

# Foreign Language Assessment Report

## Spring 2018

Author: Joseph F. van Gaalen, Ph.D., Director, 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 the 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 spring 2018 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, Director of Assessment & Effectiveness ([jfvangaalen@fsw.edu](mailto:jfvangaalen@fsw.edu); x16965).

## 2 FRENCH

---

### 2.1 FRE 1120

During the spring 2018 term, three sections of FRE 1120 were offered. A newly developed assessment was piloted using two sections of FRE 1120 accounting for 29 artifacts, or 47% 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.

#### 2.1.1 Descriptive Statistics, Learning Outcomes, and Objectives

The lead French faculty has identified five areas of interest for evaluation. Given that it the assessment is currently in pilot phase, no achievement objectives have yet been set. The Learning Outcomes (LOs) are:

- LO 1: Reading/Writing Introductions

- LO 2: Reading/Writing Daily Activities
- LO 3: Reading/Writing Hobbies
- LO 4: Reading/Writing Physical Appearances
- LO 5: Reading/Writing Indoor and Outdoor Spaces

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 2.0 to 2.9. The highest scoring area is LO 1, with a mean score of 2.9/3.0. The lowest is LO 5 at 2.0/3.0. 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. Results would suggest some level increased difficulty/complexity, although this is up to the interpretation of French faculty.

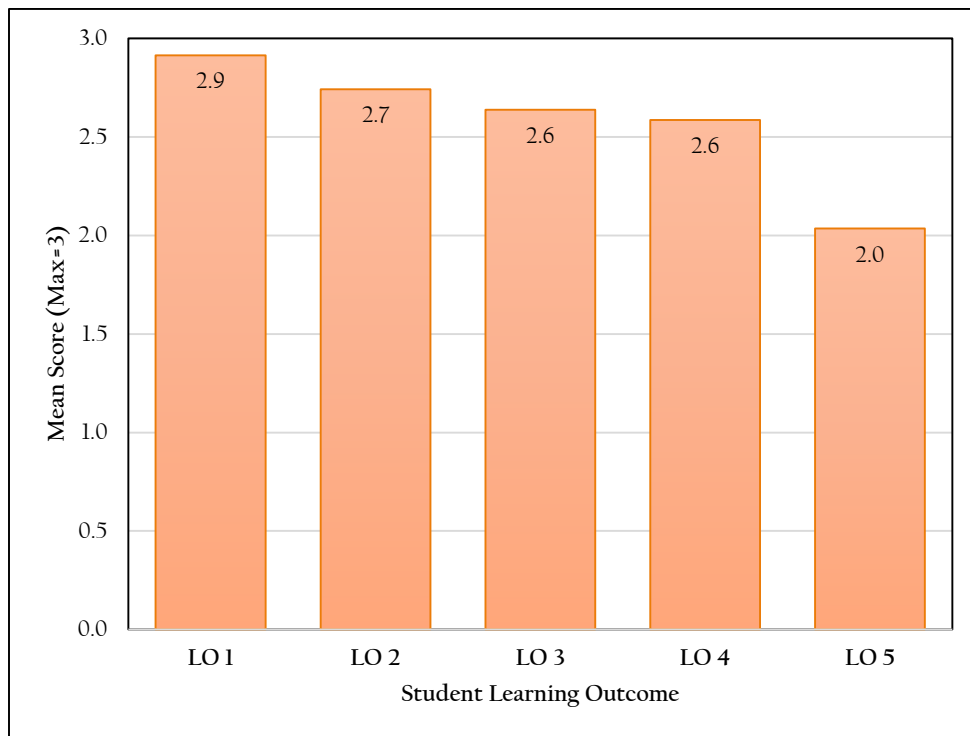


Figure 1. Mean scores by SLOs for FRE 1120.

## 2.1.2 Exploratory Analysis & Significance Testing

### 2.1.2.1 Dual Enrollment (Concurrent) to Traditional Comparison

No dual enrollment (concurrent) sections of the course were run during spring 2018 so no comparison study between dual enrollment and traditional courses could be completed.

### 2.1.2.2 Online to Traditional Comparison

While an online section of the course was offered in spring 2018, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.

### 2.1.2.3 Comparison by Campus/Site

While courses were offered at two locations in spring 2018, Thomas Edison (Lee) and FSW Online, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.

## 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 2. Scores are centered on 14/15 with a strong negative skew meaning scores are tending strongly towards higher values (Starkweather, 2010). In total, 69% of artifacts score 13/15 or higher and 9% total 8/15 or lower.

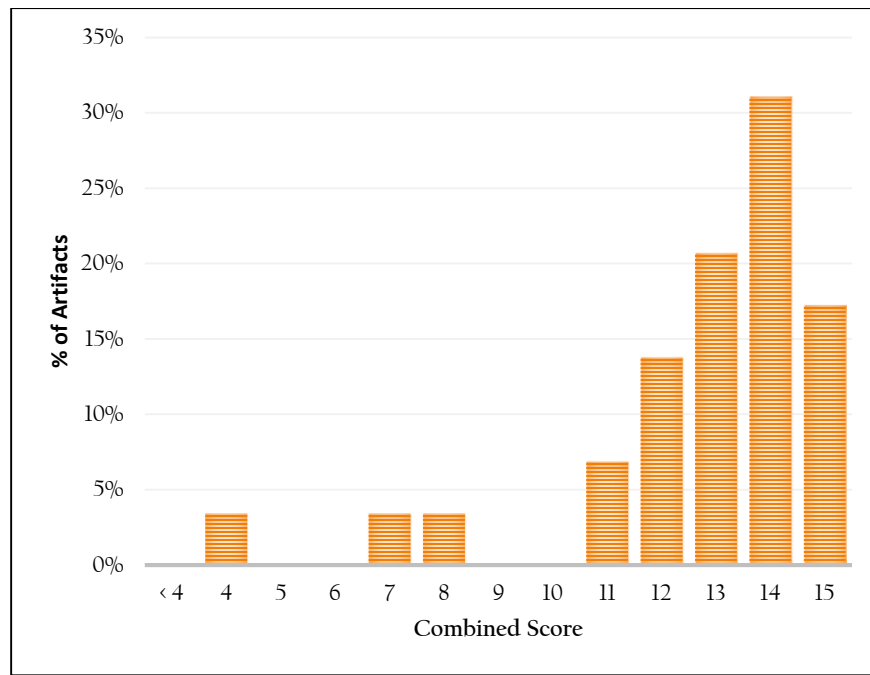


Figure 2. Distribution of combined (total) scores for assessment.

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. Figure 3 exhibits the percent of artifacts with correct answers for each question (defined as 1a - translate French by answering a question written in French using an English response, 1b - translate French by answering a question written in French using a French response, and so on) based on student response to the confidence of that particular topic. For example, LO 1a and 1b are coupled with a survey question regarding confidence in reading and writing introductions. Similarly, LO 2a and 2b are coupled with a survey question regarding confidence in reading and writing in daily activities.

Results for LOs 1, 2, 3, and 5 yield little to conclude as achievement based on “Strongly Agree” and “Agree” are fairly similar. In LO 4, those answering “Strongly Agree” outperform those answer “Agree”

by as much as 27% in the case of LO 4b. While results vary wildly with those answering “Disagree,” the sample size is never higher than n=2, and so interpretation is limited.

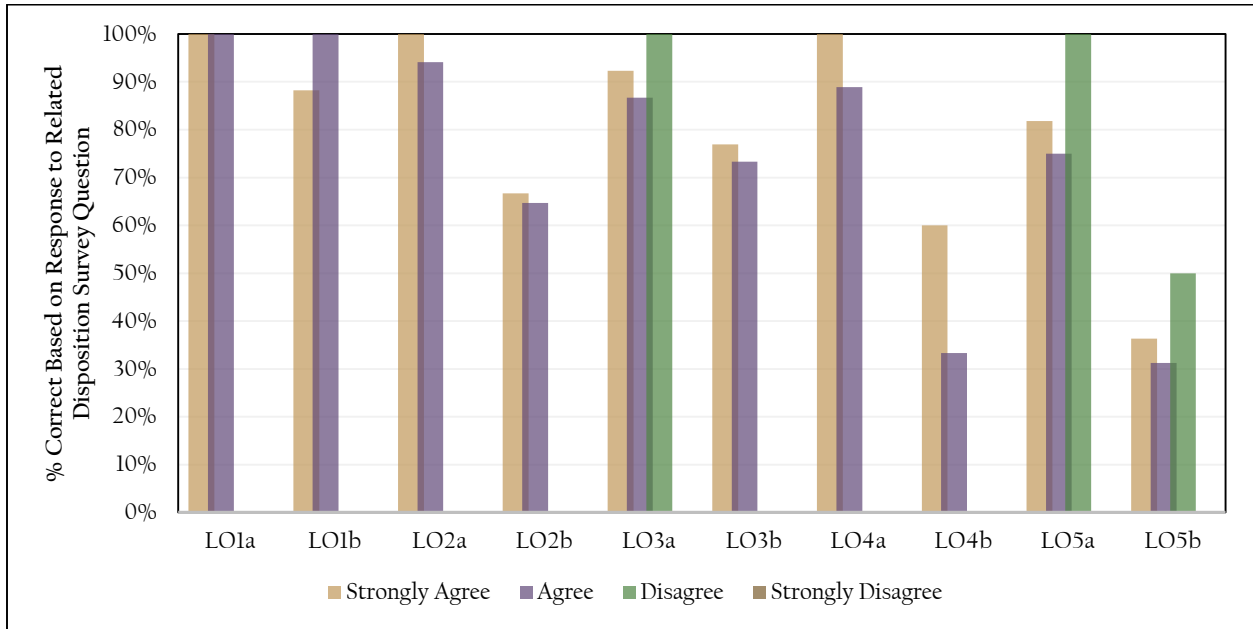


Figure 3. Comparison of % correct based on response to related disposition survey item. Sample size (Strongly Agree / Agree / Disagree / Strongly Disagree) for LO 1: 17/12/0/0, LO 2: 12/17/0/0, LO 3: 13/15/1/0, LO 4: 20/9/0/0, and LO 5: 11/16/2/0.

A similar study to that above but instead using mean scores instead of percent of artifacts with correct answers for each question is shown in Figure 4. As with above, LO 4 shows substantial differences in achievement between those answering “Strongly Agree” and those answering “Agree.” In LO 4b, mean scores drop from 1.7/3 for those answering “Strongly Agree” to 1.4/3 for those answering “Agree.”

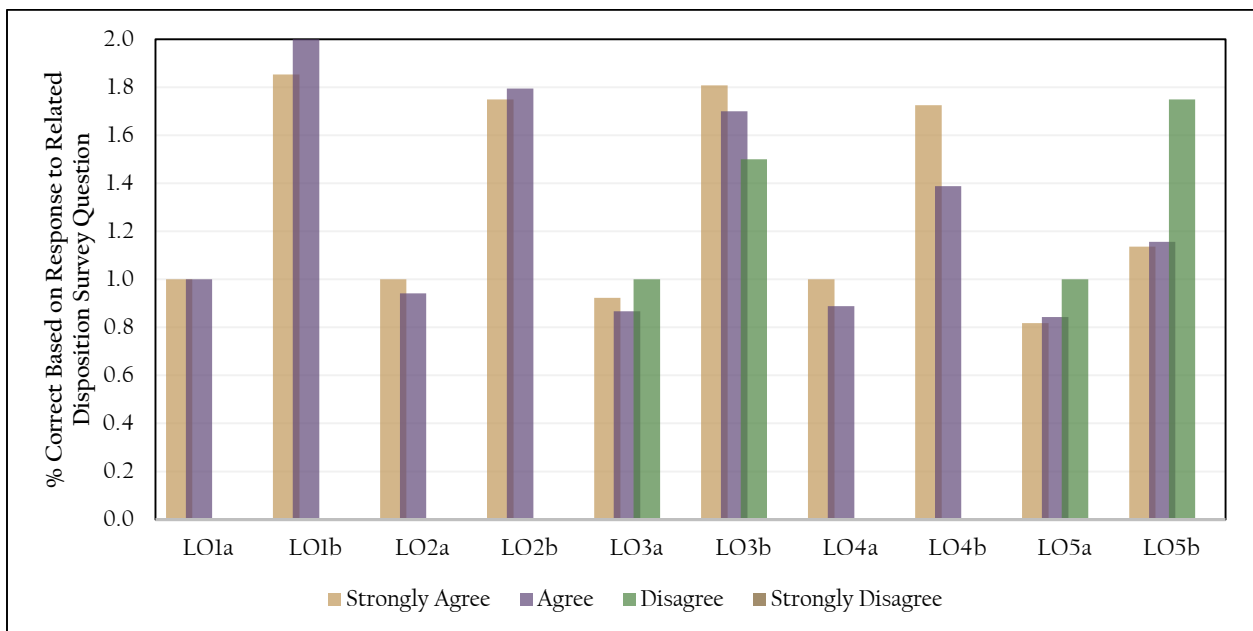


Figure 4. Comparison of mean scores based on response to related disposition survey item. Sample size (Strongly Agree / Agree / Disagree / Strongly Disagree) for LO 1: 17/12/0/0, LO 2: 12/17/0/0, LO 3: 13/15/1/0, LO 4: 20/9/0/0, and LO 5: 11/16/2/0.

Another way of gauging achievement based on disposition survey responses is through a scatter plot as shown in Figure 5 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. In this pilot study, 3 of 29 students report strong to very strong confidence in their skillset and yet did not achieve results comparable to that confidence. The highest of those three scoring 8/15, or 53%. Similarly, 1 of 29 students self-reports a lack of confidence despite scoring 13.5/15, or 90%.

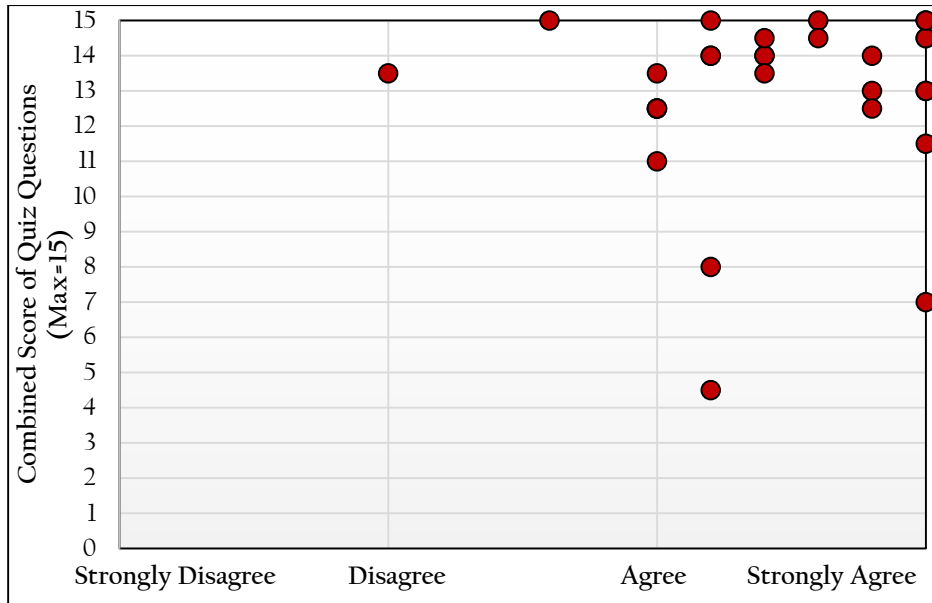


Figure 5. Scatter plot of combined (total) scores based on general response to survey questions.

### 2.1.3.2 Longitudinal Study

The assessment for FRE 1120 is in its first pilot phase for spring 2018. As further data is collected in coming terms, this section will track achievement through time and highlight strengths, weaknesses and any long term trends.

## 2.2 FRE 1121

During the spring 2018 term, three sections of FRE 1121 were offered. A newly developed assessment was piloted using one sections of FRE 1121 accounting for 20 artifacts, or 30% 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.

### 2.2.1 Descriptive Statistics, Learning Outcomes, and Objectives

The lead French faculty has identified four areas of interest for evaluation. Given that it the assessment is currently in pilot phase, no achievement objectives have yet been set. The Learning Outcomes (LOs) are:

- LO 1: Reading/Writing Daily Life
- 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 1.3 to 1.9, somewhat lower than the 2.0-2.9 range in FRE 1120 (Figure 6). The highest scoring area is LO 2, with a mean score of 1.9/3.0. The lowest is LO 4 at 1.3/3.0.

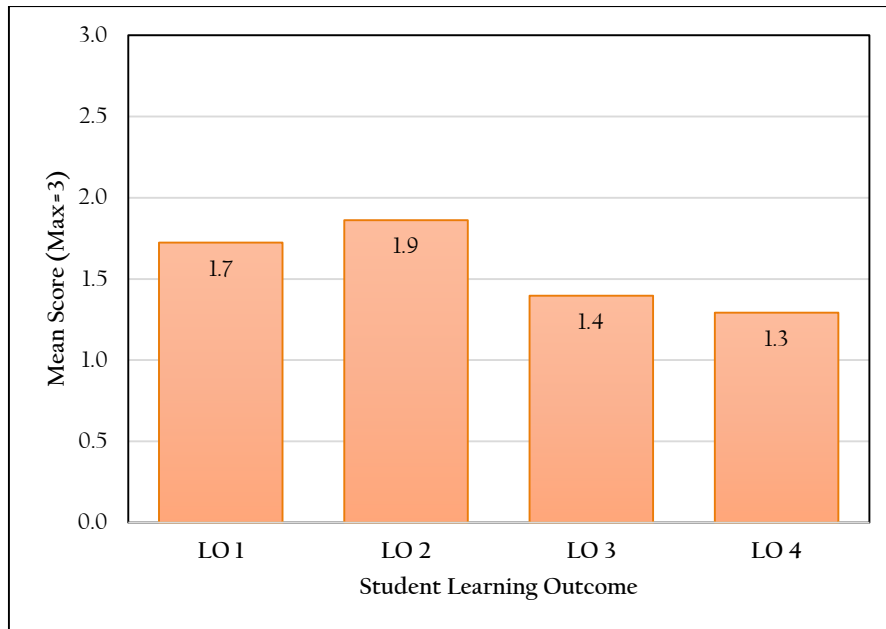


Figure 6. Mean scores by SLOs for FRE 1120.

### 2.2.2 Exploratory Analysis & Significance Testing

#### 2.2.2.1 Dual Enrollment (Concurrent) to Traditional Comparison

While a dual enrollment (concurrent) section of the course was offered during spring 2018, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.

#### 2.2.2.2 Online to Traditional Comparison

While an online section of the course was offered in spring 2018, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.

### 2.2.2.3 Comparison by Campus/Site

While courses were offered at three locations in spring 2018, Thomas Edison (Lee), FSW Online, and offsite, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.

## 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 10/12 with a moderate negative skew meaning scores are tending somewhat towards higher values (Starkweather, 2010). In total, 45% of artifacts score 10/12 or higher and 5% total 4/12 or lower.

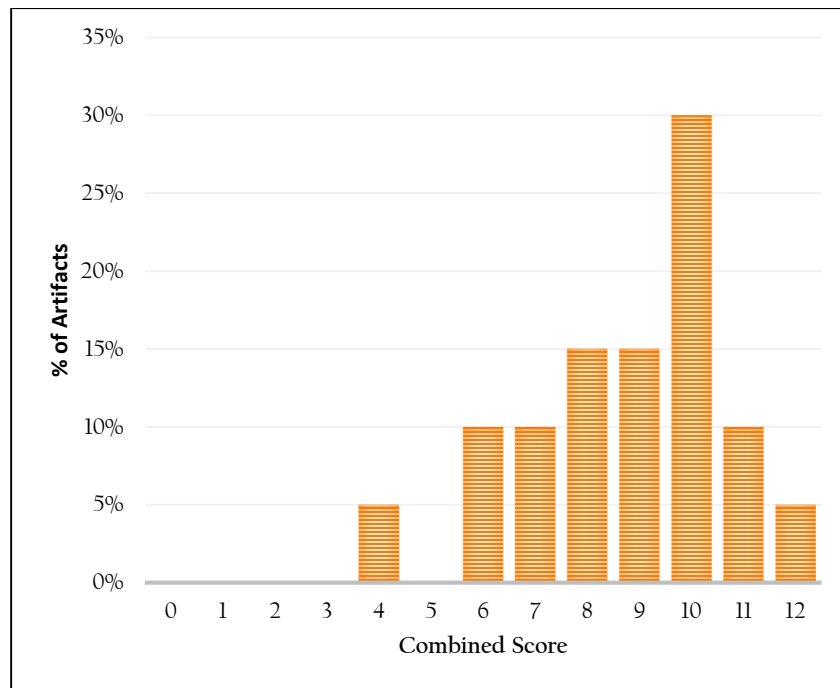


Figure 7. Distribution of combined (total) scores for assessment.

Figure 8 exhibits the percent of artifacts with correct answers for each question (defined as 1a - translate French by answering a question written in French using an English response, 1b - translate French by answering a question written in French using a French response, and so on) based on student response to the confidence of that particular topic. For example, LO 1a and 1b are coupled with a survey question regarding confidence in reading and writing introductions. Similarly, LO 2a and 2b are coupled with a survey question regarding confidence in reading and writing in daily activities.

Results for this study are expected to yield some valuable results, as shown in FRE 1120. Note that most students self-reported "Agree," and so analysis by this type of disaggregation does not yield results as fruitful as those seen in FRE 1120. Results from LO 1a appear to exhibit an indirect relationship (higher confidence coincides with lower results) although those answering "Strongly Agree" is limited to n=5.

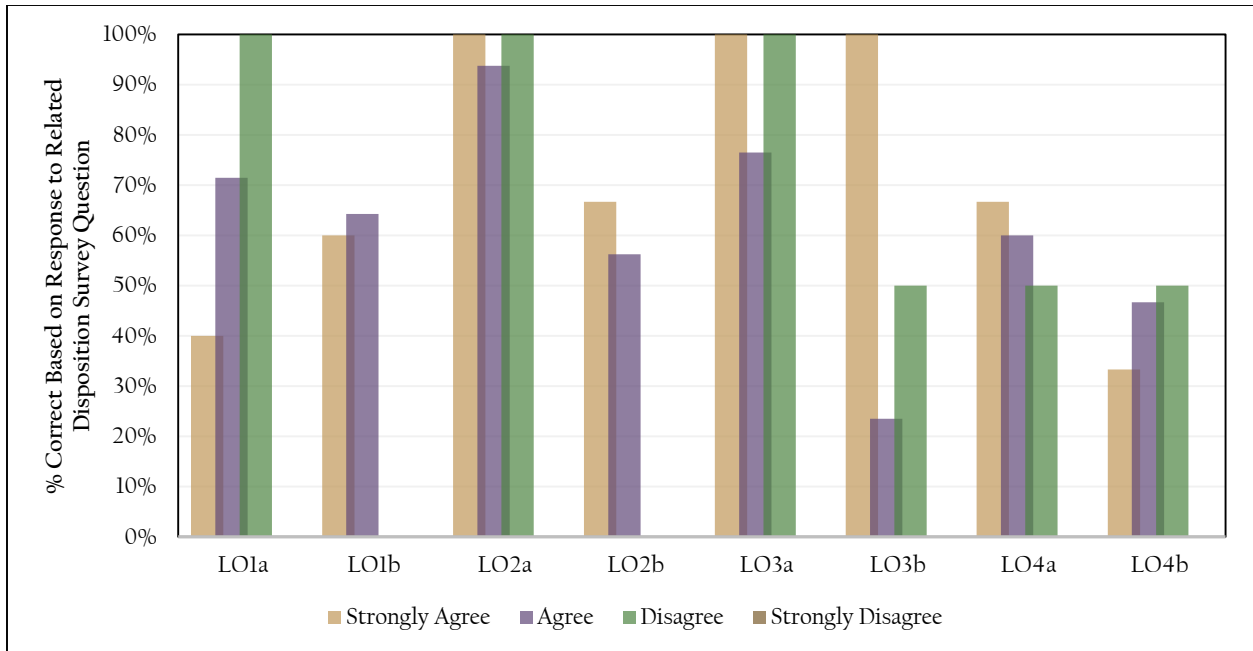


Figure 8. Comparison of % correct based on response to related disposition survey item. Sample size (Strongly Agree / Agree / Disagree / Strongly Disagree) for LO 1: 5/14/1/0, LO 2: 3/16/1/0, LO 3: 1/17/2/0, and LO 4: 3/15/2/0.

A similar study to that above but instead using mean scores instead of percent of artifacts with correct answers for each question is shown in Figure 9. As with above, most students self-reported “Agree,” and so analysis by this type of disaggregation does not yield results as fruitful as those seen in FRE 1120. Results from LO 1a again appear to exhibit an indirect relationship (higher confidence coincides with lower results) although those answering “Strongly Agree” is limited to n=5.

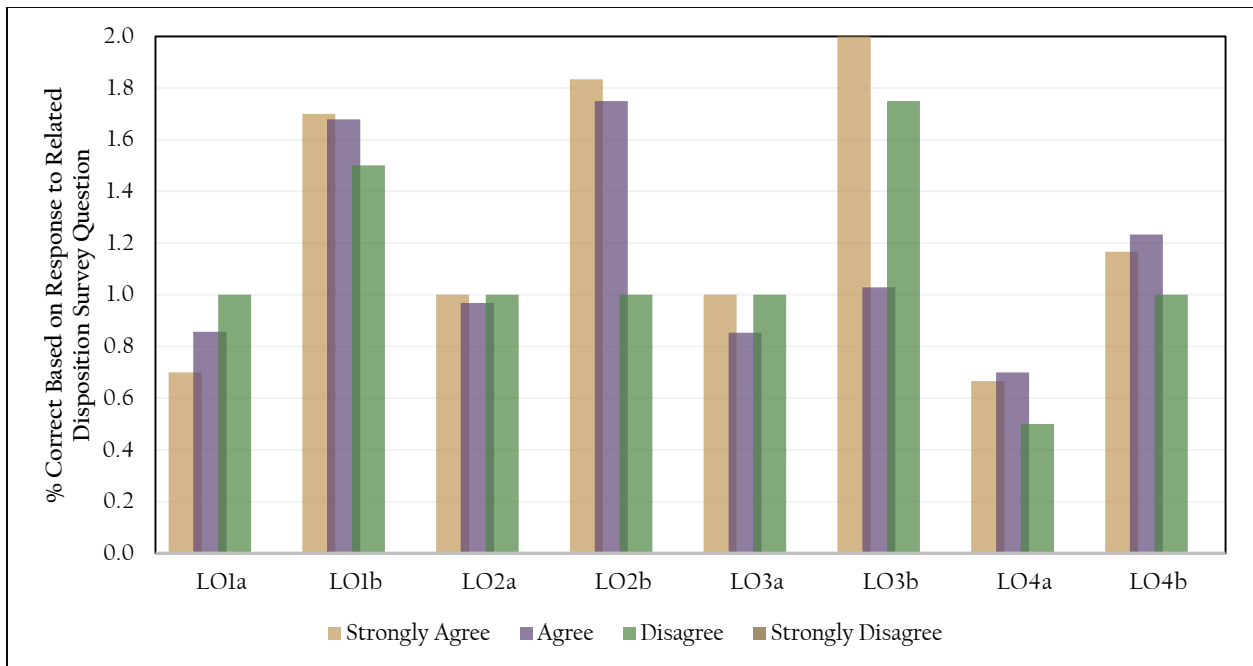


Figure 9. Comparison of mean scores based on response to related disposition survey item. Sample size (Strongly Agree / Agree / Disagree / Strongly Disagree) for LO 1: 5/14/1/0, LO 2: 3/16/1/0, LO 3: 1/17/2/0, and LO 4: 3/15/2/0.



Another way of gauging achievement based on disposition survey responses is through a scatter plot as shown in Figure 10 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. In this pilot study, most students exhibit limited confidence compared with FRE 1120, and this correlates well with lower scores on the achievement portion. Also, 2 of 20 students self-report a lack of confidence in their skillset. One achieved a 10.5/12 while the other a 6.5/12.

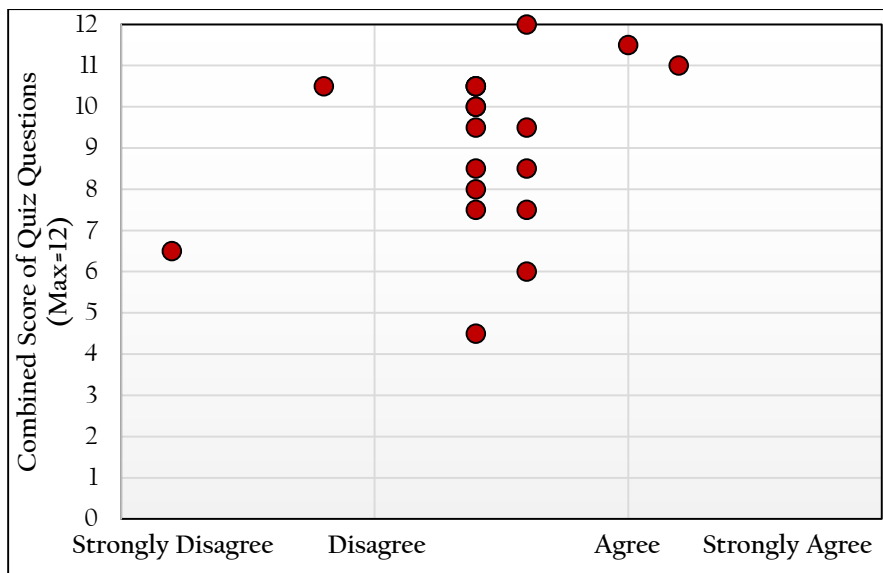


Figure 10. Scatter plot of combined (total) scores based on general response to survey questions.

### 2.2.3.2 Longitudinal Study

The assessment for FRE 1121 is in its first pilot phase for spring 2018. As further data is collected in coming terms, this section will track achievement through time and highlight strengths, weaknesses and any long term trends.

## 3 SPANISH

### 3.1 SPN 1120

#### 3.1.1 Descriptive Statistics, Learning Outcomes, and Learning Objectives

During the spring 2018 semester 9 sections of SPN 1120 Beginning Spanish I were offered. Of those, artifacts from a common final were collected from 7 sections. One section did not report data. A second section (an online section) used an altered form of the common assessment skewing any potential data comparison. A total of 339 students were enrolled in SPN 1120. Of those, 107 artifacts were collected representing a sample size of 54% 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 77% of artifacts score 70% or higher in the oral competency exam section (Section I) (Table 1). 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 56% 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 nearly met. Results exhibit 64% of artifacts scored 70% or higher in Section IV and 79% of artifacts scored 70% or higher in Section V. For a graphical representation of SLO achievement, see Figure 11.

n = 107	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%	77%	56%	90%	64%	79%	
Mean (as %)	80%	73%	86%	78%	82%	78%
Median (as %)	83%	77%	87%	80%	88%	80%
Section Score Max	15	60	15	15	20	125
Section Mean	12.0	43.7	12.8	11.6	16.4	96.6
Section Median	13	46	13	12	17.5	99

Table 1. Percentage of student achievement level as per SLOs (SPN 1120).

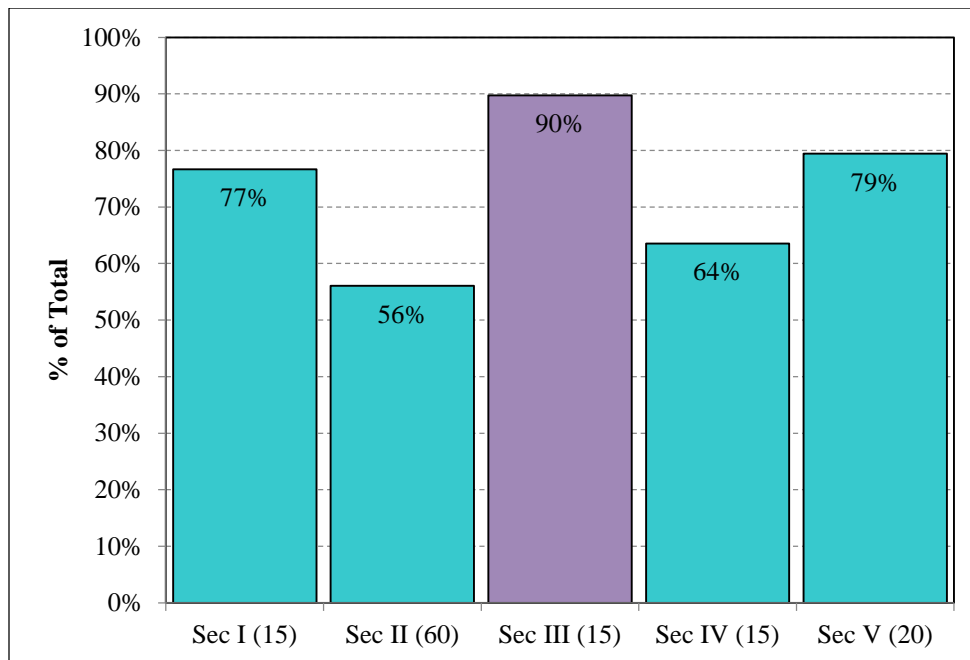


Figure 11. SLO achievement for SPN 1120 by exam section (Percentage of student achievement level as per SLOs). Purple denotes having met objective.

### 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 spring 2018 so no comparison study between dual enrollment and traditional courses could be completed.

#### 3.1.2.2 *Online to Traditional Comparison*

During the spring 2018 semester, two online sections reported data. From those course sections, 31 total online artifacts were collected from SPN 1120 and 76 traditional artifacts were collected from SPN 1120. A comparison of basic statistics is provided in Table 2. Online artifacts mean scores are 1.8 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 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 8%. For a graphical representation of this see Figure 12.

df = 105		
Online mean	78.5	
Online standard deviation	15.54	
Traditional mean	76.7	
Traditional standard deviation	15.36	
Effect size	-0.10	
p-value	0.600	

Table 2. Comparison of mean scores (as %) for online and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

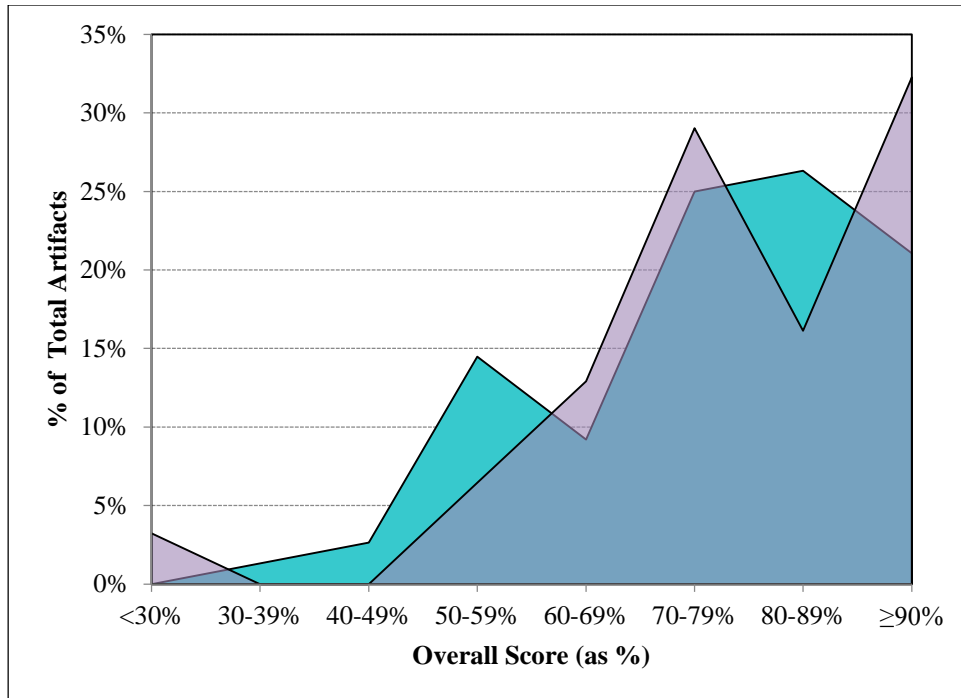


Figure 12. Score distribution for online (purple) and traditional (aqua) artifacts of SPN 1120.

### 3.1.2.3 Comparison by Campus/Site

Of the 107 artifacts collected from SPN 1120, 0 originated from the Charlotte campus, 0 from the Collier campus, 31 from FSW Online, and 76 from the Thomas Edison (Lee) campus. Scores by rubric dimension varied greatly across campuses. A comparison of mean scores by rubric dimension is provided in Table 3.

	Section I	Section II	Section III	Section IV	Section V	Combined Score
<i>Rubric Max</i>	15	60	15	15	20	125
Charlotte	~	~	~	~	~	~
Collier	~	~	~	~	~	~
FSW Online	<b>12.5</b>	<b>43.1</b>	<b>13.0</b>	<b>11.5</b>	<b>18.0</b>	98.1
Thomas Edison (Lee)	11.9	42.8	12.9	11.1	15.5	94.2

Table 3. Comparison of mean scores by site for SPN 1120. Bold denotes highest mean score in that dimension among all sites. Rubric dimensions identified in SLOs in blue.

Only two sites reported data for the spring 2018 term. FSW Online exhibits higher scores in all sections (Figure 13).

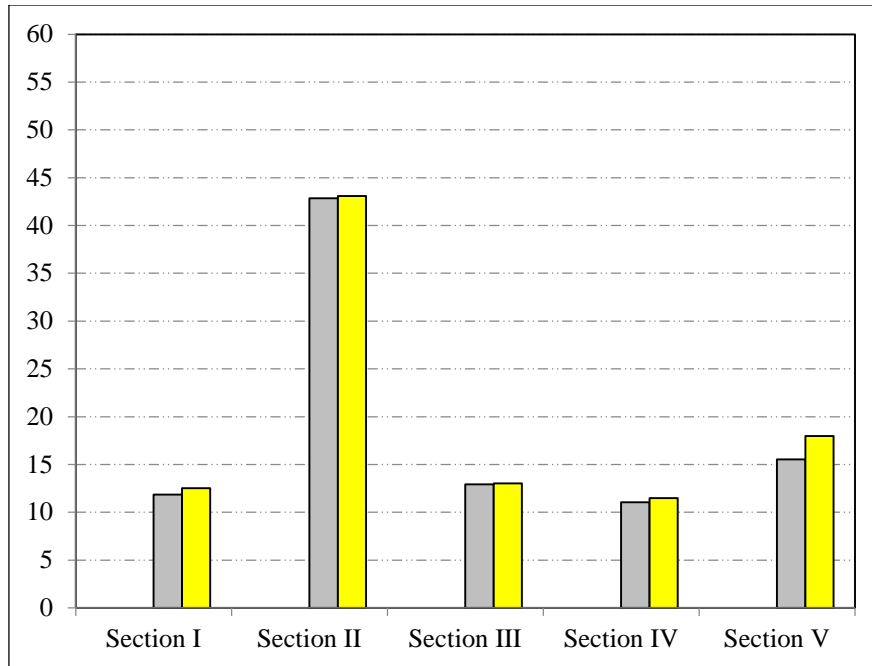


Figure 13. Comparison of mean score of exam sections by site with Charlotte in purple, Collier in teal, Thomas Edison in gray, and FSW Online in yellow.

A plot comparing descriptive statistics along with an ANOVA is typically done. However, since only two sites reported data, one of which was FSW Online, these results are exhibited in Section 3.1.2.2 above.

### 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 15. All sections exhibit peaks above 90%. Sections II and IV exhibits a bimodality with a secondary peak at 60-69% for both. Section II also continues to exhibit more widely distributed scores. Section II exhibits a peak of 24% at  $\geq 90\%$  while other sections range from 34% to 47%.

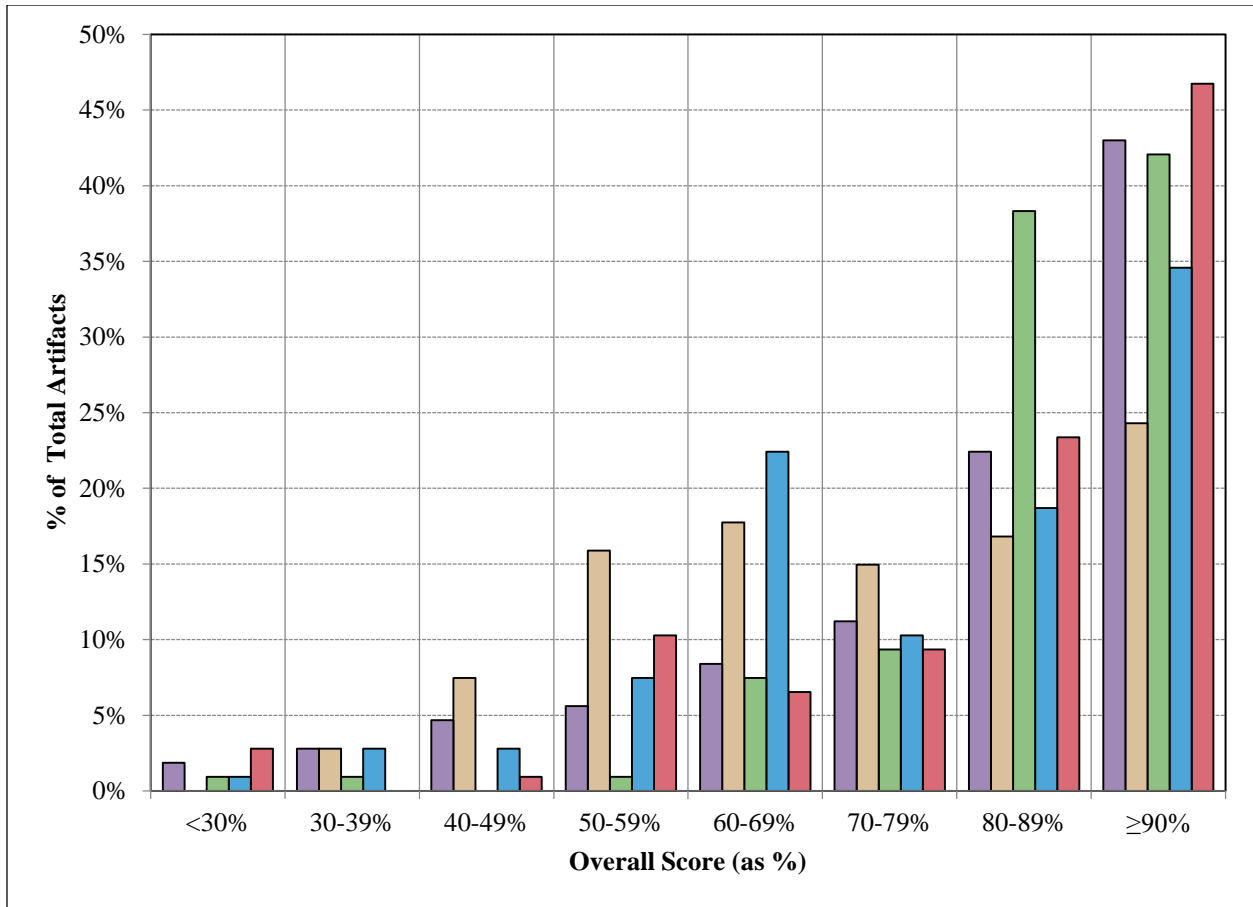


Figure 14. Histogram of SPN 1120 exam sections scores for spring 2018. Purple – Section I, Brown – Section II, Green – Section III, Blue – Section IV, and Red – Section V.

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 16). 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 16 shows that Section II remains the lowest performing compared to other sections between the ranges of 65-100%. For example, in the 75-79% range, the mean score for Section II is 69%, while the other four sections range from 74-83%. This disparity is not present at 60-64% and below. Additionally, Section III is over performing at the lowest overall scores. At the 50-54% range, the Section III mean score is 63%, whereas other sections range from 46%-52%.

	Section I	Section II	Section III	Section IV	Section V	Scale
≥ 95%	99%	94%	99%	98%	97%	100%
90-94%	94%	88%	96%	86%	93%	90%
85-89%	92%	75%	91%	87%	92%	80%
80-84%	81%	77%	88%	80%	84%	70%
75-79%	82%	69%	78%	74%	83%	60%
70-74%	73%	66%	80%	68%	78%	50%
65-69%	67%	57%	75%	64%	68%	40%
60-64%	51%	50%	79%	62%	62%	30%
55-59%	44%	50%	69%	58%	64%	
50-54%	50%	46%	63%	52%	48%	
< 50%	31%	41%	73%	26%	28%	

Figure 15. (Top) 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 17. 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). Spring 2018 exhibits the highest Section I, II, and IV since Spring 2014. Sections II and IV are the second highest mean scores since the beginning of the longitudinal study in Fall 2013.

	Section Max	Fall 2013 n=58	Spring 2014 n=90	Fall 2014 n=93	Spring 2015 n=73	Fall 2015 n=122	Spring 2016 n=141	Fall 2016 n=240	Spring 2017 n=89	Fall 2017 n=266	Spring 2018 n=107
Section I (Oral)	15	12.1	12.4	11.8	11.5	10.8	11.5	11.9	11.9	11.9	12.0
Section II (Reading)	60	40.4	45.1	40.5	39.8	41.8	38.4	39.8	41.0	41.1	43.7
Section III (Reading)	15	12.0	13.2	12.8	12.8	12.4	13.3	12.8	13.0	12.9	12.8
Section IV (Written)	15	10.6	11.7	11.1	10.9	11.5	10.5	10.9	11.0	11.5	11.6
Section V (Written)	20	16.2	16.8	16.5	16.4	15.2	15.7	15.2	16.7	15.4	16.4
<b>Combined (Overall)</b>	<b>125</b>	<b>91.3</b>	<b>99.2</b>	<b>92.8</b>	<b>91.4</b>	<b>91.6</b>	<b>89.5</b>	<b>90.6</b>	<b>93.6</b>	<b>92.8</b>	<b>96.6</b>

Table 4. Comparison of mean scores for SPN 1120 for fall 2013 through spring 2018.

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 consistently exhibits the highest mean scores over time ranging from 80% to 89%. Section II is consistently the lowest over time ranging from 64% to 75%.

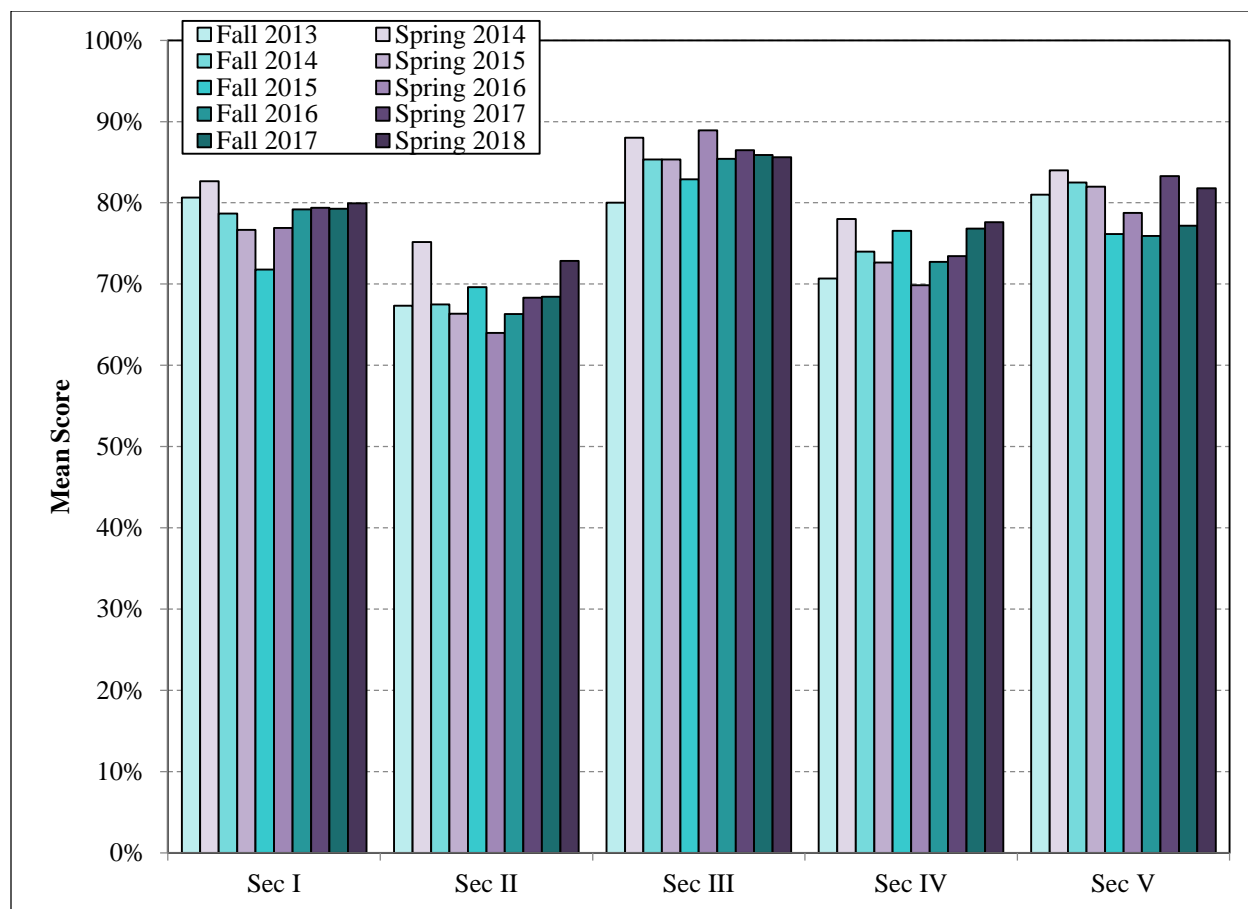


Figure 16. Comparison of mean scores (as percentage) for SPN 1120 through time from fall 2013 through spring 2018.

### 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 2018, 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 summer 2018 on the Charlotte, Collier, and Thomas Edison campuses. Results of the cumulative study are shown in Figure 17.

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 17). Results exhibit achievement levels based on previous instructor spanning as low as 11.7/30 (up from 11.5/30 following spring 2018) to as high as 21.9/30



(down from 22.2/30 following spring 2018). This is a substantially decreased range from the earliest studies where the range was 7.2/30 to 25.3/30.

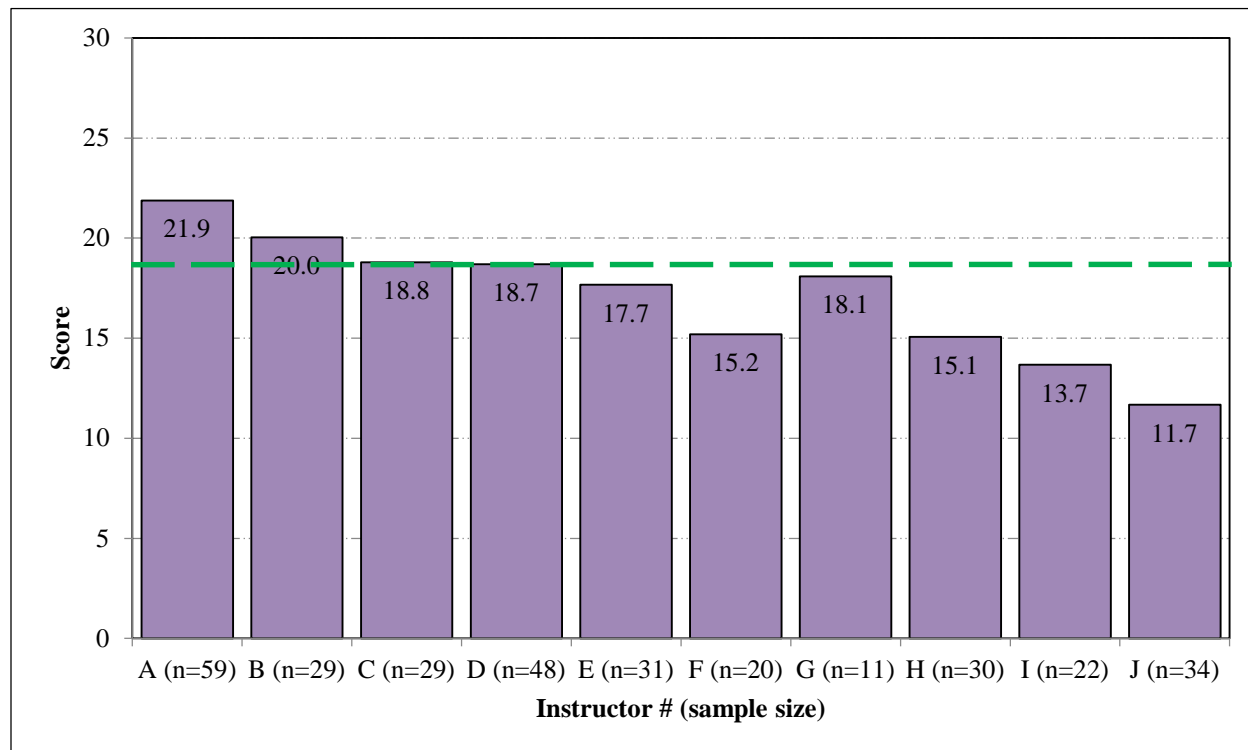


Figure 17. Comparison of achievement in entrance skills study assessment by instructor. Green dashed line denotes fall 2017 mean score for Section II of SPN 1120 (section of exam the entrance skills assessment is based).

### 3.3 SPN 1121

#### 3.3.1 Descriptive Statistics, Learning Outcomes, and Learning Objectives

Twelve sections of SPN 1121 Beginning Spanish II were offered in spring 2018. Of those, artifacts were collected from a common final from 8 of 12 sections. A total of 224 students were enrolled in SPN 1121. Of those, 136 artifacts were collected representing a sample size of 61%.

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 6 below.

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 78% of artifacts score 70% or higher in the oral competency exam section (Section I) (Table 5, Figure 18). 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 48% of artifacts scored 70% or higher in Section II and 51% 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 partially met. Results exhibit 49% of artifacts scored 70% or higher in Section V and 87% of artifacts scored 80% or higher in Section VII.

n = 136	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	78%	48%	65%	69%	49%	51%	87%	
Mean (as %)	81%	62%	72%	74%	70%	68%	84%	73%
Median (as %)	87%	67%	78%	80%	67%	73%	90%	77%
Section Score Max Possible	15	15	40	15	12	15	20	132
Section Mean	12.2	9.3	29.0	11.2	8.4	10.1	16.7	96.9
Section Median	13	10	31	12	8	11	18	101

Table 5. Percentage of student achievement level as per SLOs (SPN 1121).

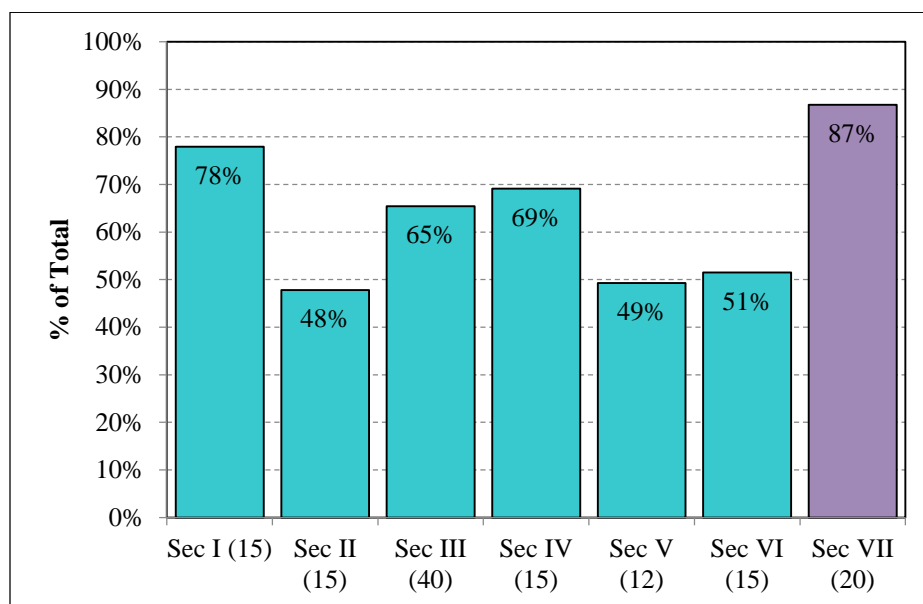


Figure 18. SLO achievement for SPN 1121 by exam section (Percentage of student achievement level as per SLOs). Purple denotes having met objective.

### 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 spring 2018 so no comparison study between dual enrollment and traditional courses could be completed.

#### 3.3.2.2 Online to Traditional Comparison

During the spring 2018 semester, three online sections reported data. From those course sections, 41 total online artifacts were collected from SPN 1121 and 95 traditional artifacts were collected from SPN 1121. A comparison of basic statistics is provided in Table 6. Online artifacts mean scores are 3.4 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 11%. For a graphical representation of this see Figure 19.

df = 134	
Online mean	71.0
Online standard deviation	19.60
Traditional mean	74.4
Traditional standard deviation	16.12
Effect size	-0.16
p-value	0.339

Table 6. Comparison of mean scores (as %) for online and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

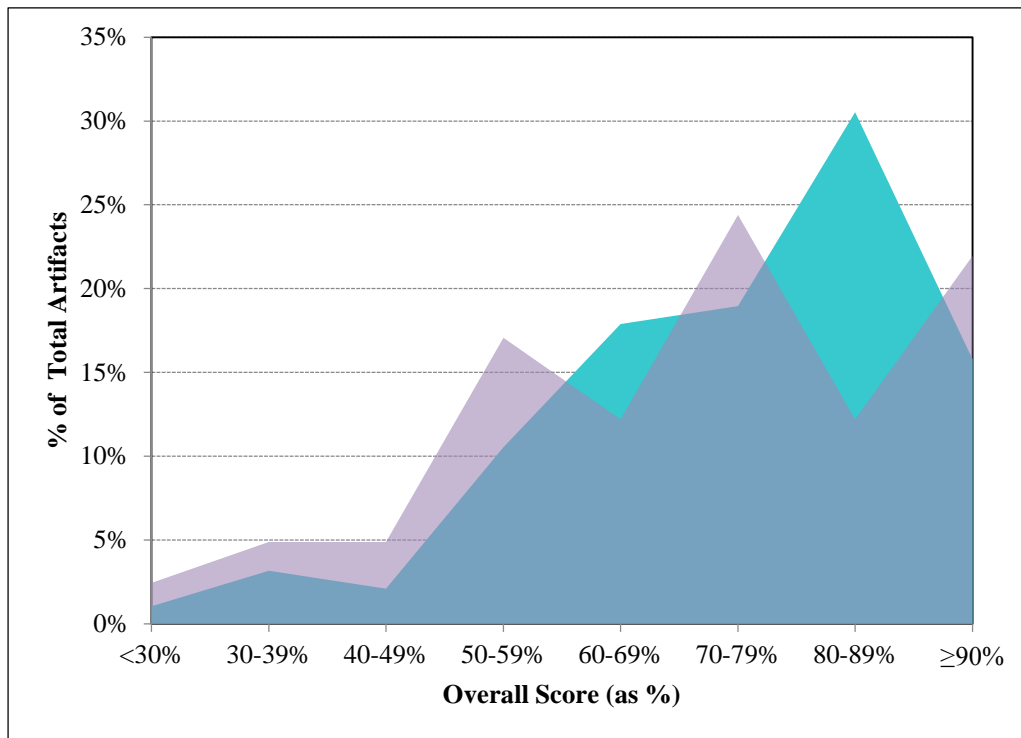


Figure 19. Score distribution for online (purple) and traditional (aqua) artifacts of SPN 1121.

### 3.3.2.3 Comparison by Campus/Site

Since the only two sites in which courses were offered was Thomas Edison (Lee) and FSW Online, results of this comparison are exhibited in 3.3.2.2 (see 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 20. All sections exhibit peaks centered on  $\geq 90\%$ . Sections II, V, and VI exhibit a bimodality. Section II exhibits a secondary peak at  $< 30\%$ . Section V exhibits a secondary peak at 60-69%. And Section VI exhibits a secondary peak centered on 70-79%.

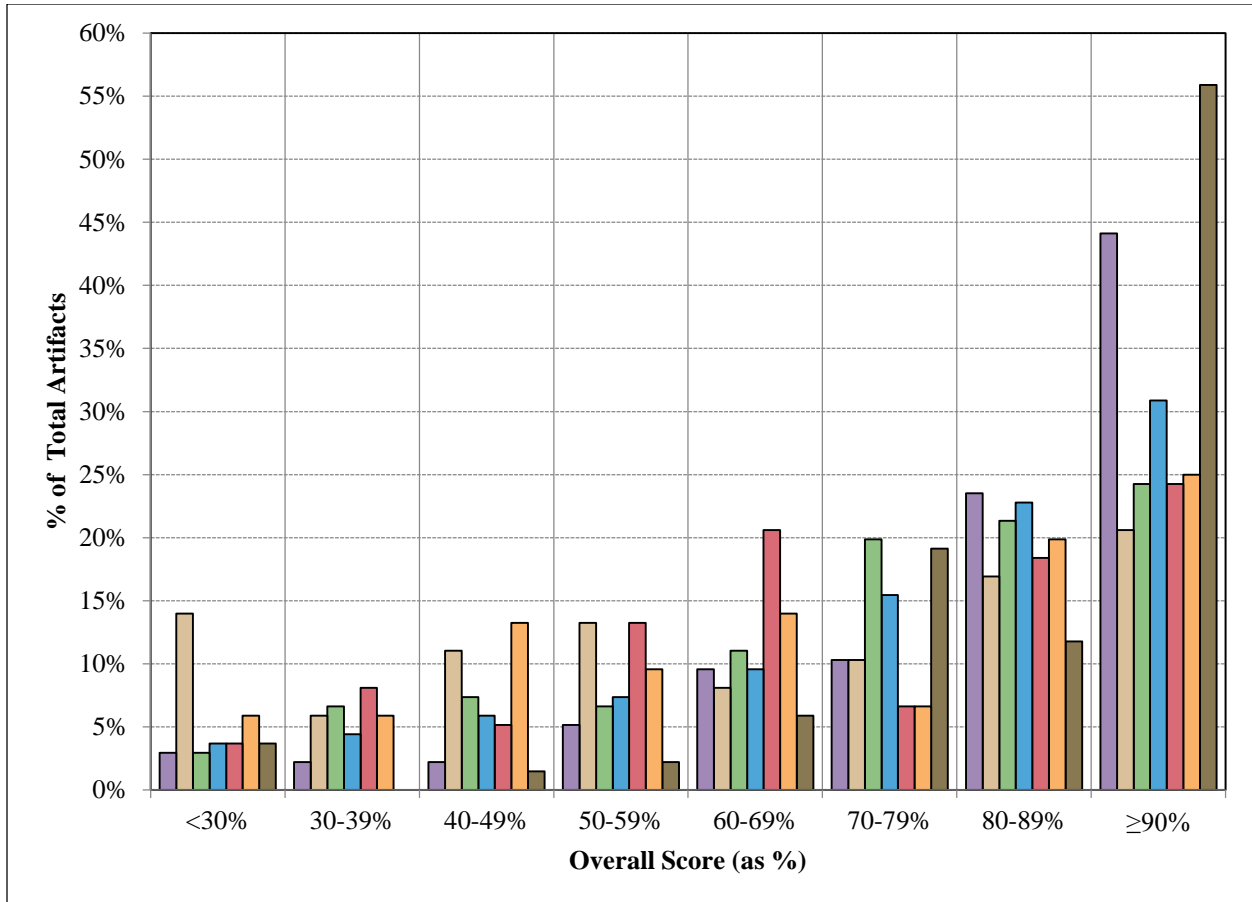


Figure 20. Histogram of SPN 1121 exam sections scores for spring 2018. Purple – Section I, Brown – Section II, Green – Section III, Blue – Section IV, Red – Section V, Orange – Section VI, and Black – Section VII.

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 21). 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 21 shows several trends. First, Sections I and VII exhibit strong performance at lower overall scores. For example, in the 70-74% range, Section I and VII exhibit scores of 82% and 90%, respectively, while other sections range from 55%-77%. Second, Sections II and VI exhibit the weakest performance at higher scores. For example, in the range of 80-84%, Sections II and VI exhibit scores of 75% and 72%, while other sections range from 76%-92%. And third, and very low overall scores, Section exhibits substantially lower scores ultimately contributing to the very low overall

scores. For example, at scores of < 50%, Section II exhibits a score of 12%, while other sections range from 27%-52%.

	Section I	Section II	Section III	Section IV	Section V	Section VI	Section VII	Scale
≥ 95%	98%	95%	95%	98%	97%	99%	99%	100%
90-94%	96%	90%	89%	89%	91%	90%	95%	90%
85-89%	92%	84%	80%	88%	83%	89%	93%	80%
80-84%	88%	75%	82%	88%	76%	72%	92%	70%
75-79%	89%	65%	69%	76%	80%	68%	90%	60%
70-74%	82%	55%	77%	73%	65%	63%	90%	50%
65-69%	81%	49%	62%	77%	62%	57%	77%	40%
60-64%	71%	49%	59%	63%	50%	67%	75%	30%
55-59%	69%	40%	67%	55%	53%	49%	70%	
50-54%	63%	36%	55%	58%	49%	42%	67%	
< 50%	47%	12%	39%	30%	42%	27%	52%	

Figure 21. (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 7 and Figure 22). 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). Spring 2018 exhibits the highest score on record for Section VII. Spring 2018 also exhibits the second highest score on record for Section I. Because some exam sections have different maximum scores (12, 15, 20, and 40), to see which of the seven sections of the exam exhibits the strongest achievement it may be best to relate them in terms of percent. There is no particular section that stands out particularly high or low compared with others. Sections II, III, IV, V, VI, and VII exhibit high variability over time.

	Section Max	Fall 2013 n=10	Sp 2014 n=115	Fall 2014 n=25	Sp 2015 n=58	Fall 2015 n=17	Sp 2016 n=109	Fall 2016 n=42	Sp 2017 n=140	Fall 2017 n=45	Sp 2018 n=136
Section I (Oral)	15	11.5	12.3	11.9	12.2	11.5	10.7	11.5	11.8	10.6	12.2
Section II (Reading)	15	9.5	9.6	10.7	10.6	10.6	9.5	9.8	8.6	9.1	9.3
Section III (Read/Write)	40	34.2	32.3	30.0	31.1	29.9	30.8	25.9	29.2	27.3	29.0
Section IV (Read/Write)	15	9.5	11.4	10.6	11.4	11.3	10.5	10.7	10.3	11.8	11.2
Section V (Writing)	12	7.5	5.7	9.5	8.5	8.1	8.3	7.2	7.9	8.1	8.4
Section VI (Reading)	15	9.6	10.3	11.6	10.7	9.9	9.0	9.7	8.7	10.2	10.1
Section VII (Writing)	20	14.2	15.4	16.1	16.4	16.4	14.6	15.6	15.1	15.7	16.7
<b>Combined (Overall)</b>	<b>132</b>	<b>96.0</b>	<b>97.0</b>	<b>100.5</b>	<b>100.9</b>	<b>97.7</b>	<b>93.5</b>	<b>90.3</b>	<b>91.6</b>	<b>92.8</b>	<b>96.9</b>

Table 7. Comparison of mean scores for SPN 1121 for fall 2013 through spring 2018.

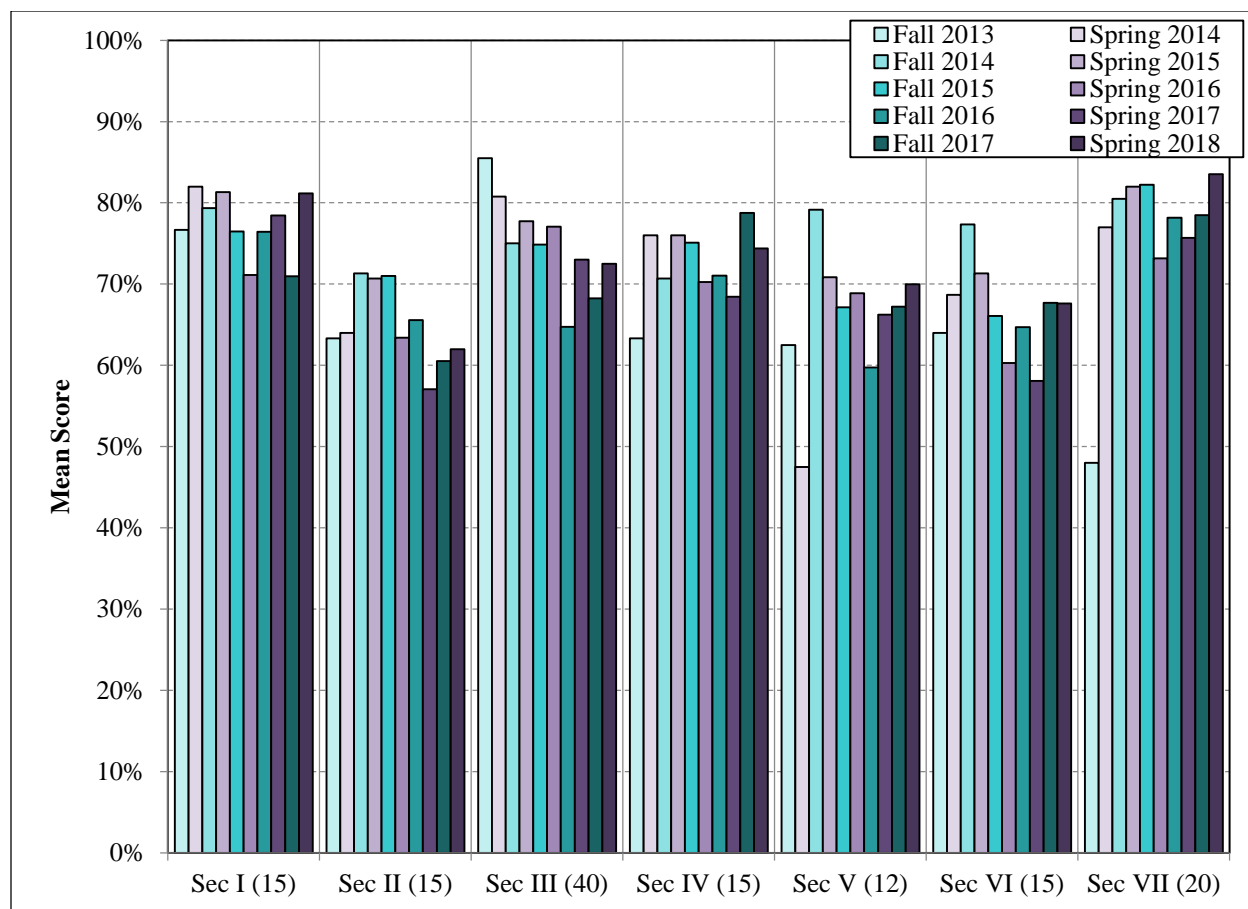


Figure 22. Comparison of mean scores (as percentage) for SPN 1121 through time from fall 2013 through spring 2018.

## 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 spring 2018 assessment.

### 4.1 FRENCH

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

1. A newly developed assessment was piloted using two sections of FRE 1120 accounting for 29 artifacts, or 47% 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.

2. In a study of Learning Outcome (LO) achievement, results exhibit mean scores across the five LOs ranging from 2.0 to 2.9. The highest scoring area is LO 1, with a mean score of 2.9/3.0. The lowest is LO 5 at 2.0/3.0. 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. Results would suggest some level increased difficulty/complexity, although this is up to the interpretation of French faculty.
3. No comparison of dual enrollment (concurrent) to traditional artifacts was completed because no dual enrollment sections were offered during spring 2018.
4. While an online section of the course was offered in spring 2018, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.
5. While courses were offered at two locations in spring 2018, Thomas Edison (Lee) and FSW Online, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.
6. In a study of score distribution by section, scores are centered on 14/15 with a strong negative skew meaning scores are tending strongly towards higher values. In total, 69% of artifacts score 13/15 or higher and 9% total 8/15 or lower.
7. In a study of achievement based on disposition survey question response, results for LOs 1, 2, 3, and 5 yield little to conclude as achievement based on "Strongly Agree" and "Agree" are fairly similar. In LO 4, those answering "Strongly Agree" outperform those answer "Agree" by as much as 27% in the case of LO 4b. While results vary wildly with those answering "Disagree," the sample size is never higher than n=2, and so interpretation is limited.
8. In a study of mean score based on disposition survey question response, as with #7 above, LO 4 shows substantial differences in achievement between those answering "Strongly Agree" and those answering "Agree." In LO 4b, mean scores drop from 1.7/3 for those answering "Strongly Agree" to 1.4/3 for those answering "Agree."
9. In an additional study of achievement based on disposition survey question response, 3 of 29 students report strong to very strong confidence in their skillset and yet did not achieve results comparable to that confidence. The highest of those three scoring 8/15, or 53%. Similarly, 1 of 29 students self-reports a lack of confidence despite scoring 13.5/15, or 90%.
10. The assessment for FRE 1120 is in its first pilot phase for spring 2018. As further data is collected in coming terms, this section will track achievement through time and highlight strengths, weaknesses and any long term trends.

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

1. A newly developed assessment was piloted using one sections of FRE 1121 accounting for 20 artifacts, or 30% 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.

2. In a study of Learning Outcome (LO) achievement, results exhibit mean scores across the four LOs ranging from 1.3 to 1.9, somewhat lower than the 2.0-2.9 range in FRE 1120. The highest scoring area is LO 2, with a mean score of 1.9/3.0. The lowest is LO 4 at 1.3/3.0.
3. No comparison of dual enrollment (concurrent) to traditional artifacts was completed because no dual enrollment sections were offered during spring 2018.
4. While an online section of the course was offered in spring 2018, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.
5. While courses were offered at three locations in spring 2018, Thomas Edison (Lee), FSW Online, and offsite, the assessment was in pilot phase. As a result, the study's focus was solely on course sections in which the assessment developer served as instructor so no comparison could be completed.
6. In a study of score distribution by section, scores are centered on 10/12 with a moderate negative skew meaning scores are tending somewhat towards higher values. In total, 45% of artifacts score 10/12 or higher and 5% total 4/12 or lower.
7. In a study of achievement based on disposition survey question response, results for this study are expected to yield some valuable results, as shown in FRE 1120. Note that most students self-reported "Agree," and so analysis by this type of disaggregation does not yield results as fruitful as those seen in FRE 1120. Results from LO 1a appear to exhibit an indirect relationship (higher confidence coincides with lower results) although those answering "Strongly Agree" is limited to n=5.
8. In a study of mean score based on disposition survey question response, as with #7 above, most students self-reported "Agree," and so analysis by this type of disaggregation does not yield results as fruitful as those seen in FRE 1120. Results from LO 1a again appear to exhibit an indirect relationship (higher confidence coincides with lower results) although those answering "Strongly Agree" is limited to n=5.
9. In an additional study of achievement based on disposition survey question response, most students exhibit limited confidence compared with FRE 1120, and this correlates well with lower scores on the achievement portion. Also, 2 of 20 students self-report a lack of confidence in their skillset. One achieved a 10.5/12 while the other a 6.5/12.
10. The assessment for FRE 1121 is in its first pilot phase for spring 2018. As further data is collected in coming terms, this section will track achievement through time and highlight strengths, weaknesses and any long term trends.

## 4.2 SPANISH

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

1. Achievement of 80% of artifacts scoring 70% or better (SLO 1): Achievement was nearly met as results exhibit 77% of artifacts score 70% or higher in the oral competency exam section (Section I).
2. Achievement of 80% of artifacts scoring 70% or better (SLO 2): Achievement was partially met. Results exhibit 56% of artifacts scored 70% or higher in Section II and 90% of artifacts scored 70% or higher in Section III.
3. Achievement of 80% of artifacts scoring 70% or better (SLO 3): Achievement was nearly met. Results exhibit 64% of artifacts scored 70% or higher in Section IV and 79% of artifacts scored 70% or higher in Section V.



4. No dual enrollment (concurrent) sections of the course were run during spring 2018 so no comparison study between dual enrollment and traditional courses could be completed.
5. In a comparison of online to traditional artifacts, online artifacts score 1.8 points higher than traditional. Results were not statistically significantly different.
6. Only two sites reported data, one of which was FSW Online, these results are exhibited in #5 above.
7. In a study of score distribution by section, sections exhibit peaks above 90%. Sections II and IV exhibits a bimodality with a secondary peak at 60-69% for both. Section II also continues to exhibit more widely distributed scores. Section II exhibits a peak of 24% at  $\geq 90\%$  while other sections range from 34% to 47%.
8. 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 65-100%. For example, in the 75-79% range, the mean score for Section II is 69%, while the other four sections range from 74-83%. This disparity is not present at 60-64% and below. Additionally, Section III is over performing at the lowest overall scores. At the 50-54% range, the Section III mean score is 63%, whereas other sections range from 46%-52%.
9. In a longitudinal study of data distribution through time, Spring 2018 exhibits the highest Section I, II, and IV since Spring 2014. Sections II and IV are the second highest mean scores since the beginning of the longitudinal study in Fall 2013.
10. In a study assessing student skills and retention of materials from SPN 1120 prior to beginning SPN 1121, results exhibit achievement levels based on previous instructor spanning as low as 11.7/30 (up from 11.5/30 following spring 2018) to as high as 21.9/30 (down from 22.2/30 following spring 2018). This is a substantially decreased range from the earliest studies where the range was 7.2/30 to 25.3/30.

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

1. Achievement of 80% of artifacts scoring 70% or better (SLO 1): Achievement was nearly met as results exhibit 78% of artifacts score 70% or higher in the oral competency exam section (Section I).
2. Achievement of 80% of artifacts scoring 70% or better (SLO 2): Achievement was not met. Results exhibit 48% of artifacts scored 70% or higher in Section II and 51% of artifacts scored 70% or higher in Section VI.
3. Achievement of 80% of artifacts scoring 70% or better (SLO 3): Achievement was partially met. Results exhibit 49% of artifacts scored 70% or higher in Section V and 87% of artifacts scored 80% or higher in Section VII.
4. No dual enrollment (concurrent) sections of the course were run during spring 2018 so no comparison study between dual enrollment and traditional courses could be completed.
5. In a comparison of online to traditional artifacts, online artifacts score 3.4 lower than traditional. Results were not statistically significantly different.
6. No cross-campus comparison was completed because only two sites offered course sections (FSW Online and Thomas Edison, a traditional site). The results of that comparison are shown in item #5 above.
7. In a study of score distribution by section, all sections exhibit peaks centered on  $\geq 90\%$ . Sections II, V, and VI exhibit a bimodality. Section II exhibits a secondary peak at  $< 30\%$ . Section V exhibits a secondary peak at 60-69%. And Section VI exhibits a secondary peak centered on 70-79%.
8. In a study of section score distribution based on overall score, several trends are present. First, Sections I and VII exhibit strong performance at lower overall scores. For example, in the 70-

74% range, Section I and VII exhibit scores of 82% and 90%, respectively, while other sections range from 55%-77%. Second, Sections II and VI exhibit the weakest performance at higher scores. For example, in the range of 80-84%, Sections II and VI exhibit scores of 75% and 72%, while other sections range from 76%-92%. And third, and very low overall scores, Section exhibits substantially lower scores ultimately contributing to the very low overall scores. For example, at scores of < 50%, Section II exhibits a score of 12%, while other sections range from 27%-52%.

9. In a longitudinal study of data distribution through time, Spring 2018 exhibits the highest score on record for Section VII. Spring 2018 also exhibits the second highest score on record for Section I.

## 5 REFERENCES

---

- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2<sup>nd</sup> 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* (2<sup>nd</sup> 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/](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.