

Foreign Language Assessment Report

Fall 2019

Author: Joseph F. van Gaalen, Ph.D., Asst. VP, IR, Assessment & Effectiveness

1 INTRODUCTION

Florida SouthWestern's Foreign Language Department employs a common course assessment to measure student progress in course level objectives, a practice shown to be effective in establishing data driven instruction (Hall, 2010). Courses included in assessment are: FRE 1120 *Elementary French I*, FRE 1121 *Elementary French II*, SPN 1120 *Beginning Spanish I*, and SPN 1121 *Beginning Spanish II*. Through achievement of these courses students will acquire and demonstrate competency in speaking, writing, reading comprehension and listening comprehension in standard Spanish or French at the beginner's level. The assessment outcomes outlined below define the method of assessment for each course assessment as well as measure current Student Learning Outcomes (SLOs) and identify areas for future SLOs to be assessed. Additionally, the plan provides information on achievement levels of dual enrollment (concurrent) artifacts compared with traditional, as well as online artifacts compared with traditional artifacts as highlighted in the course level assessment plan. This report provides achievement analysis for both fall 2019 as well as longitudinal studies, where applicable.

For additional detail or further analysis not provided in this report, please contact Dr. Joseph F. van Gaalen, Asst. VP, IR, Assessment & Effectiveness (jfvangaalen@fsw.edu; x16965).

2 FRENCH

2.1 FRE 1120

During the fall 2019 term, five sections of FRE 1120 were offered. An assessment first piloted in AY 2017-18 was recorded in four section of FRE 1120 accounting for 55 artifacts, or 73% of the population. The assessment utilizes a combination quiz and disposition survey. In the assessment, the students are asked for their relative levels of confidence in translating a given topic in conjunction with performing brief translations of the same topic. Each of five disposition survey questions are tethered to two questions in which students are asked to (1) translate French by answering a question written in French using an English response and (2) translate French by answering a question written in French using a French response. The resulting assessment consists of 15 questions comprised of five survey questions and 10 achievement related questions. In form, the assessment measures achievement as well as alignment with student confidence of a given topic. Note that several instances, the scoring was altered. It is important that scoring is consistent in an assessment such as this, so it may be important to review this area.

2.1.1 Descriptive Statistics, Learning Outcomes, and Objectives

The lead French faculty has identified five areas of interest for evaluation. No achievement metric goals have yet been set. The Learning Outcomes (LOs) are:

- LO 1: Reading/Writing Introductions
- LO 2: Reading/Writing Self, Classes & Campus
- LO 3: Reading/Writing Activity Likes & Dislikes
- LO 4: Reading/Writing Home & Rooms
- LO 5: Reading/Writing Descriptions & Family

Each LO consists of two questions. The first, worth one point, requires a translation from French using English responses. The second, worth two points, requires comprehension of French using French responses. Each question can be scored at half-point intervals. The result is a combined maximum score of 15 points. Results exhibit mean scores across the five LOs ranging from 2.2 to 2.6, about the same as that seen in fall 2018, where the range was 2.3 to 2.7. It is important to note that each LOs in this study are arranged in order of advancement for the course. LO 1 is encountered early in the term with each LO encountered sequentially to LO 5, which is studied near the term's end.

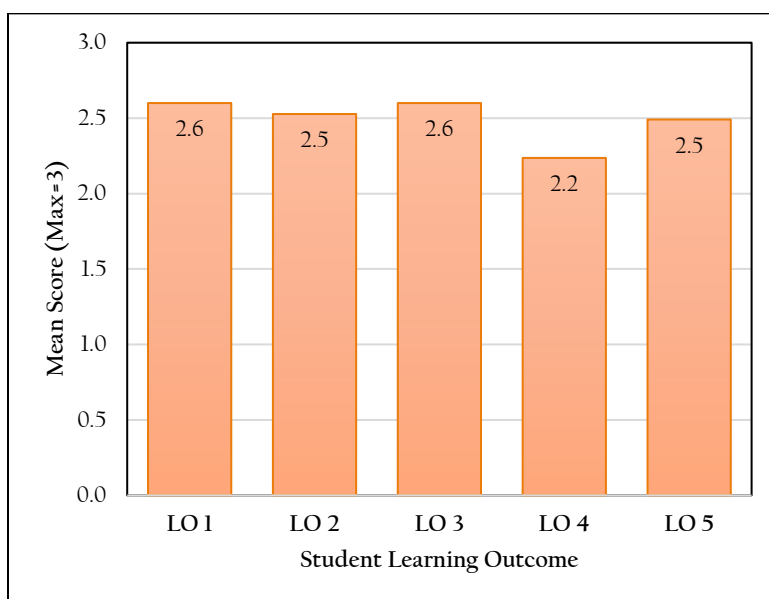


Figure 1. Mean scores by SLOs for FRE 1120.

2.1.2 Exploratory Analysis & Significance Testing

2.1.2.1 Dual Enrollment (Concurrent) to Traditional Comparison

While there was one dual enrollment (concurrent) section run during fall 2019, data was not reported so no comparison study between dual enrollment and traditional courses could be completed.

2.1.2.2 Online to Traditional Comparison

During the fall 2019 semester, one online section reported data able to be used in assessment and 3 traditional sections reported data. From those course sections, 18 total online artifacts were collected from FRE 1120 and 37 traditional artifacts were collected from FRE 1120. Both online and traditional course sections exhibit similar artifact score distributions in the assessment at 13/15 or higher (Figure 2). However, traditional artifacts to occupy a substantial number of lower scores that the online section does not. For example, approximately 24% of traditional artifacts score between 6/15 and 9/15. By comparison, none from online exhibit scores in that range.

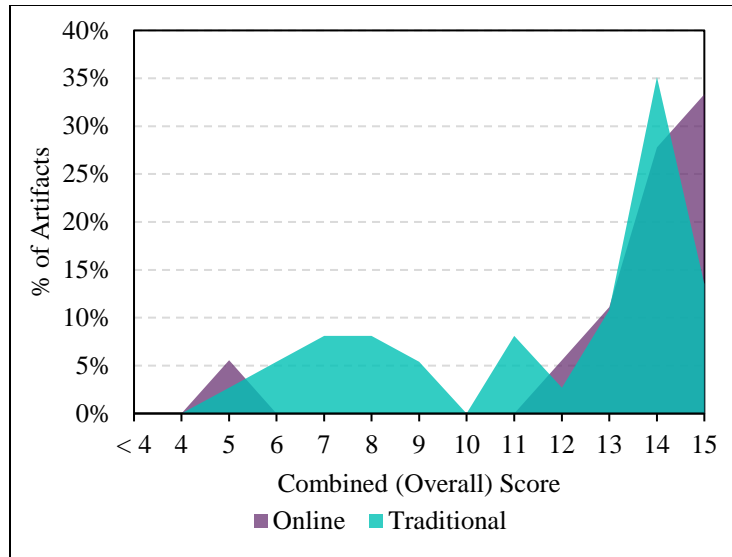


Figure 2. Score distribution for online (purple) and traditional (aqua) artifacts of FRE 1120.

2.1.2.3 Comparison by Campus/Site

While courses were offered at three locations, the third, the Collier Campus, is comprised by only eight artifacts. As a result, any meaningful comparison is best left with the online-to-traditional comparison in 2.1.2.2 above.

2.1.3 Data Distribution & Longitudinal Studies

2.1.3.1 Data Distribution

A distribution of combined (total) scores from the 10 achievement questions is shown below in Figure 3. Scores are centered on 14/15 with a strong negative skew meaning scores are tending strongly towards higher values (Starkweather, 2010). In total, 64% of artifacts score 13/15 or higher.

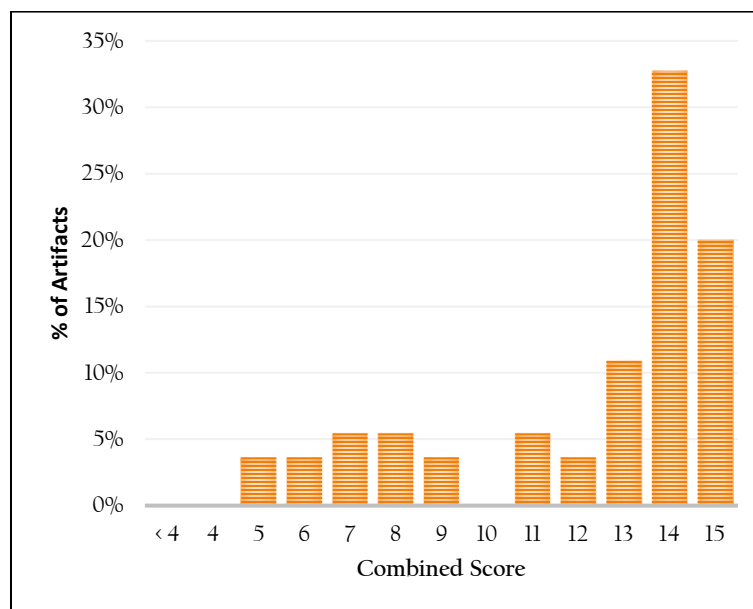


Figure 3. Distribution of combined (total) scores for assessment.

One of the strengths of a combined disposition survey and achievement assessment is the way in which the data can be analyzed. Survey questions that gauge confidence in a topic can now be compared with the results of achievement in that topic to better describe how student confidence varies with student strengths and weaknesses. This gauge is reflected in the scatterplot shown in Figure 4 below. These results yield a clearer sense of just how many students self-report a strong understanding of the topic and yet do not appear to be capable of demonstrating it. A clear trend in confidence correlating with performance is visible with increases in “Agree” and “Strongly Agree” coinciding with higher combined scores.

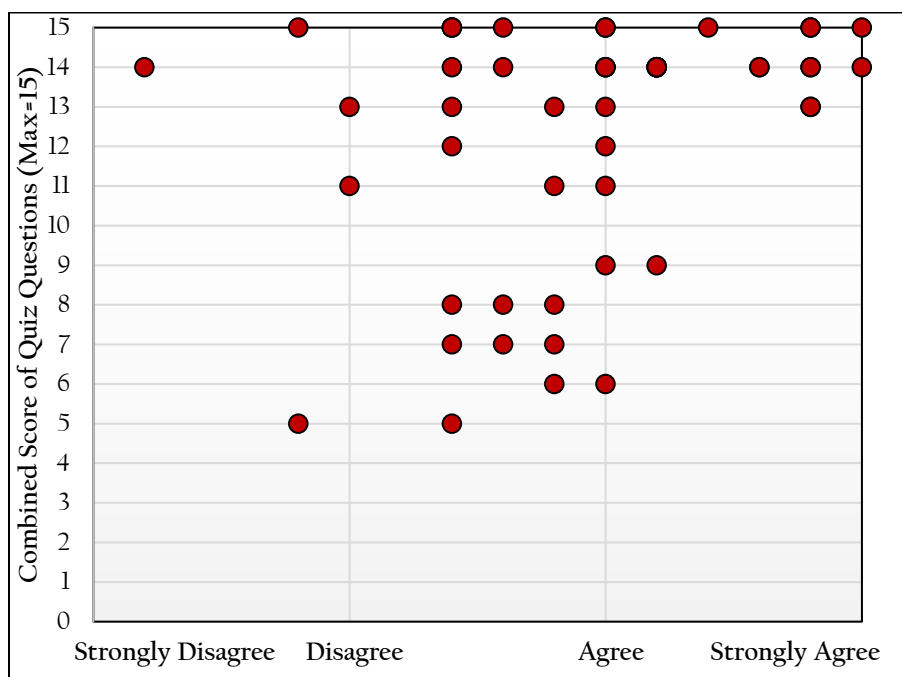


Figure 4. Scatter plot of combined (total) scores based on general response to survey questions.

2.1.3.2 Longitudinal Study

Further description of achievement over time in FRE 1120 is provided in Table 1. Both demographics of students and student count vary by semester it may be more reasonable to compare like semesters (Fall vs. Fall, Spring vs. Spring) (see <http://www.fsw.edu/facultystaff/assessment/history> for further details). In three terms, LOs 1 through 5 for fall 2019 rank 2nd, 2ⁿ, 2nd, 3rd, and 3rd, respectively.

	LO 1	LO 2	LO 3	LO 4	LO 5	Overall
Fall 2018 (n=26)	2.6	2.4	2.7	2.3	2.7	12.7
Spring 2019 (n=19)	2.9	3.0	3.0	2.9	2.9	14.7
Fall 2019 (n=55)	2.6	2.5	2.6	2.2	2.5	12.4

Table 1. Comparison of mean scores for FRE 1120 for fall 2018 through the present.

2.2 FRE 1121

During the fall 2019 term, two sections of FRE 1121 were offered. The assessment was recorded in both, however, in one section results were shuffled, so no data could be utilized. The data collection accounts for 14 artifacts, or 67% of the population. The assessment utilizes a combination quiz and disposition survey. In the assessment, the students are asked for their relative levels of confidence in translating a

given topic in conjunction with performing brief translations of the same topic. Each of four disposition survey questions are tethered to two questions in which students are asked to (1) translate French by answering a question written in French using an English response and (2) translate French by answering a question written in French using a French response. The resulting assessment consists of 12 questions comprised of four survey questions and 8 achievement related questions. In form, the assessment measures achievement as well as alignment with student confidence of a given topic. Note that several instances, the scoring was altered. It is important that scoring is consistent in an assessment such as this, so it may be important to review this area.

2.2.1 Descriptive Statistics, Learning Outcomes, and Objectives

In FRE 1121, a similar assessment seen in FRE 1120 is utilized. The lead French faculty has identified five areas of interest for evaluation. No achievement metric goals have yet been set. The Learning Outcomes (LOs) are:

- LO 1: Reading/Writing Introductions
- LO 2: Reading/Writing Completed Actions in the Past
- LO 3: Reading/Writing Repeated and Habitual Actions in the Past
- LO 4: Reading/Writing Hypothetical Situations

Each LO consists of two questions. The first, worth one point, requires a translation from French using English responses. The second, worth two points, requires comprehension of French using French responses. Each question can be scored at half-point intervals. The result is a combined maximum score of 12 points. Results exhibit mean scores across the four LOs ranging from 2.7 to 2.9. It is important to note that each LOs in this study are arranged in order of advancement for the course. LO 1 is encountered early in the term with each LO encountered sequentially to LO 4, which is studied near the term's end.

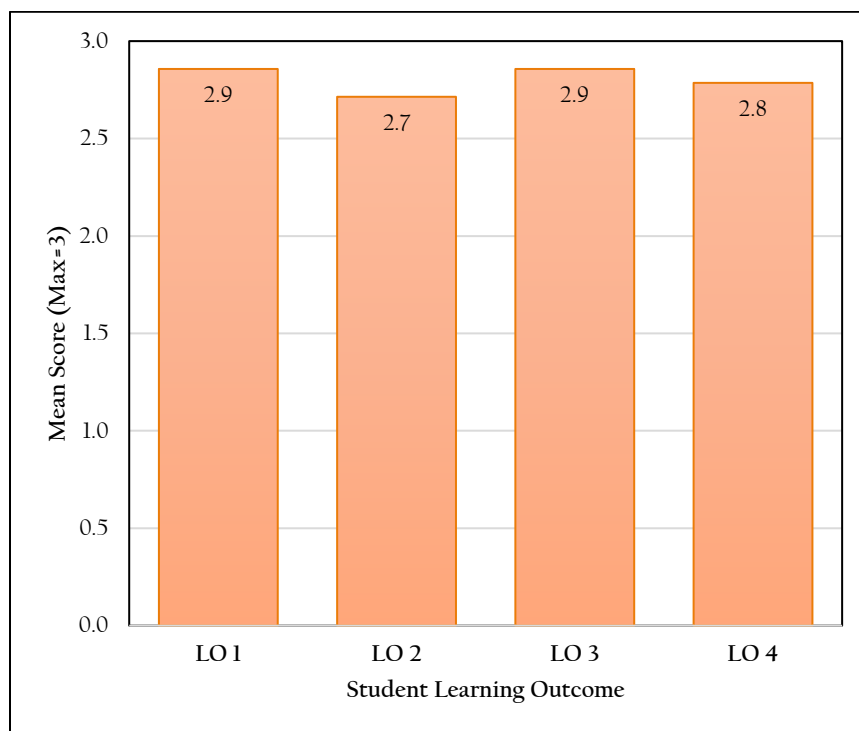


Figure 5. Mean scores by SLOs for FRE 1121.

2.2.2 Exploratory Analysis & Significance Testing

2.2.2.1 Dual Enrollment (Concurrent) to Traditional Comparison

There was no dual enrollment (concurrent) section run during fall 2019, so no comparison study between dual enrollment and traditional courses could be completed.

2.2.2.2 Online to Traditional Comparison

While both an online and a traditional section were offered, data could not be utilized from the traditional section due to data recording issues. As a result, no comparison analysis between the two could be completed.

2.2.2.3 Comparison by Campus/Site

While courses were offered at three locations, the third, the Collier Campus, is comprised by only eight artifacts. As a result, any meaningful comparison is best left with the online-to-traditional comparison in 2.1.2.2 above.

2.2.3 Data Distribution & Longitudinal Studies

2.2.3.1 Data Distribution

A distribution of combined (total) scores from the 8 achievement questions is shown below in Figure 7. Scores are centered on 12/12 with a strong negative skew meaning scores are tending strongly towards higher values (Starkweather, 2010). In total, 71% of artifacts score a perfect 12/12.

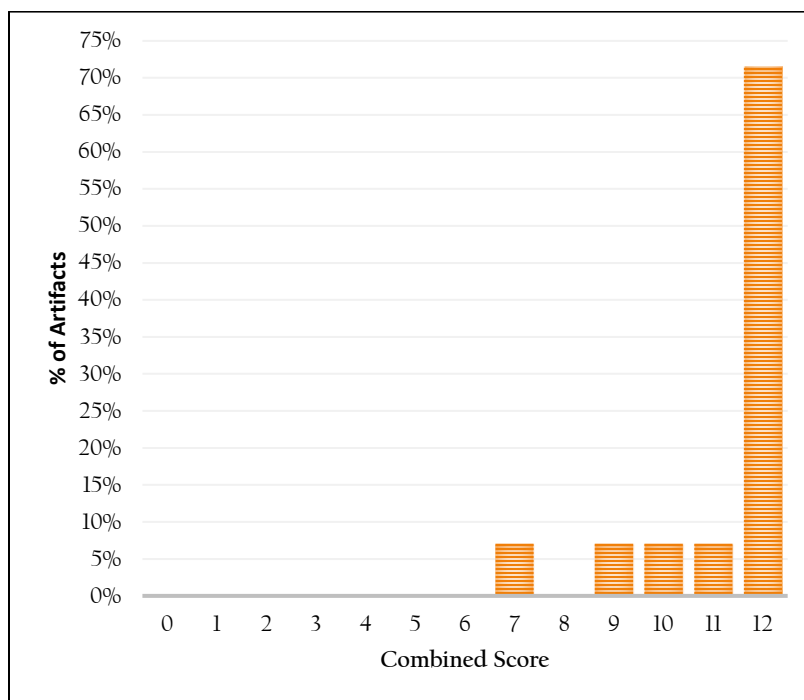


Figure 6. Distribution of combined (total) scores for assessment.

One of the strengths of a combined disposition survey and achievement assessment is the way in which the data can be analyzed. Survey questions that gauge confidence in a topic can now be compared with the results of achievement in that topic to better describe how student confidence varies with student

strengths and weaknesses. This gauge is reflected in the scatterplot shown in Figure 8 below. These results yield a clearer sense of just how many students self-report a strong understanding of the topic and yet do not appear to be capable of demonstrating it. Because most students scored a perfect 12/12, it is unclear if any correlation exists.

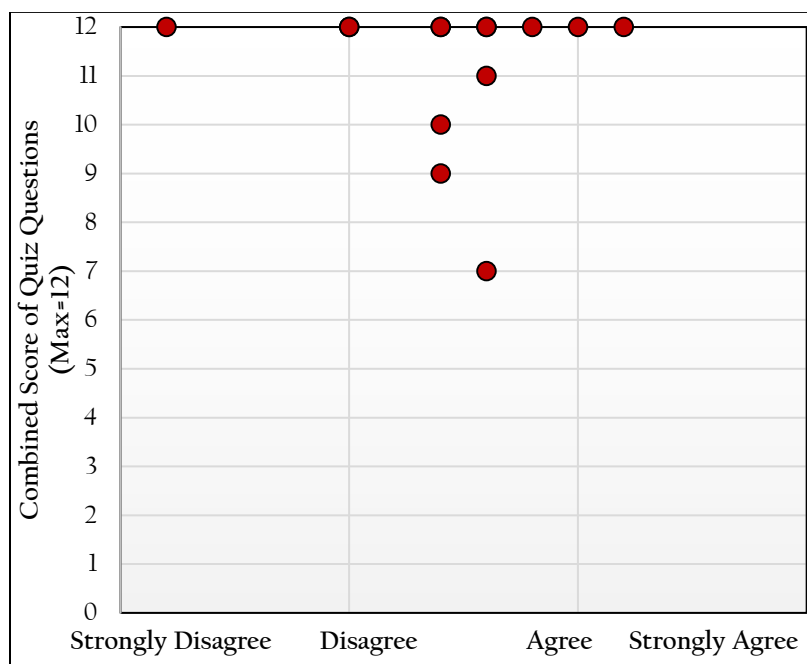


Figure 7. Scatter plot of combined (total) scores based on general response to survey questions.

2.2.3.2 Longitudinal Study

Fall 2019 is the first data collection of this assessment. A longitudinal study will begin following data collection of fall 2020 data.

3 SPANISH

3.1 SPN 1120

3.1.1 Descriptive Statistics, Learning Outcomes, and Learning Objectives

During the fall 2019 semester 14 sections of SPN 1120 Beginning Spanish I were offered. Of those, artifacts from a common final were collected from 13 sections. One section did not report data. A total of 314 students were enrolled in SPN 1120. Of those, 240 artifacts were collected representing a sample size of 76% of the population.

Using a common course assessment, the FSW Spanish faculty defined three areas of interest for evaluation that apply to SPN 1120. The Student Learning Outcomes (SLOs) and their objectives or measures of success are:

- SLO 1: Students will be able to understand spoken Spanish. The faculty established measure of success for this SLO is for 80% of students to demonstrate competency with a score of 70% or better in the oral comprehension exam sections (Section I).
- SLO 2: Students will be able to understand written Spanish. The faculty established measure of success for this SLO is for 80% of students to demonstrate competency with a score of 70% or better in the reading comprehension exam sections (Section II and III).
- SLO 3: Students will be able to write effectively in the Spanish language. The faculty established measure of success for this SLO is for 80% of students to demonstrate competency with a score of 70% or better in the writing competency exam sections (Section IV and V).

The faculty established measure of success for SLO 1, 80% of students scoring 70% or higher in Section I, was nearly met as results exhibit 76% of artifacts score 70% or higher in the oral competency exam section (Section I) (Table 2). The faculty established measure of success for SLO 2, 80% of students scoring 70% or higher in Sections II and III, was partially met. Results exhibit 40% of artifacts scored 70% or higher in Section II and 90% of artifacts scored 70% or higher in Section III. The faculty established measure of success for SLO 3, 80% of students scoring 70% or higher in Sections IV and V, was partially met. Results exhibit 63% of artifacts scored 70% or higher in Section IV and 91% of artifacts scored 70% or higher in Section V. For a graphical representation of SLO achievement, see Figure 8.

n = 240	Section I (Oral)	Section II (Reading)	Section III (Reading)	Section IV (Written)	Section V (Written)	Combined (Overall)
Goal	80% of artifacts scored $\geq 70\%$ for all sections					
% above 70%	76%	40%	90%	63%	91%	
Mean (as %)	80%	63%	88%	75%	83%	
Median (as %)	83%	61%	93%	80%	85%	
Section Score Max	15	60	15	15	20	125
Section Mean	12.0	37.7	13.2	11.3	16.7	90.8
Section Median	13	36.9	14	12	17	90.5

Table 2. Percentage of student achievement level as per SLOs (SPN 1120).

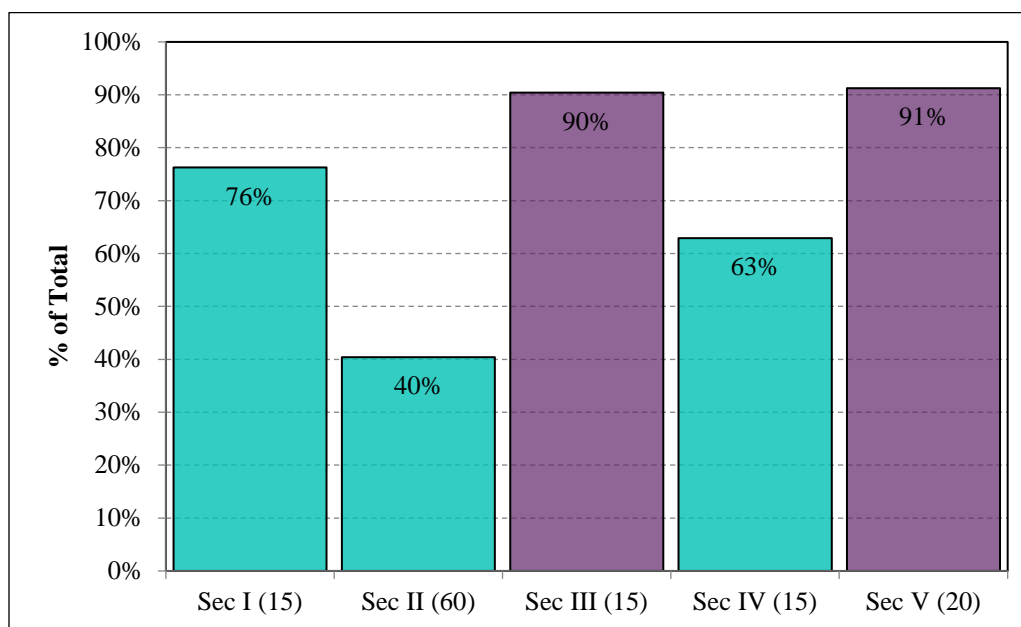


Figure 8. SLO achievement for SPN 1120 by exam section (Percentage of student achievement level as per SLOs). Purple denotes having met objective.

3.1.2 Exploratory Analysis & Significance Testing

Multiple comparisons of artifact scores across varying formats, campuses, and student types were made to more thoroughly detail the assessment data. Each course was divided into appropriate subgroups such as by campus or enrollment status to perform the analysis. Where possible, additional methods of analysis were conducted to provide a broader picture of these comparisons.

3.1.2.1 *Dual Enrollment (Concurrent) to Non-Dual Enrollment Comparison*

No dual enrollment (concurrent) sections of the course were run during fall 2019 so no comparison study between dual enrollment and traditional courses could be completed.

3.1.2.2 *Online to Traditional Comparison*

During the fall 2019 semester, two online section reported data able to be used in assessment and 11 traditional sections reported data. From those course sections, 31 total online artifacts were collected from SPN 1120 and 209 traditional artifacts were collected from SPN 1120. A comparison of basic statistics is provided in Table 3. Online artifacts mean scores are 0.5 lower than traditional artifacts. Differences in the means were tested for significance using a Welch's t-test according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999) and were found to not be statistically significantly different. Therefore, we cannot reject the null hypothesis that the differences in the means of the online and traditional artifacts are equal to 0, and we cannot conclude this with a 95% confidence that the differences in scores are not solely due to chance.

Effect size was calculated using a method devised by Rosenthal and Rosnow (1991) for meta-analytical purposes in potential comparisons with other institutions (Lipsey and Wilson, 1993). The results exhibit what Cohen (1988) would consider a small effect size. In other words, non-overlap score distribution from online artifacts to traditional artifacts is approximately 1%. For a graphical representation of this see Figure 9.

df = 238	
Online mean	72.2
Online standard deviation	19.76
Traditional mean	72.7
Traditional standard deviation	14.47
Effect size	0.021
p-value	0.875

Table 3. Comparison of mean scores (as %) for online and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

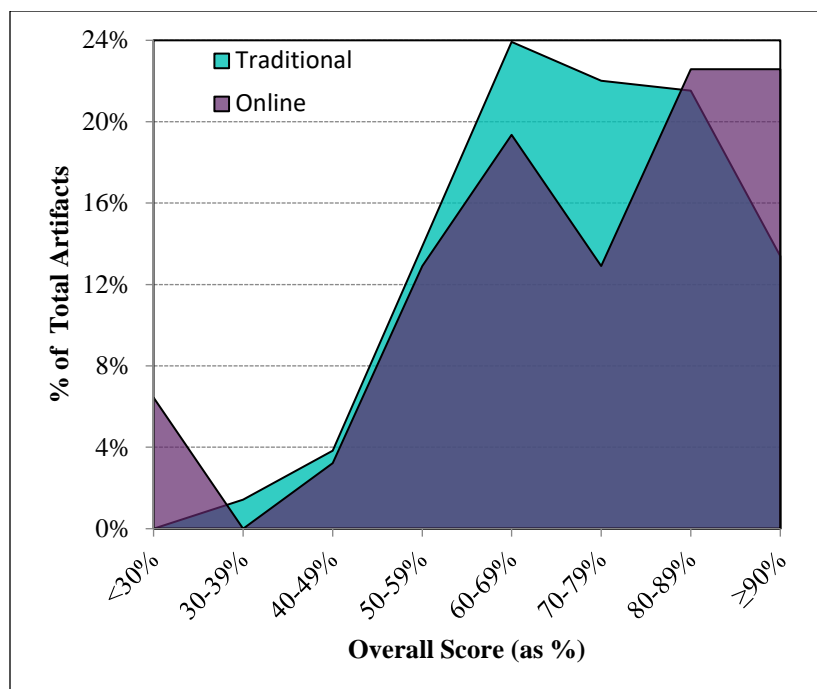


Figure 9. Score distribution for online (purple) and traditional (aqua) artifacts of SPN 1120.

3.1.2.3 Comparison by Campus/Site

Of the 240 artifacts collected from SPN 1120, 44 originated from the Charlotte campus, 37 from the Collier campus, 31 from FSW Online, and 128 from the Thomas Edison (Lee) campus. Scores by rubric dimension varied greatly across campuses. The Collier campus exhibits the highest mean scores in 3 of 5 sections and the overall score. The Charlotte campus exhibits the highest mean score in 2 of 5 sections. A comparison of mean scores by rubric dimension is provided in Table 4 and Figure 10.

	Section I	Section II	Section III	Section IV	Section V	Combined Score
<i>Rubric Max</i>	15	60	15	15	20	125
Charlotte	11.2	30.7	14.3	13.2	16.9	86.4
Collier	12.8	44.3	14.0	12.3	17.1	100.5
FSW Online	12.4	38.2	13.9	8.8	16.9	90.2
Thomas Edison (Lee)	12.2	39.1	12.3	11.1	16.6	91.3

Table 4. Comparison of mean scores by site for SPN 1120. Bold denotes highest mean score in that dimension among all sites.

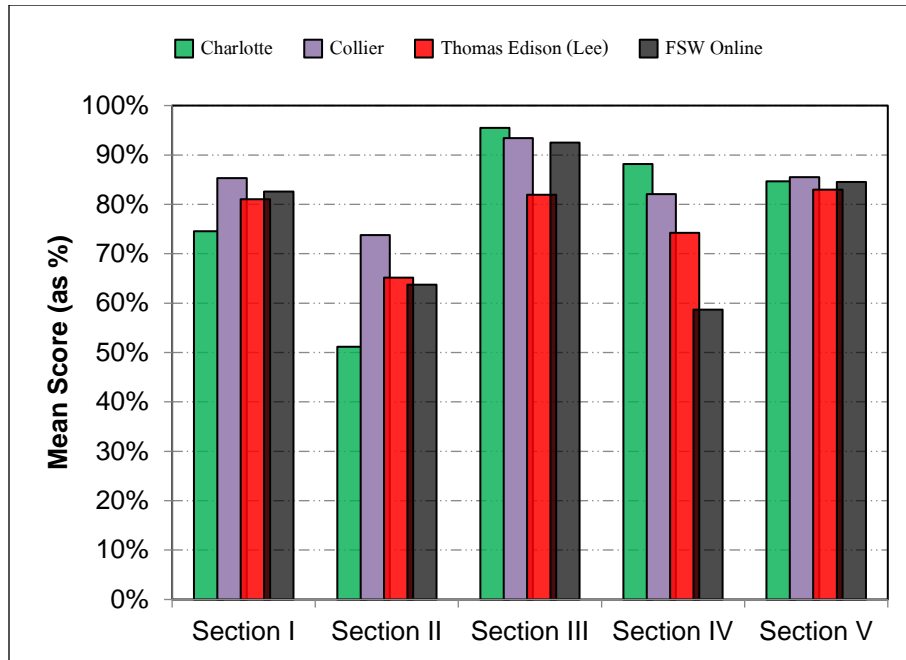


Figure 10. Comparison of mean score of exam sections by site.

3.1.3 Data Distribution & Longitudinal Study

3.1.3.1 Data Distribution

A histogram depicting the distribution of scores across each exam section is shown in Figure 11. All sections exhibit peaks above 90% except for Section II, which peaks at 50-59%. Section II also continues to exhibit more widely distributed scores. Section II exhibits a maximum of 12% at $\geq 90\%$ while other sections range from 36% to 57%.

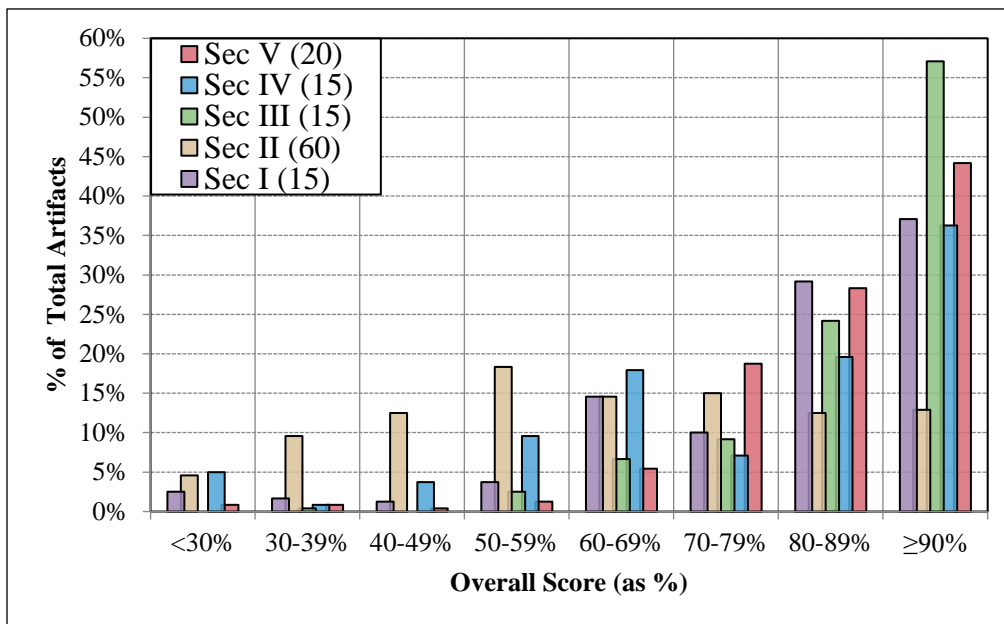


Figure 11. Histogram of SPN 1120 exam sections scores for fall 2019. Purple – Section I, Brown – Section II, Green – Section III, Blue – Section IV, and Red – Section V.

To describe the behavior of the section scores based on overall achievement, a color map, or binary raster image, was created by calculating the mean scores for each exam section as a function of combined score (Figure 9). The color represents the mean section score achieved overall score as shown in the x-axis as a percentage.

A review of the colormap in Figure 12 shows that Section II remains the lowest performing compared to other sections between the ranges of 50-100%. In the overall scoring range of 75%-79%, for example, Section II exhibits a mean score of 58% while the other four sections range from 77%-88%. At the overall scoring range at 90%-94%, Section II exhibits a mean score of 80%, while the other four sections range from 92%-95%.

	Section I	Section II	Section III	Section IV	Section V	
≥ 95%	98%	94%	99%	98%	97%	Scale
90-94%	94%	80%	95%	92%	93%	100%
85-89%	89%	76%	94%	85%	90%	90%
80-84%	83%	63%	92%	85%	85%	80%
75-79%	83%	58%	88%	77%	80%	70%
70-74%	75%	55%	85%	62%	82%	60%
65-69%	72%	45%	82%	62%	77%	50%
60-64%	61%	43%	73%	57%	78%	40%
55-59%	46%	40%	73%	51%	74%	30%
50-54%	46%	30%	74%	45%	64%	
< 50%	45%	29%	60%	23%	41%	

Figure 12. Colormap of mean scores for each exam section based on overall scoring bin for SPN 1120. An exam section with hotter colors (reds) compared with other sections means section achievement is stronger in that area than others. An exam section with colder colors (blues) compared with other sections means section achievement is weaker in that area than others.

3.1.3.2 Longitudinal Study

Further description of achievement over time in SPN 1120 is provided in Table 5 and Figure 13. Both demographics of students and student count vary by semester it may be more reasonable to compare like semesters (Fall vs. Fall, Spring vs. Spring) (see <http://www.fsw.edu/facultystaff/assessment/history> for further details). In 13 terms, Sections I through V for fall 2019 rank 3rd, 13th, 2nd, 4th, and 2nd, respectively.

	Section I	Section II	Section III	Section IV	Section V	Overall
Fall 2013 (n=58)	12.1	40.4	12.0	10.6	16.2	91.3
Spring 2014 (n=90)	12.4	45.1	13.2	11.7	16.8	99.2
Fall 2014 (n=93)	11.8	40.5	12.8	11.1	16.5	92.8
Spring 2015 (n=73)	11.5	39.8	12.8	10.9	16.4	91.4
Fall 2015 (n=122)	10.8	41.8	12.4	11.5	15.2	91.6
Spring 2016 (n=141)	11.5	38.4	13.3	10.5	15.7	89.5
Fall 2016 (n=240)	11.9	39.8	12.8	10.9	15.2	90.6
Spring 2017 (n=89)	11.9	41.0	13.0	11.0	16.7	93.6
Fall 2017 (n=266)	11.9	41.1	12.9	11.5	15.4	92.8
Spring 2018 (n=107)	12.0	43.7	12.8	11.6	16.4	96.6
Fall 2018 (n=212)	10.9	39.4	12.5	10.4	15.9	89.1
Spring 2019 (n=80)	11.6	41.3	12.4	10.3	16.3	91.8
Fall 2019 (n=240)	12.0	37.7	13.2	11.3	16.7	90.8

Table 5. Comparison of mean scores for SPN 1120 for fall 2013 through fall 2019.

Because some exam sections have different maximum scores (15, 20, and 60), to see which of the five sections of the exam exhibits the strongest achievement it may be best to relate them in terms of percent. As a percentage, Section III exhibit the highest mean scores in 12 of 13 terms including fall 2019. Section II exhibits the lowest mean scores in 12 of 13 terms, spring 2019 being the only term in which it is not the lowest. For spring 2019, Section IV exhibits the lowest mean score.

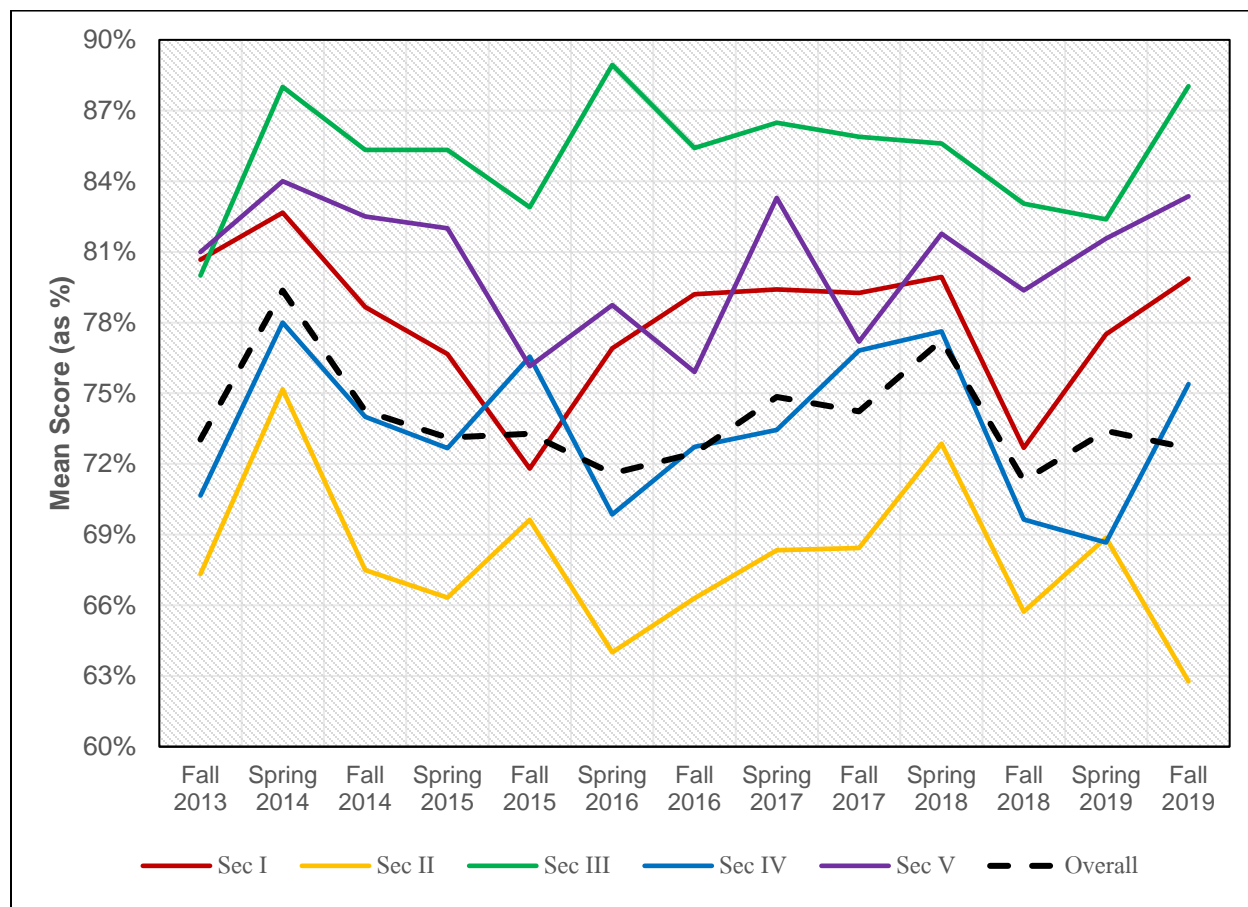


Figure 13. Comparison of mean scores (as percentage) for SPN 1120 through time from fall 2013 through fall 2019.

3.2 ENTRANCE SKILLS STUDY

In the spring 2016 term, a brief assessment was piloted in two sections of SPN 1121 on the Thomas Edison campus prior to entering into any engaged study in the course. The purpose of the assessment was to assess student skills and retention of materials from SPN 1120 prior to beginning SPN 1121. The assessment consists of a 15-minute reading comprehension quiz mirroring that which is seen in Section II of the SPN 1120 common final exam. The study compared study skill level based on the instructor they had for the previous course (SPN 1120) in an effort to align student skill level upon entry into SPN 1121. The pilot program for this study was first included in the fall 2015 assessment report. This program continues and was most recently administered at the beginning of spring 2020, before the writing of this report and is included here. The assessment has now been administered in all traditional sections of SPN 1121 beginning with summer 2016 through spring 2020 on the Charlotte, Collier, and Thomas Edison campuses. Results of the cumulative study are shown in Figure 14.

In concept, upon entry to SPN 1121, all students should have at least a passing score (or nearly so) of the final exam from SPN 1120. The entrance skills study exhibits a range of mean scores across instructor that is diminished since previous years (Figure 11). Over time, the range between highest and lowest instructors has narrowed. In fall 2016, the range was 17.8. In spring 2020, that gap has narrowed to 6.8.

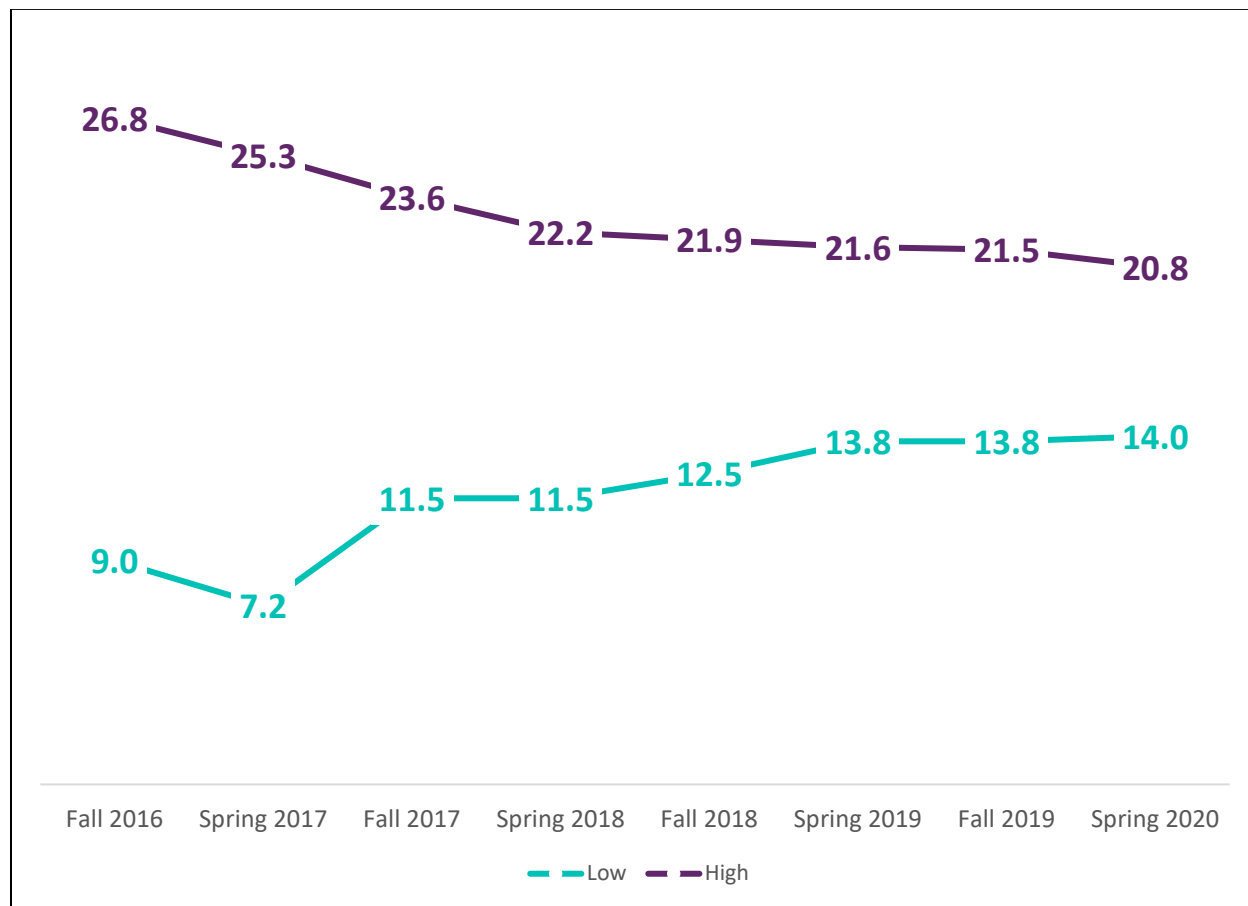


Figure 14. Comparison of achievement in entrance skills study assessment over time between highest scoring instructor (score listed is mean score) and lowest.

3.3 SPN 1121

3.3.1 Descriptive Statistics, Learning Outcomes, and Learning Objectives

Six sections of SPN 1121 Beginning Spanish II were offered in fall 2019. Of those, artifacts were collected from a common final from 5 of 6 sections. A total of 95 students were enrolled in SPN 1121. Of those, 63 artifacts were collected representing a sample size of 66%.

Using a common course assessment, the FSW Spanish faculty defined the same three areas of interest for evaluation that apply to SPN 1121 as those used for SPN 1120. For details on each SLO, see 3.1.1. The only difference between SPN 1121 and SPN 1120 in terms of measuring these outcomes is that the exam sections differ slightly and are noted in Table 5 below.

The faculty established measure of success for SLO 1, 80% of students scoring 70% or higher in Section I, was not met as results exhibit 89% of artifacts score 70% or higher in the oral competency exam section

(Section I) (Table 6, Figure 15). The faculty established measure of success for SLO 2, 80% of students scoring 70% or higher in reading only sections, Sections II, and VI, was not met. Results exhibit 46% of artifacts scored 70% or higher in Section II and 60% of artifacts scored 70% or higher in Section VI. The faculty established measure of success for SLO 3, 80% of students scoring 70% or higher in writing only sections, Sections V and VII, was nearly met. Results exhibit 56% of artifacts scored 70% or higher in Section V and 83% of artifacts scored 80% or higher in Section VII.

n = 185	Section I (Oral)	Section II (Reading)	Section III (Read/ Write)	Section IV (Read/ Write)	Section V (Writing)	Section VI (Reading)	Section VII (Writing)	Combined (Overall)
Goal	80% of artifacts scored $\geq 70\%$ for all sections							
% above Goal	59%	46%	49%	60%	56%	56%	83%	
Mean (as %)	72%	61%	61%	71%	67%	63%	78%	
Median (as %)	77%	67%	68%	80%	71%	77%	85%	
Section Score Max Possible	15	15	40	15	12	15	20	132
Section Mean	10.9	9.2	24.3	10.7	8.0	9.5	15.5	88.0
Section Median	11.5	10	27	12	8.5	11.5	17	89.5

Table 6. Percentage of student achievement level as per SLOs (SPN 1121).

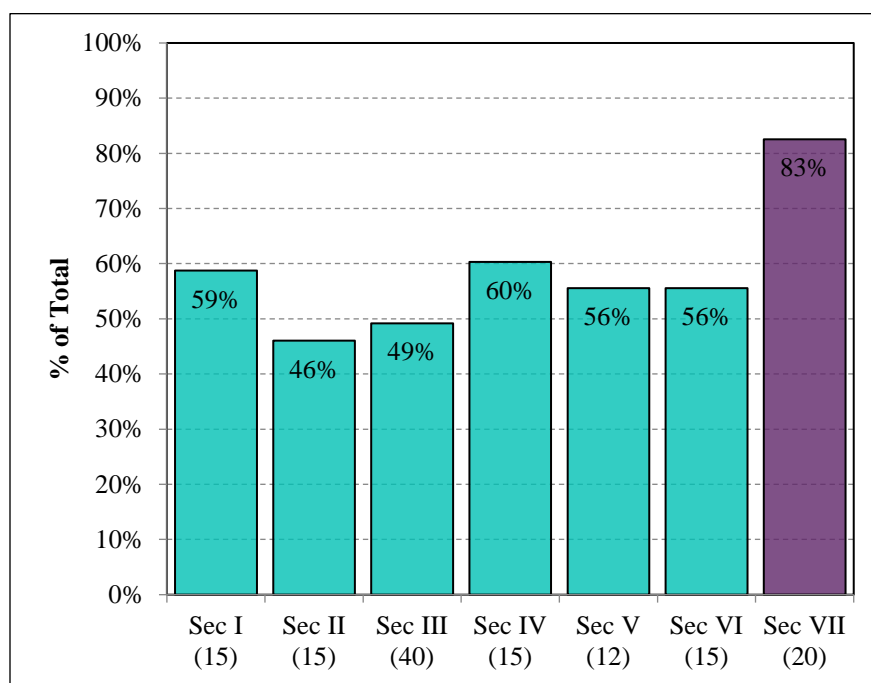


Figure 15. SLO achievement for SPN 1121 by exam section (Percentage of student achievement level as per SLOs). Purple denotes having met objective.

3.3.2 Exploratory Analysis & Significance Testing

Multiple comparisons of artifact scores across varying formats, campuses, and student types were made, where possible, in order to add depth to the causes of the distribution of the artifacts. Each course was divided into the appropriate subgroups to perform the analysis. In cases where a subgroup is not represented in the course comparisons were not conducted and are noted for comprehensiveness.

3.3.2.1 Dual Enrollment (Concurrent) to Non-Dual Enrollment Comparison

No dual enrollment (concurrent) sections of the course were run during fall 2019 so no comparison study between dual enrollment and traditional courses could be completed.

3.3.2.2 Online to Traditional Comparison

During the fall 2019 semester, two online sections reported data. From those course sections, 28 total online artifacts were collected from SPN 1121 and 35 traditional artifacts were collected from SPN 1121. A comparison of basic statistics is provided in Table 7. Online artifacts mean scores are 10.1 higher than traditional artifacts. Differences in the means were tested for significance using a Welch's t-test according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999) and were found to be statistically significantly different. Therefore, we can reject the null hypothesis that the differences in the means of the online and traditional artifacts are equal to 0, and we can conclude this with a 95% confidence that the differences in scores are not solely due to chance.

Effect size was calculated using a method devised by Rosenthal and Rosnow (1991) for meta-analytical purposes in potential comparisons with other institutions (Lipsey and Wilson, 1993). The results exhibit what Cohen (1988) would consider a medium effect size. In other words, non-overlap score distribution from online artifacts to traditional artifacts is approximately 36%. For a graphical representation of this see Figure 16.

df = 61		
Online mean	72.1	
Online standard deviation	16.00	
Traditional mean	62.3	
Traditional standard deviation	20.04	
Effect size	0.556	
p-value	0.034	

Table 7. Comparison of mean scores (as %) for online and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

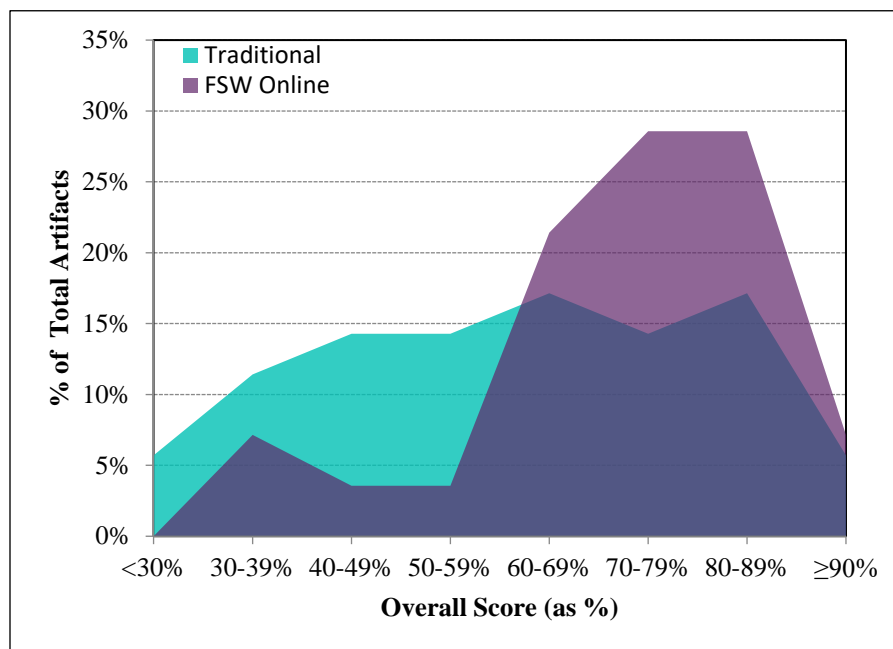


Figure 16. Score distribution for online (purple) and traditional (aqua) artifacts of SPN 1121.

3.3.2.3 Comparison by Campus/Site

Because only two sites reported data, Thomas Edison and FSW Online, a comparison by site is encompassed in 3.3.2.2 above.

3.3.3 Data Distribution & Longitudinal Study

3.3.3.1 Data Distribution

A histogram depicting the distribution of scores across each exam section is shown in Figure 17. Sections I, IV, V, and VII exhibit peaks centered on $\geq 90\%$. Section II exhibits a peak centered at 60-69%. Section III exhibits a peak centered at 70-79%. And section VI exhibits a peak centered at 80-89%.

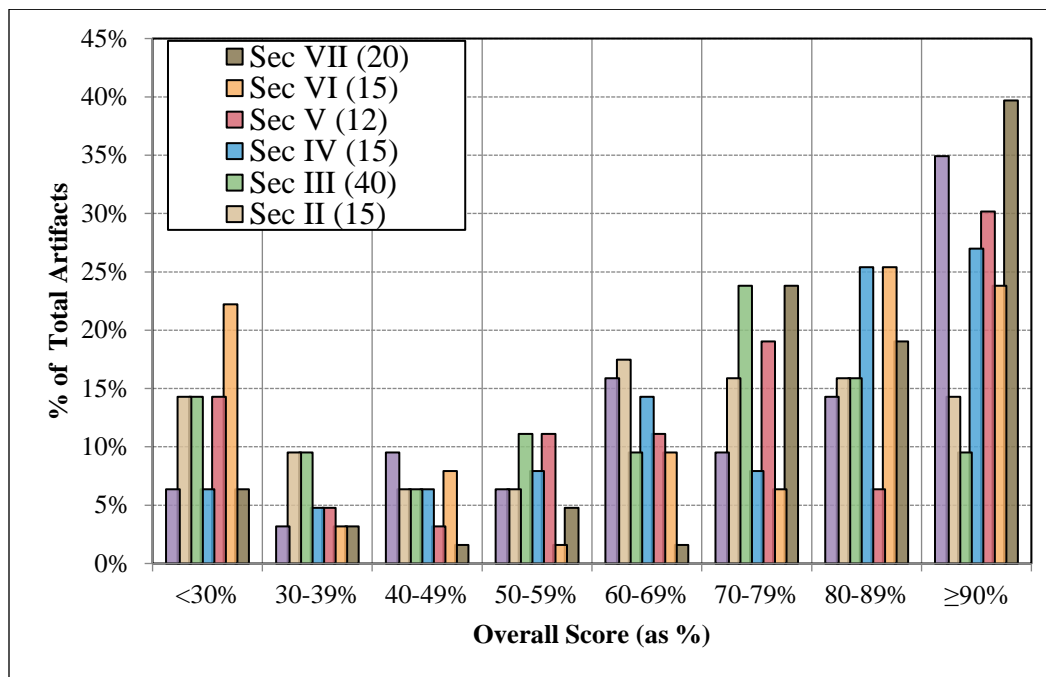


Figure 17. Histogram of SPN 1121 exam sections scores for fall 2019. Purple – Section I, Brown – Section II, Green – Section III, Blue – Section IV, Red – Section V, Orange – Section VI, and Black – Section VII.

To describe the behavior of the section scores based on overall achievement, a color map, or binary raster image, was created by calculating the mean scores for each exam section as a function of combined score (Figure 15). The color represents the mean section score achieved overall score as shown in the x-axis as a percentage.

A review of the colormap in Figure 18 shows several trends. First, Section VII exhibits a strong performance at lower overall scores. For example, in the 60-64% range, Section VII exhibits a mean score of 86%, while other sections range from 48%-81%. Second, Sections II and III, and to a lesser extent, Section VI, exhibit the weakest performance at mid-range scores. For example, in the range of 75-79%, Sections II and III exhibit scores of 66% and 54%, while other sections range from 80%-87%. And third, at high scores (90 or higher) all sections score fairly evenly, although Section II is somewhat lagging. In the $\geq 95\%$ range, Section II exhibits a mean of 88% while all other sections range from 95-100%.

	Section I	Section II	Section III	Section IV	Section V	Section VI	Section VII	
≥ 95%	100%	88%	95%	98%	98%	100%	100%	Scale
90-94%	98%	88%	85%	92%	86%	94%	92%	100%
85-89%	91%	75%	79%	90%	92%	80%	98%	90%
80-84%	78%	75%	76%	88%	84%	85%	86%	80%
75-79%	87%	66%	54%	85%	80%	83%	86%	70%
70-74%	67%	73%	61%	80%	60%	79%	79%	60%
65-69%	82%	62%	54%	77%	59%	51%	85%	50%
60-64%	62%	53%	52%	53%	81%	48%	86%	40%
55-59%	54%	54%	48%	54%	76%	33%	75%	30%
50-54%	62%	48%	73%	60%	46%	23%	65%	
< 50%	51%	29%	37%	39%	32%	34%	48%	

Figure 18. (Top) Colormap of mean scores for each exam section based on overall scoring bin for SPN 1121. An exam section with hotter colors (reds) compared with other sections means section achievement is stronger in that area than others. An exam section with colder colors (blues) compared with other sections means section achievement is weaker in that area than others.

3.3.3.2 Longitudinal Study

Further description of achievement over time in SPN 1121 is provided in Table 8 and Figure 19). Both demographics of students and student count vary by semester. It may be more reasonable to compare like semesters (Fall vs. Fall, Spring vs. Spring). (see <http://www.fsw.edu/facultystaff/assessment/history> for further details). In 13 terms, Sections I through VII for fall 2019 rank 10th, 9th, 13th, 8th, 9th, 10th, and 9th highest, respectively. In fall 2019, for the first time since the study began, Section III exhibits the lowest mean score.

	Section I Oral	Section II Reading	Section III Read/Write	Section IV Read/Write	Section V Writing	Section VI Reading	Section VII Writing	Overall
F 2013 (n=10)	11.5	9.5	34.2	9.5	7.5	9.6	14.2	96.0
Sp 2014 (n=115)	12.3	9.6	32.3	11.4	5.7	10.3	15.4	97.0
F 2014 (n=25)	11.9	10.7	30.0	10.6	9.5	11.6	16.1	100.5
Sp 2015 (n=58)	12.2	10.6	31.1	11.4	8.5	10.7	16.4	100.9
F 2015 (n=17)	11.5	10.6	29.9	11.3	8.1	9.9	16.4	97.7
Sp 2016 (n=109)	10.7	9.5	30.8	10.5	8.3	9.0	14.6	93.5
F 2016 (n=42)	11.5	9.8	25.9	10.7	7.2	9.7	15.6	90.3
Sp 2017 (n=140)	11.8	8.6	29.2	10.3	7.9	8.7	15.1	91.6
F 2017 (n=45)	10.6	9.1	27.3	11.8	8.1	10.2	15.7	92.8
Sp 2018 (n=136)	12.2	9.3	29.0	11.2	8.4	10.1	16.7	96.9
F 2018 (n=67)	10.4	8.7	27.9	10.2	8.2	8.3	15.9	89.6
Sp 2019 (n=185)	11.7	9.2	30.8	11.1	8.9	9.8	17.3	98.7
F 2019 (n=63)	10.9	9.2	24.3	10.7	8.0	9.5	15.5	88.0

Table 8. Comparison of mean scores for SPN 1121 for fall 2013 through fall 2019.

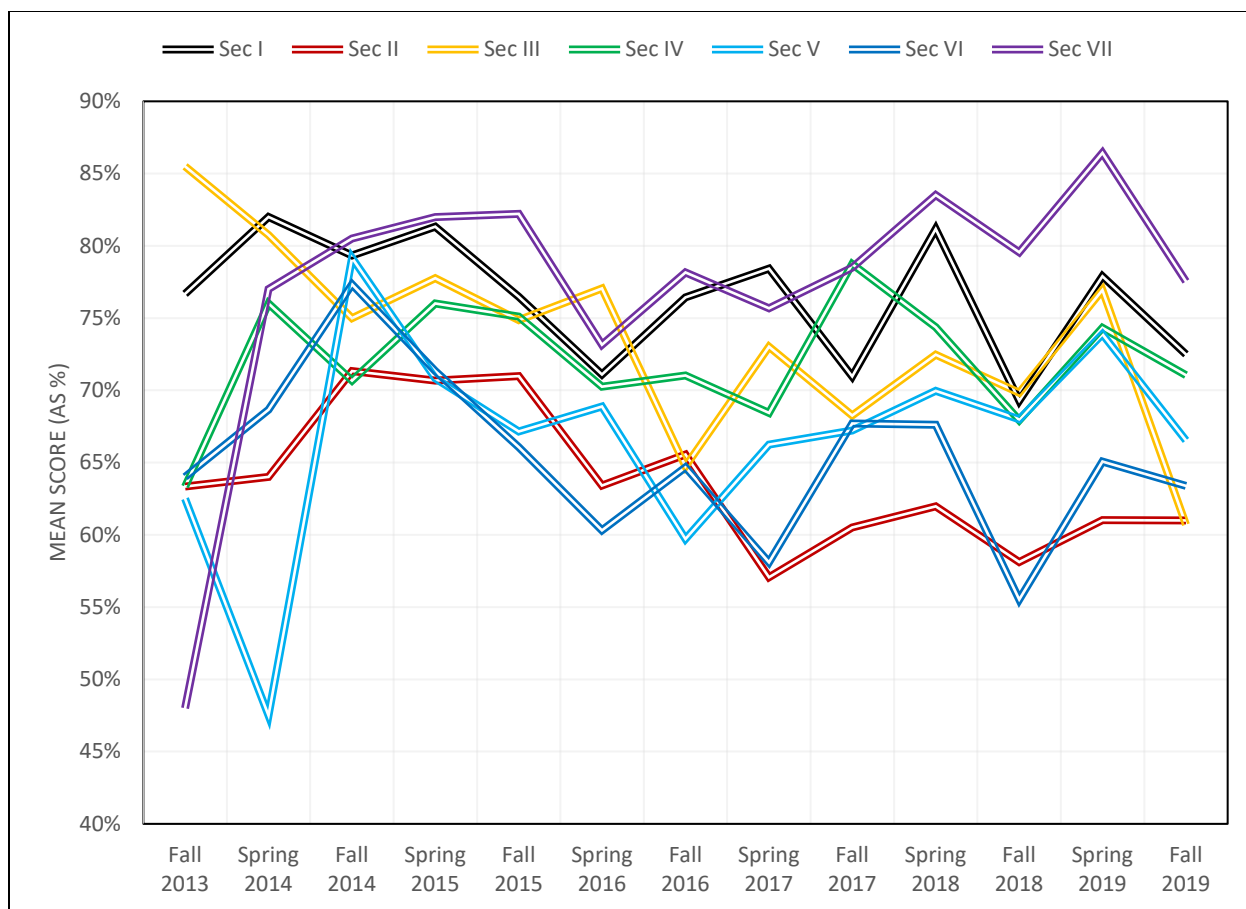


Figure 19. Comparison of mean scores (as percentage) for SPN 1121 through time.

4 CONCLUSIONS

Florida SouthWestern's Foreign Language Department employs a common course assessment in both French and Spanish courses to measure student progress in course level objectives in an effort to improve instruction. What follows is a drilldown of findings for both disciplines (French and Spanish) for the fall 2019 assessment.

4.1 FRENCH

A drill-down of FRE 1120 results are as follows:

1. During the fall 2019 term, five sections of FRE 1120 were offered. An assessment first piloted in AY 2017-18 was recorded in four section of FRE 1120 accounting for 55 artifacts, or 73% of the population.
2. In a study of Learning Outcome (LO) achievement, results exhibit mean scores across the five LOs ranging from 2.2 to 2.6, about the same as that seen in fall 2018, where the range was 2.3 to 2.7. It is important to note that each LOs in this study are arranged in order of advancement for the course. LO 1 is encountered early in the term with each LO encountered sequentially to LO 5, which is studied near the term's end.

3. While there was one dual enrollment (concurrent) section run during fall 2019, data was not reported so no comparison study between dual enrollment and traditional courses could be completed.
4. In a comparison between online and traditional artifacts, both online and traditional course sections exhibit similar artifact score distributions in the assessment at 13/15 or higher. However, traditional artifacts occupy a substantial number of lower scores that the online section does not. For example, approximately 24% of traditional artifacts score between 6/15 and 9/15. By comparison, none from online exhibit scores in that range.
5. While courses were offered at three locations, the third, the Collier Campus, is comprised by only eight artifacts. As a result, any meaningful comparison is best left with the online-to-traditional comparison.
6. In a study of achievement based on disposition survey question response, results yield a clearer sense of just how many students self-report a strong understanding of the topic and yet do not appear to be capable of demonstrating it. A clear trend in confidence correlating with performance is visible with increases in “Agree” and “Strongly Agree” coinciding with higher combined scores.
7. In a longitudinal study, in three terms, LOs 1 through 5 for fall 2019 rank 2nd, 2ⁿ, 2nd, 3rd, and 3rd, respectively.

A drill-down of FRE 1121 results are as follows:

1. During the fall 2019 term, two sections of FRE 1121 were offered. The assessment was recorded in both, however, in one section results were shuffled, so no data could be utilized. The data collection accounts for 14 artifacts, or 67% of the population.
2. In a study of Learning Outcome (LO) achievement, results exhibit mean scores across the four LOs ranging from 2.7 to 2.9. It is important to note that each LOs in this study are arranged in order of advancement for the course. LO 1 is encountered early in the term with each LO encountered sequentially to LO 4, which is studied near the term’s end.
3. There was no dual enrollment (concurrent) section run during fall 2019, so no comparison study between dual enrollment and traditional courses could be completed.
4. While both an online and a traditional section were offered, data could not be utilized from the traditional section due to data recording issues. As a result, no comparison analysis between the two could be completed.
5. While courses were offered at three locations, the third, the Collier Campus, is comprised by only eight artifacts. As a result, any meaningful comparison is best left with the online-to-traditional comparison.
6. In a study of achievement based on disposition survey question response, results yield a clearer sense of just how many students self-report a strong understanding of the topic and yet do not appear to be capable of demonstrating it. Because most students scored a perfect 12/12, it is unclear if any correlation exists.
7. Fall 2019 is the first data collection of this assessment. A longitudinal study will begin following data collection of fall 2020 data.

4.2 SPANISH

A drill-down of SPN 1120 results are as follows:

1. During the fall 2019 semester 14 sections of SPN 1120 Beginning Spanish I were offered. Of those, artifacts from a common final were collected from 13 sections. One section did not report data.

A total of 314 students were enrolled in SPN 1120. Of those, 240 artifacts were collected representing a sample size of 76% of the population.

2. Achievement of 80% of artifacts scoring 70% or better (SLO 1): Achievement was nearly met as results exhibit 76% of artifacts score 70% or higher in the oral competency exam section (Section I).
3. Achievement of 80% of artifacts scoring 70% or better (SLO 2): Achievement was partially met. Results exhibit 40% of artifacts scored 70% or higher in Section II and 90% of artifacts scored 70% or higher in Section III.
4. Achievement of 80% of artifacts scoring 70% or better (SLO 3): Achievement was partially met. Results exhibit 63% of artifacts scored 70% or higher in Section IV and 91% of artifacts scored 70% or higher in Section V.
5. No dual enrollment (concurrent) sections of the course were run during fall 2019 so no comparison study between dual enrollment and traditional courses could be completed.
6. In a comparison of online to traditional artifacts, online artifacts scores are 0.5 lower than traditional. Results were not statistically significantly different.
7. In a cross-campus comparison study, the Collier campus exhibits the highest mean scores in 3 of 5 sections and the overall score. The Charlotte campus exhibits the highest mean score in 2 of 5 sections.
8. In a study of score distribution by section, all sections exhibit peaks above 90% except for Section II, which peaks at 50-59%. Section II also continues to exhibit more widely distributed scores. Section II exhibits a maximum of 12% at $\geq 90\%$ while other sections range from 36% to 57%.
9. In a study of section score distribution based on overall score, Section II remains the lowest performing compared to other sections between the ranges of 50-100%. In the overall scoring range of 75%-79%, for example, Section II exhibits a mean score of 58% while the other four sections range from 77%-88%. At the overall scoring range at 90%-94%, Section II exhibits a mean score of 80%, while the other four sections range from 92%-95%.
10. In a longitudinal study of data distribution through time, in 13 terms, Sections I through V for fall 2019 rank 3rd, 13th, 2nd, 4th, and 2nd, respectively.
11. In a study assessing student skills and retention of materials from SPN 1120 prior to beginning SPN 1121, over time, the range between highest and lowest instructors has narrowed. In fall 2016, the range was 17.8. In spring 2020, that gap has narrowed to 6.8.

A drill-down of SPN 1121 results are as follows:

1. Six sections of SPN 1121 Beginning Spanish II were offered in fall 2019. Of those, artifacts were collected from a common final from 5 of 6 sections. A total of 95 students were enrolled in SPN 1121. Of those, 63 artifacts were collected representing a sample size of 66%.
2. Achievement of 80% of artifacts scoring 70% or better (SLO 1): Achievement was not met as results exhibit 89% of artifacts score 70% or higher in the oral competency exam section (Section I).
3. Achievement of 80% of artifacts scoring 70% or better (SLO 2): Achievement was not met. Results exhibit 46% of artifacts scored 70% or higher in Section II and 60% of artifacts scored 70% or higher in Section VI.
4. Achievement of 80% of artifacts scoring 70% or better (SLO 3): Achievement was nearly met. Results exhibit 56% of artifacts scored 70% or higher in Section V and 83% of artifacts scored 80% or higher in Section VII.
5. No dual enrollment (concurrent) sections of the course were run during fall 2019 so no comparison study between dual enrollment and traditional courses could be completed.

6. In a comparison of online to traditional artifacts, online artifacts score 10.1 higher than traditional. Results were statistically significantly different.
7. Because only two sites reported data, Thomas Edison and FSW Online, no cross-campus comparison was completed because that study was encompassed by the online-to-traditional study.
8. In a study of score distribution by section, Sections I, IV, V, and VII exhibit peaks centered on $\geq 90\%$. Section II exhibits a peak centered at 60-69%. Section III exhibits a peak centered at 70-79%. And section VI exhibits a peak centered at 80-89%.
9. In a study of section score distribution based on overall score, several trends are present. First, Section VII exhibits a strong performance at lower overall scores. For example, in the 60-64% range, Section VII exhibits a mean score of 86%, while other sections range from 48%-81%. Second, Sections II and III, and to a lesser extent, Section VI, exhibit the weakest performance at mid-range scores. For example, in the range of 75-79%, Sections II and III exhibit scores of 66% and 54%, while other sections range from 80%-87%. And third, at high scores (90 or higher) all sections score fairly evenly, although Section II is somewhat lagging. In the $\geq 95\%$ range, Section II exhibits a mean of 88% while all other sections range from 95-100%.
10. In a longitudinal study of data distribution through time, in 13 terms, Sections I through VII for fall 2019 rank 10th, 9th, 13th, 8th, 9th, 10th, and 9th highest, respectively. In fall 2019, for the first time since the study began, Section III exhibits the lowest mean score.

5 REFERENCES

- Cohen, J. 1988. Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Earlbaum Associates, Hillsdale, NJ.
- Davis, J.C. 1973. Statistics and Data Analysis in Geology. John Wiley & Sons, New York, New York, 564 pp.
- Lipsey, M.W. and Wilson, D.B. 1993. The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.
- McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.
- Rosenthal, R. and Rosnow, R.L. 1991. Essentials of behavioral research: Methods and data analysis (2nd ed.). McGraw Hill, New York, NY.
- Starkweather, J. D. 2010. Introduction to Statistics for the Social Sciences. In: Research and Statistical Support. Retrieved from http://www.unt.edu/rss/class/Jon/ISSS_SC/.
- Wilkinson, L. 1999. APA Task Force on Statistical Inference. Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist* 54 (8), 594–604.