

English Assessment Report – Fall 2015

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

1 INTRODUCTION

Fall 2014 marked the beginning of a new assessment plan for the English Department of Florida SouthWestern State College (FSW) in three courses: ENC 0022 *Writing for College Success*, ENC 1101 *Composition I*, and ENC 1102 *Composition II*. The planned assessment practice continues in fall 2015 in which instructors use a common rubric with seven identified rubric dimensions in the case of ENC 0022, and five dimensions for both ENC 1101 and ENC 1102. The assessment plan uses a random sample of 20% of all course sections offered in ENC 1101 and ENC 1102. In the case of ENC 0022, because it is a course being assessed by assessment plans in addition to the English Department (developmental assessment) all course sections for ENC 0022 are assessed.

The standard assessment plan highlighted above is designed to evaluate each course and inform faculty on Student Learning Objectives (SLOs) for future assessment plans. Additionally, the plan provides information on achievement levels of Dual Enrollment artifacts compared with non-Dual Enrollment, as well as Online artifacts compared with traditional artifacts as highlighted in the QEP course level assessment plan. Other analyses such as comparison by term length (standard vs. mini-term) and longitudinal studies are included.

In response to AY 2014-2015 assessment results, a revised “Thesis” and “Evidence” rubric criteria will be piloted with select instructors in an effort to improve rater reliability. The pilot program includes three instructors across seven course sections. The pilot study will be assessed using the same common rubric as the standard assessment with the exception of the piloted “Thesis” and “Evidence” criteria.

For additional detail or 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 ENC 0022

2.1 LEARNING OBJECTIVES & DESCRIPTIVE STATISTICS

Using common rubric criterion as an assessment method, the FSW English faculty defined multiple areas of interest for evaluation based on core outcomes for the course. Those outcomes include:

- Plan and write paragraphs and essays reflecting styles and tones appropriate for their audience and use adequate support, coherence, and unity that demonstrate understanding of content for expository and persuasive purposes.
- Establish a substantive claim, link claims to relevant evidence, and acknowledge competing arguments, gather information needed, and accurately incorporate source material into their own writing to avoid plagiarism.
- Identify and correctly use proper conventions for sentence grammar and avoid illogical shifts in pronouns and verbs in their own writing and on tests.

- Identify and use proper conventions for spelling, capitalization, and punctuation in their own writing and on tests.
- Identify and correctly use the conventions of a variety of sentence structures and will be able to avoid sentence fragments, comma splices, and fused sentences in their own writing and on tests.
- Identify and write effective topic sentences and thesis statements that address task and audience and use logical structure, support, and transitional devices for expository and persuasive purposes.

2.1.1 Learning Objectives

ENC 0022 is scored using a rubric with seven dimensions: Introductory Paragraph, Support Paragraphs, Organization, Concluding Paragraph, Grammar, Mechanics, and Research. Each dimension is scored on a scale of 1 to 4 (1-Unacceptable, 2-Needs work, 3-Average, 4-Above average), with 0s if the baseline of 'Unacceptable' is not met. The English department has identified a target statistic for measurement purposes (SLO1) of measuring the percentage of artifacts scoring a 2 or greater.

For the fall 2015 assessment, 141 artifacts were collected for ENC 0022 from 11 of 14 course sections. The lowest scoring rubric dimension for percentage of artifacts scoring a 2 or greater is Research at 87%. All other dimensions exhibit percentage of 96% or higher (Table 1). For a visual comparison of scores by dimension, see Figure 1.

Rubric Score	Introductory Paragraph	Support Paragraphs	Organization	Concluding Paragraph	Grammar	Mechanics	Research
Developing or higher	98%	99%	100%	100%	98%	96%	87%
4	33%	32%	33%	29%	16%	14%	32%
3	58%	59%	57%	55%	59%	58%	44%
2	7%	9%	10%	15%	23%	25%	11%
1	2%	1%	0%	0%	2%	3%	13%
0	0%	0%	0%	0%	0%	0%	0%

Table 1. Percentage of student achievement level by rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 0022.

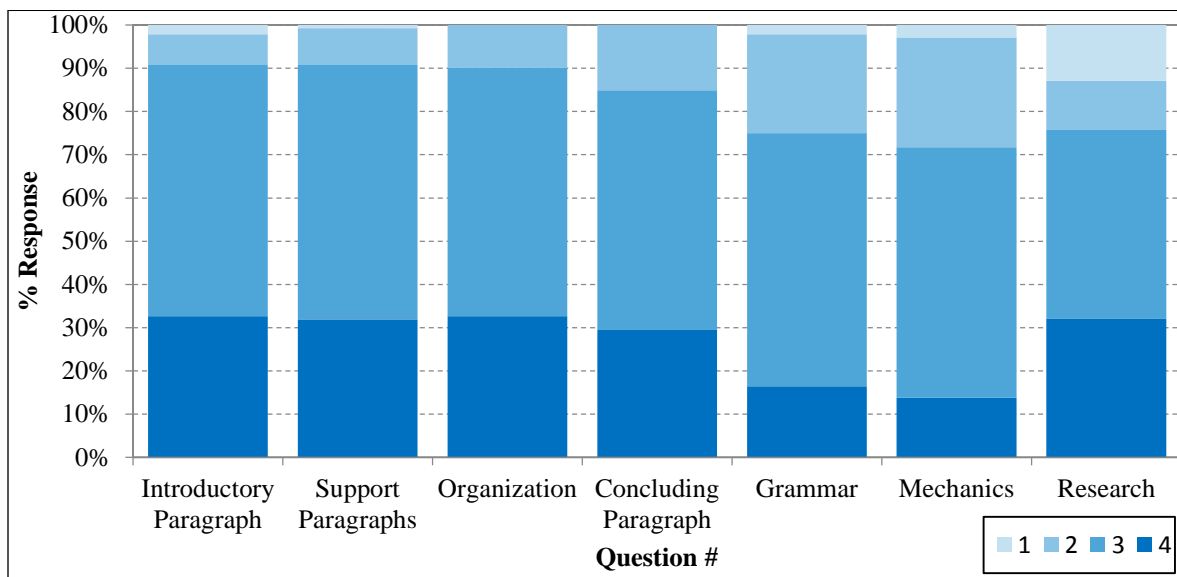


Figure 1. ENC 0022 distribution of rubric scores by dimension.

2.1.2 Descriptive Statistics & Longitudinal Studies

Descriptive statistics for ENC 0022 artifacts can be found in Table 2. A histogram of artifact scores for all 141 artifacts is shown in Figure 2. Distribution of artifact scores is bimodal centered on 21/28 and 28/28, and is moderately negatively skewed, meaning scores are shifted towards the higher range. To describe the behavior of the rubric dimensions based on overall achievement a color map, or binary raster image was created by calculating the mean scores for each dimension as a function of combined score (Figure 3). To create this image the rubric scores (4, 3, 2, 1, or 0) for each artifact was grouped based on combined raw rubric score (7 dimensions x maximum rubric level of 4 = 28 overall points). The color represents the mean rubric score achieved in each dimension based on the combined score as shown in the x-axis.

	Introductory Paragraph	Support Paragraphs	Organization	Concluding Paragraph	Grammar	Mechanics	Research	TOTAL
n	141	141	141	139	140	139	140	137
Max	4	4	4	4	4	4	4	28
Min	1	1	2	2	1	1	1	10
Median	3	3	3	3	3	3	3	21
Mode	3	3	3	3	3	3	3	21
Mean	3.2	3.2	3.2	3.1	2.9	2.8	3.0	21.5
Standard deviation	0.66	0.62	0.61	0.65	0.69	0.69	0.98	3.89
Skewness	-0.71	-0.37	-0.17	-0.16	-0.27	-0.30	-0.74	-0.31
Kurtosis	1.22	0.31	-0.52	-0.67	0.10	0.13	-0.36	0.16

Table 2. Descriptive statistics for ENC 0022 common course assessment.

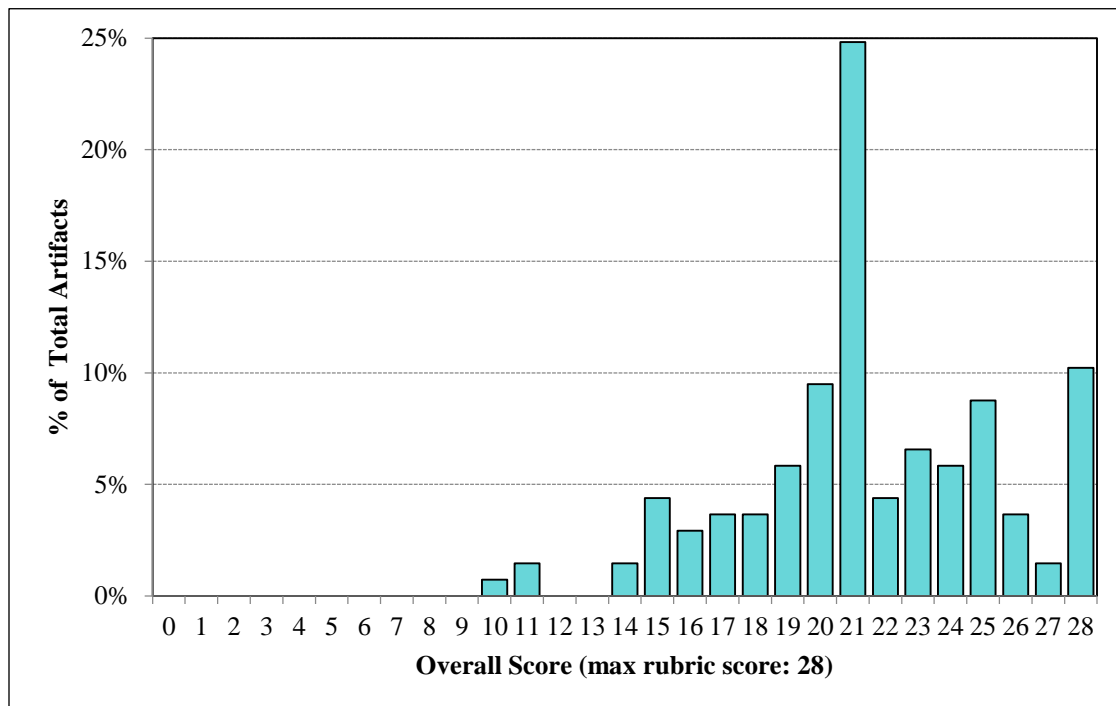


Figure 2. Overall score distribution for ENC 0022 artifacts (fall 2015 term).

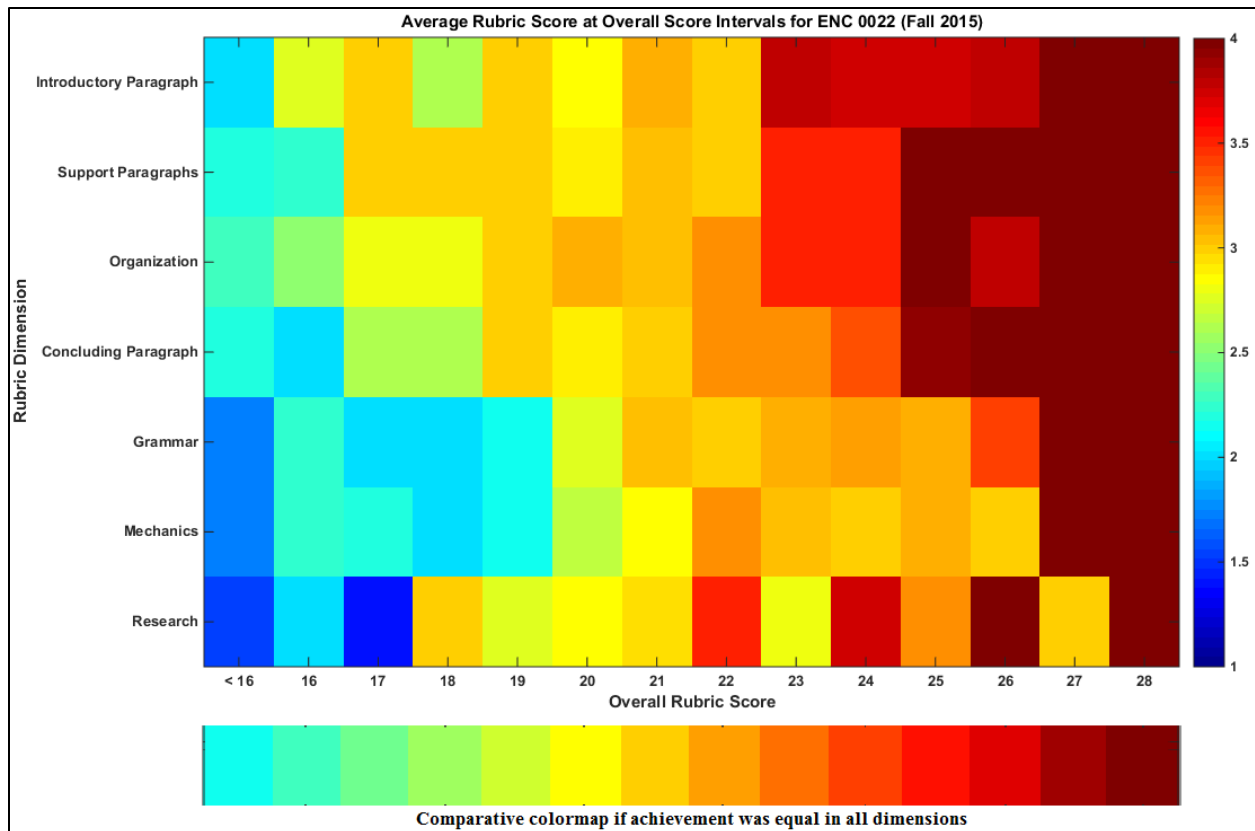


Figure 3. (Top) Colormap of mean scores for each rubric dimension (range: 0-4) based on overall rubric score (combined rubric score of all dimensions, max=28) for ENC 0022. (Bottom) Comparison rubric dimension if dimension score is the same as overall (i.e. artifact overall score is equally distributed across all sections). A rubric dimension with hotter colors (reds/yellows) means that dimension 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 2 shows that between 20/28 and 22/28 (approximately 75% overall score) all dimensions fair relatively equally (hot colors fairly evenly distributed). When overall rubric scores range below 20/28 (below 75%) achievement in the Grammar and Mechanics dimensions lags substantially behind all other dimensions. For example, at an overall score of 18/28, Grammar and Mechanics both exhibit average scores of 2.0/4 while the other five dimensions range from 2.6/4 to 3.0/4. This lag is also visible at higher scores (ranging 25/28 and 26/28) in which Grammar and Mechanics exhibit average scores ranging from 3.0/4 to 3.4/4 while other dimensions range from 3.8/4 to 4/4. From a student performance perspective, under achieving students tend to be exceedingly lagging in Grammar and Mechanics, average students tend to be equally strong in all dimensions, and over achieving students tend to again lag in Grammar and Mechanics. In short, there is an upper limit to which even the best students do not attain for Grammar and Mechanics.

A comparison of fall 2015 results with past results is shown in Figure 4 below. Results exhibit improvement across all areas. However, some artifacts reported in fall 2014 may have been reported as 0s rather than non-reporting. As a result, scores for fall 2014 may be slightly artificially dampened. Discussion of the issue following fall 2014 results is expected to have rectified the issue.

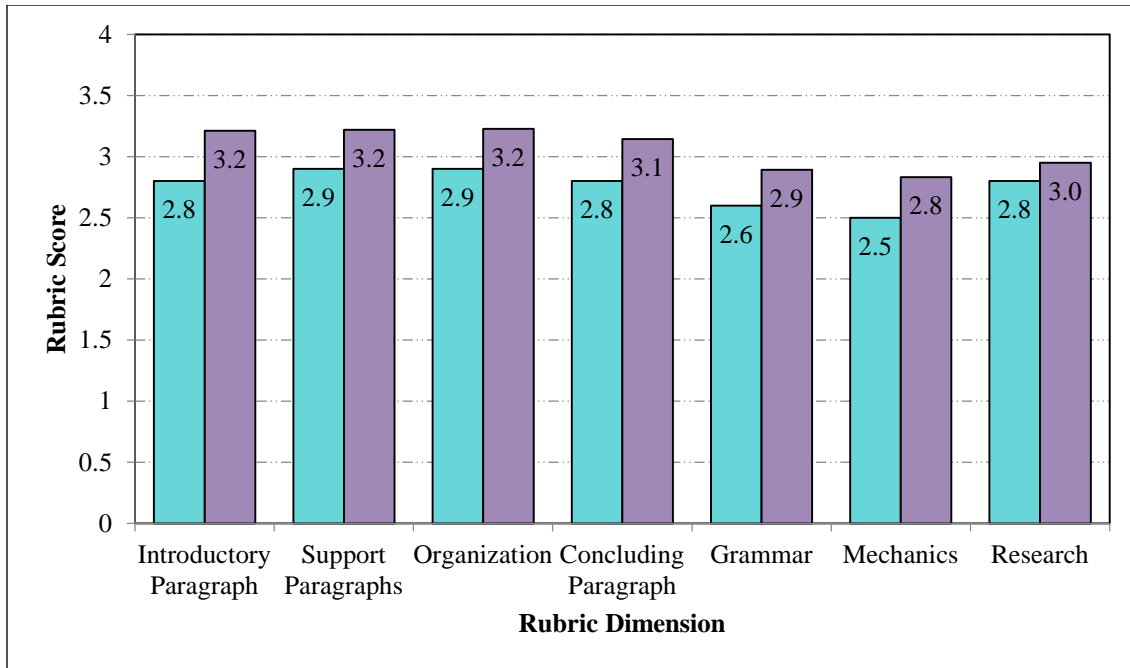


Figure 4. Comparison of mean scores for ENC 0022 through time beginning fall 2014 (teal) and fall 2015 (purple).

2.2 COMPARISONS BY SITE, FORMAT, AND STUDENT TYPE

2.2.1 Dual Enrollment to non-Dual Enrollment Comparison

ENC 0022 is not offered as a dual enrollment (offsite) course nor is it offered to dual enrollment students onsite and so no comparison study between dual enrollment artifacts and traditional artifacts can be made.

2.2.2 Online to Traditional Comparison

ENC 0022 is not offered as an online course and so no comparison study between online artifacts and traditional artifacts can be made.

2.2.3 Comparison by Site/Campus

Of the 141 artifacts collected from ENC 0022, 11 originated from the Charlotte campus, 39 from the Collier campus, 7 from the Hendry-Glade Center, and 80 from the Thomas Edison (Lee) campus. Scores by rubric dimension varied greatly across campuses. A comparison of mean scores by rubric dimension is provided in Table 3.

	Introductory Paragraph	Support Paragraphs	Organization	Concluding Paragraph	Grammar	Mechanics	Research
Charlotte	3.5	3.4	3.2	3.2	3.2	2.6	2.7
Collier	3.0	3.1	3.2	3.1	2.8	2.7	3.2
Hendry-Glades	3.1	3.3	3.7	3.4	3.0	3.0	2.6
Thomas Edison (Lee)	3.3	3.2	3.2	3.2	2.9	2.9	2.9

Table 3. Comparison of mean scores by site for ENC 0022. Bold denotes highest mean score in that dimension among all sites.

No site is consistently higher compared to other sites, however, the Thomas Edison (Lee) campus is consistently the lowest. Charlotte campus exhibits higher scores in 3 of 7 dimensions. Hendry-Glades also exhibits higher scores in 3 of 7 dimensions while Collier exhibits higher scores in 1 of 7 dimensions. A plot comparing descriptive statistics of the combined (overall) scores by site is presented in Figure 5. There is extensive overlap between sites with both Charlotte and Hendry-Glades exhibiting a smaller range of scores, not uncommon given the smaller sample size for both.

A one-way analysis of variance was used to compare means of the combined rubric scores at each site. Results of the ANOVA exhibit no statistically significant difference between sites (see Table 4). Therefore, we cannot reject the null hypothesis that the mean rubric scores at each site are equal to each other and we cannot conclude with a 95% confidence that the differences in scores are not solely due to chance.

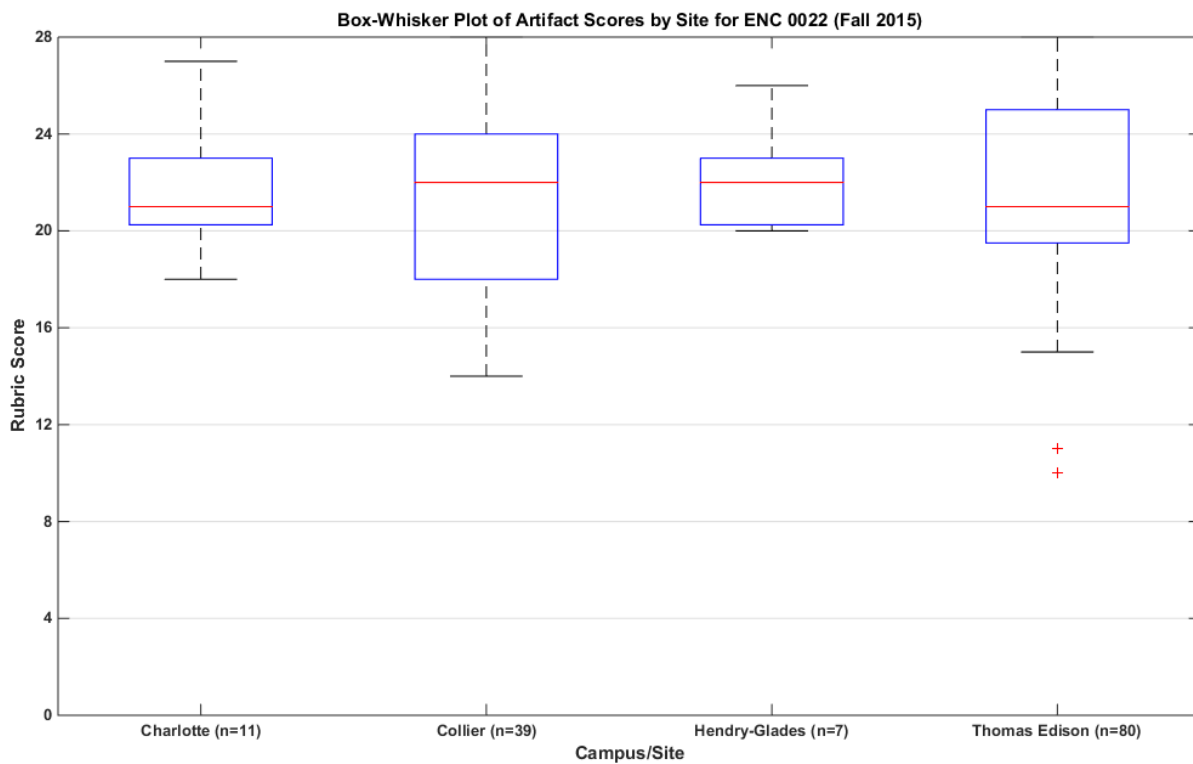


Figure 5. Box-Whisker plot of scores distributed by site for ENC 0022. Red line depicts median score. Upper and lower box boundaries indicate 75% quartile and 25% quartile (box represents central 50% of the scores). Vertical lines represent remaining scores outside central 50% that are not outliers. Red '+'s denote outliers.

Source of Variation	Sum of squared differences	df	Mean Squares	F _{obs}	p-value	F _{crit}
Between Sites	14.9	3	5.0	0.32	0.810	2.67
Within Sites	2048.1	133	15.4			
Total	2063.0	136				

Table 4. Results of one-way ANOVA of combined rubric scores at each site for ENC 0022.

2.2.4 Mini-term to Full-term Comparison

ENC 0022 was not offered as a mini-term course and so no comparison study between mini-term artifacts and full-term artifacts can be made.

3 ENC 1101

3.1 LEARNING OBJECTIVES & DESCRIPTIVE STATISTICS

Using common rubric criterion as an assessment method, the FSW English faculty defined multiple areas of interest for evaluation based on core outcomes for the course. Those outcomes include:

- Students must demonstrate the ability to write essays following various rhetorical modes, strategies, and purposes.
- Students must demonstrate effective research skills, and incorporate documented direct quotations and paraphrases from a variety of sources, using MLA format.

3.1.1 Learning Objectives

ENC 1101 is scored using a rubric with five dimensions: Thesis, Evidence, Organization / Style, Grammar / Mechanics, and Documentation. Each scored on a scale of 1 to 4 (1-Does not meet standards, 2-Approaching standards, 3-Meets standards, 4-Exceeds standards), with 0s if the benchmark is not met. The English department has identified a target statistic for measurement purposes (SLO1) of measuring the percentage of artifacts scoring a 2 or greater.

For the fall 2015 assessment, 413 artifacts were collected for ENC 1101 from 22 of 30 course sections sampled from 150 course sections offered. The resultant sample represents 14.7% of the population. The lowest scoring rubric dimension for percentage of artifacts scoring a 2 or greater is Grammar / Mechanics at 93% (Table 5). For a visual comparison of scores by dimension, see Figure 6.

Rubric Score	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation
Developing or higher	95%	95%	95%	93%	94%
4	46%	40%	38%	24%	30%
3	31%	31%	41%	49%	39%
2	18%	24%	16%	21%	25%
1	5%	5%	5%	7%	6%
0	0%	0%	0%	0%	0%

Table 5. Percentage of student achievement level by rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1101.

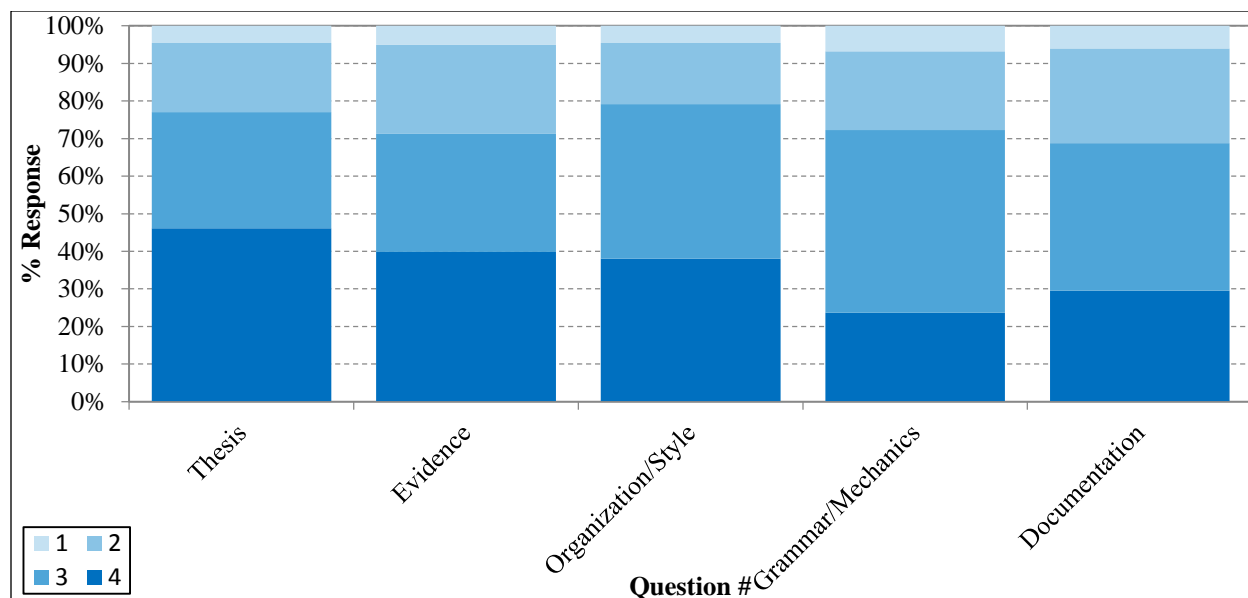


Figure 6. ENC 1101 distribution of rubric scores by dimension.

3.1.2 Descriptive Statistics

Descriptive statistics for ENC 1101 artifacts can be found in Table 6. A histogram of artifact scores for all 413 artifacts is shown in Figure 7. Distribution of artifact scores is centered on 17/20 and is moderately negatively skewed, meaning scores are shifted towards the higher range. To describe the behavior of the rubric dimensions based on overall achievement a color map, or binary raster image was created by calculating the mean scores for each dimension as a function of combined score (Figure 8). To create this image the rubric scores (4, 3, 2, 1, or 0) for each artifact was grouped based on combined raw rubric score (5 dimensions x maximum rubric level of 4 = 20 overall points). The color represents the mean rubric score achieved in each dimension based on the combined score as shown in the x-axis.

	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation	TOTAL
n	414	414	413	414	413	413
Max	4	4	4	4	4	20
Min	1	1	1	1	1	4
Median	3	3	3	3	3	16
Mode	4	4	3	3	3	17
Mean	3.2	3.1	3.1	2.9	2.9	15.2
Standard deviation	0.89	0.91	0.84	0.84	0.89	3.53
Skewness	-0.76	-0.52	-0.70	-0.48	-0.37	-0.66
Kurtosis	-0.43	-0.81	-0.17	-0.27	-0.70	-0.23

Table 6. Descriptive statistics for ENC 1101 common course assessment.

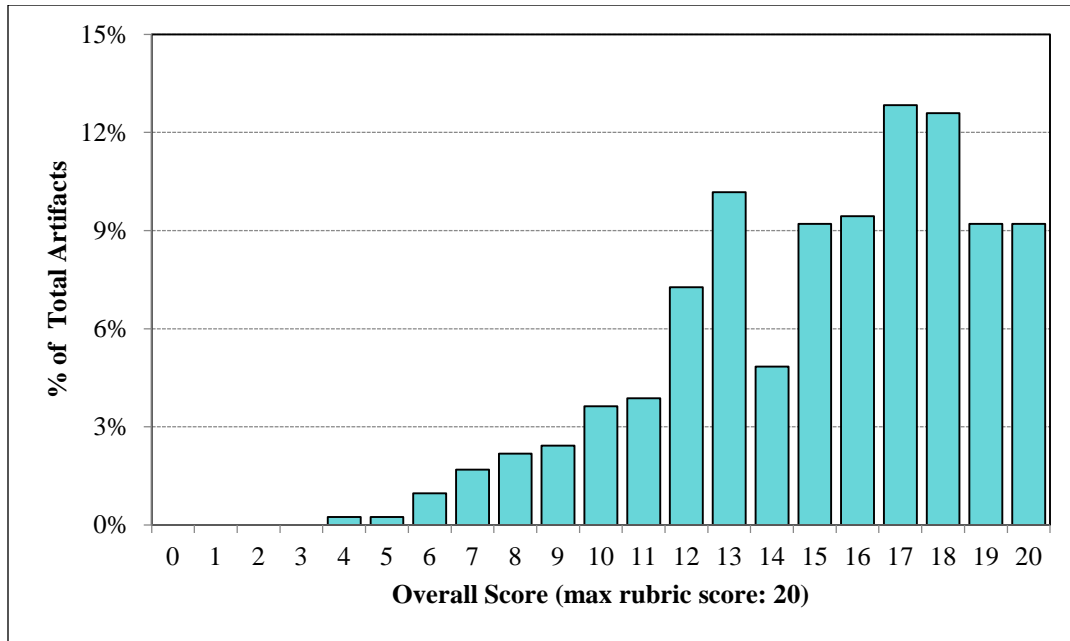


Figure 7. Overall score distribution for ENC 1101 artifacts (fall 2015 term).

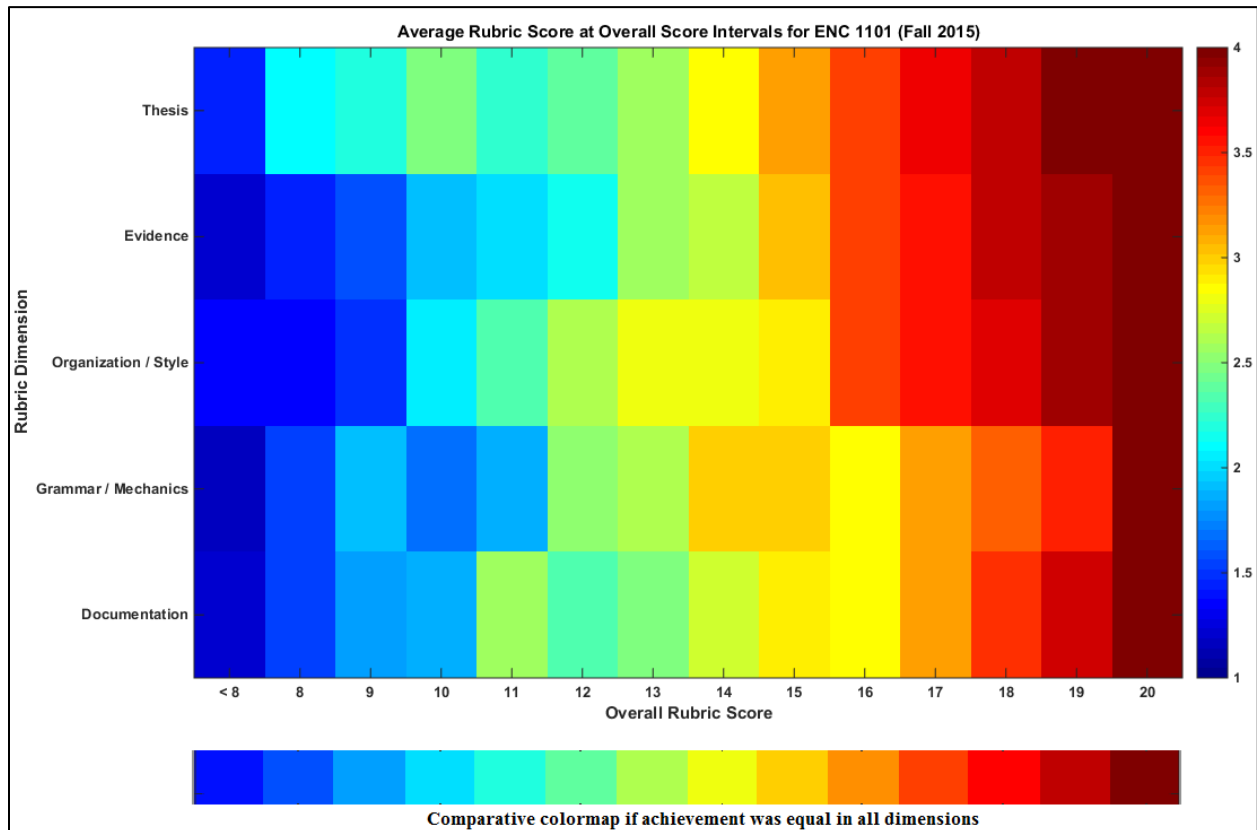


Figure 8. (Top) Colormap of mean scores for each rubric dimension (range: 0-4) based on overall rubric score (combined rubric score of all dimensions, max=20) for ENC 1101. (Bottom) Comparison rubric dimension if dimension score is the same as overall (i.e. artifact overall score is equally distributed across all sections). A rubric dimension with hotter colors (reds/yellows) means that dimension 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 8 shows that from 15/20 and lower (approximately 75% overall score) all dimensions fair relatively equally (hot/cool colors fairly evenly distributed). When overall rubric scores range 16/20 or above (above 75%) achievement, the Grammar / Mechanics dimension lags slightly behind all other dimensions. For example, at an overall score of 18/20, Grammar / Mechanics exhibits average scores of 3.1/4 while the other four dimensions range from 3.4/4 to 3.7/4. From a student performance perspective, average and under achieving students tend to be equal in all dimensions while over achieving students never extend above average students in the Grammar / Mechanics dimension.

A comparison of fall 2015 results with past results is shown in Figure 9 below. Results exhibit consistency across all areas. The Thesis dimension continues to be the dimension with the highest mean score with a mean score of 3.2/4 in both years.

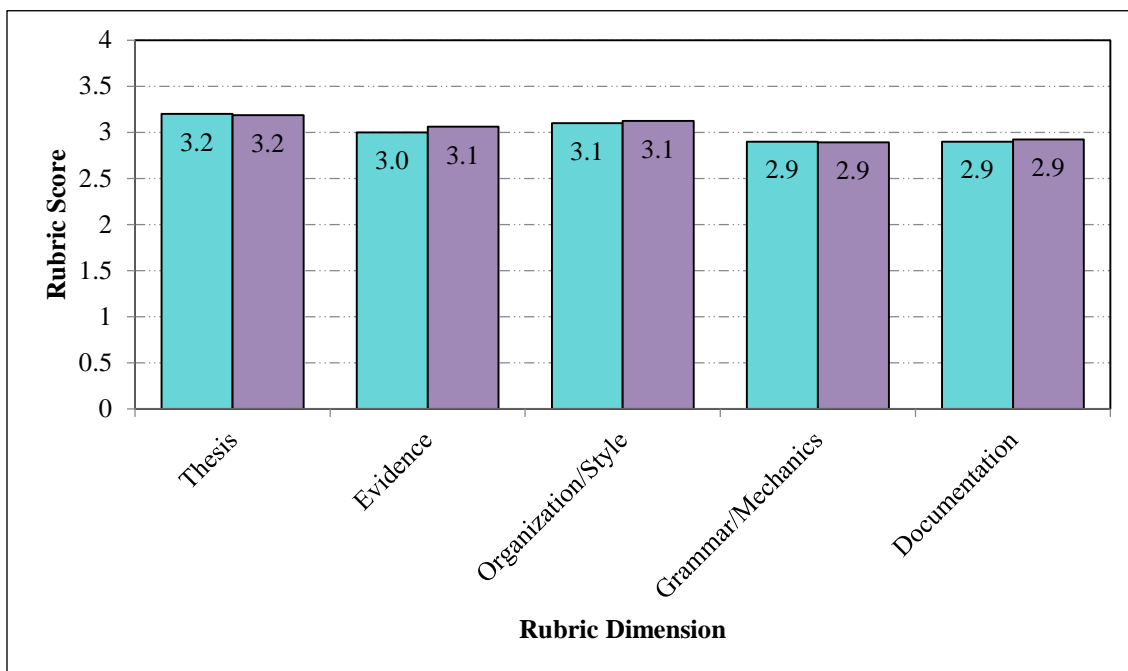


Figure 9. Comparison of mean scores for ENC 1101 through time beginning fall 2014 (teal) and fall 2015 (purple).

3.2 COMPARISONS BY SITE, FORMAT, AND STUDENT TYPE

3.2.1 Dual Enrollment to non-Dual Enrollment Comparison

During the fall 2015 semester, 35 dual enrollment artifacts were collected in ENC 1101 and 361 traditional (non-online) artifacts were collected in ENC 1101. A comparison of mean scores is provided in Table 7. The dual enrollment mean score is 0.2 higher than traditional artifacts. The difference in the means was 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 difference in the mean scores of dual enrollment and traditional artifacts can be a result of chance.

df = 394	
Dual enrollment mean	15.3
Dual enrollment standard deviation	2.13
Traditional mean	15.1
Traditional standard deviation	3.66
Effect size	0.06
p-value	0.518

Table 7. Comparison of mean scores for dual enrollment and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

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 statistically significant 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 10.

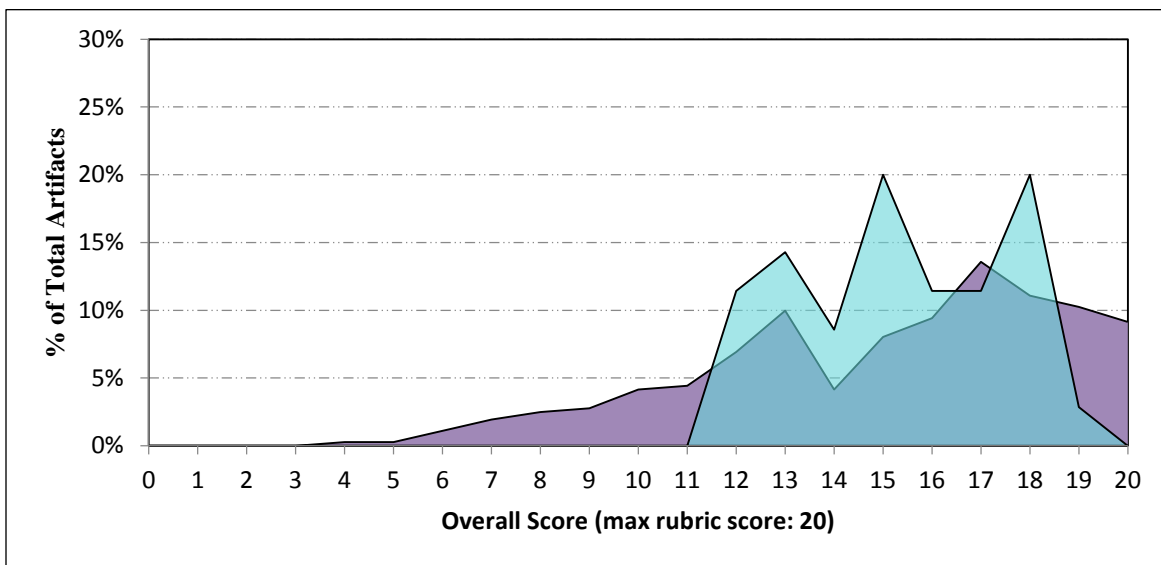


Figure 10. Score distribution for dual enrollment (purple) and traditional (teal) artifacts for ENC 1101.

3.2.2 Online to Traditional Comparison

During the fall 2015 semester, 17 total online artifacts were collected in ENC 1101 and 361 traditional artifacts were collected in ENC 1101. A comparison of mean scores is provided in Table 8. The online artifact mean score is 1.9 higher than traditional artifacts. The difference in the means was tested for significance using a Welch's t-test according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999) and was found to be statistically significantly different. Therefore, we can reject the null hypothesis that the difference in the mean scores of online and traditional artifacts can be a result of 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 statistically significant results exhibit what Cohen (1988) would consider a medium effect size. In other words, non-overlap score distribution from online artifacts to traditional artifacts is approximately 21%. For a graphical representation of this see Figure 11.

df = 376	
Online mean	17.0
Online standard deviation	2.72
Traditional mean	15.1
Traditional standard deviation	3.66
Effect size	0.29
p-value	0.011

Table 8. Comparison of mean scores for online and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

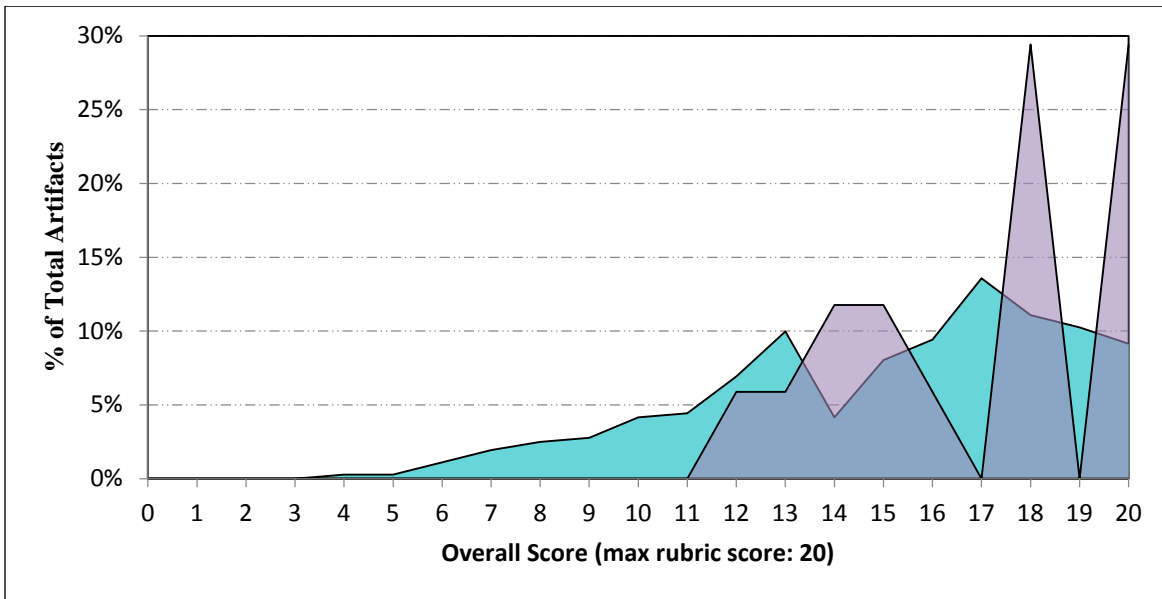


Figure 11. Score distribution for online (purple) and traditional (teal) artifacts of ENC 1101.

3.2.3 Comparison by Site/Campus

Of the 413 artifacts collected from ENC 1101, 67 originated from the Collier campus, 17 from FSW Online, 27 from the Hendry-Glade Center, 267 from the Thomas Edison (Lee) campus, and 35 from offsite. Scores by rubric dimension varied greatly across campuses. A comparison of mean scores by rubric dimension is provided in Table 9.

	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation
Collier	3.1	2.8	3.1	2.8	2.6
FSW Online	3.2	3.5	3.5	3.6	3.2
Hendry-Glades	3.2	3.1	3.1	3.0	3.2
Thomas Edison (Lee)	3.2	3.1	3.1	2.8	2.9
Offsite	3.0	2.9	3.1	3.1	3.3

Table 9. Comparison of mean scores by site for ENC 1101. Bold denotes highest mean score in that dimension among all sites.

FSW Online is consistently higher compared to other sites. FSW Online exhibits the highest scores in 4 of 5 dimensions. A plot comparing descriptive statistics of the combined (overall) scores by site is presented in Figure 12. There is extensive overlap between sites with both FSW Online and Offsite (dual enrollment) exhibiting a smaller range of scores, not uncommon given the smaller sample size for both.

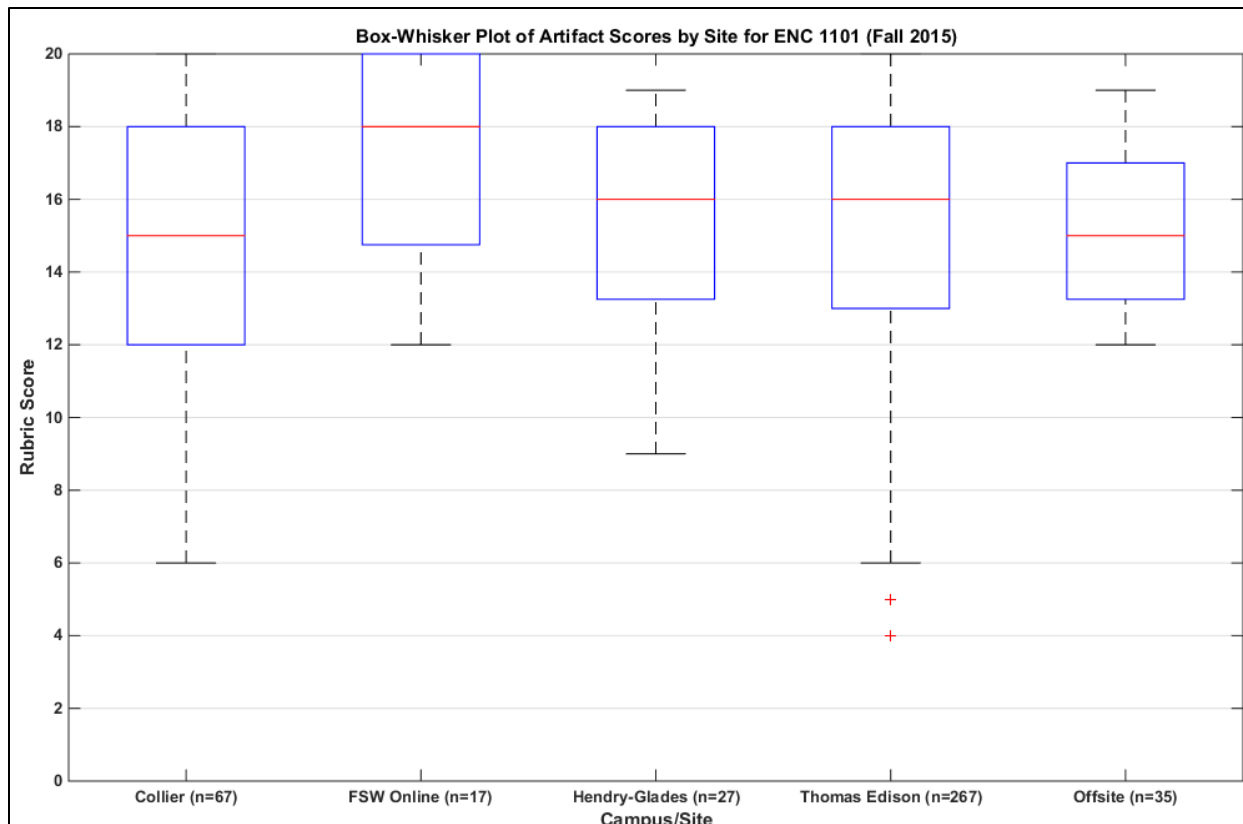


Figure 12. Box-Whisker plot of scores distributed by site for ENC 1101. Red line depicts median score. Upper and lower box boundaries indicate 75% quartile and 25% quartile (box represents central 50% of the scores). Vertical lines represent remaining scores outside central 50% that are not outliers. Red '+'s denote outliers.

A one-way analysis of variance was used to compare means of the combined rubric scores at each site. Results of the ANOVA exhibit no statistically significant difference between sites (see Table 10). Therefore, we cannot reject the null hypothesis that the mean rubric scores at each site are equal to each other and we cannot conclude with a 95% confidence that the differences in scores are not solely due to chance.

Source of Variation	Sum of squared differences	df	Mean Squares	F _{obs}	p-value	F _{crit}
Between Sites	103.2	4	25.8	2.09	0.082	2.39
Within Sites	5043.6	408	12.4			
Total	5146.7	412				

Table 10. Results of one-way ANOVA of combined rubric scores at each site for ENC 1101.

3.2.4 Mini-term to Full-term Comparison

The only mini-term artifacts received originated from a course section that was also the only online section received. As a result, for comparison of mini-term to full-term course sections, please see results from Section 3.2.2.

4 ENC 1102

4.1 LEARNING OBJECTIVES & DESCRIPTIVE STATISTICS

Using common rubric criterion as an assessment method, the FSW English faculty defined multiple areas of interest for evaluation based on core outcomes for the course. Those outcomes include:

- Students must demonstrate the ability to write essays following various rhetorical modes, strategies, and purposes.
- Students must demonstrate effective research skills, and incorporate documented direct quotations and paraphrases from a variety of sources, using MLA format.

4.1.1 Learning Objectives

ENC 1102 is scored using a rubric with five dimensions: Thesis, Evidence, Organization / Style, Grammar / Mechanics, and Documentation. Each scored on a scale of 1 to 4 (1-Does not meet standards, 2-Approaching standards, 3-Meets standards, 4-Exceeds standards), with 0s if the benchmark is not met. The English department has identified a target statistic for measurement purposes (SLO1) of measuring the percentage of artifacts scoring a 2 or greater.

For the fall 2015 assessment, 161 artifacts were collected for ENC 1102 from 9 of 12 course sections sampled from 61 course sections offered. As with ENC 1101, the resultant sample represents 14.7% of the population. The lowest scoring rubric dimension for percentage of artifacts scoring a 2 or greater is Grammar / Mechanics at 88% (Table 11). For a visual comparison of scores by dimension, see Figure 13.

Rubric Score	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation
Developing or higher	93%	89%	95%	88%	93%
4	39%	22%	22%	9%	16%
3	42%	41%	55%	48%	47%
2	12%	26%	19%	31%	30%
1	7%	11%	4%	12%	7%
0	0%	0%	1%	0%	0%

Table 11. Percentage of student achievement level by rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1102.

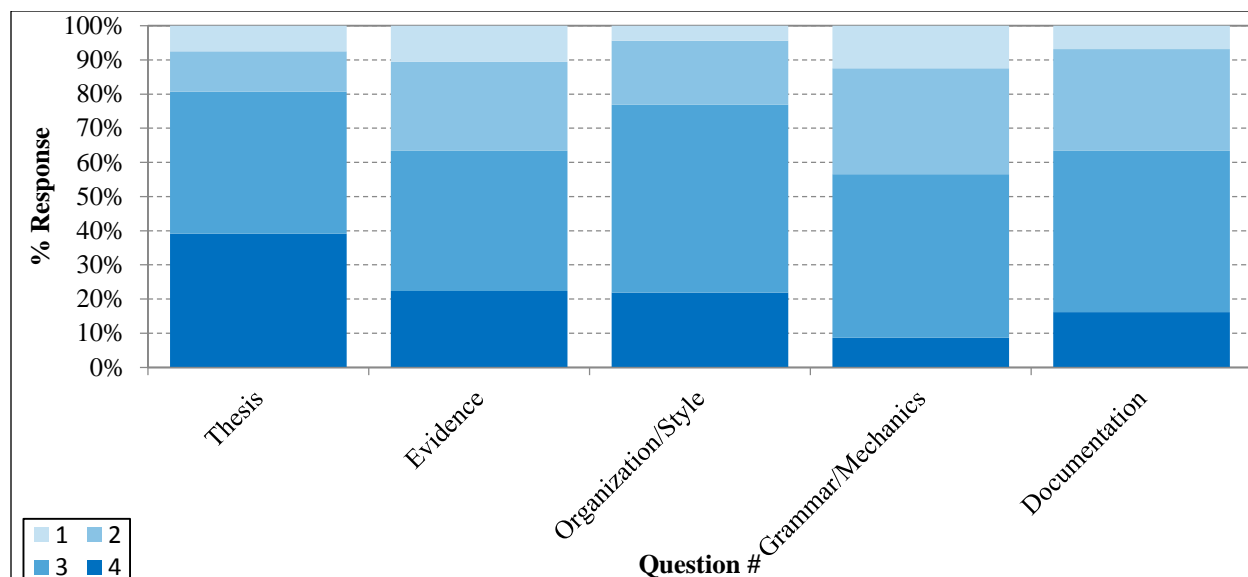


Figure 13. ENC 1102 distribution of rubric scores by dimension.

4.1.2 Descriptive Statistics & Longitudinal Studies

Descriptive statistics for ENC 1102 artifacts can be found in Table 12. A histogram of artifact scores for all 161 artifacts is shown in Figure 14. Distribution of artifact scores is centered on 15/20 and is moderately negatively skewed, meaning scores are shifted towards the higher range. To describe the behavior of the rubric dimensions based on overall achievement a color map, or binary raster image was created by calculating the mean scores for each dimension as a function of combined score (Figure 15). To create this image the rubric scores (4, 3, 2, 1, or 0) for each artifact was grouped based on combined raw rubric score (5 dimensions x maximum rubric level of 4 = 20 overall points). The color represents the mean rubric score achieved in each dimension based on the combined score as shown in the x-axis.

	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation	TOTAL
n	161	161	161	161	161	161
Max	4	4	4	4	4	20
Min	1	1	0	1	1	5
Median	3	3	3	3	3	14
Mode	3	3	3	3	3	15
Mean	3.1	2.8	2.9	2.5	2.7	14.1
Standard deviation	0.89	0.92	0.79	0.82	0.81	3.32
Skewness	-0.89	-0.31	-0.70	-0.30	-0.23	-0.43
Kurtosis	0.13	-0.71	0.81	-0.46	-0.40	-0.04

Table 12. Descriptive statistics for ENC 1102 common course assessment.

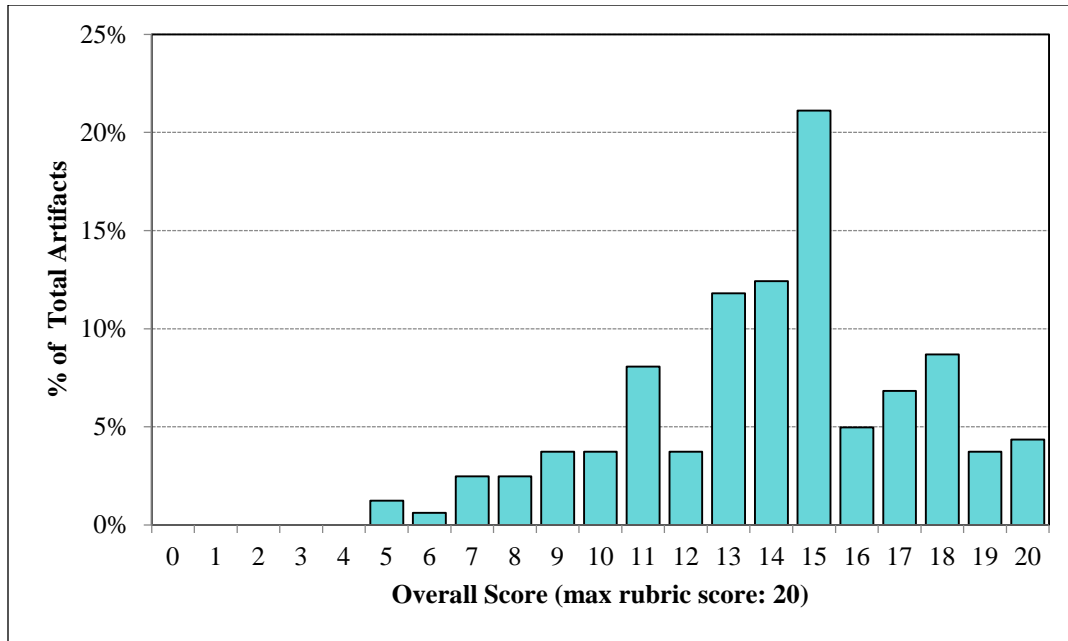


Figure 14. Overall score distribution for ENC 1102 artifacts (fall 2015 term).

A review of the colormap in Figure 15 shows that the Thesis dimension is exceptionally strong even at low overall scores. For example, at an overall score of 11/20, Thesis mean score is 2.9/4 where the other four dimensions range from 1.2/4 to 2.3/4. (colors fairly evenly distributed). The Grammar / Mechanics dimension lags behind all other dimensions when overall scores are 14/20 or higher. For example, at an overall score of 19/20, Grammar / Mechanics exhibits average scores of 3.2/4 while the other four dimensions range from 3.8/4 to 4/4. From a student performance perspective, average and under achieving students tend to remain strong in the Thesis dimension while over achieving students never extend above average students in the Grammar / Mechanics dimension.

A comparison of fall 2015 results with past results is shown in Figure 16 below. Results exhibit slight declines across all areas. The Thesis dimension continues to be the dimension with the highest mean score with a mean score of 3.2/4 in both years. Further, the Grammar/Mechanics dimension exhibits a substantial decline from 2014 to 2015. This difference is the only that is statistically significant, meaning it isn't likely that it is due to chance.

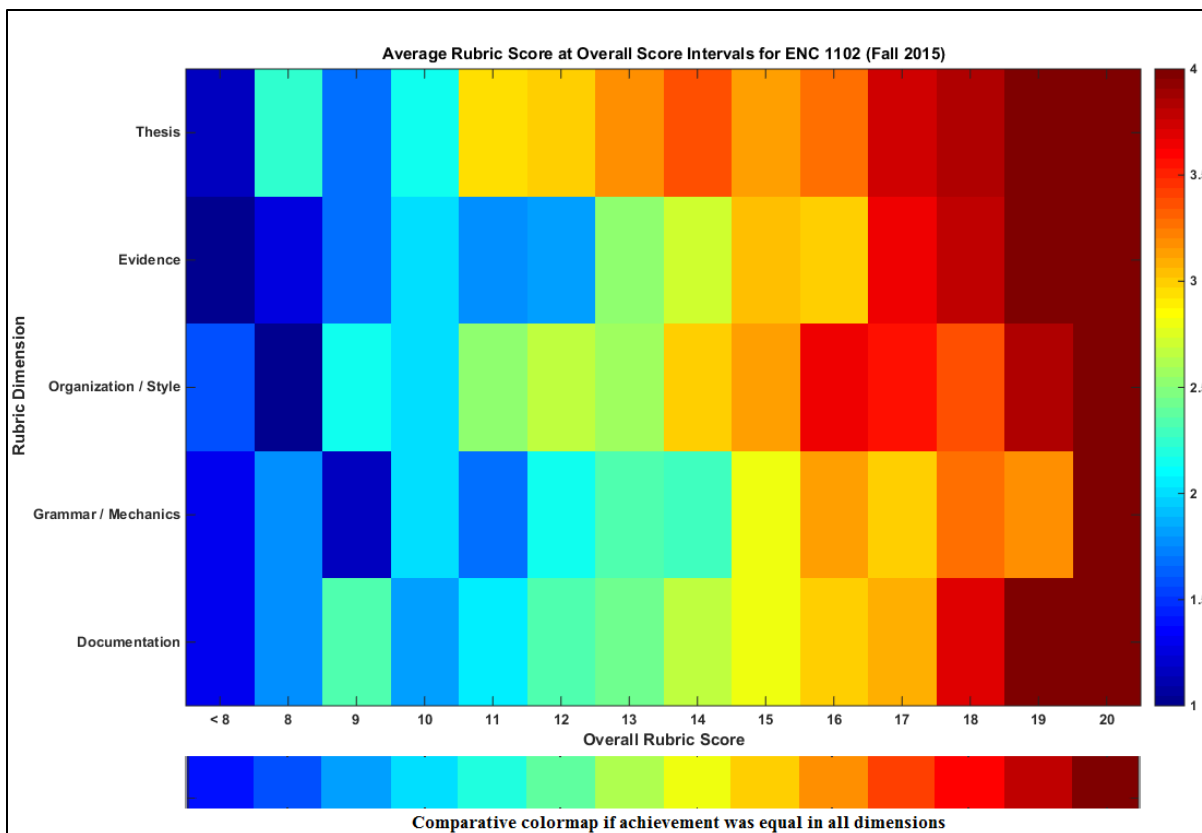


Figure 15. (Top) Colormap of mean scores for each rubric dimension (range: 0-4) based on overall rubric score (combined rubric score of all dimensions, max=20) for ENC 1102. (Bottom) Comparison rubric dimension if dimension score is the same as overall (i.e. artifact overall score is equally distributed across all sections). A rubric dimension with hotter colors (reds/yellows) means that dimension 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.

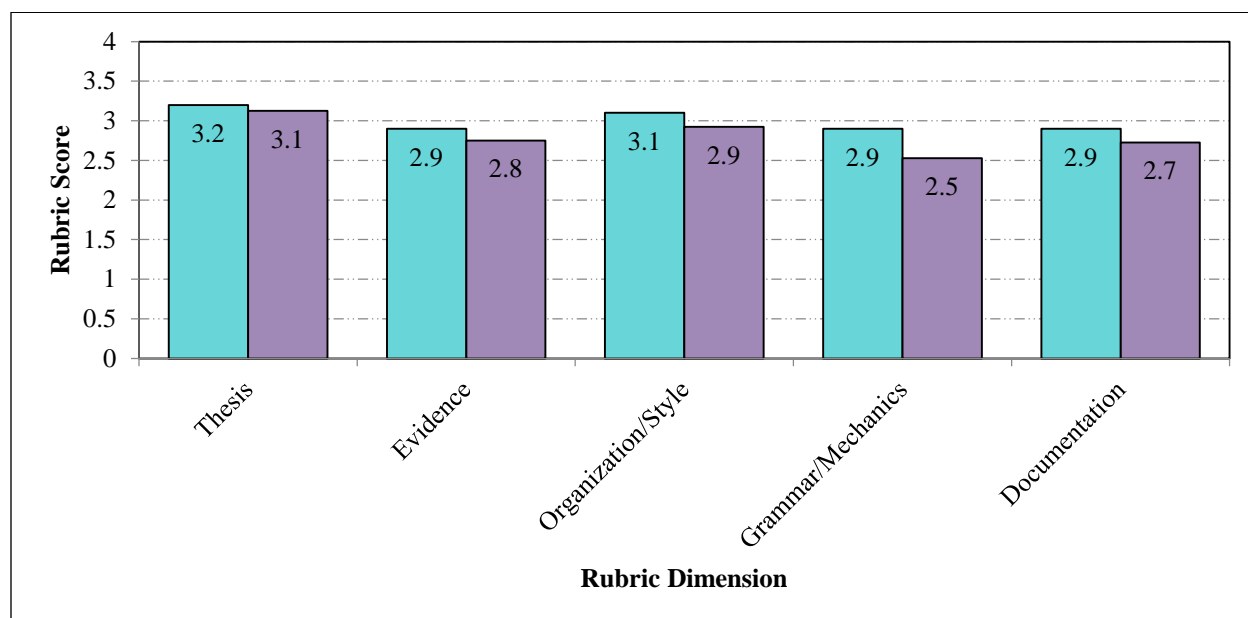


Figure 16. Comparison of mean scores for ENC 1102 through time beginning fall 2014 (teal) and fall 2015 (purple).

4.2 COMPARISON BY SITE, FORMAT, AND STUDENT TYPE

4.2.1 Dual Enrollment to non-Dual Enrollment Comparison

Two dual enrollment course sections were targeted in the sample. Only one of those two supplied results. In that course, only five artifacts were collected. A sample size this small is insufficient for any significance testing or comparisons yielding meaningful results (de Winter, 2013).

4.2.2 Online to Traditional Comparison

During the fall 2015 semester, 36 total online artifacts were collected in ENC 1101 and 120 traditional artifacts were collected in ENC 1102. A comparison of mean scores is provided in Table 13. The online artifact mean score is 2.6 lower than traditional artifacts. The difference in the means was tested for significance using a Welch's t-test according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999) and was found to be statistically significantly different. Therefore, we can reject the null hypothesis that the difference in the mean scores of online and traditional artifacts can be a result of chance.

df = 154		
Online mean		12.0
Online standard deviation		3.79
Traditional mean		14.6
Traditional standard deviation		2.93
Effect size		0.61
p-value		4.1×10^{-4}

Table 13. Comparison of mean scores for online and traditional artifacts. Positive effect sizes indicate a higher mean score for traditional artifacts.

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 statistically significant results exhibit what Cohen (1988) would consider a medium-to-large effect size. In other words, non-overlap score distribution from online artifacts to traditional artifacts is approximately 39%. For a graphical representation of this see Figure 17.

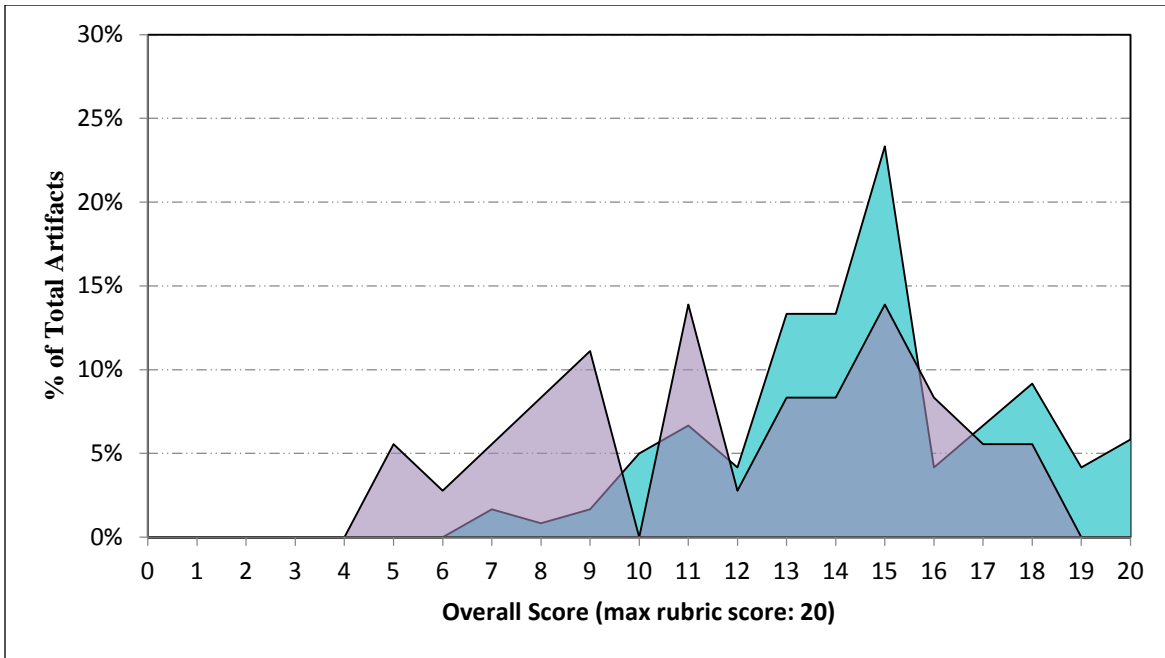


Figure 17. Score distribution for online (purple) and traditional (teal) artifacts of ENC 1102.

4.2.3 Comparison by Site/Campus

Of the 161 artifacts collected from ENC 1102, 41 originated from the Collier campus, 36 from FSW Online, 79 from the Thomas Edison (Lee) campus, and 5 from offsite. Scores by rubric dimension are consistently higher offsite (dual enrollment). However, since there are only five Offsite artifacts, this may not be an accurate representation of the population. A comparison of mean scores by rubric dimension is provided in Table 14. A plot comparing descriptive statistics of the combined (overall) scores by site is presented in Figure 18. There is extensive overlap between sites although the Thomas Edison (Lee) campus exhibits an upper 20% of artifacts beyond the other sites (dual enrollment notwithstanding with only 5 samples).

	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation
Collier	3.2	3.0	2.9	2.8	2.9
FSW Online	2.8	2.1	2.8	1.9	2.4
Thomas Edison (Lee)	3.2	2.9	3.0	2.7	2.8
Offsite	4.0	3.2	3.2	2.8	3.4

Table 14. Comparison of mean scores by site for ENC 1102. Bold denotes highest mean score in that dimension among all sites.

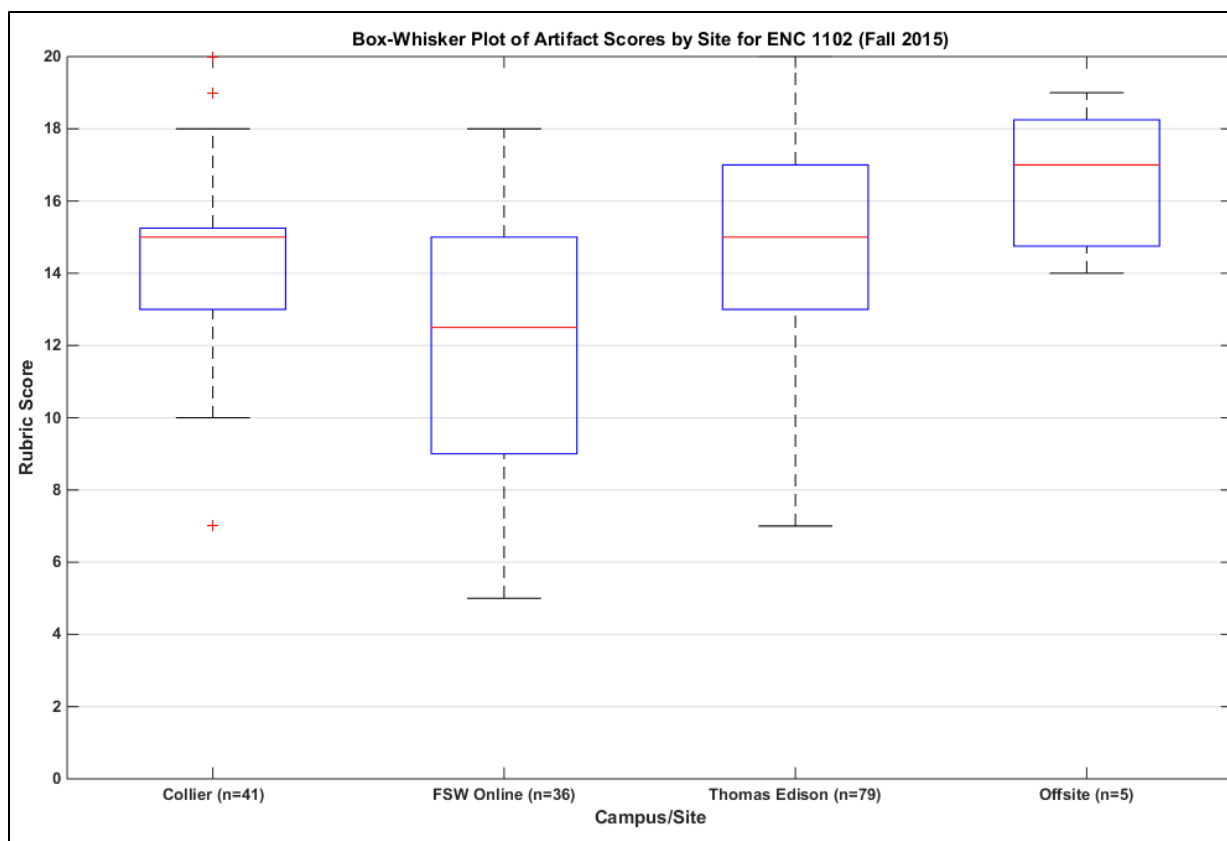


Figure 18. Box-Whisker plot of scores distributed by site for ENC 1102. Red line depicts median score. Upper and lower box boundaries indicate 75% quartile and 25% quartile (box represents central 50% of the scores). Vertical lines represent remaining scores outside central 50% that are not outliers. Red '+'s denote outliers.

4.2.4 Mini-term to Full-term Comparison

The only mini-term artifacts received originated from a course section that was also the only online section received. As a result, for comparison of mini-term to full-term course sections, please see results from Section 4.2.2.

5 PILOT STUDY RESULTS

A revised "Thesis" and "Evidence" rubric criteria was piloted during the fall 2015 term. Seven course sections across both ENC 1101 and ENC 1102 were targeted for the sample. Due to unforeseen circumstances, only three sections were recorded. The pilot was assessed using the same common rubric as the standard assessment with the exception of the piloted criteria. The dimension was scored using the same scale of 1 to 4 (1-Unacceptable, 2-Needs work, 3-Average, 4-Above average), with 0s if the baseline of 'Unacceptable' is not met.

5.1 ENC 1101 PILOT COMPARISON

The fall 2015 ENC 1101 assessment included 413 artifacts. The pilot study using the new Thesis and Evidence criteria included 21 artifacts. A comparison of achievement percentages in the Thesis rubric dimension is shown below in Table 15. Results exhibit a slight decrease in percentage of '4's achieved,

but a substantial increase in the percentage of artifacts scoring 3 or higher. All 100% of artifacts in the piloted Thesis criterion score 3 or higher. By comparison, only 77% of those in the current Thesis criterion score 3 or higher. The mean score for the Thesis pilot also increased (3.2/4 up to 3.5/4) although the change is not statistically significant.

Rubric Score	Thesis	Thesis Pilot
Developing or higher	94%	100%
4	51%	48%
3	26%	52%
2	17%	0%
1	6%	0%
0	0%	0%

Table 15. Comparison of percentage of student achievement level by Thesis rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1101.

A comparison of achievement percentages in the Evidence rubric dimension is shown below in Table 16. Results exhibit a decrease in percentage of '4's achieved, but an increase in the percentage of artifacts scoring 2 or higher (the SLO benchmark). The mean score for the Evidence pilot also increased (3.1/4 up to 3.4/4) although the change is not statistically significant.

Rubric Score	Evidence	Evidence Pilot
Developing or higher	94%	100%
4	51%	44%
3	26%	50%
2	17%	6%
1	6%	0%
0	0%	0%

Table 16. Comparison of percentage of student achievement level by Evidence rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1101.

5.2 ENC 1102 PILOT COMPARISON

The fall 2015 ENC 1102 assessment included 161 artifacts. The pilot study using the new Thesis and Evidence criteria included 34 artifacts. A comparison of achievement percentages in the Thesis rubric dimension is shown below in Table 17. Results exhibit a substantial increase in percentage of '4's achieved (39% to 64%). The mean score for the Thesis pilot also increased (3.1/4 up to 3.5/4) and the change is statistically significant ($t(192)=2.05$, $p=0.046$).

Rubric Score	Thesis	Thesis Pilot
Developing or higher	93%	97%
4	39%	64%
3	42%	21%
2	12%	12%
1	7%	3%
0	0%	0%

Table 17. Comparison of percentage of student achievement level by Thesis rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1102.

A comparison of achievement percentages in the Evidence rubric dimension is shown below in Table 18. Results exhibit a substantial increase in percentage of '4's achieved as well as an increase in the percentage of artifacts scoring 2 or higher (the SLO benchmark). The mean score for the Evidence pilot also increased (2.8/4 up to 3.6/4) and the change is statistically significant ($t(187)=5.72$, $p=3.58 \times 10^{-7}$).

Rubric Score	Evidence	Evidence Pilot
Developing or higher	89%	97%
4	22%	65%
3	41%	29%
2	26%	3%
1	11%	3%
0	0%	0%

Table 18. Comparison of percentage of student achievement level by Evidence rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1102.

6 CONCLUSIONS

FSW's English Department assessment plan includes three courses: ENC 0022 *Writing for College Success*, ENC 1101 *Composition I*, and ENC 1102 *Composition II*. Instructors use a common rubric with seven identified rubric dimensions in the case of ENC 0022, and five dimensions for both ENC 1101 and ENC 1102. The assessment plan uses a random sample of 20% of all course sections offered in ENC 1101 and ENC 1102 and a 100% collection of ENC 0022 courses. The department has historically used a benchmark of percentage of students scoring 2 or higher in rubric dimensions as a means to measure achievement in the courses. Additionally, a pilot study was conducted to assess a new rubric criterion for "Evidence" to replace the current one.

A drilldown of ENC 0022 results are as follows:

1. All seven rubric dimensions had > 87% achievement at level 2 or higher. The lowest dimension was Research while all other dimensions exceeded 96%.
2. Distribution of artifact scores is bimodal centered on 21/28 and 28/28, and is moderately negatively skewed, meaning scores are shifted towards the higher range.
3. In a study comparing rubric achievement based on overall score, under achieving students tend to be exceedingly lagging in Grammar and Mechanics, average students tend to be equally strong in all dimensions, and over achieving students tend to again lag in Grammar and Mechanics.
4. In a longitudinal study, results exhibit improvement across all areas. However, some artifacts reported in fall 2014 may have been reported as 0s rather than non-reporting so 2014 scores may be slightly lower than actual.
5. No comparison of dual enrollment to traditional artifacts was completed because no dual enrollment sections of the course were offered.
6. No comparison of online to traditional artifacts was completed because no online sections of the course were offered.
7. In a cross-campus comparison, scores varied greatly across rubric dimensions. No site is consistently higher compared to other sites, however, the Thomas Edison (Lee) campus is consistently the lowest.

8. No comparison of mini-term artifacts and full-term artifacts was completed because no mini-term sections of the course were offered.

A drilldown of ENC 1101 results are as follows:

1. All seven rubric dimensions had > 93% achievement at level 2 or higher. The lowest dimension was Grammar / Mechanics.
2. Distribution of artifact scores is centered on 17/20 and is moderately negatively skewed, meaning scores are shifted towards the higher range.
3. In a study comparing rubric achievement based on overall score, average and under achieving students tend to be equal in all dimensions while over achieving students never extend above average students in the Grammar / Mechanics dimension.
4. In a longitudinal study, results exhibit consistency across all areas. The Thesis dimension continues to be the dimension with the highest mean score with a mean score of 3.2/4 in both years.
5. In a study comparing dual enrollment to traditional (non-online) artifacts, the dual enrollment mean score is 0.2 higher than traditional artifacts but was not statistically significant.
6. In a study comparing online to traditional artifacts, the online artifact mean score is 1.9 higher than traditional artifacts and was statistically significant.
7. In a cross-campus comparison, scores varied greatly across rubric dimensions. FSW Online is consistently higher compared to other sites.
8. The only mini-term artifacts received originated from a course section that was also the only online section received. As a result, for comparison of mini-term to full-term course sections, please see results from #6 above.

A drilldown of ENC 1102 results are as follows:

1. All seven rubric dimensions had > 88% achievement at level 2 or higher. The lowest dimension was Grammar / Mechanics.
2. Distribution of artifact scores is centered on 15/20 and is moderately negatively skewed, meaning scores are shifted towards the higher range.
3. In a study comparing rubric achievement based on overall score, average and under achieving students tend to remain strong in the Thesis dimension while over achieving students never extend above average students in the Grammar / Mechanics dimension.
4. In a longitudinal study, results exhibit slight declines across all areas. The Thesis dimension continues to be the dimension with the highest mean score with a mean score (3.2/4) in both years while Grammar / Mechanics exhibits a substantial statistically significant decline from 2014 to 2015.
5. In a study comparing dual enrollment to traditional (non-online) artifacts, only five artifacts were collected and so a meaningful study could not be conducted.
6. In a study comparing online to traditional artifacts, the online artifact mean score is 2.6 lower than traditional artifacts and was found to be statistically significant.
7. In a cross-campus comparison, scores varied greatly across rubric dimensions. Thomas Edison (Lee) campus exhibits an upper 20% of artifacts beyond the other sites.
8. The only mini-term artifacts received originated from a course section that was also the only online section received. As a result, for comparison of mini-term to full-term course sections, please see results from #6 above.

A drilldown of Thesis and Evidence Pilot study results are as follows:

1. In a study comparing ENC 1101 Thesis dimension with that of the pilot study using the new Thesis criterion, results exhibit a substantial increase in the percentage of artifacts scoring 3 or higher. The mean score for the Thesis pilot also increased (3.2/4 up to 3.5/4) although the change is not statistically significant.
2. In a study comparing ENC 1101 Evidence dimension with that of the pilot study using the new Evidence criterion, results exhibit a decrease in percentage of '4's achieved, but an increase in the percentage of artifacts scoring 2 or higher (the SLO benchmark). The mean score for the Evidence pilot also increased (3.1/4 up to 3.4/4) although is not statistically significant.
3. In a study comparing ENC 1102 Thesis dimension with that of the pilot study using the new Thesis criterion, results exhibit a substantial increase in percentage of '4's achieved (39% to 64%). The mean score for the Evidence pilot also increased (3.1/4 up to 3.5/4) and the change is statistically significant ($t(192)=2.05$, $p=0.046$).
4. In a study comparing ENC 1102 Evidence dimension with that of the pilot study using the new Evidence criterion, results exhibit a substantial increase in percentage of '4's achieved as well as an increase in the percentage of artifacts scoring 2 or higher (the SLO benchmark). The mean score for the Evidence pilot also increased (2.8/4 up to 3.6/4) and the change is statistically significant ($t(187)=5.72$, $p=3.58 \times 10^{-7}$).

7 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.
- 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.
- 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.