Foreign Language Assessment Report
Spring 2020

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 spring 2020 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 Spring 2020 term, three sections of FRE 1120 were offered. An assessment first piloted in AY 2017-18 was recorded in three section of FRE 1120 accounting for 35 artifacts, or 74% 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 2.7 to 3.0 (Figure 1), compared with 2.2-2.6 in fall 2019, 2.3 to 2.7 in fall 2018. 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.
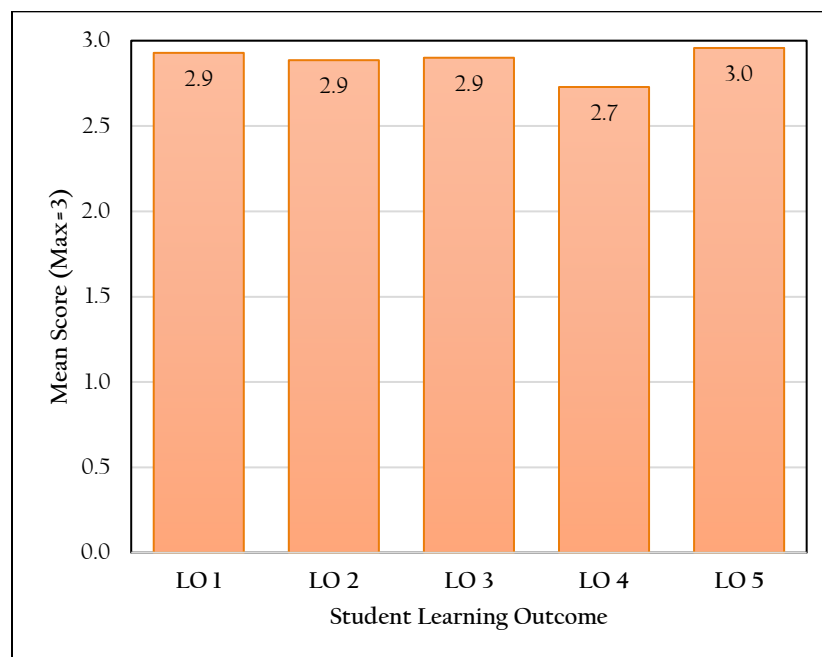

Figure 1. Mean scores by SLOs for FRE 1120.

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 spring 2020, 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 Spring 2020 semester, one online section reported data able to be used in assessment and two traditional sections reported data. From those course sections, 20 total online artifacts were collected from FRE 1120 and 15 traditional artifacts were collected from FRE 1120. Both online and traditional course sections exhibit similar artifact score distributions in the assessment at 14/15 or higher (Figure 2).
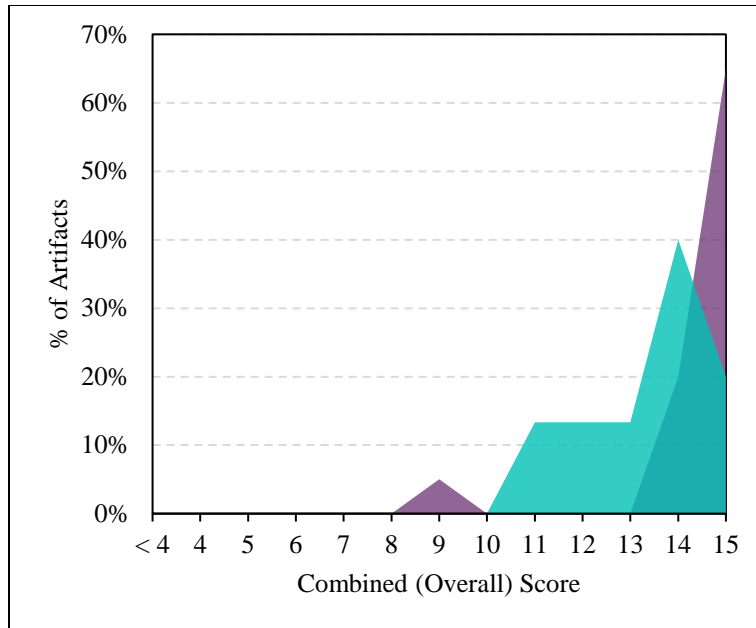
Figure 2. Score distribution for online (purple) and traditional (aqua) artifacts of FRE 1120.

2.1.2.3   Comparison by Campus/Site

Courses were only offered at two sites as comprised in the online to traditional comparison above. 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 15/15 with a strong negative skew meaning scores are tending strongly towards higher values (Starkweather, 2010). In total, 75% of artifacts score 14/15 or higher.
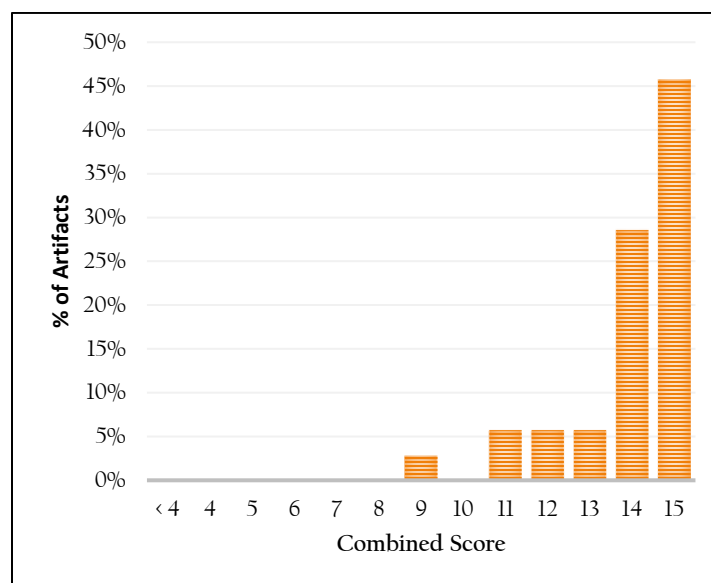

Figure 3. Distribution of combined (total) scores for assessment.

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, unlike in the past, do not yield a very clear 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. The scores appear to reflect high achievement despite self-reporting by students that they don't understand the material.
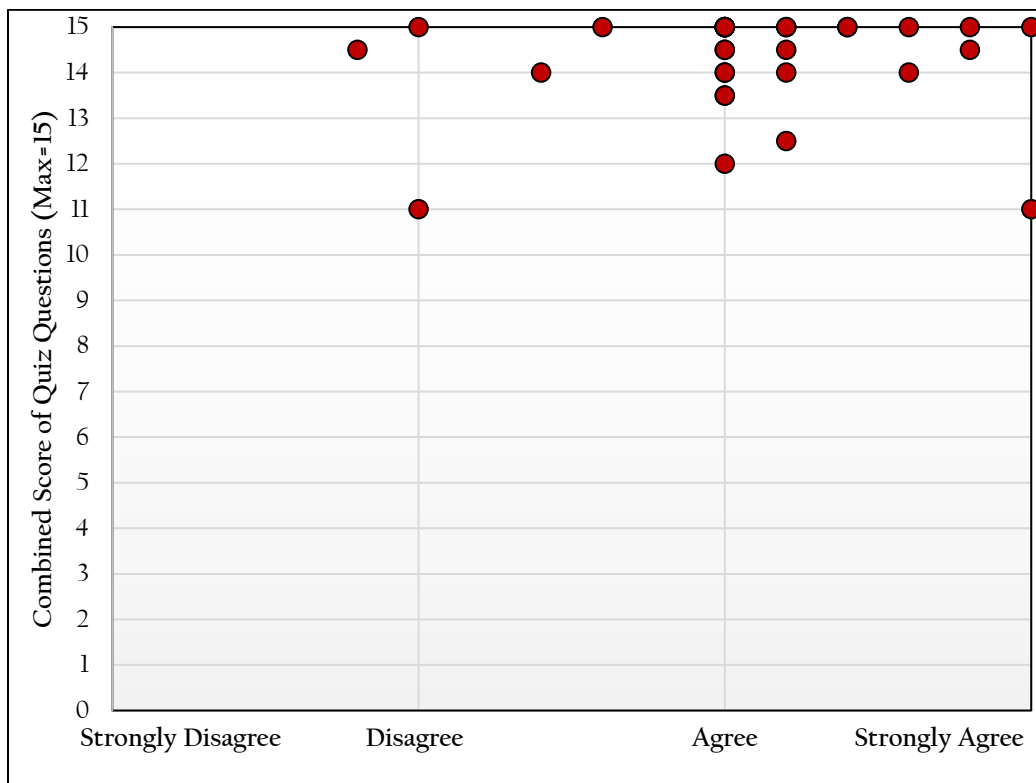

Figure 4. Scatter plot of combined (total) scores based on general response to survey questions.

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 four terms, LOs 1 through 5 for the Spring 2020 rank 1st, 2nd, 2nd, 2nd, and 1st, 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
Spring 2020 (n=35)	2.9	2.9	2.9	2.7	3.0	2.9

Table 1. Comparison of mean scores for FRE 1120 over time.

2.2   FRE 1121

During the Spring 2020 term, four sections of FRE 1121 were offered. The assessment was recorded in three sections, however, in one section results and scoring did not match that expected from the assessment plan, so no data could be utilized. The data collection accounts for 17 artifacts, or 32% of the population. The assessment utilizes a combination quiz and disposition survey (five in total in spring 2020 compared with four in previous terms). 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 four 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.

2.2.1   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
- ➢ LO 5: Future Events and Projects

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 2.1 to 2.5. 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.
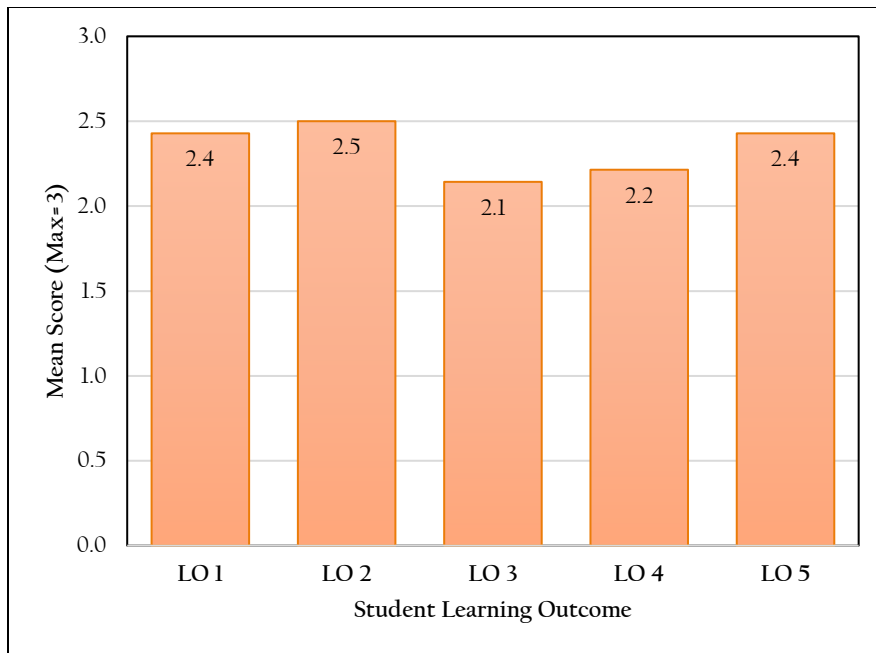
Figure 5. Mean scores by SLOs for FRE 1121.

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 spring 2020, 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 online section due to the data not matching the expected for the assessment. 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 three 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 17 achievement questions is shown below in Figure 6. Scores are centered on 14/15 with a strong negative skew meaning scores are tending strongly towards higher values (Starkweather, 2010). In total, 6% of artifacts score a perfect 15/15 and 47% scored 14/15.
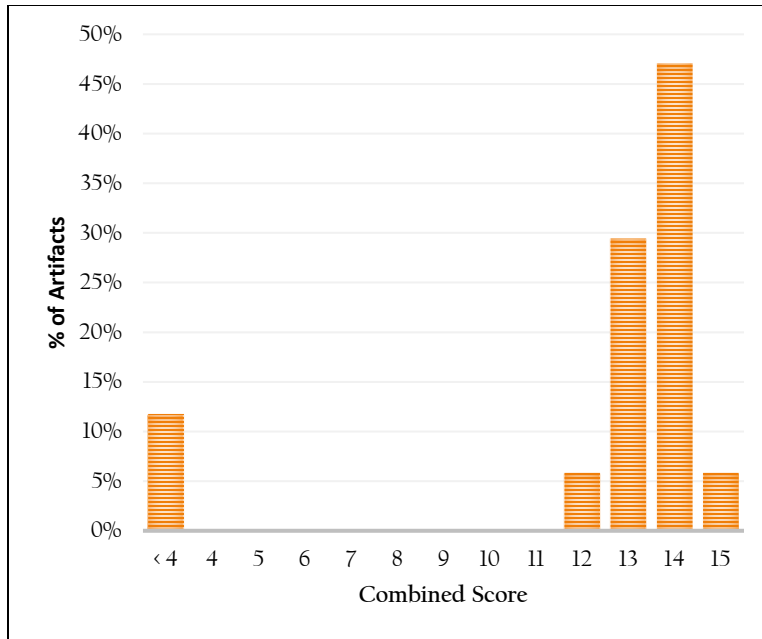
Figure 6. Distribution of combined (total) scores for assessment.

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 7 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. There appears to be the beginnings of a trend of confidence with achievement, although sample size is limited (n=17).
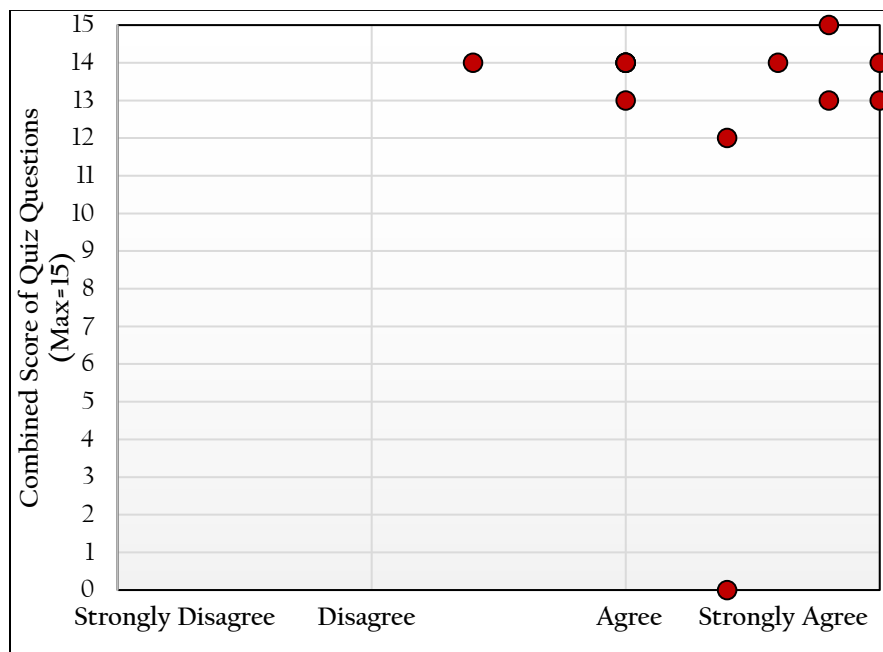

Figure 7. Scatter plot of combined (total) scores based on general response to survey questions.

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 Spring 2020 semester 10 sections of SPN 1120 *Beginning Spanish I* were offered. Of those, artifacts from a common final were collected from 10 sections. A total of 200 students were enrolled in SPN 1120. Of those, 123 artifacts were collected representing a sample size of 62% 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 70% 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 46% of artifacts scored 70% or higher in Section II and 85% 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 67% of artifacts scored 70% or higher in Section IV and 78% of artifacts scored 70% or higher in Section V. For a graphical representation of SLO achievement, see Figure 8.

n = 123	Section I (Oral)	Section II (Reading)	Section III (Reading)	Section IV (Written)	Section V (Written)	Combined (Overall)
Goal	*80% of artifacts scored ≥70% for all sections*					
% above 70%	70%	46%	85%	67%	78%	
Mean (as %)	76%	61%	86%	73%	79%	
Median (as %)	80%	65%	87%	80%	85%	
Section Score Max	15	60	15	15	20	125
Section Mean	11.4	36.7	12.8	11.0	15.8	87.7
Section Median	12	39.0	13	12	17	92

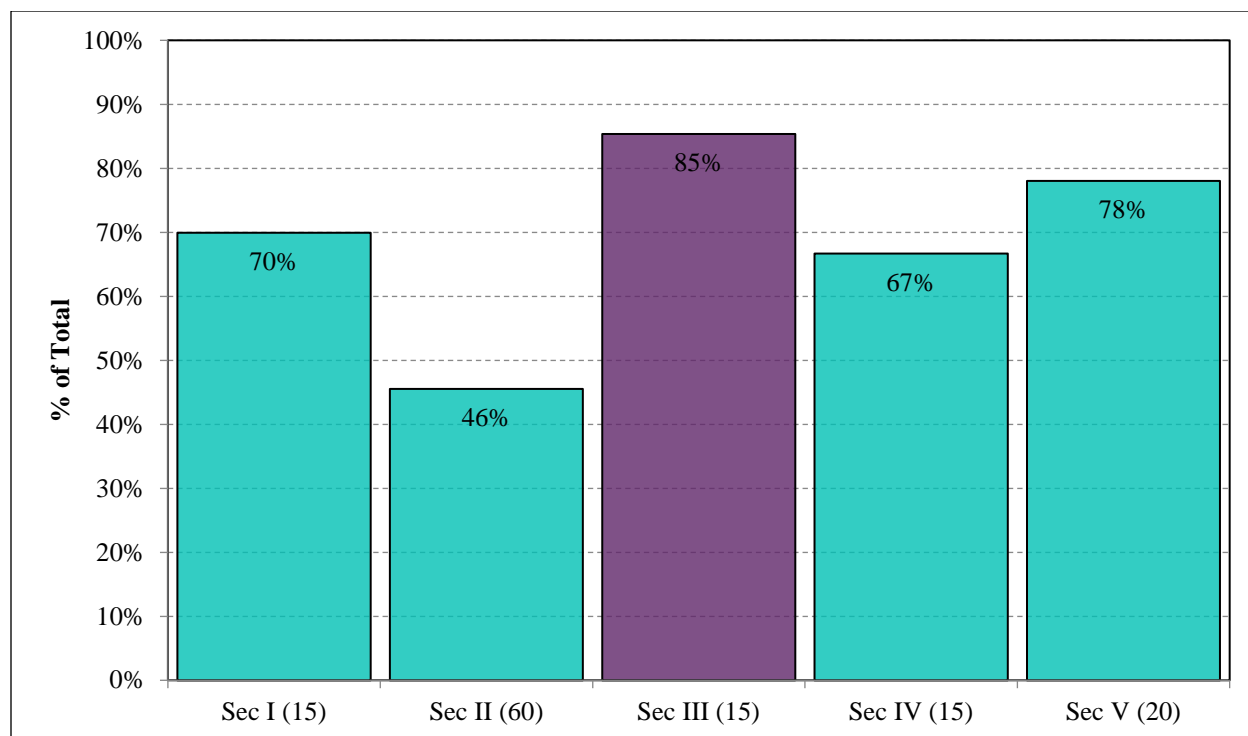*Table 2. Percentage of student achievement level as per SLOs (SPN 1120).*

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 spring 2020 so no comparison study between dual enrollment and traditional courses could be completed.

3.1.2.2   Online to Traditional Comparison

During the Spring 2020 semester, three online section reported data able to be used in assessment and eight traditional sections reported data. From those course sections, 42 total online artifacts were collected from SPN 1120 and 81 traditional artifacts were collected from SPN 1120. A comparison of basic statistics is provided in Table 3. Online artifacts mean scores are 11.9 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 large effect size. In other words, non-overlap score distribution from

online artifacts to traditional artifacts is approximately 40%. For a graphical representation of this see Figure 9.

df = 121	
Online mean	78.0
Online standard deviation	17.46
Traditional mean	66.1
Traditional standard deviation	18.54
Effect size	-0.639
p-value	0.001

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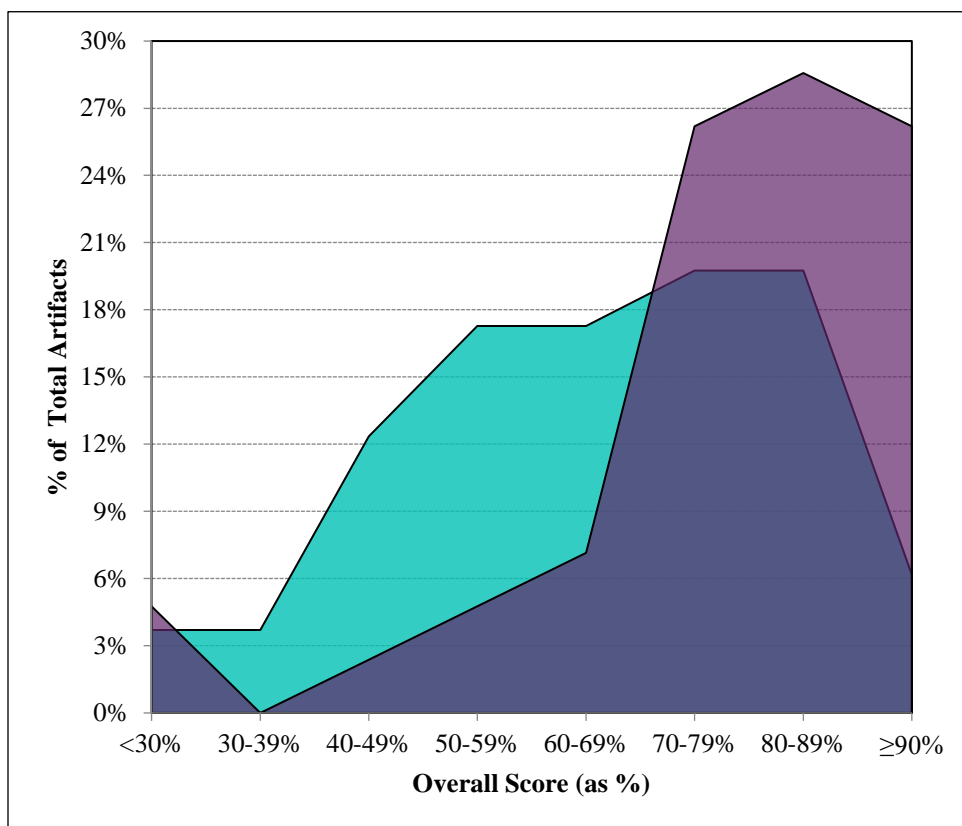

Figure 9. Score distribution for online (purple) and traditional (aqua) artifacts of SPN 1120.

3.1.2.3   Comparison by Campus/Site

Of the 123 artifacts collected from SPN 1120, 10 originated from the Charlotte campus, 0 from the Collier campus, 42 from FSW Online, and 71 from the Thomas Edison (Lee) campus. Scores by rubric dimension varied greatly across campuses. FSW Online exhibits the highest mean scores in 5 of 5 sections and the overall score. 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
Rubric Max	*15*	*60*	*15*	*15*	*20*	*125*
Charlotte	8.0	24.6	13.4	10.9	15.3	72.2
Collier	~	~	~	~	~	~
FSW Online	**12.7**	**41.2**	**14.0**	**11.9**	**17.7**	**97.5**
Thomas Edison (Lee)	11.0	35.7	12.1	10.5	14.7	84.1

Table 4. Comparison of mean scores by site for SPN 1120. Bold denotes highest mean score in that dimension among all sites.
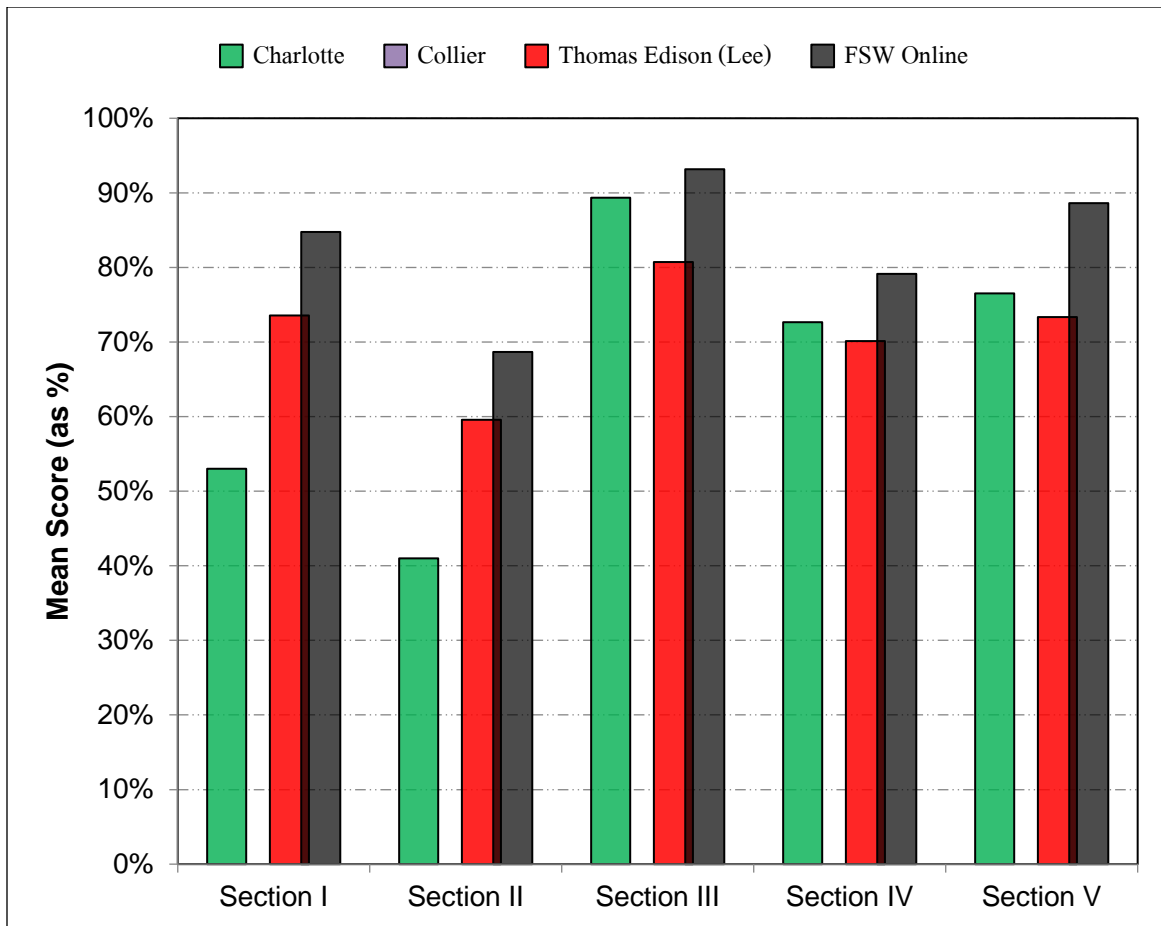
Figure 10. Comparison of mean score of exam sections by site.

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 80-89%. Section II also continues to exhibit more widely distributed scores. Section II exhibits a maximum of 8% at ≥ 90% while other sections range from 29% to 47%.

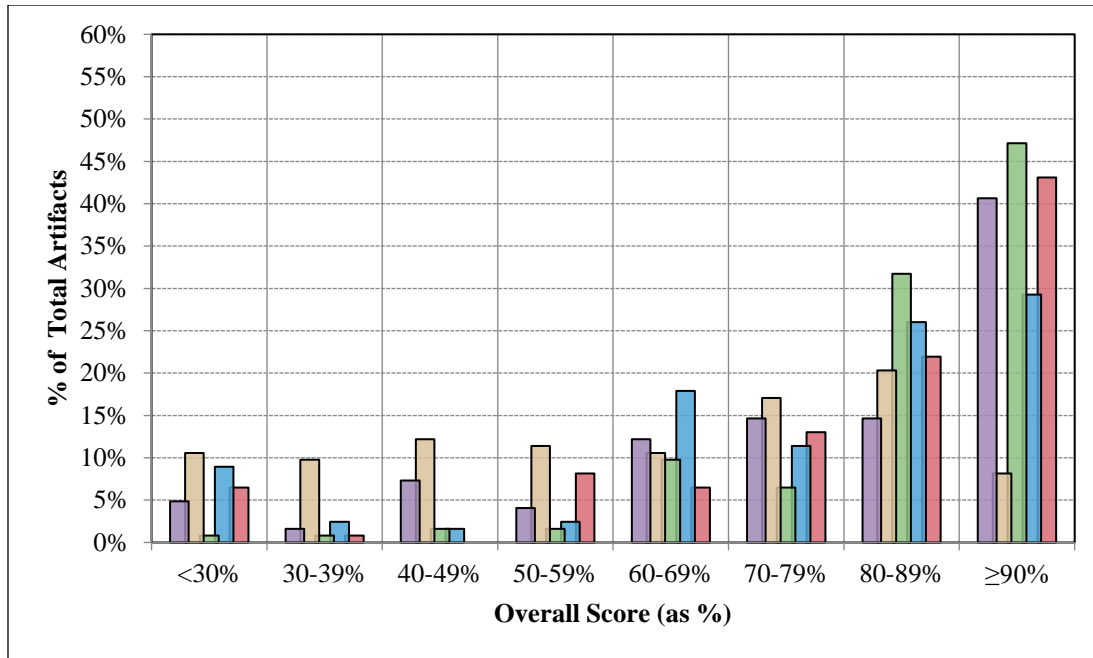*Figure 11. Histogram of SPN 1120 exam sections scores for spring 2020. Purple – Section I, Brown – Section II, Green – Section III, Blue – Section IV, and Red – Section V.*

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 80%-84%, for example, Section II exhibits a mean score of 64% while the other four sections range from 80%-92%. At the overall scoring range at 90%-94%, Section II exhibits a mean score of 84%, while the other four sections range from 90%-98%.

	Section I	Section II	Section III	Section IV	Section V		Scale
≥ 95%	100%	89%	99%	98%	100%		100%
90-94%	96%	84%	98%	90%	95%		90%
85-89%	91%	76%	94%	83%	91%		80%
80-84%	85%	64%	92%	80%	91%		70%
75-79%	73%	63%	85%	79%	85%		60%
70-74%	73%	53%	84%	78%	74%		50%
65-69%	62%	46%	79%	76%	69%		40%
60-64%	64%	40%	77%	67%	64%		30%
55-59%	27%	33%	80%	53%	85%		
50-54%	56%	41%	66%	41%	53%		
< 50%	37%	28%	64%	17%	40%		

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. Demographics and sample size vary by semester so it may be more reasonable to compare like semesters (Fall vs. Fall, Spring vs. Spring) (see http://www.fsw.edu/facultystaff/assessment/history for examples). In 14 terms, Sections I through V for the Spring 2020 term rank 12th, 14th, 6th, 7th, and 10th, 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
Spring 2020 (n=123)	11.4	36.7	12.8	11.0	15.8	87.7

Table 5. Comparison of mean scores for SPN 1120 for fall 2013 through spring 2020.

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 13 of 14 terms including the Spring 2020 term. Section II exhibits the lowest mean scores in 13 of 14 terms, spring 2019 being the only term in which it is not the lowest. For spring 2019, Section IV exhibits the lowest mean score.
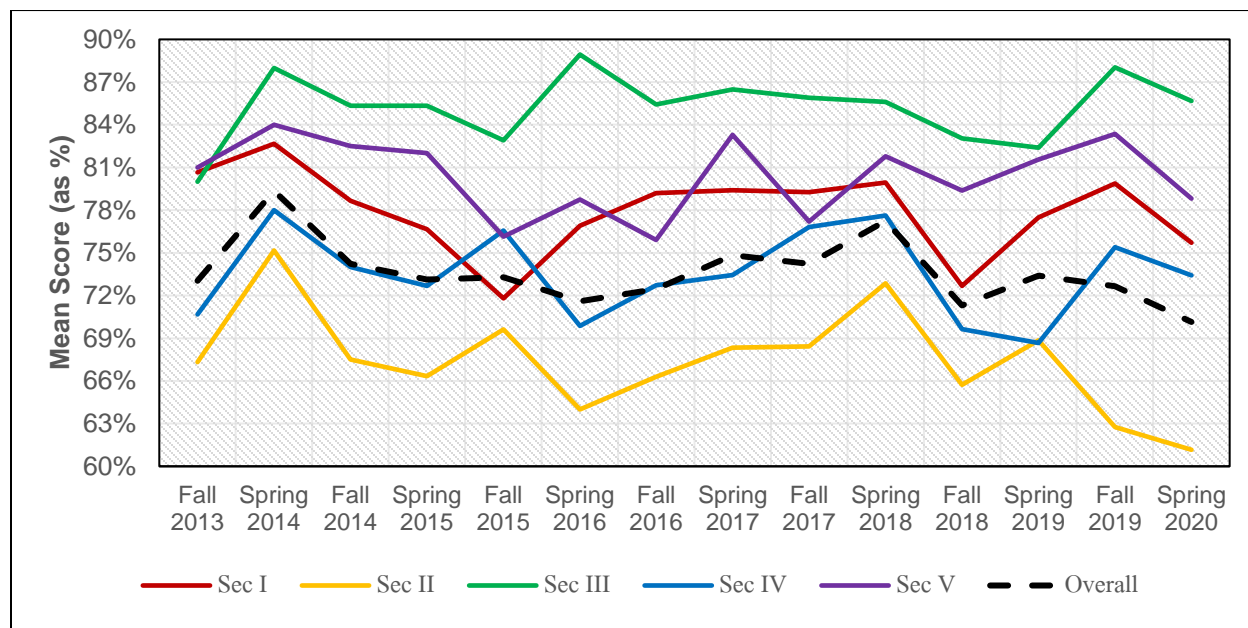

Figure 13. Comparison of mean scores (as percentage) for SPN 1120 through time from fall 2013 through spring 2020.

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. No study was conducted in summer 2020 to include in this report due to the Covid-19 pandemic. 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.
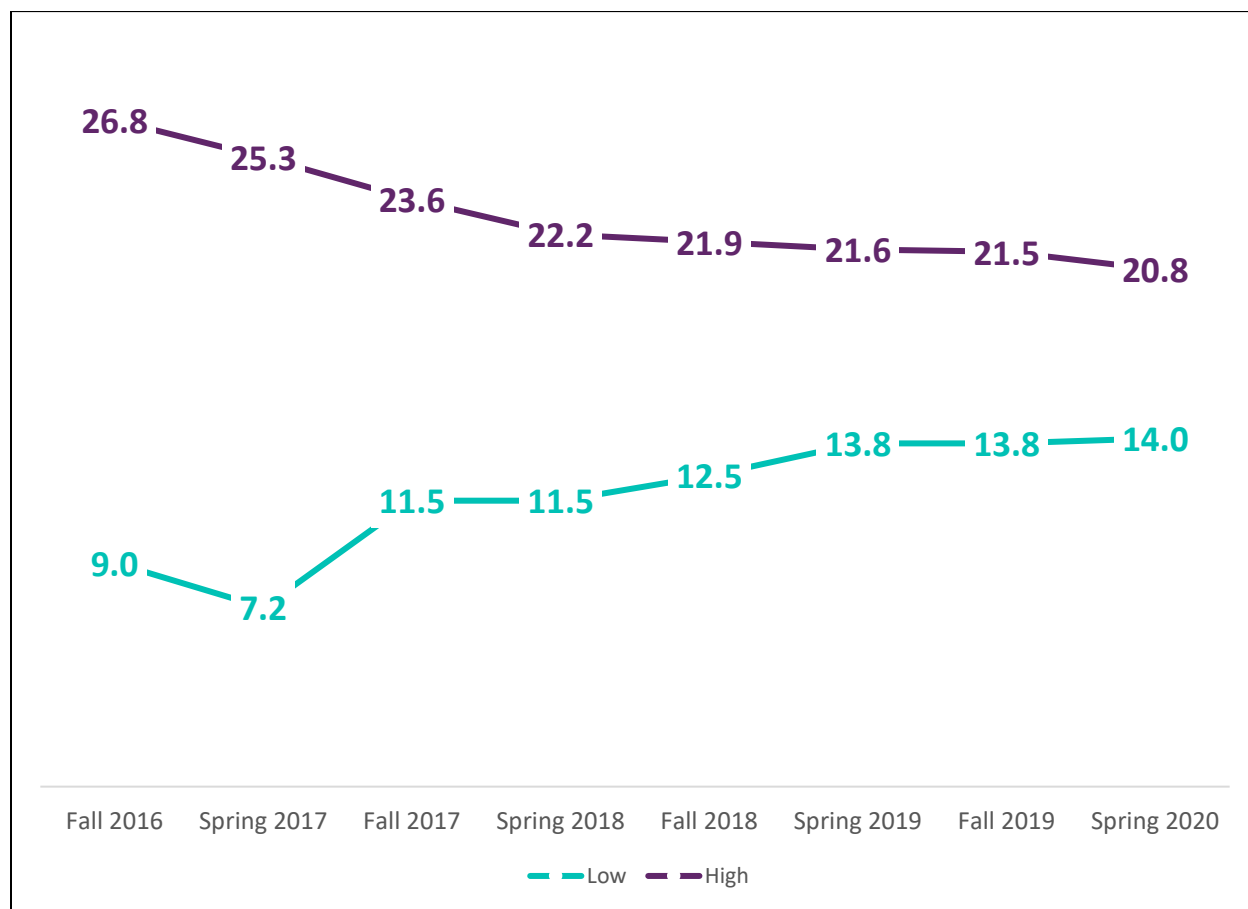

Figure 14. Comparison of achievement in entrance skills study assessment over time between highest scoring instructor (score listed is mean score) and lowest.

3.3 SPN 1121

3.3.1 Descriptive Statistics, Learning Outcomes, and Learning Objectives

Ten sections of SPN 1121 Beginning Spanish II were offered in spring 2020. Of those, artifacts were collected from a common final from 10 of 10 sections. A total of 200 students were enrolled in SPN 1121. Of those, 168 artifacts were collected representing a sample size of 84%.

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 58% 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 32% of artifacts scored 70% or higher in Section II and 48% 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 48% of artifacts scored 70% or higher in Section V and 88% of artifacts scored 80% or higher in Section VII.

n = 168	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 ≥70% for all sections*							
% above Goal	58%	32%	48%	72%	45%	48%	88%	
Mean (as %)	70%	51%	62%	76%	62%	62%	81%	
Median (as %)	77%	53%	68%	80%	67%	67%	85%	
Section Score Max Possible	*15*	*15*	*40*	*15*	*12*	*15*	*20*	*132*
Section Mean	10.5	7.7	24.9	11.5	7.4	9.3	16.3	87.4
Section Median	11.5	8	27	12	8	10	17	89.5

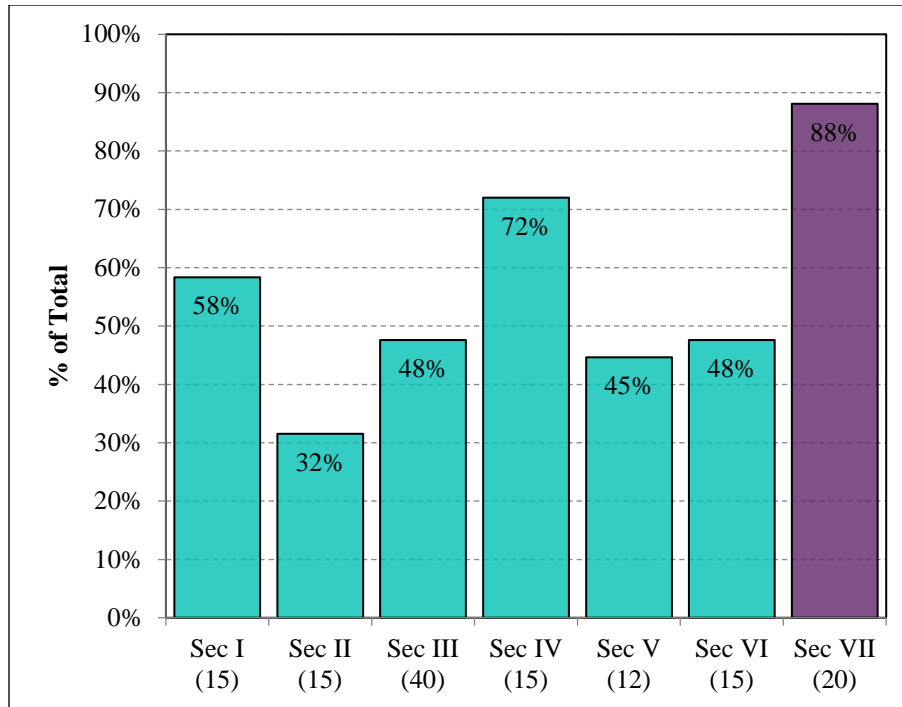*Table 6. Percentage of student achievement level as per SLOs (SPN 1121).*

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 spring 2020 so no comparison study between dual enrollment and traditional courses could be completed.

3.3.2.2 *Online to Traditional Comparison*

During the Spring 2020 semester, two online sections reported data. From those course sections, 34 total online artifacts were collected from SPN 1121 and 134 traditional artifacts were collected from SPN 1121. A comparison of basic statistics is provided in Table 7. Online artifacts mean scores are 8.8 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 5%. For a graphical representation of this see Figure 16.

df = 166	
Online mean	67.3
Online standard deviation	12.52
Traditional mean	76.1
Traditional standard deviation	19.35
Effect size	0.078
p-value	0.618

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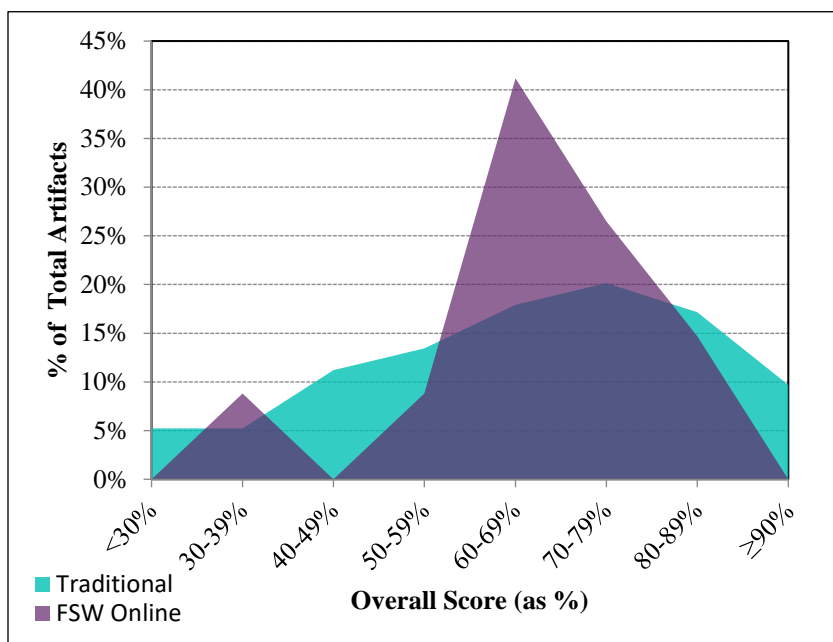

Figure 16. Score distribution for online (purple) and traditional (aqua) artifacts of SPN 1121.

3.3.2.3 Comparison by Campus/Site

Of the 168 artifacts collected from SPN 1121, 13 originated from the Charlotte campus, 26 from the Collier campus, 34 from FSW Online, and 95 from the Thomas Edison (Lee) campus. Scores by rubric dimension varied greatly across campuses. FSW Online exhibits the highest mean scores in 6 of 7 sections. A comparison of mean scores by rubric dimension is provided in Table 8 and Figure 17.

	Section I	Section II	Section III	Section IV	Section V	Section VI	Section VII	Combined Score
Rubric Max	*15*	*15*	*40*	*15*	*12*	*15*	*20*	*132*
Charlotte	8.8	6.6	10.7	8.5	6.1	9.5	14.6	64.7
Collier	9.1	9.4	26.8	11.5	7.1	7.3	13.7	84.8
FSW Online	**10.9**	**7.8**	21.9	**12.5**	**7.6**	**10.6**	**17.6**	88.9
Thomas Edison (Lee)	10.9	7.3	**27.3**	11.5	7.6	9.4	16.8	**90.7**

Table 8. Comparison of mean scores by site for SPN 1120. Bold denotes highest mean score in that dimension among all sites.
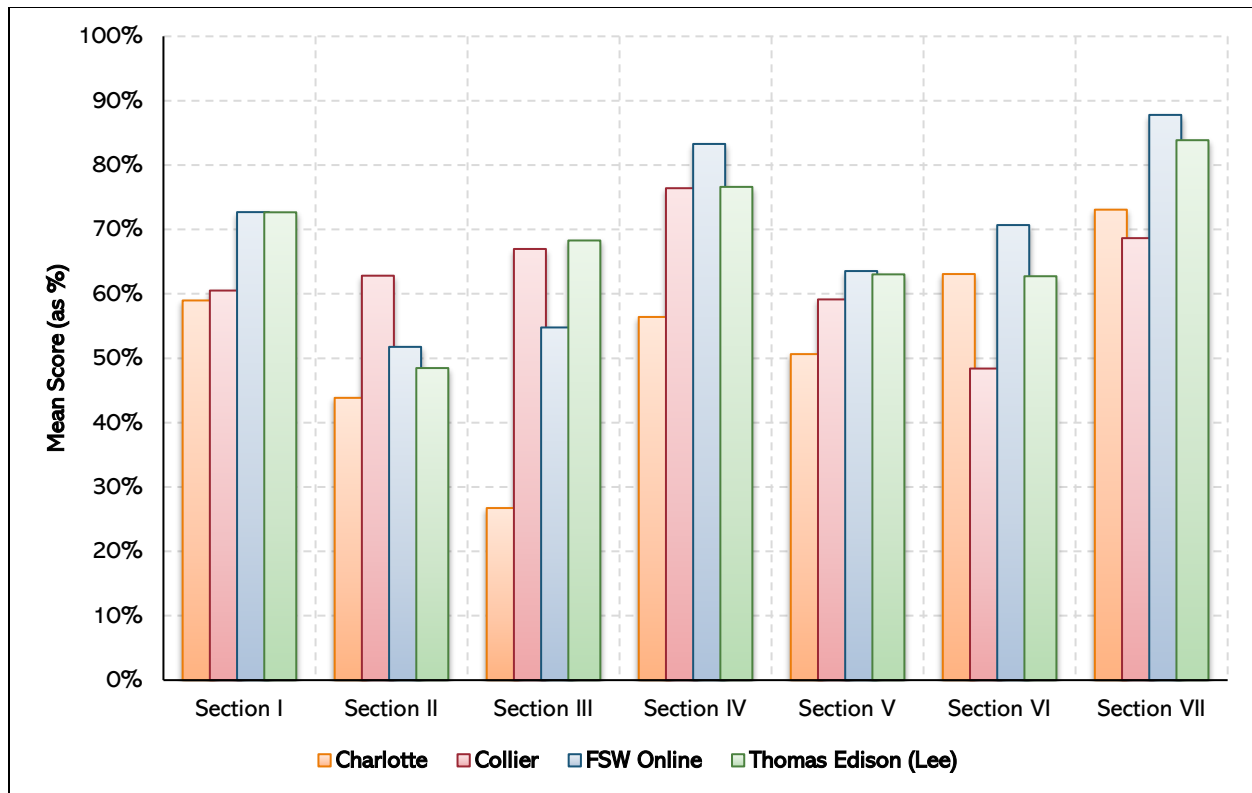
Figure 17. Comparison of mean score of exam sections by site.

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 18. Sections I, IV, V, VI, and VII exhibit peaks centered on ≥90%. Section II exhibits a peak centered at <30%. Section III exhibits a peak centered at 70-79%.
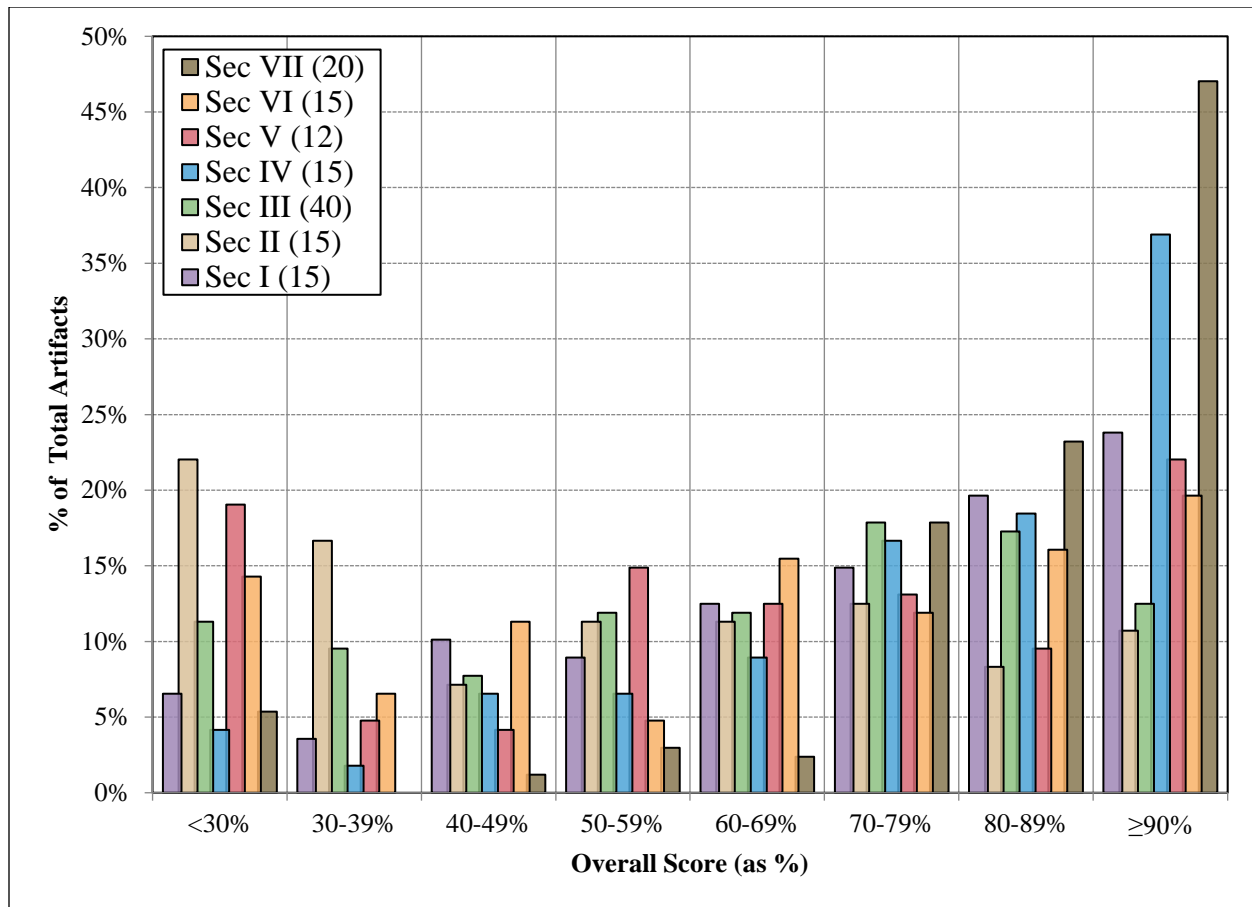
Figure 18. Histogram of SPN 1121 exam sections scores for spring 2020. Purple – Section I, Brown – Section II, Green – Section III, Blue – Section IV, Red – Section V, Orange – Section VI, and Black – Section VII.

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 19). 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 19 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 42%-72%. Second, Sections II, III, and V, exhibit the weakest performance at mid-range scores. For example, in the range of 75-79%, Sections II, III, and V exhibit scores of 65-70%, while other sections range from 77%-87%. And third, at high scores (90 or higher) all sections score fairly evenly, although Sections II and VI are somewhat lagging. In the ≥95% range, Sections II and VI exhibit mean of 86% and 84%, respectively, while all other sections range from 90-97%.

	Section I	Section II	Section III	Section IV	Section V	Section VI	Section VII		Scale
≥ 95%	96%	100%	93%	100%	97%	93%	100%		Scale
90-94%	92%	86%	90%	95%	97%	84%	96%		100%
85-89%	92%	79%	82%	94%	87%	77%	95%		90%
80-84%	88%	61%	78%	94%	77%	82%	93%		80%
75-79%	77%	65%	73%	84%	70%	85%	87%		70%
70-74%	73%	63%	66%	85%	65%	67%	82%		60%
65-69%	70%	48%	59%	82%	62%	62%	89%		50%
60-64%	64%	42%	60%	72%	54%	59%	86%		40%
55-59%	60%	27%	53%	74%	57%	46%	83%		30%
50-54%	52%	33%	51%	59%	43%	43%	82%		
< 50%	46%	19%	34%	46%	32%	31%	53%		

Figure 19. (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 9 and Figure 20). 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 14 terms, Sections I through VII for spring 2020 rank 13th, 14th, 13th, 2nd, 12th, 12th, and 5th highest, respectively.

	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
Sp 2020 (n=168)	10.5	7.7	24.9	11.5	7.4	9.3	16.3	87.4

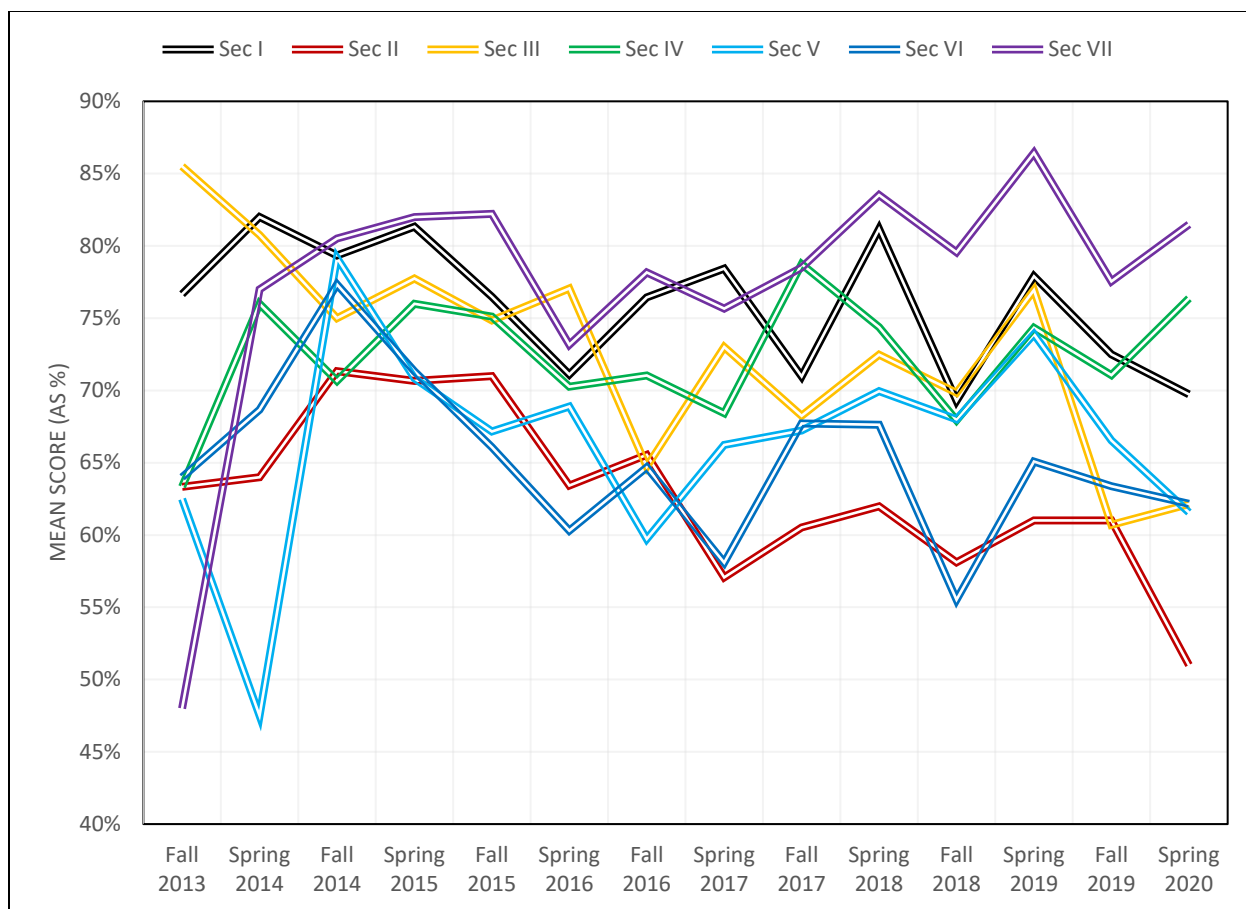*Table 9. Comparison of mean scores for SPN 1121 for fall 2013 through spring 2020.*

Figure 20. 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 spring 2020 assessment.

4.1 FRENCH

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

1. During the Spring 2020 term, three sections of FRE 1120 were offered. An assessment first piloted in AY 2017-18 was recorded in three section of FRE 1120 accounting for 35 artifacts, or 74% of the population.
2. In a study of Learning Outcome (LO) achievement, results exhibit mean scores across the five LOs ranging from 2.7 to 3.0 (Figure 1), compared with 2.2-2.6 in fall 2019, 2.3 to 2.7 in fall 2018. 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 spring 2020, 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 14/15 or higher.
5. Courses were only offered at two sites as comprised in the online to traditional comparison above. As a result, any meaningful comparison is best left with the online-to-traditional comparison above.
6. In a study of achievement based on disposition survey question response, results, unlike in the past, do not yield a very clear 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. The scores appear to reflect high achievement despite self-reporting by students that they don't understand the material.
7. In a longitudinal study, in four terms, LOs 1 through 5 for the Spring 2020 rank 1st, 2nd, 2nd, 2nd, and 1st, respectively.

A drill-down of FRE 1121 results are as follows:
1. During the Spring 2020 term, four sections of FRE 1121 were offered. The assessment was recorded in three sections, however, in one section results and scoring did not match that expected from the assessment plan, so no data could be utilized. The data collection accounts for 17 artifacts, or 32% of the population.
2. In a study of Learning Outcome (LO) achievement, exhibit mean scores across the five LOs ranging from 2.1 to 2.5. 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. There was no dual enrollment (concurrent) section run during spring 2020, 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 online section due to the data not matching the expected for the assessment. 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 three 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. There appears to be the beginnings of a trend of confidence with achievement, although sample size is limited (n=17).
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 Spring 2020 semester 10 sections of SPN 1120 *Beginning Spanish I* were offered. Of those, artifacts from a common final were collected from 10 sections. A total of 200 students were enrolled in SPN 1120. Of those, 123 artifacts were collected representing a sample size of 62% of the population.

2. Achievement of 80% of artifacts scoring 70% or better (SLO 1): Achievement was nearly met as results exhibit 70% 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 46% of artifacts scored 70% or higher in Section II and 85% of artifacts scored 70% or higher in Section III.
4. Achievement of 80% of artifacts scoring 70% or better (SLO 3): Achievement was nearly met. Results exhibit 67% of artifacts scored 70% or higher in Section IV and 78% of artifacts scored 70% or higher in Section V.
5. No dual enrollment (concurrent) sections of the course were run during spring 2020 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 11.9 higher than traditional artifacts. Results were statistically significantly different.
7. In a cross-campus comparison study FSW Online exhibits the highest mean scores in 5 of 5 sections and the overall score.
8. In a study of score distribution by section, all sections exhibit peaks above 90% except for Section II, which peaks at 80-89%. Section II also continues to exhibit more widely distributed scores. Section II exhibits a maximum of 8% at ≥ 90% while other sections range from 29% to 47%.
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 80%-84%, for example, Section II exhibits a mean score of 64% while the other four sections range from 80%-92%. At the overall scoring range at 90%-94%, Section II exhibits a mean score of 84%, while the other four sections range from 90%-98%.
10. In a longitudinal study of data distribution through time, in 14 terms, Sections I through V for the Spring 2020 term rank 12th, 14th, 6th, 7th, and 10th, 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. No study was conducted in summer 2020 to include in this report due to the Covid-19 pandemic.

A drill-down of SPN 1121 results are as follows:
1. Ten sections of SPN 1121 Beginning Spanish II were offered in spring 2020. Of those, artifacts were collected from a common final from 10 of 10 sections. A total of 200 students were enrolled in SPN 1121. Of those, 168 artifacts were collected representing a sample size of 84%.
2. Achievement of 80% of artifacts scoring 70% or better (SLO 1): Achievement was not met as results exhibit 58% 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 32% of artifacts scored 70% or higher in Section II and 48% 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 48% of artifacts scored 70% or higher in Section V and 88% of artifacts scored 80% or higher in Section VII.
5. No dual enrollment (concurrent) sections of the course were run during spring 2020 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 8.8 lower than traditional artifacts. Results were not statistically significantly different.
7. In cross-campus comparison, FSW Online exhibits the highest mean scores in 6 of 7 sections.

8. In a study of score distribution by section, Sections I, IV, V, VI, and VII exhibit peaks centered on ≥90%. Section II exhibits a peak centered at <30%. Section III exhibits a peak centered at 70-79%.
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 42%-72%. Second, Sections II, III, and V, exhibit the weakest performance at mid-range scores. For example, in the range of 75-79%, Sections II, III, and V exhibit scores of 65-70%, while other sections range from 77%-87%. And third, at high scores (90 or higher) all sections score fairly evenly, although Sections II and VI are somewhat lagging. In the ≥95% range, Sections II and VI exhibit mean of 86% and 84%, respectively, while all other sections range from 90-97%.
10. In a longitudinal study of data distribution through time, in 14 terms, Sections I through VII for spring 2020 rank 13th, 14th, 13th, 2nd, 12th, 12th, and 5th highest, respectively.

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.