

Foreign Language Assessment Report – Spring 2015

Author: Joseph F. van Gaalen, Ph.D., Director, Academic Assessment

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). 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. The assessment plan also provides comparisons between dual Enrollment and non-dual enrollment students, online versus traditional students, and by site, where possible. Where data is sufficient, additional analyses are provided including data distribution studies, longitudinal studies, and section-by-section comparisons.

For additional detail on further analysis not provided in this report, please contact Dr. Joseph F. van Gaalen, Director of Academic Assessment, Academic Affairs (jfvangaalen@fsw.edu; x16965).

2 FRENCH

During the spring 2015 semester four sections of FRE 1120 Elementary French I were offered. Of those, artifacts were collected from one section. Sections were taught by two different instructors, one of which was represented in the artifacts collected. A total of 70 students were enrolled in FRE 1120. Of those, 15 artifacts were collected representing a sample size of 21.4% of the population. Five sections of FRE 1121 Elementary French II were offered. Of those, artifacts were collected from one section. Sections were taught by two different instructors, one of which was represented in the artifacts collected. A total of 41 students were enrolled in FRE 1121. Of those, 11 artifacts were collected representing a sample size of 26.8%.

2.1 DESCRIPTIVE STATISTICS & LEARNING OBJECTIVES

2.1.1 FRE 1120

Using a common course assessment the FSW French faculty defined three areas of interest for evaluation that apply to FRE 1120, oral comprehension, reading, and writing. It should be noted that in previous years, Section III was split into two sections totaling 6 points and 20 points, respectively. Beginning spring 2015, Section III combines the two and so the overall score is comprised of 9 sections as opposed to the previous 10. The Student Learning Outcomes (SLOs) and their measure of success are:

SLO1: Students will be able to understand spoken French. 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 (not yet included in data for spring 2015 exam).

SLO2: Students will be able to understand written French. 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 IX).

SLO3: Students will be able to write effectively in the French 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 I).

The faculty established measure of success for SLO2, 80% of students scoring 70% or higher in Section IX, was met as results exhibit 93% of artifacts score 70% or higher in the reading competency exam section (Section IX) (Table 1). The faculty established measure of success for SLO3, 80% of students scoring 70% or higher in Section I, was met. Results exhibit 93% of artifacts scored 70% or higher in Section I. For a graphical representation of SLO achievement, see Figure 1. The lowest achieving area of the assessment exam are Section VIII at 47% achieving 70% or greater.

n = 15	Sec I	Sec II	Sec III	Sec IV	Sec V	Sec VI	Sec VII	Sec VIII	Sec IX	Combined (Overall)
Goal	<i>80% of artifacts scored $\geq 70\%$ for all sections</i>									
% above Goal	100%	87%	87%	73%	47%	93%	60%	47%	93%	87%
Mean	93%	92%	80%	85%	65%	92%	77%	66%	83%	81%
Median	100%	100%	81%	90%	63%	100%	80%	68%	85%	86%
Section Score Max	10	6	26	20	8	10	10	14	20	124
Section Mean	9.3	5.5	20.8	17.0	5.2	9.2	7.7	9.3	16.5	100.5
Section Median	10	6	21	18	5	10	8	9.5	17	107

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

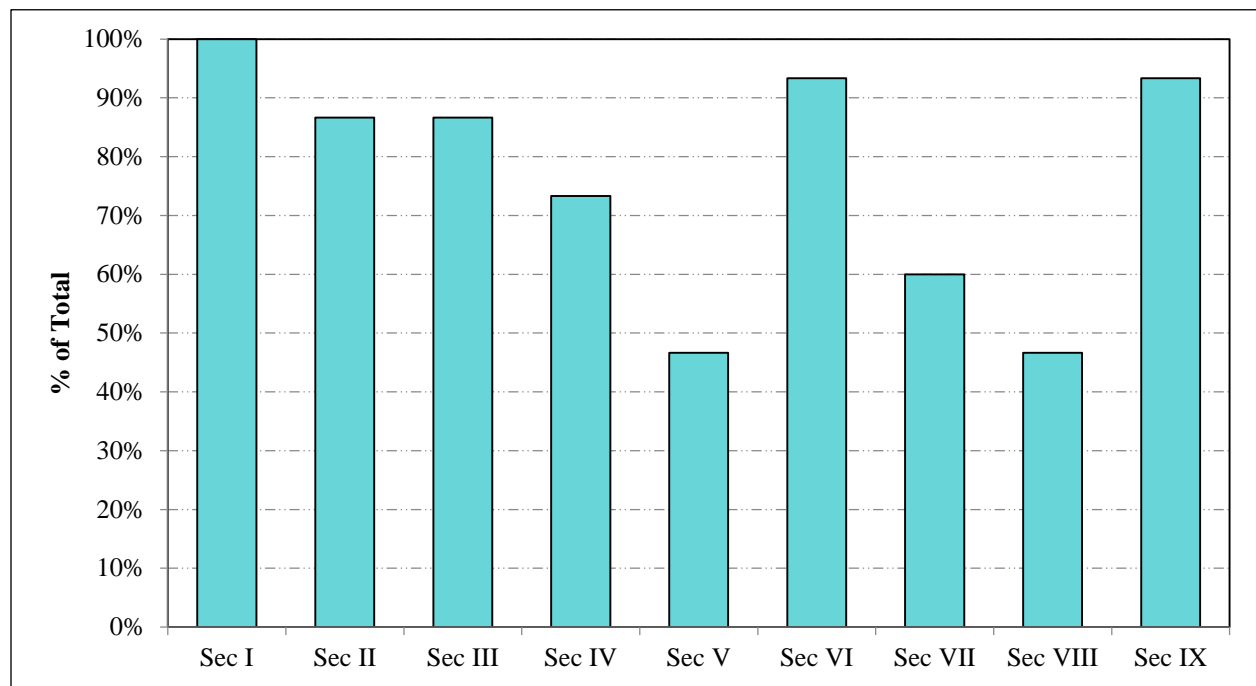


Figure 1. SLO achievement for FRE 1120 by exam section (Percentage of student achievement level as per SLOs).

2.1.2 FRE 1121

Using a common course assessment the FSW French faculty defined the same three areas of interest for evaluation that apply to FRE 1121 as those used for FRE 1120. For details on each SLO, see 2.1.1. The only difference between FRE 1121 and FRE 1120 in terms of measuring these outcomes is that the exam sections differ slightly and are noted below.

The faculty established measure of success for SLO2, 80% of students scoring 70% or higher in Section XI, was not met as results exhibit 73% of artifacts score 70% or higher in the reading competency exam section (Section XI) (Table 2). The faculty established measure of success for SLO3, 80% of students scoring 70% or higher in Section I, was met. Results exhibit 91% of artifacts scored 70% or higher in Section I. For a graphical representation of SLO achievement, see Figure 2.

n = 10	Sec I	Sec II	Sec III	Sec IV	Sec V	Sec VI	Sec VII	Sec VIII	Sec IX	Sec X	Sec XI	Combined (Overall)
Goal	<i>80% of artifacts scored $\geq 70\%$ for all sections</i>											
% above Goal	91%	73%	45%	36%	45%	55%	91%	27%	45%	64%	73%	
Mean	91	86	60	65	70	63	89	46	63	73	77	
Median	100	90	63	63	67	75	100	40	50	92	80	
Section Score Max	10	10	8	8	12	4	8	5	4	12	20	101
Section Mean	9.1	8.6	4.8	5.2	8.4	2.5	7.1	2.3	2.5	8.8	15.3	74.5
Section Median	10	9	5	5	8	3	8	2	2	11	16	75.5

Table 2. Percentage of student achievement level as per SLOs (FRE 1121).

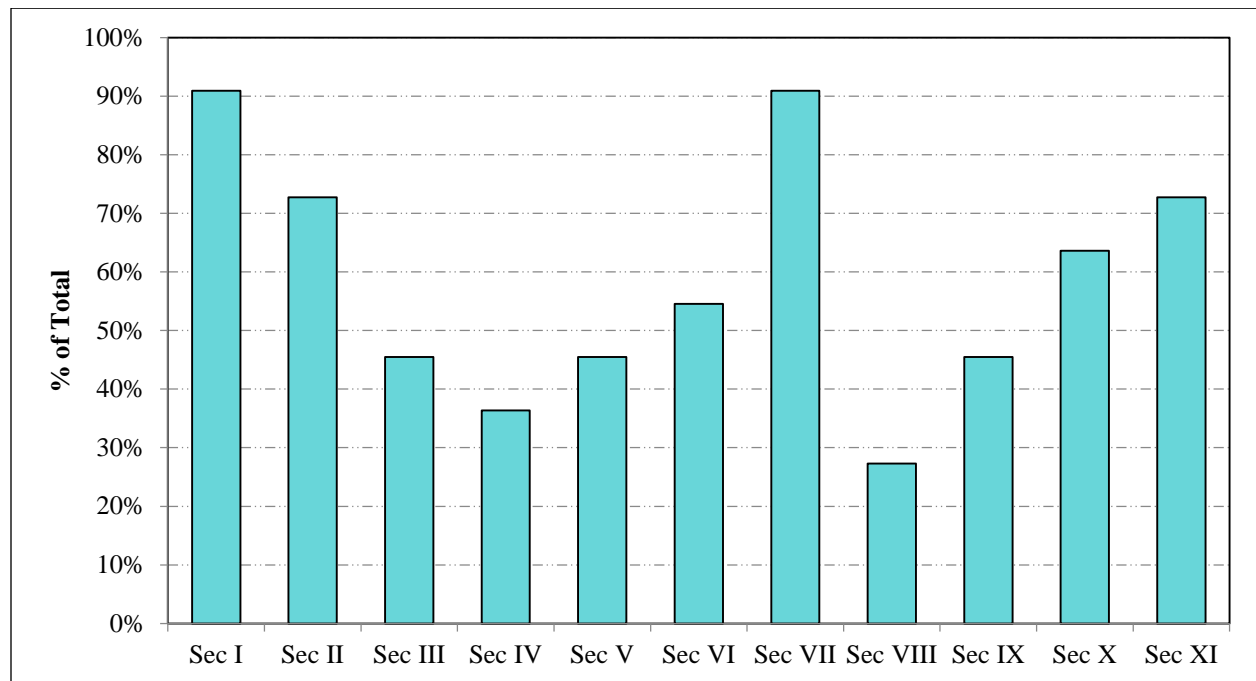


Figure 2. SLO achievement for FRE 1121 by exam section (Percentage of student achievement level as per SLOs).

2.2 EXPLORATORY ANALYSIS AND 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.

2.2.1 FRE 1120

The FRE 1120 Elementary French I course was offered in both online and traditional format, however, no artifacts from online sections of the course were collected during spring 2015. Additionally, a small proportion of artifacts originated from Dual Enrollment (DE) students. All of the DE artifacts used in the analysis originate from onsite locations as opposed to offsite, strictly DE sections. Finally, FRE 1120 was offered on FSW Online and the Thomas Edison (Lee) campus.

2.2.1.1 Comparison by Site, Format, or Student Type

2.2.1.1.1 Dual Enrollment to non-Dual Enrollment Comparison

Only four Dual Enrollment (DE) artifacts were collected from a total collection of just 15 so no comparison study between online and traditional could be completed.

2.2.1.1.2 Online to Traditional Comparison

No artifacts from online sections of the course were collected during spring 2015 so no comparison study between online and traditional could be completed.

2.2.1.1.3 Comparison by Campus/Site

Of the artifacts collected, all originated from the Thomas Edison (Lee) campus so no cross site comparison could be completed.

2.2.1.2 Data Distribution (& Longitudinal Study)

A histogram depicting the distribution of scores across each exam section as well as overall (combined) score is shown in Figure 3. Most sections exhibit moderate S-curves where the largest percentages of artifacts occur in the highest scoring bins although exceptions exist. Mathematically, a multi-section exam with S-curve score distributions in each section results in the typical bell-shaped curve for overall scores. Sections III and IX exhibit increased percentages of artifacts in the lower scoring bins compared with other sections.

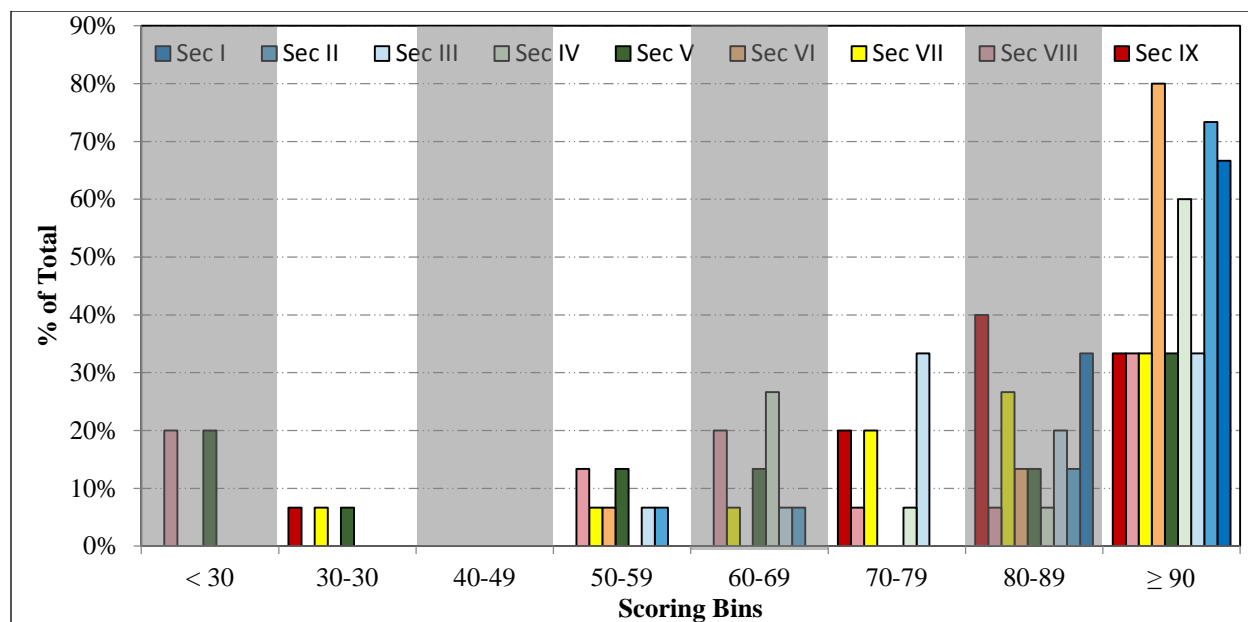


Figure 3. Histogram of FRE 1120 exam sections and combined (overall) score for spring 2015.

A series of histograms comparing year-to-year are typical of this section of the report. However, since no study was completed for spring 2014 no comparisons are reported here.

2.2.2 FRE 1121

The FRE 1121 Elementary French II course was offered in both online and traditional format, however, no artifacts from online sections of the course were collected during spring 2015. Additionally, a small proportion of artifacts originated from Dual Enrollment (DE) students. All of the DE artifacts used in the analysis originate from onsite locations as opposed to offsite, strictly DE sections. Finally, FRE 1121 was offered on the Collier campus, FSW Online and the Thomas Edison (Lee) campus.

2.2.2.1 Comparison by Site, Format, or Student Type

2.2.2.1.1 Dual Enrollment to non-Dual Enrollment Comparison

Only one Dual Enrollment (DE) artifacts was collected from a total collection of just 11 so no comparison study between online and traditional could be completed.

2.2.2.1.2 Online to Traditional Comparison

No artifacts from online sections of the course were collected during spring 2015 so no comparison study between online and traditional could be completed.

2.2.2.1.3 Comparison by Campus/Site

Of the artifacts collected, all originated from the Thomas Edison (Lee) campus so no cross site comparison could be completed. Five sections of FRE 1121 were run during spring 2015 with one at Collier, two on FSW Online, and two at Thomas Edison (Lee).

2.2.2.2 Data Distribution (& Longitudinal Study)

A histogram depicting the distribution of scores across each exam section as well as overall (combined) score is shown in Figure 4. Most sections exhibit moderate S-curves where the largest percentages of

artifacts occur in the highest scoring bins, although with such a small sample size, interpretation is limited. Mathematically, a multi-section exam with S-curve score distributions in each section results in the typical bell-shaped curve for overall scores. Section VIII exhibits increased percentages of artifacts in the lower scoring bins compared with other sections.

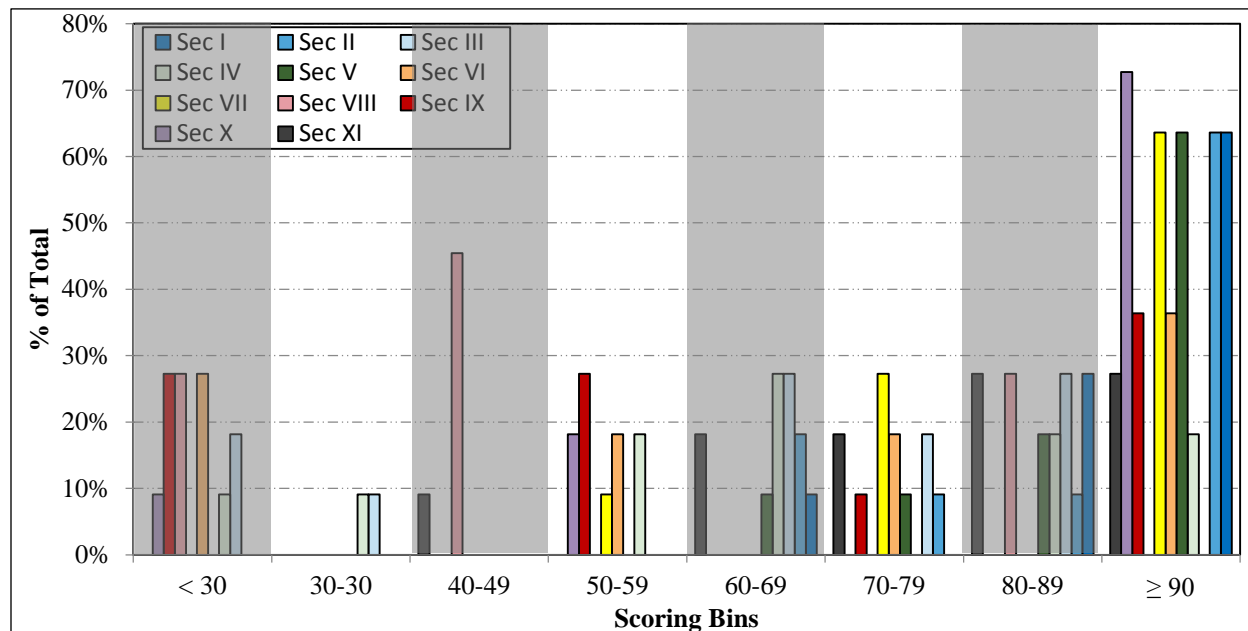


Figure 4. Histogram of FRE 1120 exam sections and combined (overall) score for spring 2015.

A series of histograms comparing year-to-year are typical of this section of the report. However, since no study was completed for spring 2014 no comparisons are reported here.

3 SPANISH

During the spring 2015 semester 11 sections of SPN 1120 Beginning Spanish I were offered. Of those, artifacts were collected from six sections. Sections were taught by seven different instructors, five of which were represented in the artifacts collected. A total of 249 students were enrolled in SPN 1120. Of those, 73 artifacts were collected representing a sample size of 29.3% of the population. Ten sections of SPN 1121 Beginning Spanish II were offered. Of those, artifacts were collected from three sections. Sections were taught by seven different instructors, four of which were represented in the artifacts collected. A total of 240 students were enrolled in SPN 1121. Of those, 58 artifacts were collected representing a sample size of 24.1%.

3.1 DESCRIPTIVE STATISTICS & LEARNING OBJECTIVES

3.1.1 SPN 1120

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 measure of success are:

SLO1: 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).

SLO2: 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).

SLO3: 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 SLO1, 80% of students scoring 70% or higher in Section I, was not met as results exhibit only 68.5% of artifacts score 70% or higher in the oral competency exam section (Section I) (Table 3). The faculty established measure of success for SLO2, 80% of students scoring 70% or higher in Sections II and III, was partially met. Results exhibit 50.7% of artifacts scored 70% or higher in Section II and 87.7% of artifacts scored 70% or higher in Section III. Similarly, the faculty established measure of success for SLO3, 80% of students scoring 70% or higher in Sections IV and V, was partially met. Results exhibit 67.1% of artifacts scored 70% or higher in Section IV and 84.9% of artifacts scored 70% or higher in Section V. For a graphical representation of SLO achievement, see Figure 5.

n = 73	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%	68.5%	50.7%	87.7%	67.1%	84.9%	
Mean	76.8%	66.3%	85.3%	72.9%	81.8%	73.1%
Median	86.7%	70.0%	90.0%	83.3%	87.5%	77.2%
Section Score Max	15	60	15	15	20	125
Section Mean	11.5	39.8	12.8	10.9	16.4	91.4
Section Median	13	42	13.5	12.5	17.5	96.5

Table 3. Percentage of student achievement level as per SLOs (SPN 1120).

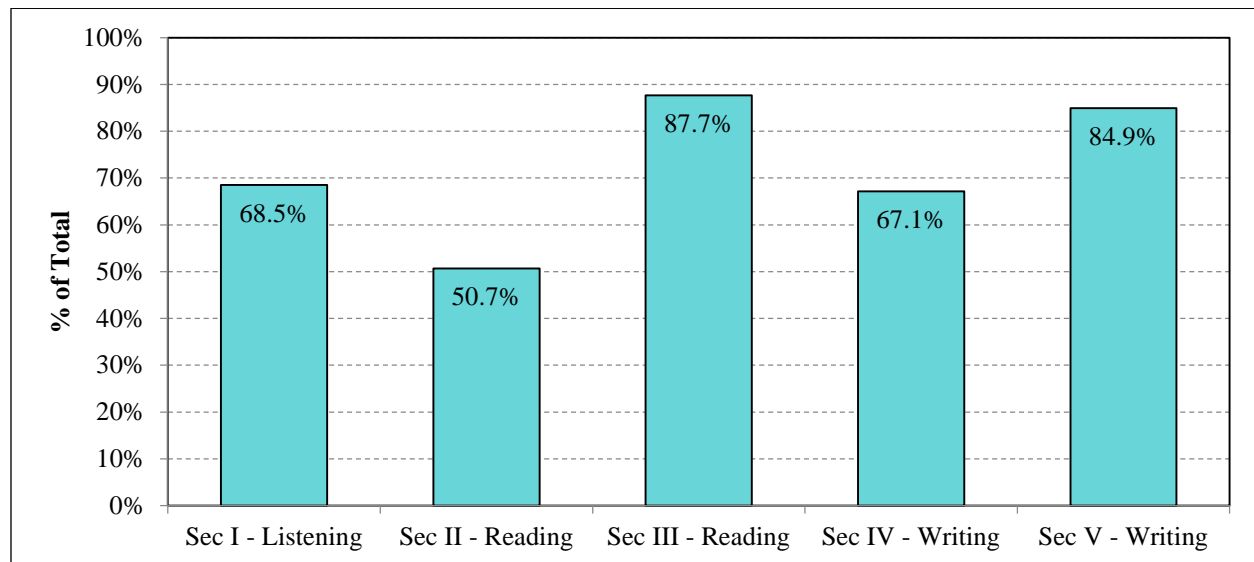


Figure 5. SLO achievement for SPN 1120 by exam section (Percentage of student achievement level as per SLOs).

3.1.2 SPN 1121

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 4 below.

The faculty established measure of success for SLO1, 80% of students scoring 70% or higher in Section I, was nearly met as results exhibit 77.6% of artifacts score 70% or higher in the oral competency exam section (Section I) (Table 4). The faculty established measure of success for SLO2, 80% of students scoring 70% or higher in reading only sections, Sections II, and VI, was not met. Results exhibit 60.3% of artifacts scored 70% or higher in Section II and 55.2% of artifacts scored 70% or higher in Section VI. The faculty established measure of success for SLO3, 80% of students scoring 70% or higher in writing only sections, Sections V and VII, was partially met. Results exhibit 51.7% of artifacts scored 70% or higher in Section V and 82.8% of artifacts scored 70% or higher in Section VII. For a graphical representation of SLO achievement, see Figure 6.

n = 58	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	77.6%	60.3%	74.1%	70.7%	51.7%	55.2%	82.8%	
Mean	81.6%	70.7%	77.9%	75.9%	70.5%	71.2%	81.8%	77.6%
Median	88.3%	80.0%	86.3%	80.0%	75.0%	76.7%	86.3%	82.1%
Section Score Max	15	15	40	15	12	15	20	132
Section Mean	12.2	10.6	31.1	11.4	8.5	10.7	16.4	100.9
Section Median	13.25	12.0	34.5	12.0	9.0	11.5	17.25	106.75

Table 4. Percentage of student achievement level as per SLOs (SPN 1121).

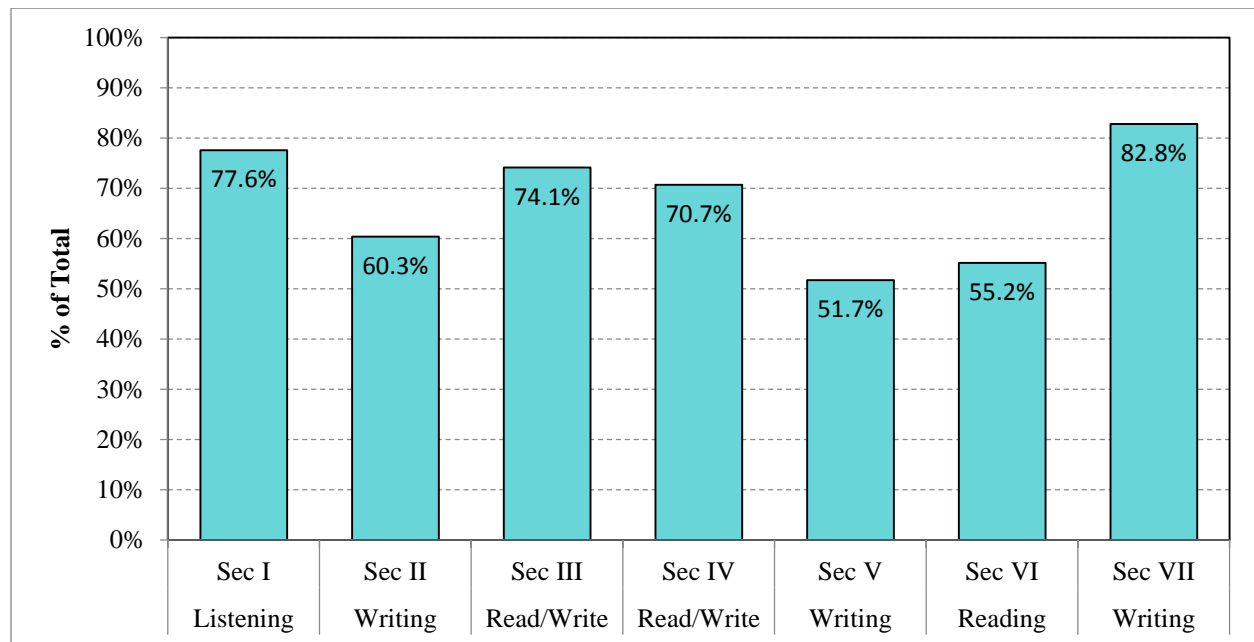


Figure 6. SLO achievement for SPN 1121 by exam section (Percentage of student achievement level as per SLOs).

3.2 EXPLORATORY ANALYSIS AND 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.

3.2.1 SPN 1120

The SPN 1120 Beginning Spanish I course was offered in both online and traditional format, however, no artifacts from online sections of the course were collected during spring 2015. Additionally, a small proportion of artifacts originated from Dual Enrollment (DE) students. All of the DE artifacts used in the analysis originate from onsite locations as opposed to offsite, strictly DE sections. Finally, SPN 1120 was offered at the Collier campus, FSW Online, and the Thomas Edison (Lee) campus.

3.2.1.1 Comparison by Site, Format, or Student Type

3.2.1.1.1 Dual Enrollment to non-Dual Enrollment Comparison

During the spring 2015 semester, 5 total Dual Enrollment (DE) artifacts were collected in the study of SPN 1120 and 68 non-Dual Enrollment (nonDE) artifacts were collected in the study of SPN 1120. No DE artifacts originated from offsite so this comparison is strictly a demographic comparison and not a comparison of instruction at multiple locations. A comparison of mean scores by exam section is provided in Table 5. Mean scores between DE and nonDE were analyzed for statistical significance using standard methods (t-test) however, sample size are too low to yield meaningful results (de Winter, 2013).

df = 71	Section I (Oral)	Section II (Reading)	Section III (Reading)	Section IV (Written)	Section V (Written)	Combined (Overall)
<i>Section Max</i>	15	60	15	15	20	125
DE mean*	13.3	44.0	14.4	11.5	18.0	101.2
nonDE mean	11.4	39.5	12.7	10.9	16.2	90.7
Effect Size	-0.380	-0.193	-0.576	-0.129	-0.392	-0.280
p-value	0.167	0.456	0.043	0.606	0.147	0.291

Table 5. Comparison of mean scores for DE and nonDE. Bold denotes statistically significant difference. Positive effect sizes indicate a higher mean score for nonDE artifacts. *Only 5 DE students (none offsite) and sample sizes are too low to yield meaningful results (de Winter, 2013).

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 small-to-medium effect sizes ranging from 0.13 to 0.57 (Table 5). In other words, non-overlap from DE artifacts to nonDE artifacts range from approximately 10% to 36%.

3.2.1.1.2 Online to Traditional Comparison

No artifacts from online sections of the course were collected during spring 2015 so no comparison study between online and traditional could be completed. A new version of the online course is in development for fall 2015 which will implement the oral portion of the assessment.

3.2.1.1.3 Comparison by Campus/Site

Of the 73 artifacts collected from SPN 1120, 16 originated from the Collier Campus and 57 from the Thomas Edison (Lee) Campus. Scores by section are consistently higher at the Collier campus which

results in a 10.0 point margin between sites for the overall score. A comparison of mean scores by exam section is provided in Table 6.

	Section I (Oral)	Section II (Reading)	Section III (Reading)	Section IV (Written)	Section V (Written)	Combined (Overall)
<i>Section Max</i>	15	60	15	15	20	125
Collier	13.8	42.4	13.3	13.3	17.0	99.9
Thomas Edison (Lee)	10.9	39.1	12.7	10.3	16.2	89.0

Table 6. Comparison of mean scores by site for SPN 1120. Bold denotes highest mean score in that dimension among all sites.

A plot comparing the distributions of the scores by site is presented in Figure 7. Because only two sites included scores, a Welch’s t-test can be used to determine statistical significance between sites according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999). The difference in overall mean scores was found to be marginally statistically significantly different (Table 7). Therefore, we can reject the null hypothesis that the difference in the means of the two artifacts from the two sites is equal to 0, and we can conclude this with a 95% confidence that the differences in scores are not solely due to chance. However, based on the work of Johnson (2013), there is a 17-25% chance that the marginally significant result may be false positives (i.e. Type I errors).

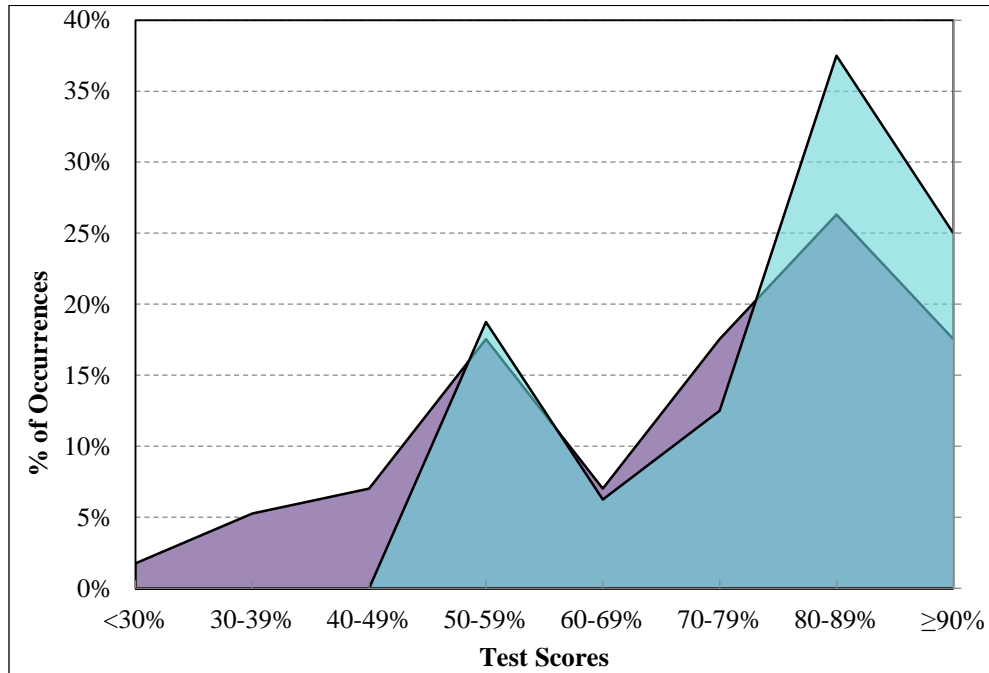


Figure 7. Distribution of SPN 1120 scores by site for Thomas Edison (purple) and Collier campus (teal).

Welch’s t-test results	
t_{obs}	-2.03
t_{crit}	1.99
p-value	0.05
Effect size	-0.48

Table 7. Results of Welch’s t-test of overall scores at Thomas Edison compared with Collier for SPN 1120. Statistically significant results p-values in bold.

3.2.1.2 Data Distribution (& Longitudinal Study)

A histogram depicting the distribution of scores across each exam section is shown in Figure 8. Sections I, III, IV, and V exhibit S-curves where the largest percentages of artifacts occur in the highest scoring bins with progressively smaller percentages in lower scoring bins. Mathematically, an multi-section exam with S-curve score distributions in each section results in the typical bell-shaped curve for overall scores.

Section II, however, stands apart from this typical pattern. For Section II, the distribution is not an S-curve but rather bimodal (two peaked). Section II exhibits 10% or more of artifacts in the 50-59 and 40-49 scoring bins where Sections I, III, IV, and V fall well below that (see Figure 8).

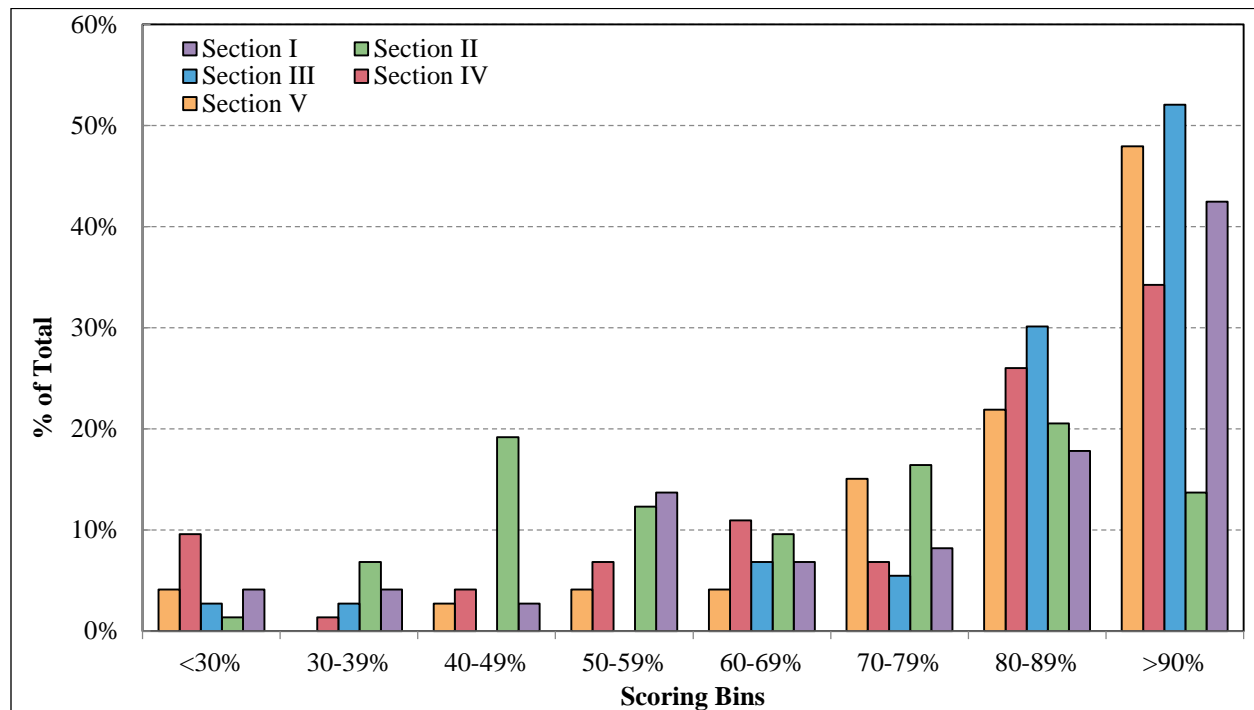


Figure 8. Histogram of SPN 1120 exam sections scores for spring 2015.

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 9 shows that Section II is underperforming compared to other sections between the range of 65%-84%. For example, in the 80-84% range, the mean score for Section II is 68%, while the other four sections range from 82-90%. Meanwhile, Section III, and to a lesser extent Section V, is over performing at the lowest overall scores. At the 45-49% range, the Section IV and V mean scores are 80% and 70%, respectively, whereas other sections range from 10%-40%.

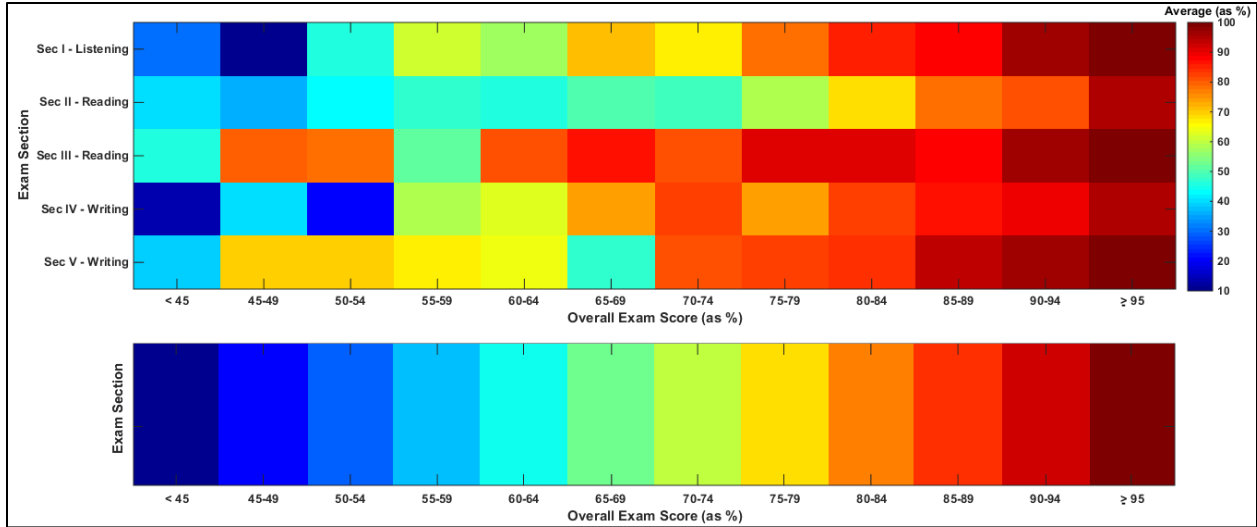


Figure 9. (Top) Colormap of mean scores for each exam section based on overall scoring bin for SPN 1120. (Bottom) Comparison exam section if section score percentage is the same as overall (i.e. artifact score is equally distributed across all sections). An exam section with hotter colors (reds/yellows) means that section achievement exceeds the overall score and is an area of strength. An exam section with colder colors (blues/greens) means that section achievement is lower than the overall score and is therefore an area of weakness.

A series of histograms comparing spring 2015 data with spring 2014 data for Overall (Combined) scores along with each exam section (Sections I-V) are shown in Figures 10 through 15. Overall scores exhibit a smaller percentage of As ($\geq 90\%$) as well as an increase in the 50-59% scoring bin. The percentage of artifacts scoring ≥ 90 dropped by 12% from year-to-year while the 50-59% increased by 10%.

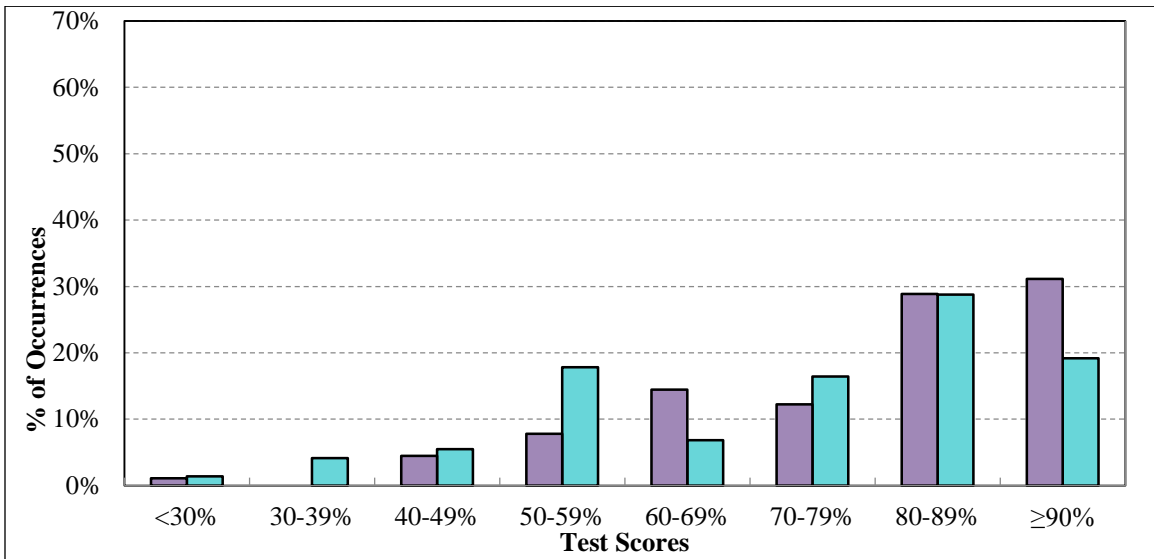


Figure 10. Histogram of SPN 1120 Combined (Overall) scores comparing spring 2014 (purple) with spring 2015 (teal).

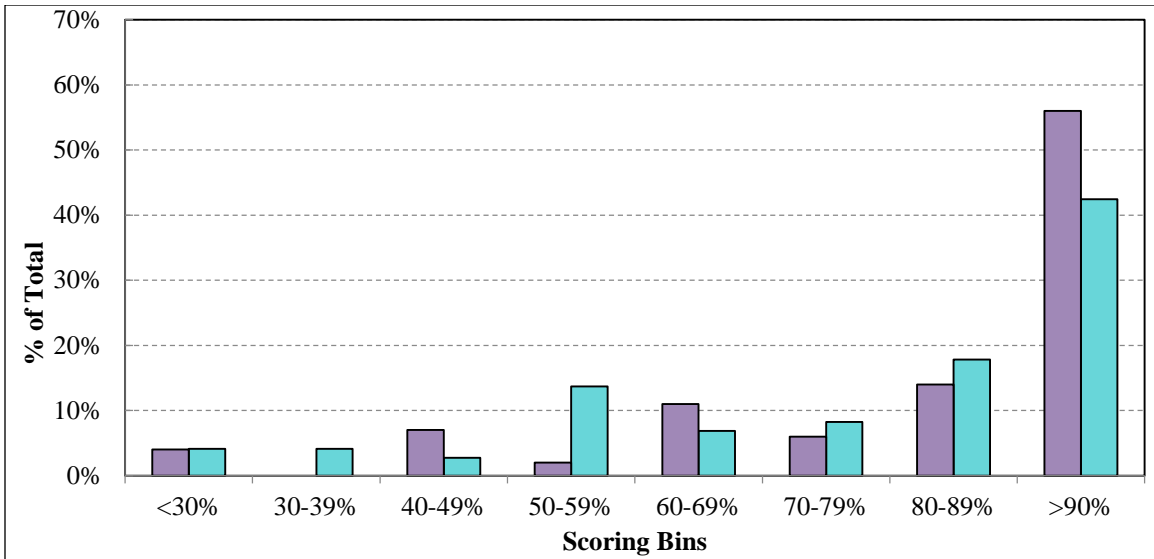


Figure 11. Histogram of SPN 1120 Section I scores comparing spring 2014 (purple) with spring 2015 (teal).

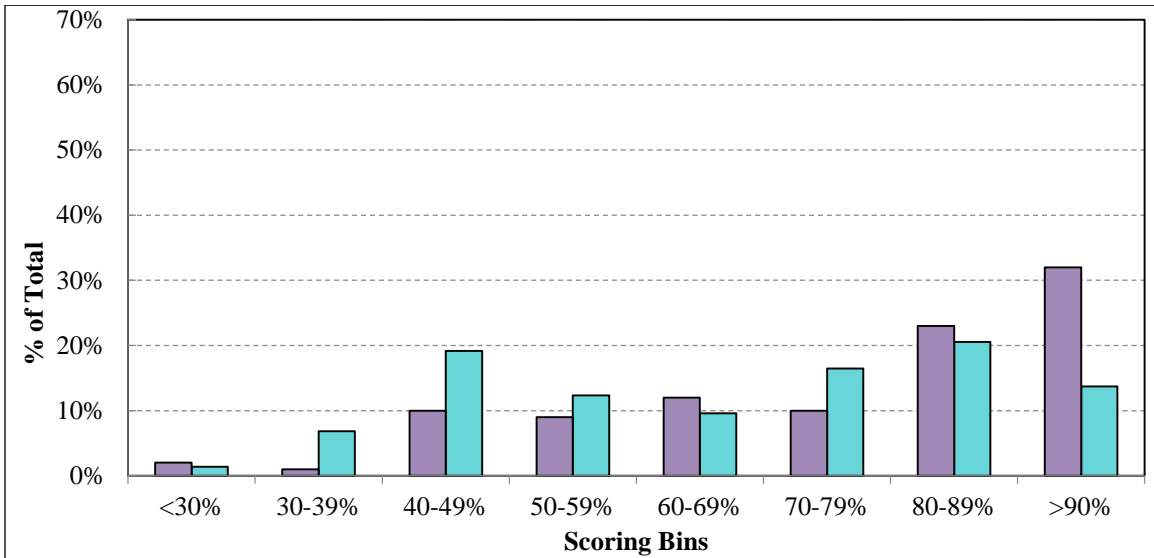


Figure 12. Histogram of SPN 1120 Section II scores comparing spring 2014 (purple) with spring 2015 (teal).

Section I data for spring 2015 exhibit a fairly typical S-Curve although not as pronounced as spring 2014 (see Figure 11). The wide distribution of Section II noted earlier is common in both spring 2014 and spring 2015 artifacts. All sections exhibit similar distribution curves to the previous academic year. In all sections there is a decrease in the percentage of artifacts scored ≥ 90 .

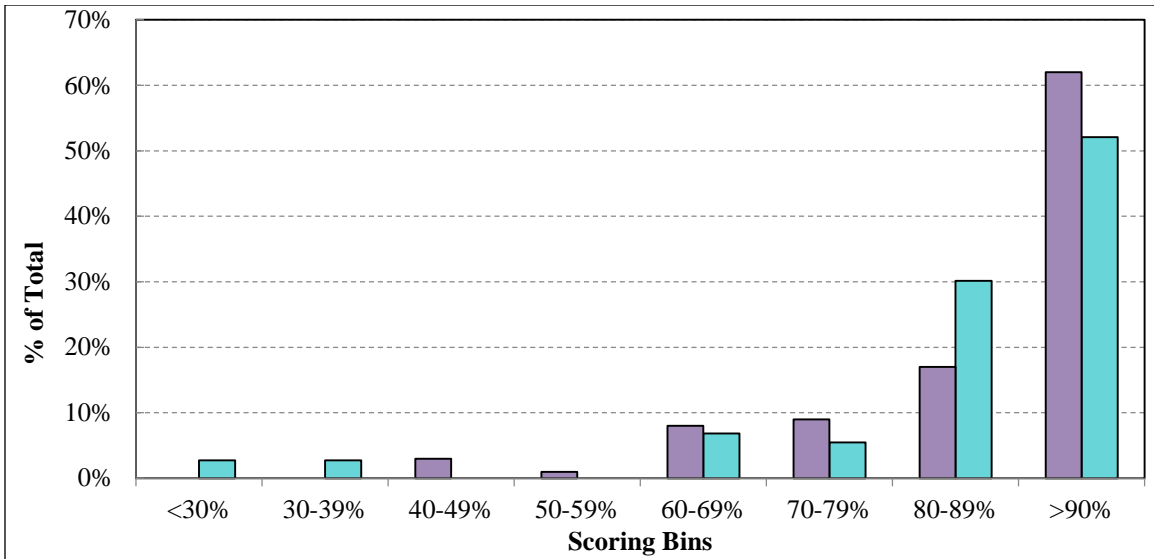


Figure 13. Histogram of SPN 1120 Section III scores comparing spring 2014 (purple) with spring 2015 (teal).

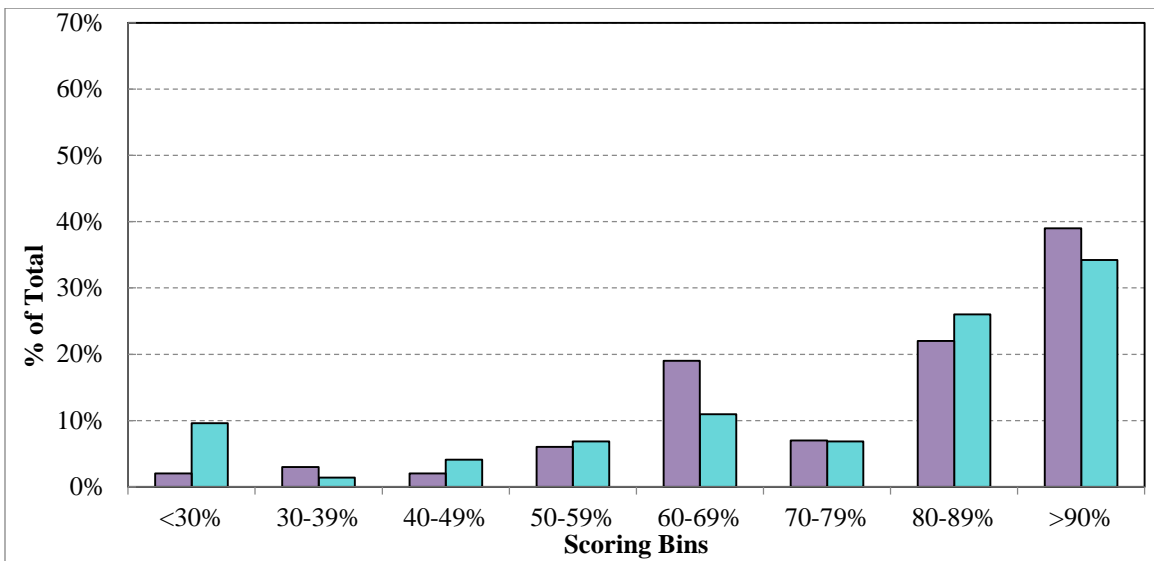


Figure 14. Histogram of SPN 1120 Section IV scores comparing spring 2014 (purple) with spring 2015 (teal).

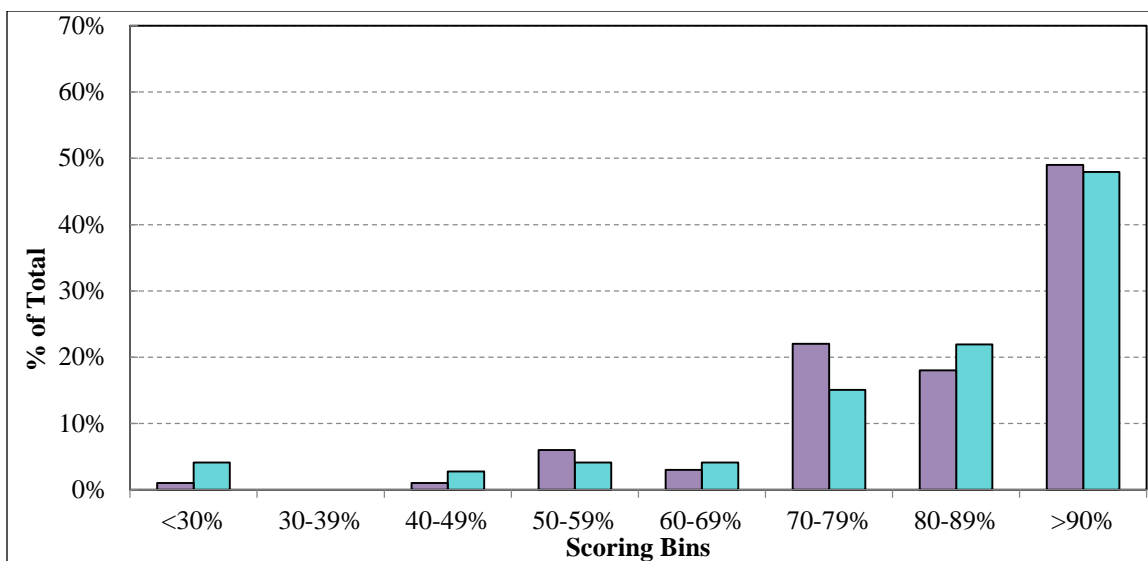


Figure 15. Histogram of SPN 1120 Section V scores comparing spring 2014 (purple) with spring 2015 (teal).

A comparison of mean scores for each exam section by semester is reported in Table 8. 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). When comparing spring-to-spring, all sections show declines from spring 2014 to spring 2015. The declines, when tested using a Welch's t-test according to standard methods (Davis, 1973; de Winter 2013; McDonald, 2009; Wilkinson, 1999), were found to be statistically significantly in one of six cases (Sections II). Results for the significant results are $t(161)=2.821$, $p=0.005$, with an effect size of 0.45.

	<i>Section Max</i>	Fall 2013 n=58	Spring 2014 n=90	Fall 2014 n=93	Spring 2015 n=73
Section I (Oral)	15	12.1	12.4	11.8	11.5
Section II (Reading)	60	40.4	45.1	40.5	39.8
Section III (Reading)	15	12.0	13.2	12.8	12.8
Section IV (Written)	15	10.6	11.7	11.1	10.9
Section V (Written)	20	16.2	16.8	16.5	16.4
Combined (Overall)	125	91.3	99.2	92.8	91.4

Table 8. Comparison of mean scores for SPN 1120 for fall 2013 through fall 2014. Bold denotes statistically significant improvements from fall 2013 to fall 2014 or spring 2014 to spring 2015. Red bold denotes statistically significant declines.

3.2.2 SPN 1121

The SPN 1121 Beginning Spanish II course was offered in both online and traditional format, however, no artifacts from online sections of the course were collected during spring 2015. Additionally, a small proportion of artifacts originated from Dual Enrollment (DE) students. All of the DE artifacts used in the analysis originate from onsite locations as opposed to offsite, strictly DE sections. Finally, SPN 1121 was offered at the Charlotte campus, Collier campus, FSW Online, and the Thomas Edison (Lee) campus.

3.2.2.1 Comparison by Site, Format, or Student Type

3.2.2.1.1 Dual Enrollment to non-Dual Enrollment Comparison

During the spring 2015 semester, 12 total Dual Enrollment (DE) artifacts were collected in the study of SPN 1121 and 46 non-Dual Enrollment (nonDE) artifacts were collected in the study of SPN 1121. No DE

artifacts originated from offsite so this comparison is strictly a demographic comparison and not a comparison of instruction at multiple locations.

A comparison of mean scores by exam section is provided in Table 9. Dual enrollment artifacts consistently score higher in all sections. The overall score for dual enrollment artifacts is 13.1 points higher than non-dual enrollment.

df = 71	Section I (Oral)	Section II (Reading)	Section III (Reading)	Section IV (Written)	Section V (Written)	Section VI (Reading)	Section VII (Writing)	Combined (Overall)
<i>Section Max</i>	15	15	40	15	12	15	20	132
DE mean	13.0	12.4	34.0	13.0	10.3	10.8	17.8	111.3
nonDE mean	12.0	10.1	30.4	11.0	8.0	10.6	16.0	98.2
Effect Size	.267	.572	.337	.586	.895	.043	.431	.468
p-value	0.333	0.044	0.225	0.040	0.002	0.874	0.119	0.097

Table 9. Comparison of mean scores for DE and nonDE. Bold denotes statistically significant difference. Positive effect sizes indicate a higher mean score for nonDE 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 (Table 10). Therefore, we cannot reject the null hypothesis that the difference in the means of the dual enrollment and non-dual enrollment traditional artifacts is equal to 0, and we cannot conclude this with a 95% confidence that the differences in scores are not solely due to chance.

Welch’s t-test results	
t _{obs}	-1.75
t _{crit}	2.00
p-value	0.097
Effect size	-0.47

Table 10. Results of Welch’s t-test of overall scores at Thomas Edison compared with Collier for SPN 1121. Statistically significant results p-values in bold.

3.2.2.1.2 Online to Traditional Comparison

No artifacts from online sections of the course were collected during spring 2015 so no comparison study between online and traditional could be completed. A new version of the online course is in development for fall 2015 which will implement the oral portion of the assessment.

3.2.2.1.3 Comparison by Campus/Site

Of the 58 artifacts collected from SPN 1121, 5 originated from the Charlotte campus, 12 from the Collier campus and 41 from the Thomas Edison (Lee) Campus. Scores by section are consistently higher at the Collier campus which results in a 20.1 point margin between sites for the overall. A comparison of mean scores by exam section is provided in Table 11.

A plot comparing the distributions of the scores by site is presented in Figure 16. Because only two sites included sufficient sample sizes to conduct significant testing (de Winter, 2013), a Welch’s t-test was used to determine statistical significance between those two sites according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999). The difference in overall mean scores was found to be statistically significantly different (Table 12). Therefore, we can reject the null hypothesis that the difference in the means of the two artifacts from the two sites is equal to 0, and we can conclude this with a 95% confidence that the differences in scores are not solely due to chance.

	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)
<i>Section Max</i>	15	15	40	15	12	15	20	132
Charlotte	7.0	5.8	17.4	8.2	6.0	6.9	8.4	59.7
Collier	14.0	12.2	37.3	13.5	11.5	13.5	18.3	120.3
Thomas Edison (Lee)	12.4	10.7	31.0	11.2	7.9	10.3	16.8	100.2

Table 11. Comparison of mean scores by site for SPN 1121. Bold denotes highest mean score in that dimension among all sites.

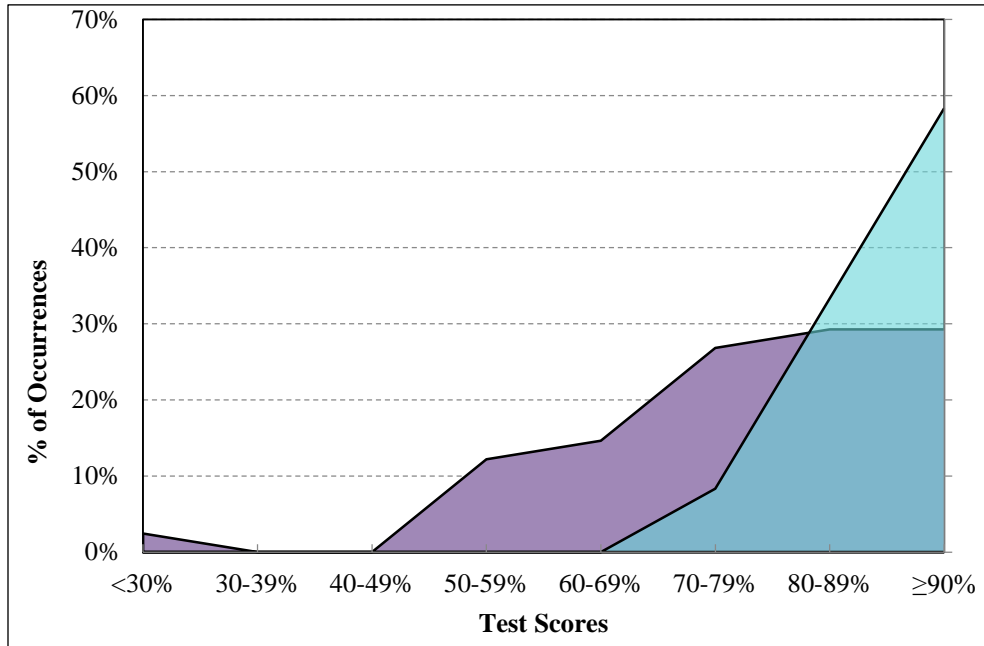


Figure 16. Distribution of SPN 1121 scores by site for Thomas Edison (purple) and Collier campus (teal).

Welch's t-test results	
t_{obs}	-4.87
t_{crit}	2.01
p-value	1.60×10^{-5}
Effect size	-1.364

Table 12. Results of Welch's t-test of overall scores at Thomas Edison compared with Collier for SPN 1121. Statistically significant results p-values in bold.

3.2.2.2 Data Distribution (& Longitudinal Study)

A histogram depicting the distribution of scores across each exam section as well as overall (combined) score is shown in Figure 17. Sections I, II, II, IV, and VII exhibit S-curves typical of sectional score distribution with the mode centered on the highest scoring bin. Sections V and VI exhibit a mode centered on the 80-89 scoring bin. Moreover, Section IV exhibits a secondary bimodal distribution with peaks centered on ≥ 90 and 50-59 in the case of Section V and 60-69 in the case of Section VI.

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 18). The color represents the mean section score achieved overall score as shown in the x-axis as a percentage.

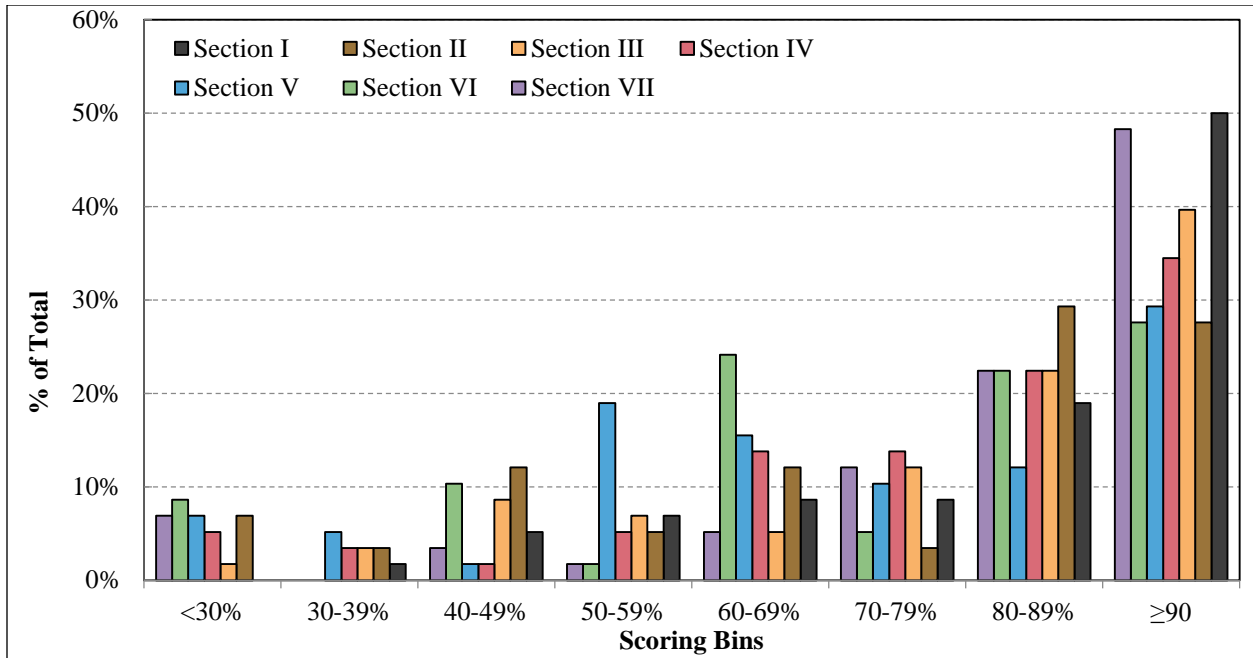


Figure 17. Histogram of SPN 1121 exam sections and combined (overall) score for spring 2015.

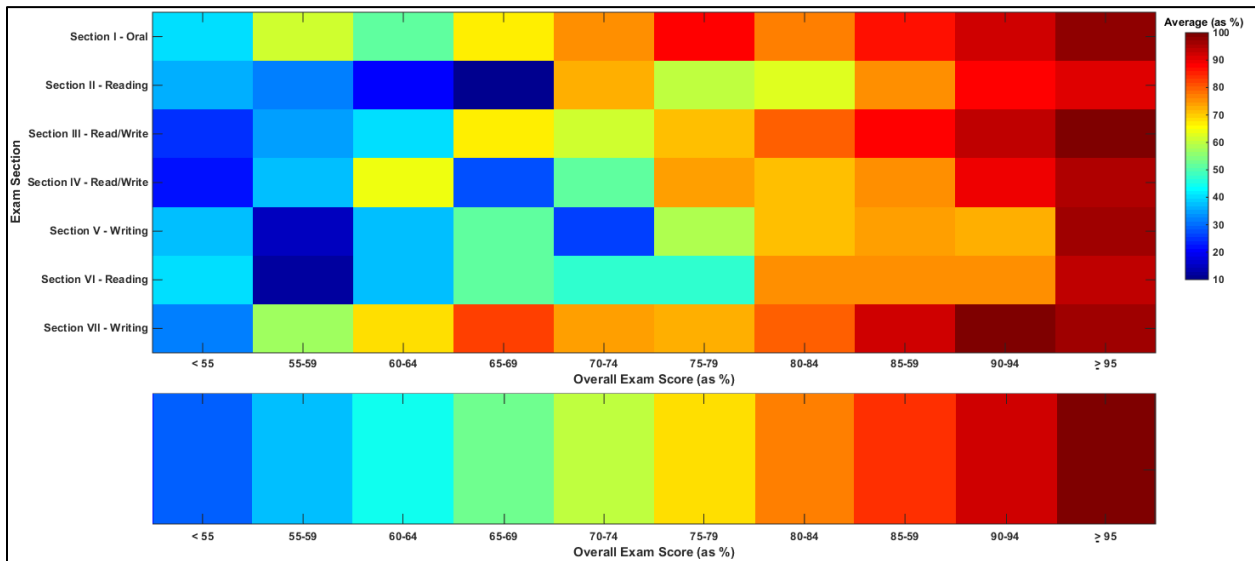


Figure 18. (Top) Colormap of mean scores for each exam section based on overall scoring bin for SPN 1121. (Bottom) Comparison exam section if section score percentage is the same as overall (i.e. artifact score is equally distributed across all sections). An exam section with hotter colors (reds/yellows) means that section achievement exceeds the overall score and is an area of strength. An exam section with colder colors (blues/greens) means that section achievement is lower than the overall score and is therefore an area of weakness.

A review of the colormap in Figure 18 shows that Sections II through VI are slightly underperforming compared to other Sections I and VII between the range of 65%-79%. Both Section I and VI show strength at lower overall scores. For example, at an overall score of 55-59%, Section I and VII averages are 56% and 53%, respectively. Both scores are expected for the overall mean. However, Sections I through VI in this scoring bin range from only 30-40%.

As with SPN 1121, a series of histograms comparing spring 2014 data with spring 2015 data for Overall (Combined) scores along with each exam section (Sections I-VII) are shown in Figures 19 through 26. Overall scores exhibit substantial improvement in number of artifacts scoring an A from year-to-year. The distribution for 2015 is less peaked, however, with fewer Bs and Cs compared to spring 2014.

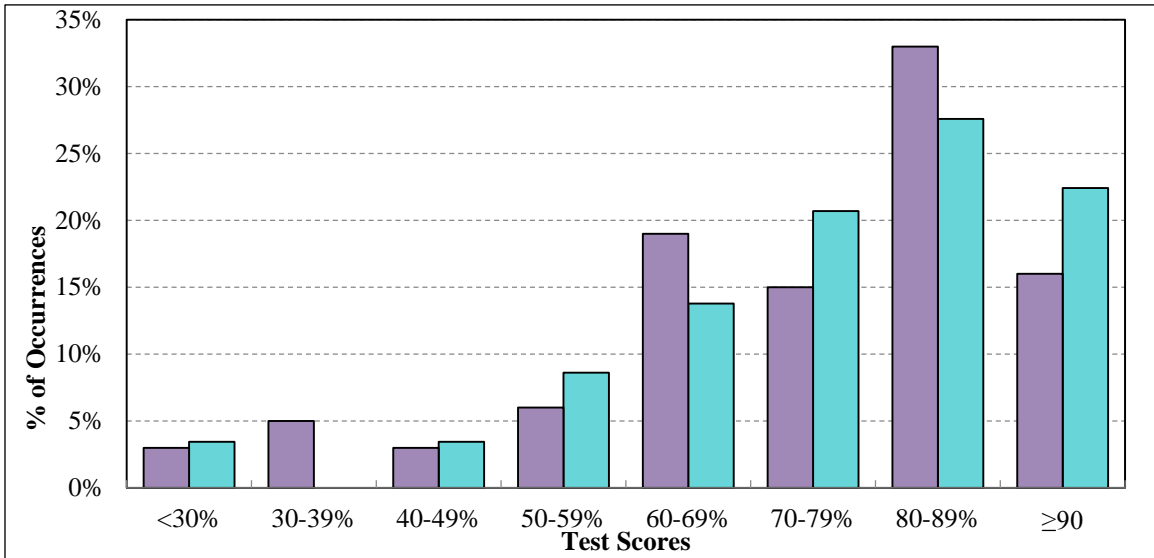


Figure 19. Histogram of SPN 1121 Combined (Overall) scores comparing spring 2014 (purple) with spring 2015 (teal).

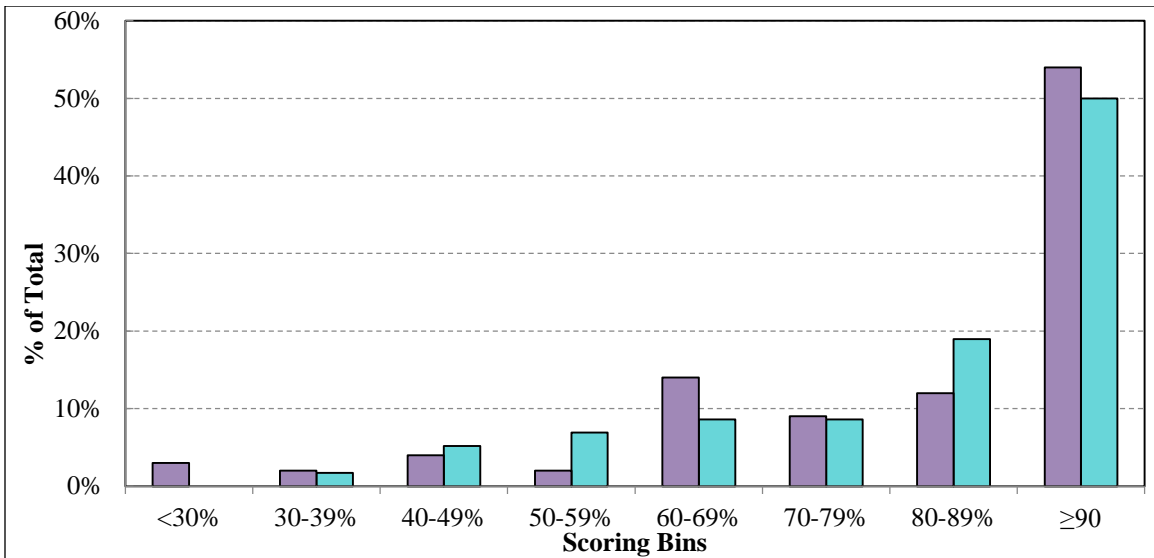


Figure 20. Histogram of SPN 1121 Section I scores comparing spring 2014 (purple) with spring 2015 (teal).

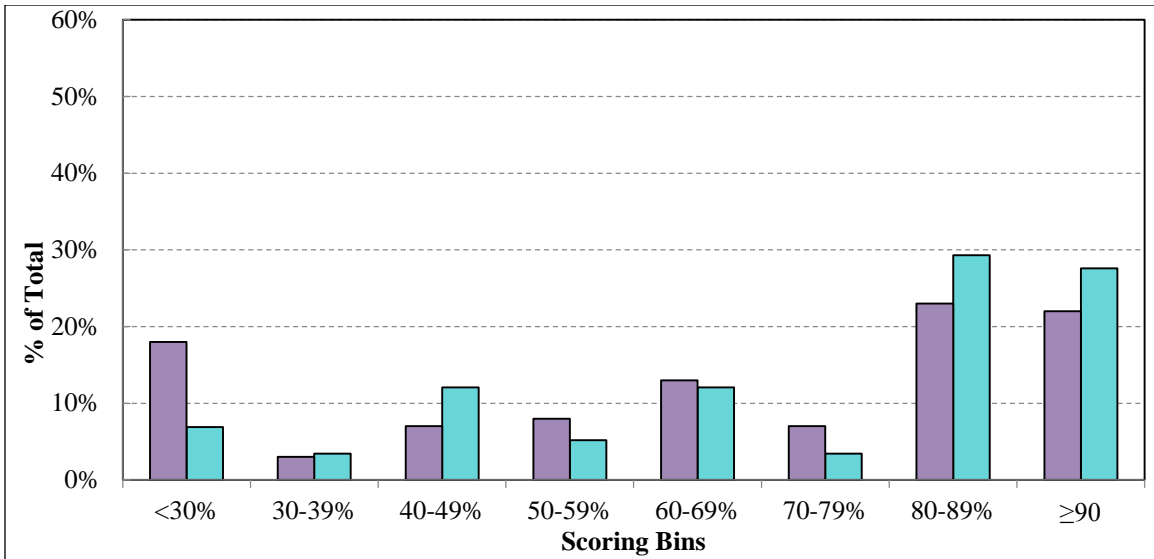


Figure 21. Histogram of SPN 1121 Section II scores comparing spring 2014 (purple) with spring 2015 (teal).

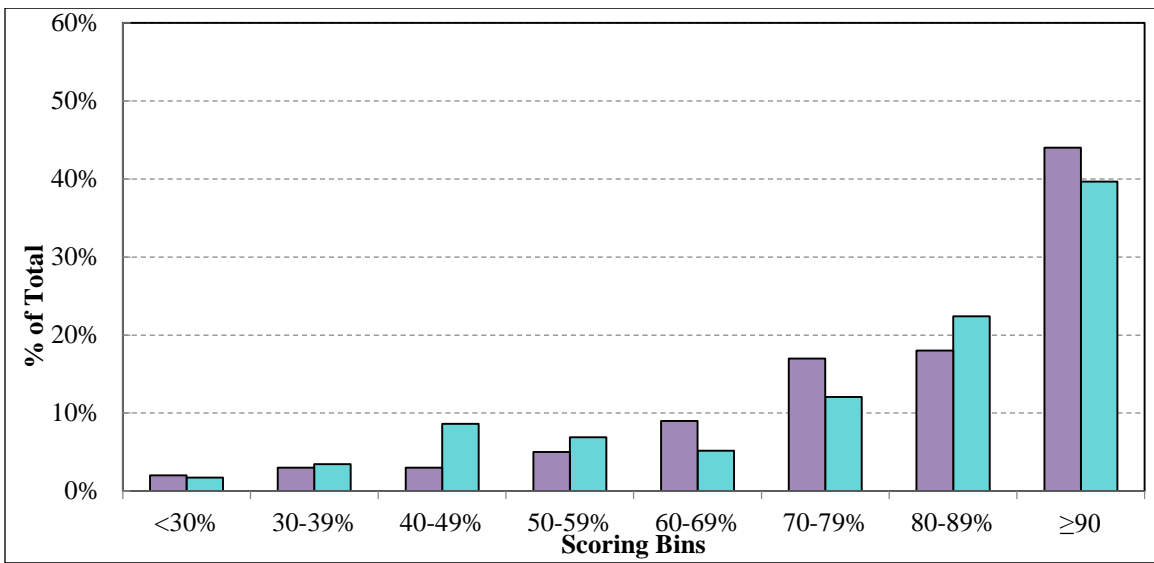


Figure 22. Histogram of SPN 1121 Section III scores comparing spring 2014 (purple) with spring 2015 (teal).

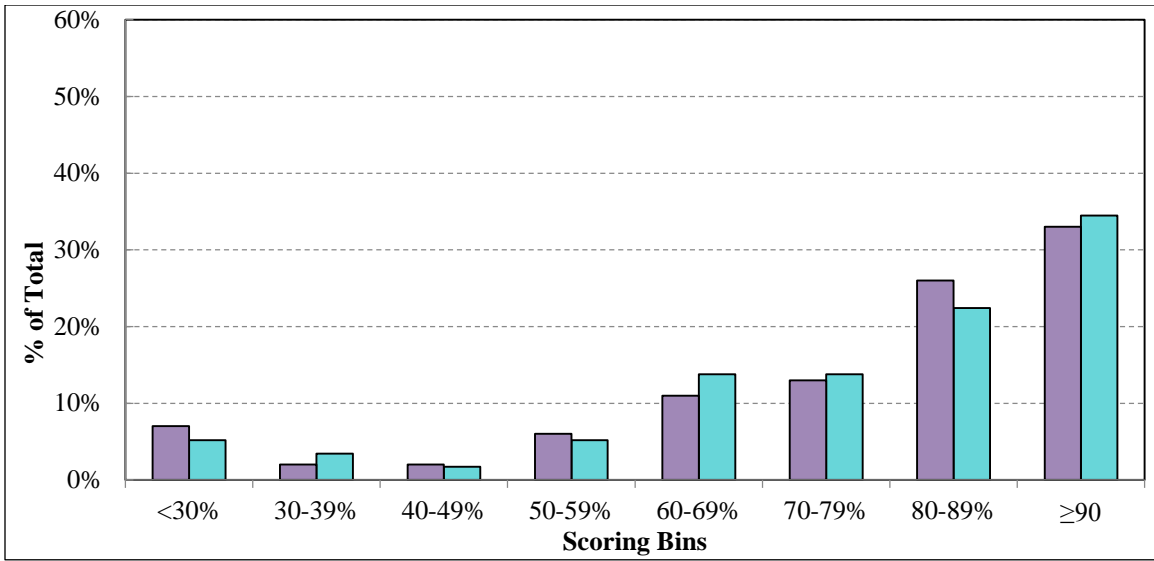


Figure 23. Histogram of SPN 1121 Section IV scores comparing spring 2014 (purple) with spring 2015 (teal).

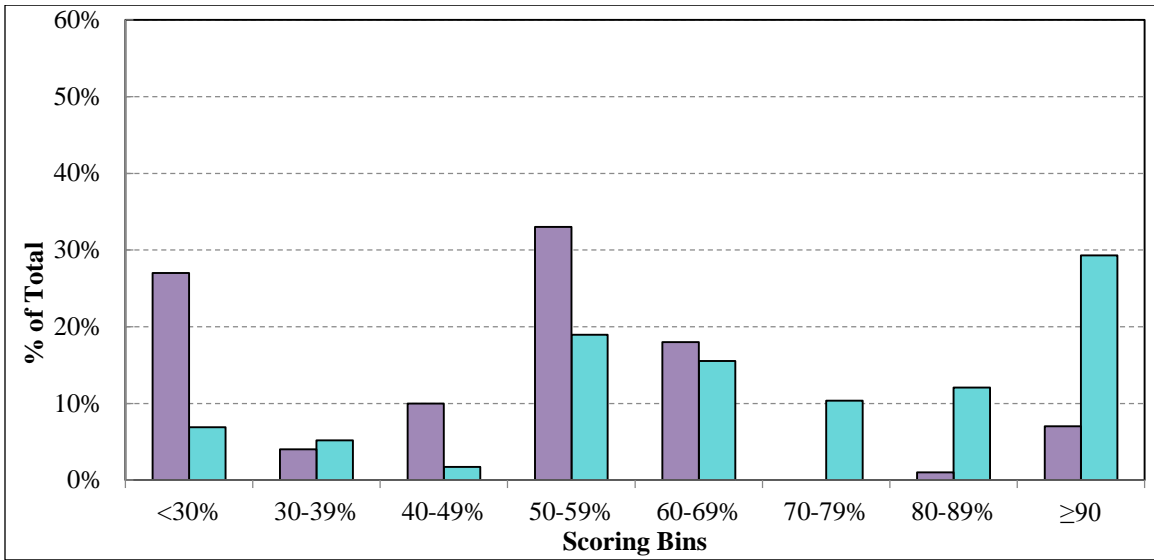


Figure 24. Histogram of SPN 1121 Section V scores comparing spring 2014 (purple) with spring 2015 (teal).

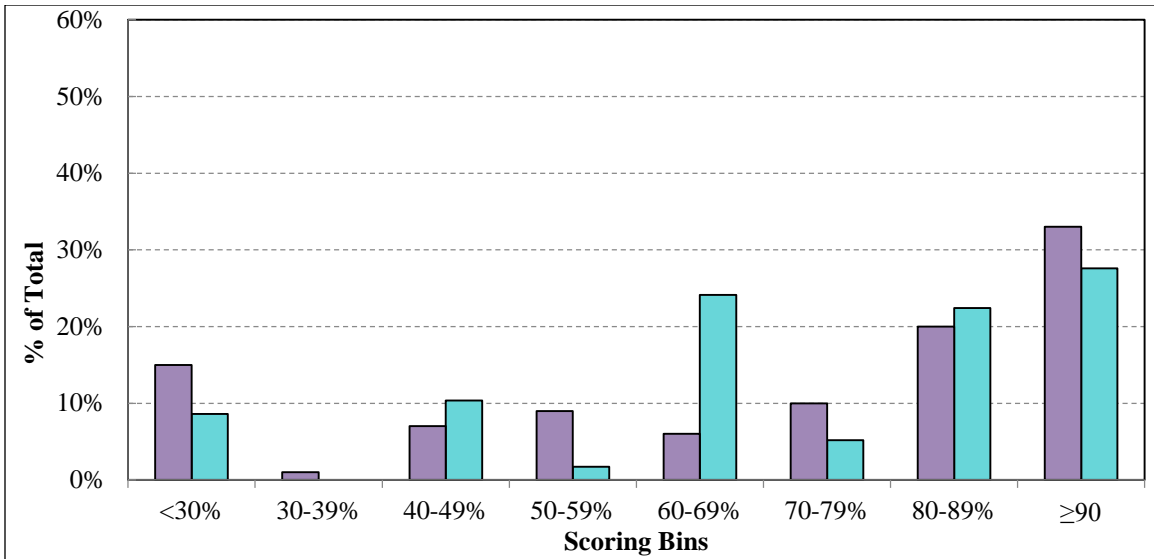


Figure 25. Histogram of SPN 1121 Section VI scores comparing spring 2014 (purple) with spring 2015 (teal).

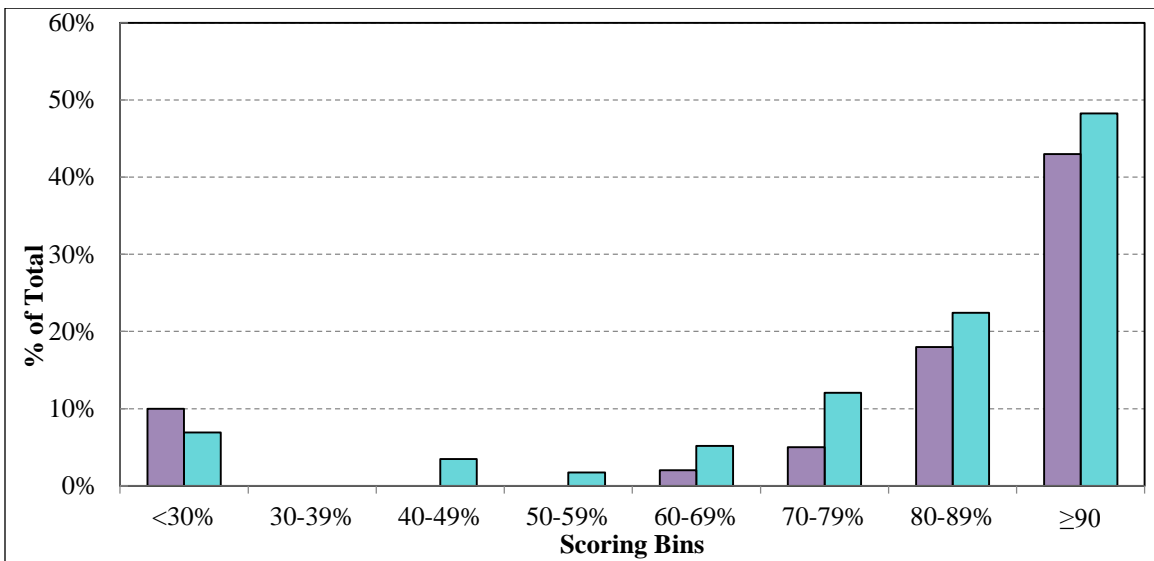


Figure 26. Histogram of SPN 1121 Section VII scores comparing spring 2014 (purple) with spring 2015 (teal).

Sections II, IV, V, and VII exhibit a positive shift (negative skew) of the data when compared with spring 2014 (see Figures 21, 23, 24, 26). Specifically, Section V exhibits substantial improvement with the percentage of As earned jumping from 7% to 29%.

A comparison of mean scores for each exam section by semester is reported in Table 12. 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). When comparing spring-to-spring, five of seven sections show improvement from spring 2014 to spring 2015. The declines and improvements across sections, when tested using a Welch's t-test according to standard methods (Davis, 1973; de Winter 2013;

McDonald, 2009; Wilkinson, 1999), were found to be statistically significantly different in Section V only ($t(171)=5.818, p=5.28 \times 10^{-8}$) with an effect size of 0.89.

	<i>Section Max</i>	Fall 2013 n=10	Spring 2014 n=115	Fall 2014 n=25	Spring 2015 n=58
Section I (Oral)	15	11.5	12.3	11.9	12.2
Section II (Reading)	15	9.5	9.6	10.7	10.6
Section III (Read/Write)	40	34.2	32.3	30.0	31.1
Section IV (Read/Write)	15	9.5	11.4	10.6	11.4
Section V (Writing)	12	7.5	5.7	9.5	8.5
Section VI (Reading)	15	9.6	10.3	11.6	10.7
Section VII (Writing)	20	14.2	15.4	16.1	16.4
Combined (Overall)	128	96.0	97.0	100.5	100.9

Table 13. Comparison of mean scores for SPN 1121 for fall 2013 through spring 2015. Bold denotes statistically significant differences from spring 2014 to spring 2015. Red denotes decrease from spring 2014 to spring 2015.

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 2015 assessment.

4.1 FRENCH

A drilldown of FRE 1120 results are as follows:

1. Achievement of 80% of artifacts scoring 70% or better (SLO1): Achievement was not analyzed in this report as a new section to be included in assessment was in development during the spring 2015 semester.
2. Achievement of 80% of artifacts scoring 70% or better (SLO2): Achievement goal was met as results exhibit 93% of artifacts score 70% or higher in the reading competency section (Section X) of the exam.
3. Achievement of 80% of artifacts scoring 70% or better (SLO3): Achievement goal was met as results exhibit 93% of artifacts score 70% or higher in the writing effectiveness section (Section I) of the exam.
4. In a complete study of all sections of the assessed exam, Section V and VIII, at 47% and 47% achieving 70% or greater, are the lowest scoring sections.
5. No comparisons of dual enrollment to non-dual enrollment students could be made because only four Dual Enrollment (DE) artifacts were collected from a total collection of just 15 for the spring 2015 semester and statistical comparisons would not be reliable or meet criteria for significance (de Winter, 2013). No DE artifacts originated from offsite so this comparison is strictly a demographic comparison and not a comparison of instruction at multiple locations.
6. No artifacts from online sections of the course were collected during spring 2015 so no comparison study between online and traditional could be completed.
7. No campus/site comparison could be made because of all the artifacts collected, all originated from the Thomas Edison (Lee) campus.

8. In a comparison of exam sections by scoring distribution, most sections exhibit S-curves where the largest percentages of artifacts occur in the highest scoring bins with progressively smaller percentages in lower scoring bins, a trait that is normal for a multi-section exam.
9. In the same study, Sections III and IX exhibit increased percentages of artifacts in the lower scoring bins compared with other sections.
10. Typically, a study comparing results year-to-year would be completed in this report, however, since no study was completed for spring 2014 no comparisons are reported here.

Below is a drilldown of the FRE 1121 assessment results:

1. Achievement of 80% of artifacts scoring 70% or better (SLO1): Achievement was not analyzed in this report as a new section to be included in assessment was in development during the spring 2015 semester.
2. Achievement of 80% of artifacts scoring 70% or better (SLO2): Achievement goal was not met as results exhibit 73% of artifacts score 70% or higher in the reading competency section (Section XI) of the exam.
3. Achievement of 80% of artifacts scoring 70% or better (SLO3): Achievement goal was met as results exhibit 91% of artifacts score 70% or higher in the writing effectiveness section (Section I) of the exam.
4. In a complete study of all sections of the assessment exam, Section VIII, at 27% achieving 70% or greater, is the lowest scoring section. Other low areas include Section IV (36%), Section III (45%) and Section IX (45%).
5. No comparisons of dual enrollment to non-dual enrollment students could be made because only one Dual Enrollment (DE) artifact was collected from a total collection of just 11 for the spring 2015 semester and statistical comparisons would not be reliable or meet criteria for significance (de Winter, 2013). No DE artifacts originated from offsite so this comparison is strictly a demographic comparison and not a comparison of instruction at multiple locations.
6. No comparisons of online to traditional students could be made because no artifacts from online sections of the course were collected during spring 2015.
7. No campus/site comparison could be made because of all the artifacts collected, all originated from the Thomas Edison (Lee) campus.
8. Typically, a study comparing results year-to-year would be completed in this report, however, since no study was completed for spring 2014 no comparisons are reported here.

4.2 SPANISH

A drilldown of SPN 1120 results are as follows:

1. Achievement of 80% of artifacts scoring 70% or better (SLO1): Achievement was not met as results exhibit only 68.5% of artifacts score 70% or higher in the oral competency exam section (Section I).
2. Achievement of 80% of artifacts scoring 70% or better (SLO2): Achievement was partially met as results exhibit 50.7% of artifacts scored 70% or higher in the first reading section, Section II and 87.7% of artifacts scored 70% or higher in the second reading section, Section III.
3. Achievement of 80% of artifacts scoring 70% or better (SLO3): Achievement was partially met as results exhibit 67.1% of artifacts scored 70% or higher in the first writing section, Section IV, and 84.9% of artifacts scored 70% or higher in the second writing section, Section V.

4. No comparisons of dual enrollment to non-dual enrollment students could be made because sample sizes were too low to yield meaningful results and statistical comparisons would not be reliable or meet criteria for significance. No DE artifacts originated from offsite so this comparison is strictly a demographic comparison and not a comparison of instruction at multiple locations.
5. No comparisons of online to traditional students could be made because no artifacts from online sections of the course were collected during spring 2015.
6. In a cross-campus comparison, of the 73 artifacts collected from SPN 1120, 16 originated from the Collier Campus and 57 from the Thomas Edison (Lee) Campus, scores by section are consistently higher at the Collier campus which results in a 10.0 point margin between sites for the overall score. The difference in overall mean scores was found to be marginally statistically significantly different.
7. In a comparison of exam sections by scoring distribution, Sections I, III, IV, and V exhibit S-curves where the largest percentages of artifacts occur in the highest scoring bins with progressively smaller percentages in lower scoring bins, a trait that is normal for a multi-section exam.
8. In the same study, the S-curve distribution for Section II is not an S-curve but rather bimodal (two peaked). Section II exhibits 10% or more of artifacts in the 50-59 and 40-49 scoring bins where Sections I, III, IV, and V fall well below that .
9. In a comparison of section scores based on overall score, Section II is underperforming compared to other sections between the range of 65%-84%. In other words, artifacts with overall scores ranging from 65%-84% tend to exhibit the lowest scores in Section II while other areas perform at or above average. This is, of course, in part due to the fact that a large percentage of the assessment is weighed in Section II and so scores in that section can tend to reflect heavily in the overall score.
10. In that same study, at the 45-49% range, the Section IV and V mean scores are 80% and 70%, respectively, whereas other sections range from 10%-40%.
11. In a study comparing spring 2015 achievement with spring 2014, overall scores ≥ 90 dropped by 12% from year-to-year while the 50-59% increased by 10%.
12. In a similar year-to-year comparison study for each assessment exam section, Section I data exhibit a fairly typical S-Curve although not as pronounced as spring 2014.
13. In a study comparing mean scores for each exam section by semester, all sections show a drop from spring 2014 to spring 2015. Of those, only Section II results are statistically significantly different from year-to-year.

A drilldown of SPN 1121 results are as follows:

1. Achievement of 80% of artifacts scoring 70% or better (SLO1): Achievement was nearly met as results exhibit 77.6% of artifacts score 70% or higher in the oral competency exam section (Section I).
2. Achievement of 80% of artifacts scoring 70% or better (SLO2): Achievement was not met in the reading only sections as results exhibit 60.3% of artifacts scored 70% or higher in Section II and 55.2% of artifacts scored 70% or higher in Section VI.
3. Achievement of 80% of artifacts scoring 70% or better (SLO3): Achievement was partially met in the writing only sections as results exhibit 51.7% of artifacts scored 70% or higher in Section V and 82.8% of artifacts scored 70% or higher in Section VII.

4. In a comparison of dual enrollment to non-dual enrollment, scores are consistently higher for dual enrollment artifacts in all sections with an overall score margin of +13.1. No DE artifacts originated from offsite so this comparison is strictly a demographic comparison and not a comparison of instruction at multiple locations.
5. No comparisons of online to traditional students could be made because no artifacts from online sections of the course were collected during spring 2015.
6. In a cross-campus comparison, of the 58 artifacts collected from SPN 1121, 5 originated from the Charlotte campus, 12 from the Collier campus and 41 from the Thomas Edison (Lee) Campus and scores by section are consistently higher at the Collier campus which results in a 20.1 point margin between sites for the overall. The difference in overall mean scores was found to be statistically significantly different.
7. In a comparison of exam sections by scoring distribution, Sections I, II, III, IV, and VII exhibit S-curves typical of sectional score distribution with the mode centered on the highest scoring bin, a trait that is normal for a multi-section exam. Sections V and VI exhibit a mode centered on the 80-89 scoring bin.
8. In the same study, Section IV exhibits a secondary bimodal distribution with peaks centered on ≥ 90 and 50-59 in the case of Section V and 60-69 in the case of Section VI.
9. In a comparison of section scores based on overall score, Sections II through VI are slightly underperforming compared to other Sections I and VII between the range of 65%-79%. In other words, artifacts with overall scores ranging from 65%-79% tend to exhibit the lowest scores in Sections II through VI while Sections I and VII perform at or above average.
10. In a study comparing spring 2015 achievement with spring 2014, overall scores exhibit substantial improvement in number of artifacts scoring an A from year-to-year, however, with fewer Bs and Cs compared to spring 2014.
11. In a similar year-to-year comparison study for each assessment exam section, Sections II, IV, V, and VII exhibit a positive shift (negative skew) of the data when compared with spring 2014.
12. In the same study, Section V exhibits substantial improvement with the percentage of As earned jumping from 7% to 29%.
13. In a study comparing mean scores for each exam section by semester, five of seven sections show improvement from spring 2014 to spring 2015. Section V results (an improvement from 5.7/12 to 8.5/12) were found to be statistically significantly different.

5 REFERENCES

- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Hillsdale, NJ.
- Davis, J.C. 1973. *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York, New York, 564 pp.
- de Winter, J.C.F. 2013. Using the Student's T-Test with Extremely Small Sample Sizes. *Practical Assessment, Research, and Evaluation*, 18(10), 1-12.
- Hall, J.C. 2010. *Doing More with Less: Making Colleges Work Better*. Springer, New York, New York, 250 pp.

- Johnson, V. 2013. Revised Standards for Statistical Evidence. *Proceedings of the National Academy of Science*, 110(48), 19313-19317.
- 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.
- Wilkinson, L. 1999. APA Task Force on Statistical Inference. *Statistical Methods in Psychology Journals: Guidelines and Explanations*. *American Psychologist* 54 (8), 594–604.