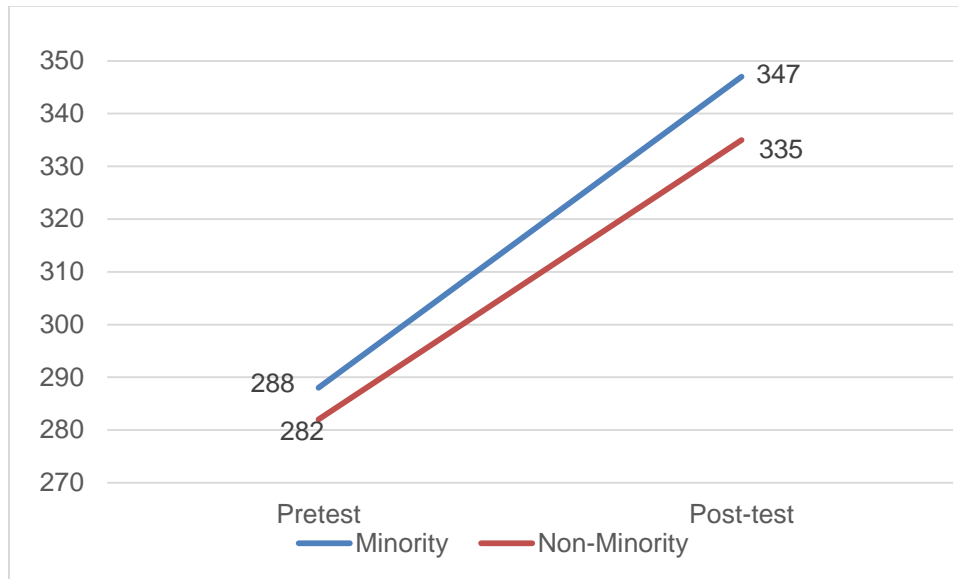


Figure 8
Gain Scores Comparing
Home School Students and Public School Students

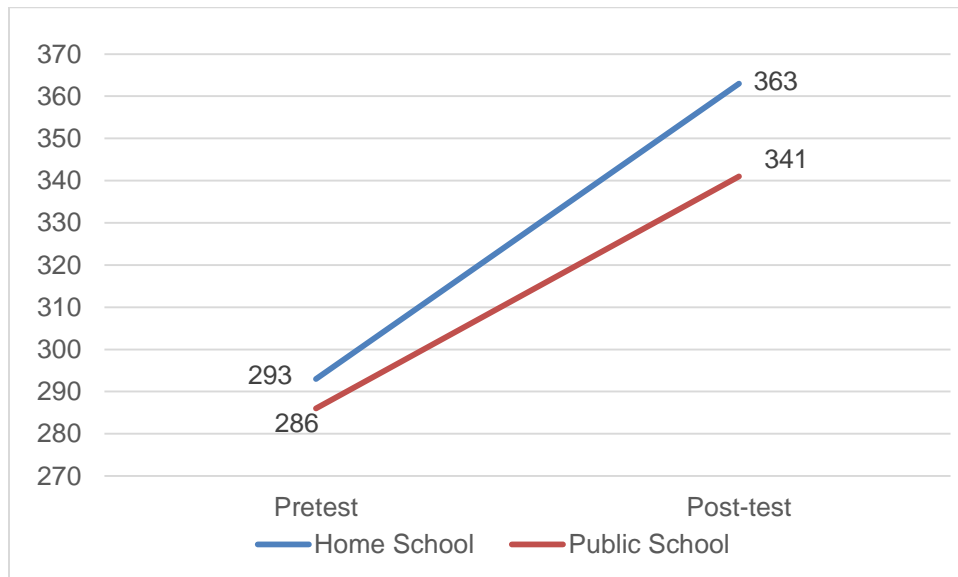


Figure 9
Gain Scores Comparing
Male Students and Female Students

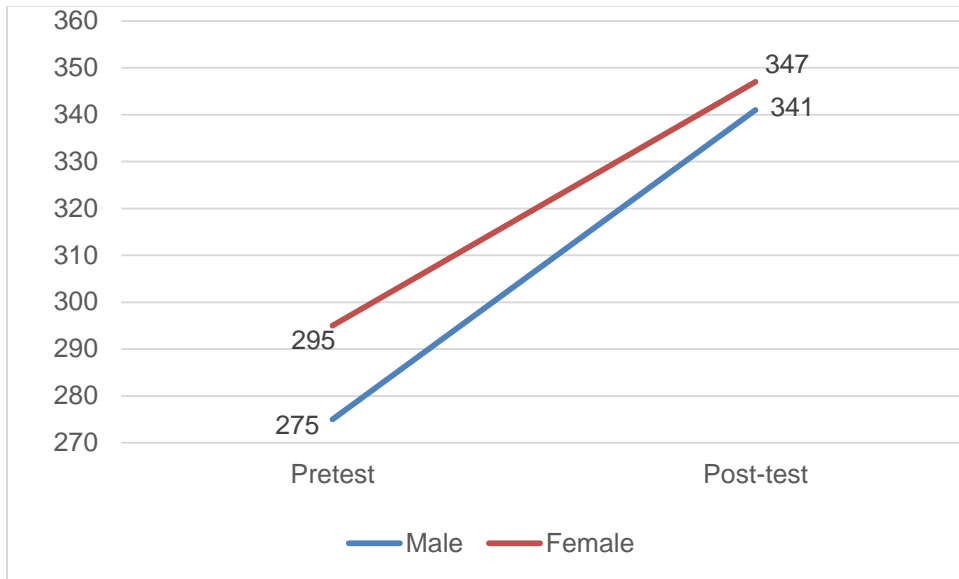
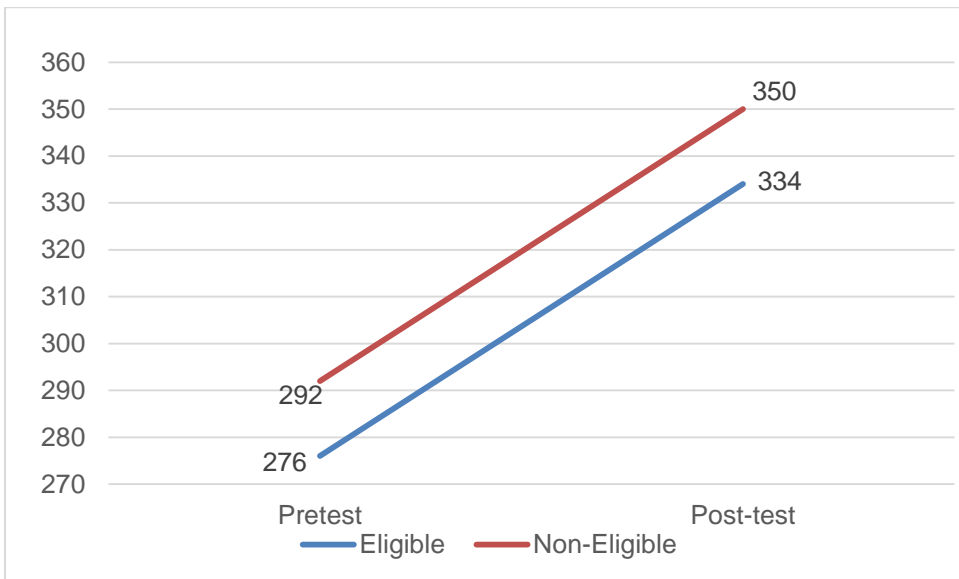


Figure 10
Gain Scores Comparing
Free/Reduced Lunch Eligible Students to Non-Eligible Students



Conclusions

1. Is the FLVS English 4: Florida College Prep 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 the FLVS English 4: FCP modules. The increase in scores was statistically significant from pretesting to posttesting for each of the four modules included in the study. In addition, the increase for all four modules combined was statistically significant. The effect sizes were medium for Module 1 and large for Modules 2, 3, and 4 and also large for the four modules combined.

The efficacy study supports the conclusion that the module instruction for English 4 is effective and gain scores are large for the combined module scores.

2. Is the FLVS English 4: Florida College Prep course equally effective in increasing the skills and strategies of students in various demographic subgroups enrolled in the course?

The first question for the five demographic characteristics was to determine if the sub-groups of each of the demographic groups increased statistically significantly. The results show that not only did each of the sub-groups increase statistically significantly, the effect sizes were large for every question.

The second question was to determine whether the gain scores of the sub-groups were different. A *Repeated Measures ANOVA* was used to compare the sub-group gain score differences. The results showed there were no statistically significant differences between the sub-groups for each demographic group. Moreover, the effect sizes were very small.

In summary, the results clearly show that the sub-groups increased statistically significantly and there were no differences between the gain scores for the various sub-groups. The English 4 proved to be effective with all demographic groups and no sub-group made larger gains than any other sub-group.

The efficacy study supports the conclusion that the module instruction for English 4 is effective and is equally effective for all demographic groups.