Educational Research Institute of America

An Efficacy Study of the Biology I Program (Standard Setting)

Florida Virtual School

[Report 461, January 2013]

Contents

Executive Summary

Florida Virtual School® contracted with the Educational Research Institute of America to conduct a study to analyze the end-of-course test score data for students enrolled in the Biology I version 15 course. The FLVS Biology I course is mapped to national standards and designed to meet Common Core and Florida Sunshine State Standards.

For this study, an end-of-course test was used. The end-of-course was made up of two parts. The first part was an assessment administered when the first half of the course was completed. The second part was an assessment administered when the entire course was completed. The study included the setting of cut-scores to student performance. The cut-scores were then applied to the test score data for students who had completed the Biology I course and both the mid-course and end-of-course assessments.

Florida Virtual School (FLVS®) is 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 is the only public school with funding tied directly to student performance.

Each course has a real-time teacher who guides each student through the coursework, which is broken down into modules. 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.

In a previous pretest/posttest study, *Report 447, An Efficacy Study of the Biology I Program* statistical comparisons of students' performance were conducted and reported. The study was based on six module tests which covered the total content for the course. The results showed statistically significant gains from pretesting to posttesting for all six modules.

For this study the focus was on setting standards so that cut scores could be used to evaluate students' performance. The development of standards is a common practice used in most states as well as a number of school districts. The forthcoming assessments being developed on a national level for the assessment of the Common Core State Standards will follow this procedure.

For this procedure the agency using the assessments produces a set of Performance Level Descriptors (PLDs) that describe the level of performance they would like to use to identify students' achievement. This was done by FLVS and a 4 level set of PLDs was developed.

A committee of curriculum experts was then assembled to review the test items using a format in which the test items are ranked from easiest to most difficult. The committee met for several days to determine their best judgments as to the number of items a student should answer correctly to be placed at each of the four levels. The performance levels and the scores determined for each level are listed below. Also listed is the percentage of students who actually scored at each of the four levels.

Performance Level	Assessment Raw Score	Percent of Students Scoring at this Level
Needs Improvement	0 to 17	0
Novice	18 to 30	2%
Capable	31 to 68	86%
Advanced	69 to 82	12%

Research Design

Carefully constructed studies are needed to determine the efficacy of online courses as these types of courses continue to expand and provide an important education opportunity to students. The enrichment of a student's educational opportunities through online courses can help to prepare students for the demands of post-secondary education and the workplace. FLVS has developed a unique approach to online course instruction which combines excellent online resources accompanied by significant support and guidance from teachers.

This study entailed the development of Performance Level Descriptors (PLDs), the establishment of cut-scores to determine the scores that would determine each students level of performance, and the application of the PLDs to students' end-of-course level of performance.

To develop the PLDs and the cut-scores, FLVS contracted with two national assessment agencies with particular expertise in educational measurement and standard setting activities. These organizations, *Educational Research Institute of America (ERIA)* and *Beck Evaluation & Testing Associates, Inc. (BETA*), worked with FLVS curriculum and management staff to plan, to carry out, and to implement the standard-setting activities.

Research Questions

The following questions guided the design of the study and the data analyses:

1. Do students enrolled in the **Florida Virtual School Biology I** course achieve success as assessed by their end-of-course test scores?
2. Do students enrolled in basic or honors courses achieve similar gains in the **Florida Virtual School Biology I** program?
3. Do students with differing demographic characteristics (gender, ethnic background, and socio-economic status) achieve similar performance levels when enrolled in the **Florida Virtual School Biology I** program?

Course Description

The course is designed with a total of seven instructional modules. These modules include instructional activities to meet a specific set of standards for each module.

More specifically, the Biology I course guides students through the study of living and non-living systems and how they interact with one another. Students explore the world they live in by posing questions and seeking answers through scientific inquiry. Discovery takes place through

observation and data collection. The students are introduced to the structure, function, diversity, and evolution of living matter. This course aims to make the content applicable to real-life for students. It encourages curiosity and provides opportunity for students to work on hands-on lab activities and develop relationships through collaboratively learning. By providing these opportunities, FLVS engages students in the study of biological science to ultimately broaden the picture of the world around them.

Segment I:

Module 1: Foundations of Biology

Module 2: Life's Origin

Module 3: Cell Reproduction

Module 4: Earth's Diversity

Segment II:

Module 5: Scientific Connections

Module 6: Classification and Diversity

Module 7: Human Systems

Besides engaging students in challenging curriculum, FLVS guides students to reflect on their learning and to evaluate their progress through a variety of assessments. Assessments can be in the form of self-checks, collaboration activities, practice lessons, multiple-choice questions, writing assignments, projects, research papers, essays, discussion-based assessments, and student discussions. State and nationally-recognized educational standards and frameworks guide assessment design. Instructors evaluate progress and provide interventions through the variety of assessments built into the course, as well as through contact with the student in other venues.

Description of the Research Sample

The study included students enrolled in the Biology I course between *August 23, 2010 and October 30, 2012.*

Tables 1 to 3 provide a description of the demographic characteristics of the students included in the analysis.

Table 1: Grade Levels of Students Comprising the Research Sample

Grade Levels				
8	*9*	*10*	*11*	*12*
3%	23%	36%	25%	13%

Table 2: Ethnicity of Students Comprising the Research Sample

Ethnicity		
Non-Minority	*Minority*	*Multi-Ethnic*
46%	*24%*	*30%*

Table 3: Gender, Course, and Free Lunch Eligibility for Free/Reduced Lunch Program of Students Comprising the Research Sample

Gender		*Course*		*Eligible for Free Reduced Lunch Program*	
Males	*Females*	*Basic*	*Honors*	*Yes*	*No*
46%	54%	73%	27%	33%	67%

Description of the Assessments

For this Biology I study, the end-of-course test included the end of the first half of the course assessment, Segment 1 test, and the end of the second half of the course assessment, Segment 2 test. The assessments focused on the skills, strategies, and knowledge necessary for effective understanding of Biology I. The Segment 1 assessment included a total of 132 banked test items and the Segment 2 assessment included a total of 114 banked items for a total of 246 test items available to generate unique tests for each student to promote academic integrity. The test items included a total of 82 groups of 3 items each. The items in each group were developed to measure a specific biology curriculum standard and were also developed to be of approximately equal difficulty. When the tests were administered, each student was randomly assigned to one test item for each of the 82 groups: 44 items from segment 1 exam and 38 items from segment 2 exam.

Standard Setting Process

The standard setting process has become a fairly common procedure for developing cut-scores for the interpretation of test information. Most national tests and a large number of states currently use the standard setting process such as the 5 points on the FCAT or an AP exam. A brief overview of standard setting is provided to help those who are unfamiliar with the process.

Determining Cut-Scores

A standard-setting study is conducted to first decide on appropriate descriptors for student performance levels. The test items are then arranged from easiest to most difficult and subject matter and curriculum specialists then review the test items to decide how many test items a student should answer correctly to achieve at each of the performance levels.

In Kindergarten to Grade 12 educational achievement testing, the concept of student performance levels is well established and serves as the design structure for course instruction and assessment. Student test performance is designed to be interpreted in terms of the content standards that the student, given his or her test score, has attained.

A cut-score cannot be arbitrarily determined, it must be empirically justified. For example, the organization cannot merely decide that the cut-score will be 70% correct. Instead, a study is conducted to determine what score best differentiates the classifications of examinees, such as competent vs. incompetent.

In the early history of educational assessments, cut-scores were based on accepted percentages correct somewhere between 60 and 75 percent. Conceptually, these percentages correct can be considered the proportion of perfection a minimally proficient examinee, for example, must achieve. Since then, however, there have been many systematic methods proposed and implemented that involve experts making decisions or judgments based on the items included on the exam and/or the examinees actually taking the exam.

Rather than relying on what experts simply think is a reasonable percentage correct to justify, for example, minimal proficiency, the standard setting method requires a standard setting panelist to judge each item individually as to whether or not he/she would expect a minimally proficient examinee to answer it correctly. It is the aggregation of these item judgments by a panel of curriculum experts that result in a cut score.

In summary, standard setting is the methodology used to define levels of achievement or proficiency and the cut-scores corresponding to those levels. A cut-score is simply the score

that serves to classify the students whose score is below the cut-score into one level and the students whose score is at or above the cut-score into the next and higher level.

Cut-Score Panelists

Panelists review the full set of test items that was arranged from the easiest items to the most difficult. They then decide how many items a student must answer correctly to be at a certain performance level. The panelists make their recommendations to FLVS administrators. These administrators review the recommendations and make the final determination of the cut scores for each performance level.

The selection of panelists in standard setting is of great importance. Panelists must be:

- Experts in the related field of examination
- Familiar with the examination methods being used
- Good problem solvers and able to work within a group
- Familiar with levels of student performance
- Have a good understanding of education, particularly for the age/grade levels of the course being studied

Panelists who are not part of the instructional staff for the assessment being studied should be included on the panel that recommends standards. External panelists offer insights, experience, and ideas that may not have arisen otherwise and which usually improve the quality of the standard-setting procedures and the defensibility of the results. The external panelists provide essential validity to the process of setting standards.

Performance Level Descriptors

Performance level descriptors are set by the administration of the group seeking to determine cut-scores for the PLDs. Most state departments of education when assessing educational progress, such as those required under the *No Child Left Behind Act* (NCLB), typically classify student test scores into four achievement levels, such as:

- below basic
- basic
- proficient
- advanced

The PLDs are always written with a several-sentence description as to what is meant by each of the levels. For example, the *Florida Department of Education* established a five level set of PLDs for state end-of-course assessments (including Biology I) as follows:

- Level 5: Students at this level demonstrate mastery of the most challenging content of the *Next Generation Sunshine State Standards.*
- Level 4: Students at this level demonstrate an above satisfactory level of success with the challenging content of the *Next Generation Sunshine State Standards*.
- Level 3: Students at this level demonstrate a satisfactory level of success with the challenging content of the *Next Generation Sunshine State Standards*.
- Level 2: Students at this level demonstrate a below satisfactory level of success with the challenging content of the *Next Generation Sunshine State Standards*.
- Level 1: Students at this level demonstrate an inadequate level of success with the challenging content of the *Next Generation of Sunshine State Standards*.

FLVS could have merely adopted the *Florida Department of Education* PLDs for their end-of-course assessments. However, the length of the tests, the specific content of the assessments, the difficulty of the test items all differ so the adoption of those PLDs would not have been valid. More importantly, the *Florida Department of Education* PLDs have a focus on graduation requirement while FLVS has a focus on curriculum and course improvement.

Evaluating Cut-Scores

How is it determined if the cut-scores for a given assessment are set appropriately? The "right" cut-scores should be both consistent with the intended educational policy *and* psychometrically sound. In short, standard setting matters; it is not simply a methodological procedure but rather an opportunity to incorporate educational policy into an assessment system. Ideally, the standard-setting process elicits educational policy and incorporates it into the test development process to ensure that the cut-scores that a test eventually produces not only reflect a state's policy but also are well-supported psychometrically.

Far from being a purely methodological process, standard setting ideally involves policy makers, test developers and measurement specialists to ensure that the test results will be useful and defensible.

Standard Setting for the Biology I Assessment

The Florida Virtual School (FLVS) Biology I end-of-course assessment is designed to gauge a student's achievement of the content standards that form the structure of the FLVS Biology I course. Again, for this study, the end-of-course test included the end of the first half of the course assessment, Segment 1 test, and the end of the second half of the course assessment, Segment 2 test. The assessments are administered at or near the completion of a student's Biology I segment in the course. A student's score on each test, when combined with other assessment results and other direct and indirect measures of accomplishment of course goals, is used to determine the student's grade in the course.

FLVS determined that it would be helpful to students, their parents, and course instructors to develop cut-scores to provide a criterion-referenced indication of student performance on the FLVS Biology I assessments. In order to develop the cut-scores, it was first necessary to establish PLDs for the assessments. These PLDs were intended to assist in the interpretation of the end-of-course assessment results.

Cut-score development, or standard setting as the process is commonly labeled, was carried out by ERIA and BETA in conjunction with staff from FLVS. The researchers used a modification of the widely used and extensively researched item-mapping ("Bookmark") method of determining standards. This method was selected for two reasons. First, it is the most commonly applied methodology used in determining student performance standards for educational assessment in the United States. Second, it is a procedure that appeared to lend itself easily to carrying out the activities virtually – that is, by conducting the necessary panel sessions via telephone and the Internet.

Summary of Major Activities of the Standard Setting Process

The standard-setting activities began with a plan developed by ERIA and BETA to carry out the various stages of the project. This plan included a timeframe for the project, a general outline of activities to be carried out, a proposed number of panelists to carry out the process of setting performance standards and the desired credentials of the panelists, a description of the anticipated relevant data resulting from the project, and an outline of how the developed standards would be used in interpreting student performance on the assessments.

FLVS staff had the responsibility of deciding upon the number of categories for segmenting student performance. That is, should scores simply be categorized as "pass/fail" or should finer discriminations of performance be used? FLVS staff considered several alternatives, guided by the desire that the same number of categories, and the same labels for these levels, be used

across several FLVS standard setting processes. This was considered important to increase the eventual usefulness of the descriptors by FLVS clients, both internal and external.

Following extensive discussions both internally and with their ERIA and BETA consultants, FLVS decided to use four categories of performance on the assessments. The labels applied to these four categories and the general performance level descriptors for each of these categories were:

- *Level 1 – Needs Improvement*: The student did not meet most standards, and significant instructional intervention is needed.
- *Level 2 – Novice*: The student meets some standards but needs instructional intervention to achieve a level of competence.
- *Level 3 – Capable*: The student meets most standards and demonstrates competency.
- *Level 4 – Advanced*: The student has mastered the standards and demonstrates exceptional ability.

While the standard-setting procedures and descriptors were being finalized, panelists who would be tasked to recommend the standards were solicited. ERIA conducted this process using recommendations from FLVS as a starting point. A total of 10 panelists agreed to take part in the activities, although one panelist was unable to attend due to a professional commitment. The panelists included four FLVS teachers or curriculum specialists, four Florida public school biology teachers, and one judge from private industry with a professional specialty in science. All judges were able to participate in the complete standard-setting activities –the training/orientation, the discussion of the performance-level descriptors, and three rounds of recommendations.

The actual standard setting was conducted virtually on October 29 and November 1, 2012; sessions were held for a total of approximately 10 ½ hours on these two days, not including time the judges spent independently making their recommendations. Sessions were planned and facilitated by Michael Beck of BETA, a researcher who has planned and conducted such activities for more than 25 state-level assessments. Sessions were conducted using a FLVS-approved PowerPoint presentation through which panelists, who were trained in the general and specific process of determining student performance standards, had an opportunity to review actual FLVS Biology I exam items, and make their recommendations. Extensive time during the sessions was provided for panelists to "translate" the above generic performance-level descriptors into language that was specifically tailored to the Biology I course on which the standards were to be based.

The general methodology for the standard setting was the item-mapping procedure, a modification of the "Bookmark method" of setting standards. This procedure was chosen primarily due to its overwhelming popularity for determining performance standards for educational tests of this type and its ease of use by panelists. Consistent with typical applications of the procedures, panelists had three opportunities to recommend standards. Following each round of judgments, panelists had an opportunity to share their perspectives and – as desired – individual recommendations with their peer judges. Panelists were shown the (anonymous) recommendations of their peers following each round. Extensive discussions of individual test items, especially those around the interim cut-score recommendations, took place. As is typically the case with item-mapping applications, panelists made their first recommendations without the benefit of seeing student performance data; these results, presented in terms of item-difficulty (*p*-value) data, were shared between the first and second rounds of judgments and were available for panelists to consider when they made Rounds 2 and 3 recommendations.

For making their judgments, panelists used an ordered-item booklet prepared by ERIA. This booklet was composed of 82 items arranged in increasingly difficult sequence. Only multiple choice items were used for this analysis. The FLVS Biology I tests that are actually administered to each student at the completion of each segment are assembled automatically and randomly from an extensive item pool, composed of 3 items assessing each of 82 slots on the test blueprint (44 from segment 1 and 38 from segment 2). Thus, each student is presented with a somewhat-unique set of 82 items which, as a set, assess the identical test blueprint as the test administered to any other student. The items used for the ordered-item booklet were the median-difficulty items for each of these 82 test-blueprint slots. That is, assume that the 3 test-pool items assessing a given test blueprint objective had item difficulty values of .61, .68, and .79. In this case, the item presented to the panelists to represent this blueprint position would have been the item with a difficulty value of .68 – the median-difficulty item. This decision was made jointly by ERIA, BETA, and FLVS as being the best representation of each objective in the Biology I item pool.

Data Analyses and Results

The results are presented in two sections. The first describes the results of the standard setting process and the cut-scores determined by the panelists and endorsed by the administration of the FLVS.

Performance Level Descriptors and Cut-scores

The table below presents the results of the session, by judge and by round of recommendations. Summary data by round are also shown in the table. The tabled numbers show the *highest* raw score for each indicated performance category. For example, in Round 3, Judge 1 recommended that scores between 0 and 15 be categorized as Needs Improvement (see the 15 under this label for Round 3, Judge 1). Similarly, this judge recommended that scores between 16 and 28 be categorized as Novice, that scores between 29 and 68 be labeled Capable, and 69 through 82 be labeled Advanced.

Inspection of the tabled data indicates that – as is typical of standard-setting activities – judges varied somewhat significantly among themselves across all three rounds of the process, although they agreed somewhat more in Round 3 than in the first round. Similarly, while individual judges changed their recommendations between rounds, often fairly significantly, the overall central tendency of the recommendations (especially for the Novice-Capable and Capable-Advanced cut points) remained fairly constant. The table also provides standard errors (SE) of the mean and median panel recommendations; these data indicate that the degree of statistical error in the mean/median panel recommendations is on the order of only one raw-score point.

The median judgments highlighted in the table are considered to be the best representation of the judges' recommendations. Medians are typically preferred over means for such work as medians are less affected by extreme or "outlier" recommendations. In the present case, of course, medians do not differ significantly from means as the tabled data demonstrate.

The final, Round 3 recommendations of the judges were that FLVS establish the following standards for the Biology I compiled end-of-course exam:

Performance Category	**Exam Raw Scores**
Needs Improvement	0 through 17
Novice	18 through 30
Capable	31 through 68
Advanced	69 through 82

An empirical, research-based activity to establish student performance standards for the FLVS Biology I end-of-course exam was planned and carried out. The activity was conducted via established, extensively validated procedures involving a panel of nine professional experts in the area, both teachers and other professionally credentialed personnel. After training in the standard-setting methodology, discussion of the performance-level descriptors, and extensive interaction among panelists of two stages of interim recommendations, the panel recommended cut-scores for each of the PLDs as described above.

Table 4: Summary of Results of Panel Recommendations for Standards for the FLVS Biology I Exam – by Judge, by Round of Judgments, and by Performance Category

	Round 1			**Round 2**			Round 3		
Judge	Needs Improvement	Novice	Capable	Needs Improvement	Novice	Capable	Needs Improvement	Novice	Capable
1	7	26	70	11	27	62	15	28	68
2	9	22	47	9	21	62	11	21	60
3	9	26	68	9	30	66	9	30	61
4	11	28	63	12	27	63	17	34	67
5	13	51	62	13	34	62	17	30	70
6	15	34	61	20	30	62	20	34	64
7	14	29	63	18	28	64	19	28	64
8	16	33	64	11	28	64	13	30	63
9	9	22	63	11	27	62	17	34	69
Median	**11**	**28**	**63**	**11**	**28**	**62**	**17**	**30**	**64**
Mean	11.4	30.1	62.3	12.7	28.0	63.0	15.3	29.9	65.1
S.D.*	3.2	8.9	6.4	3.8	3.5	1.4	3.7	4.1	3.6
SE Mean**	1.1	3.0	2.1	1.3	1.2	0.5	1.2	1.4	1.2
SE Median**	1.3	3.7	2.7	1.6	1.4	0.6	1.5	1.7	1.5

**S.D. is Standard Deviation*
***SE Mean is the Standard Error of the Mean*
****SED Median is the Standard Error of the Median*

Student Performance Results Using the Standard Setting Criteria

The standard setting criteria was applied to the total group of students and to the groups comprising those students enrolled in honors and basic levels of the course. That comparison is provided in Table 5.

The average percent correct for each of the various sub-groups was then analyzed to determine if there were any statistically significant differences between the groups. The only difference

that achieved statistical significance was the comparison of the honors students to the basic students. For the gender, socio-economic status, and ethnic groups the differences were statistically non-significant.

Tables 5 to 9 provide the mean scores for each group and the percent of students scoring in each group. While there were some differences in the distribution of the students across a group, there were no statistically significant differences in the performance of one group compared to another with the exception of the honors/basic comparison reported in Table 6.

All Students

Table 5 provides the percent of students scoring at each of the four levels on the end-of-course assessment in Biology I.

Table 5: Percent of Students Scoring at Each Performance Level

Performance Levels	*Needs Improvement*	*Novice*	*Capable*	*Advanced*
All Students	0%	2%	86%	12%

Basic and Honors Students

An independent sample *t*-test was used to determine if there was a significant difference between the total end-of-course test for the students enrolled in the basic course and the students enrolled in the honors course. The results indicated there was a statistically significant difference ($t = 5.949$, $p = \leq .0001$). The comparison of the two groups is shown in Table 6.

Table 6: Percent of Students Scoring at Each Performance Level

Performance Levels	*Mean Score*	*Needs Improvement*	*Novice*	*Capable*	*Advanced*
Basic Students	54.7	0%	4%	92%	4%
Honors Students	61.2	0%	1%	78%	21%

Male and Female Students

An independent sample *t*-test was used to determine if there was a significant difference between the total end-of-course test for male and female students. The results indicated there was no statistically significant difference in the performance of the two groups. The comparison of the two groups is shown in Table 7.

Table 7: Percent of Students Scoring at Each Performance Level

Performance Levels	*Mean Score*	*Needs Improvement*	*Novice*	*Capable*	*Advanced*
Male Students	57.5	0%	2%	89%	9%
Female Students	57.8	0%	3%	83%	14%

Lower Socio-economic and Higher Socio-economic Students

An independent sample *t*-test was used to determine if there was a significant difference between the total end-of-course test for lower socio-economic students and higher socio-economic students. The results indicated there was no statistically significant difference. The comparison of the two groups is shown in Table 8.

Table 8: Percent of Students Scoring at Each Performance Level

Performance Levels	*Mean Score*	*Needs Improvement*	*Novice*	*Capable*	*Advanced*
Lower SES Students	57.0	0%	3%	86%	10%
Higher SES Students	57.9	0%	2%	86%	12%

Ethnic Group Students

A one-way Analysis of Variance was used to determine if there was a significant difference between the total end-of-course scores for the various ethnic groups. The results indicated there was no statistically significant difference. The comparison of the three groups is shown in Table 9.

Table 9: Percent of Students Scoring at Each Performance Level

Performance Levels	*Mean Score*	*Needs Improvement*	*Novice*	*Capable*	*Advanced*
Non-Minority Students	56.8	0%	4%	87%	9%
Minority Students	60.6	0%	0%	77%	23%
Multi-Ethnic Students	57.7	0%	1	88%	11%

Conclusions

There are two major conclusions to the study. First, the process of developing Performance Level Descriptions and curriculum valid cut-scores to determine levels of performance was very successful. The panel of biology experts was able to come to a very strong consensus on the levels of performance needed to achieve each of the performance levels. The expertise of the panel in terms of their understanding of biology and their understanding of student performance guided the panel to make specific recommendations regarding the cut-scores which were then endorsed by the FLVS administration.

The second conclusion is that the cut-scores could then be used to determine student success on the end-of-course assessments. The results of that analysis are presented in response to the three questions that guided the study.

1. Do students enrolled in the **Florida Virtual School Biology I** course achieve success as assessed by their end-of-course test scores?

The results show that no students scored at the *Needs Improvement* level and 86% of the students scored at the second highest level, *Capable*. In addition, 12% of the students scored at the *Advanced* level

2. Do students enrolled in basic or honors courses achieve similar gains in the **Florida Virtual School Biology I** program?

The honors students scored at a statistically significantly higher level than the basic students. While the difference overall was statistically significant, the major difference was at the Advanced level with 21% of the honors students scoring at that level while only 4% of the Basic students did so.

3. Do students with differing demographic characteristics (gender, ethnic background, and socio-economic status) achieve similar performance levels when enrolled in the **Florida Virtual School Biology I** program?

There were no statistically significant differences based on gender, socio-economic status, or ethnic background. There were some differences between the percentages of students in each of the sub-groups, none of those differences achieved significance.

In summary, the study demonstrates the effective use of a standard setting process with a FLVS end-of-course assessment and the effectiveness of the use of the Performance Level Descriptors to assess success on the end-of-course assessments.