English Assessment Report
Fall 2016

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 2016 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 30% 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 Accountability Plan) 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. Other analyses such as comparison by term length (standard vs. mini-term) and longitudinal studies are included.

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 2016 assessment, 152 artifacts were collected for ENC 0022 from 10 of 10 course sections. The lowest scoring rubric dimension for percentage of artifacts scoring a 2 or greater is Research at 80%. All other dimensions exhibit percentage of 97% 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	99%	99%	99%	98%	98%	97%	80%
4	38%	34%	41%	32%	23%	14%	8%
3	42%	51%	43%	52%	57%	55%	30%
2	18%	14%	14%	14%	18%	27%	42%
1	1%	0%	0%	1%	1%	2%	10%
0	1%	1%	1%	1%	1%	1%	10%

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.

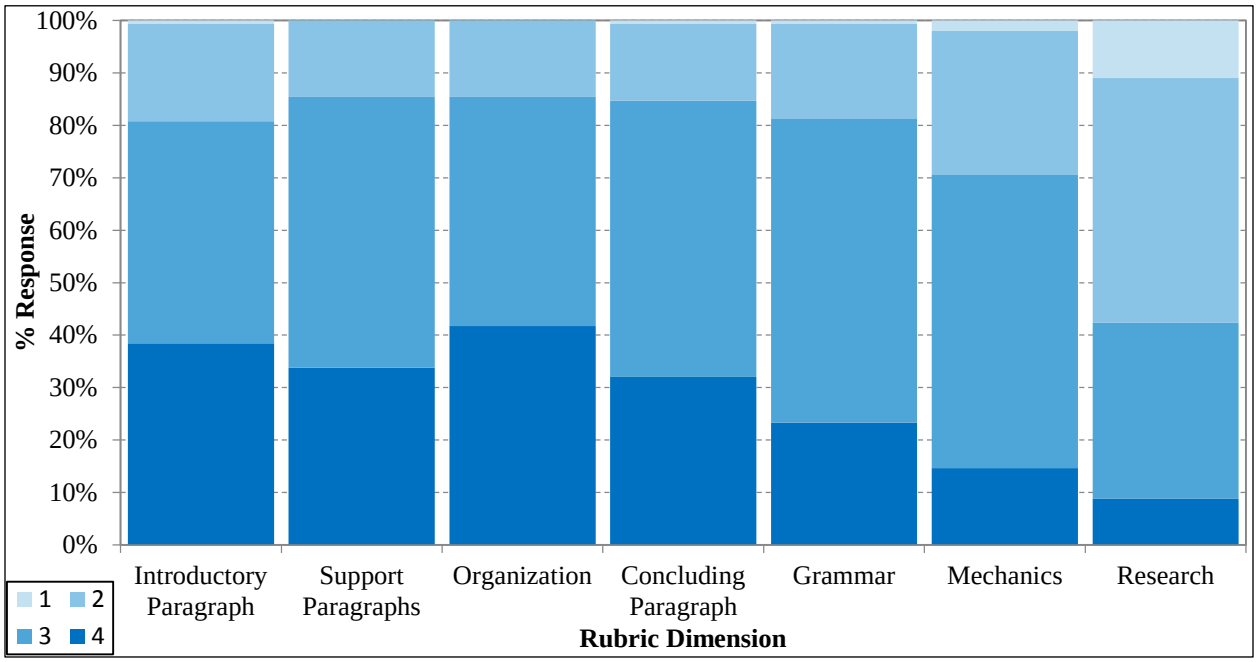

Figure 1. ENC 0022 distribution of rubric scores by dimension.

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 152 artifacts is shown in Figure 2. Distribution of artifact scores is bimodal centered on 20/28 and 27/28, 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	152	152	152	152	152	152	152	**152**
Max	4	4	4	4	4	4	4	**28**
Min	0	0	0	0	0	0	0	**2**
Median	3	3	3	3	3	3	2	**20**
Mode	3	3	3	3	3	3	2	**20**
Mean	3.2	3.2	3.3	3.1	3.0	2.8	2.2	**20.7**
Standard deviation	0.79	0.72	0.75	0.77	0.75	0.76	1.04	**4.23**
Skewness	-0.71	-0.70	-0.83	-1.00	-0.87	-0.66	-0.44	**-0.65**
Kurtosis	0.46	1.29	1.02	2.24	2.28	1.46	-0.08	**1.57**

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

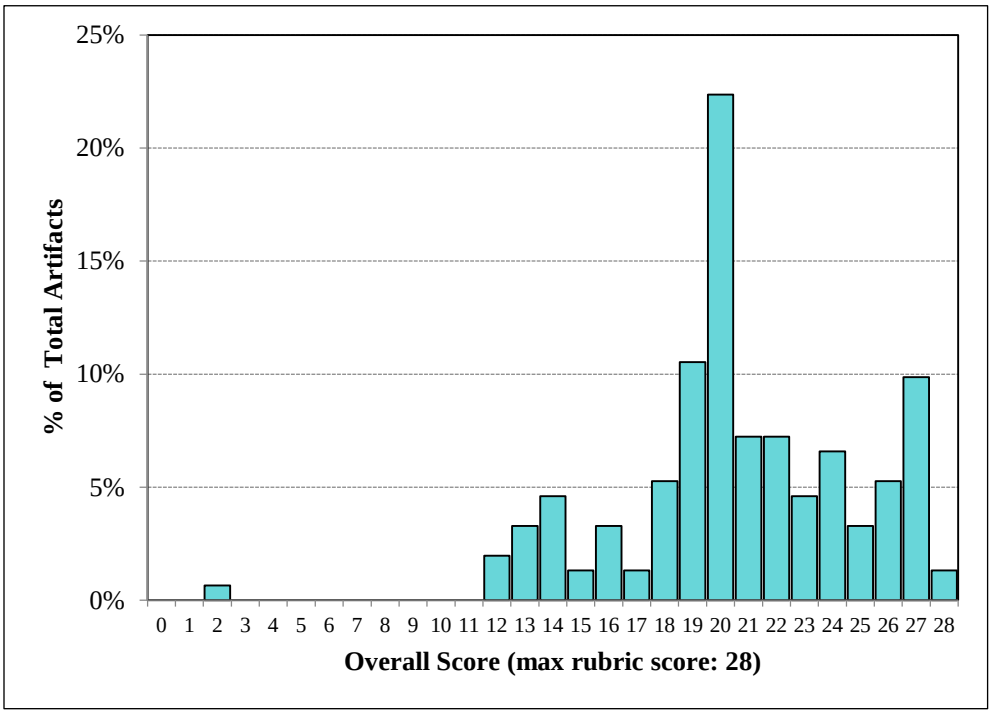

Figure 2. Overall score distribution for ENC 0022 artifacts (fall 2016 term).

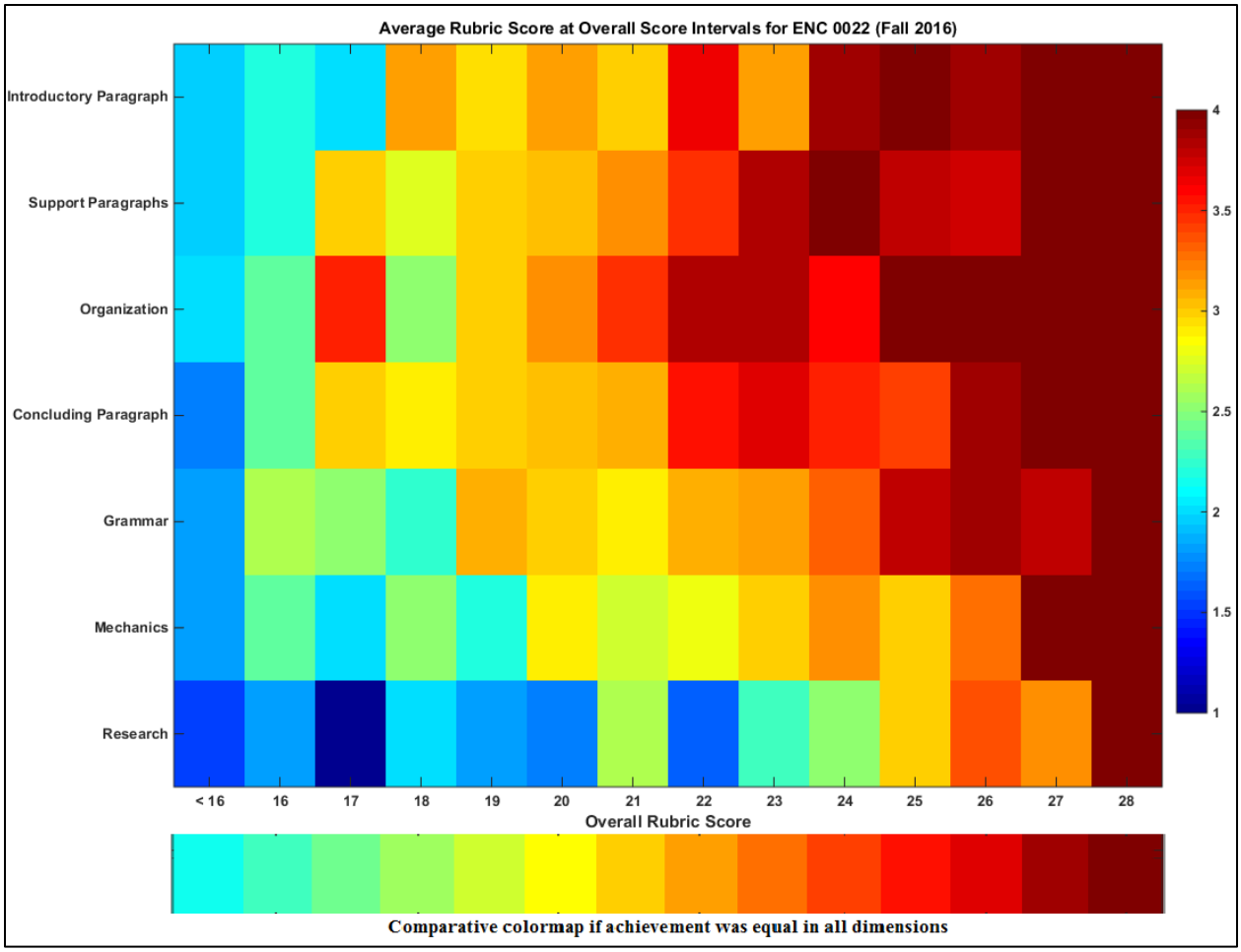

Figure 3. (Top) Colormap of mean scores for each rubric dimension (range: 0-4) based on overall rubric score (combined rubric score of all dimensions, max=28) for ENC 0022. (Bottom) Comparison rubric dimension if dimension score is the same as overall (i.e. artifact overall score is equally distributed 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 3 above shows that Research achievement consistently lags behind all other dimensions when overall scores are 16/28 or higher. For example, at 19/28, the Research mean score is 1.8/4 while others range from 2.2/4 to 3.0/4. Similarly, at 24/28, the Research mean score is 2.5/4 while others range from 3.2/4 to 4/4. From a student performance perspective, all students are weak in the Research dimension compared with others.

The colormap also exhibits strong Organization scores compared with other dimensions at higher overall scores (20/28 or higher). For example, at 22/28, the Organization dimension mean score is 3.8/4 while others range from 1.6/4 to 3.6/4. From a student performance perspective, high moderate-to-high achieving students are strongest in Organization compared with other dimensions. This is also the case, but to a lesser extent, with Supporting Paragraphs and Concluding Paragraphs.

A comparison of fall 2016 results with past results is shown in Figure 4 below. Results exhibit consistency across all areas except for Research, which exhibits a sharp decline in the most recent term. Fall 2016 data do exhibit an extensive percentage of 0s reported for Research (10%) compared with

previous years (0% for fall 2015 and 0% for fall 2014). And while one course section does exhibit 0s universally for all reported scores in Research, 0s are reported in other sections as well, so it does appear to be a real, if less substantial, characteristic of the data.

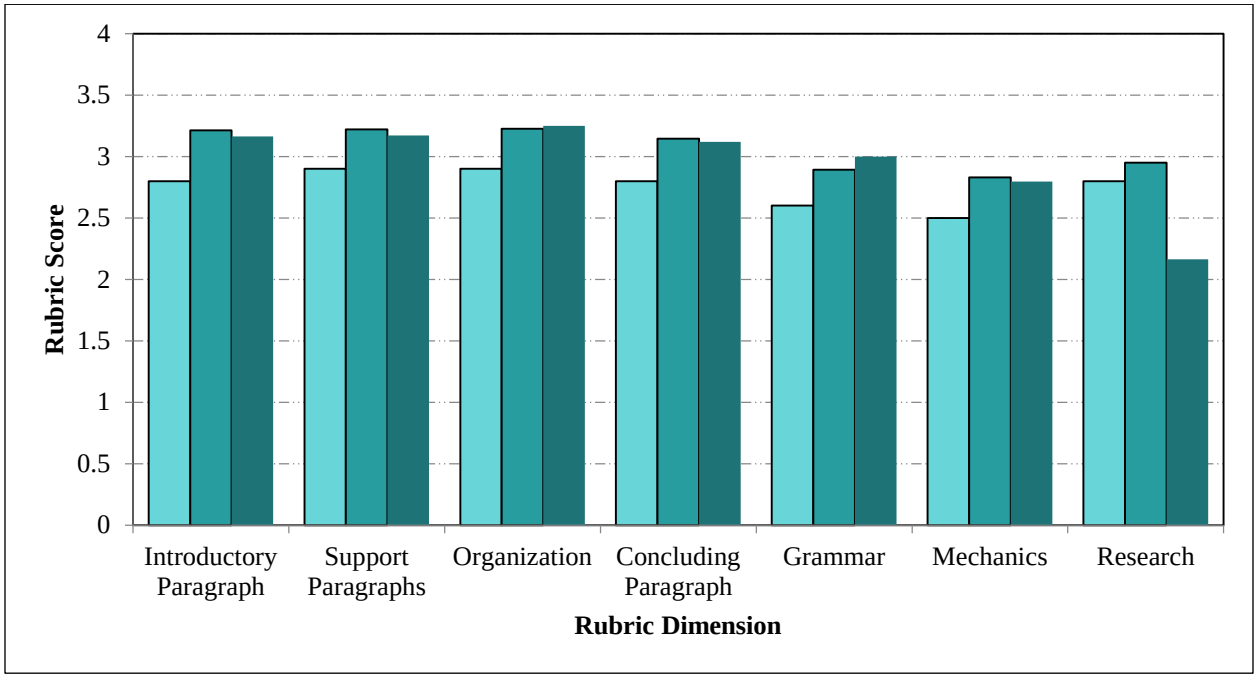

Figure 4. Comparison of mean scores for ENC 0022 through time for fall 2014 (teal), fall 2015 (darker teal), and fall 2016 (darkest teal).

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 152 artifacts collected from ENC 0022, 13 originated from the Charlotte campus, 13 from the Collier campus, 4 from the Hendry Glades Center, and 122 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	2.8	2.9	2.6	2.9	2.5	2.6	2.6
Collier	**3.9**	**3.5**	3.7	**3.8**	2.8	**3.1**	0.0
Hendry Glades	3.3	3.0	**3.8**	3.3	3.0	2.8	**3.5**
Thomas Edison (Lee)	3.1	3.2	3.3	3.1	**3.1**	2.8	2.3

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 Charlotte campus is the lowest in 6 of 7 dimensions. Collier campus exhibits the highest scores in 4 of 7 dimensions. Hendry Glades exhibits the highest scores in 2 of 7 dimensions, and Thomas Edison (Lee) exhibits the highest 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 Collier and Hendry-Glades exhibiting a smaller range of scores. Recall that Hendry Glades data includes only four records.

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.

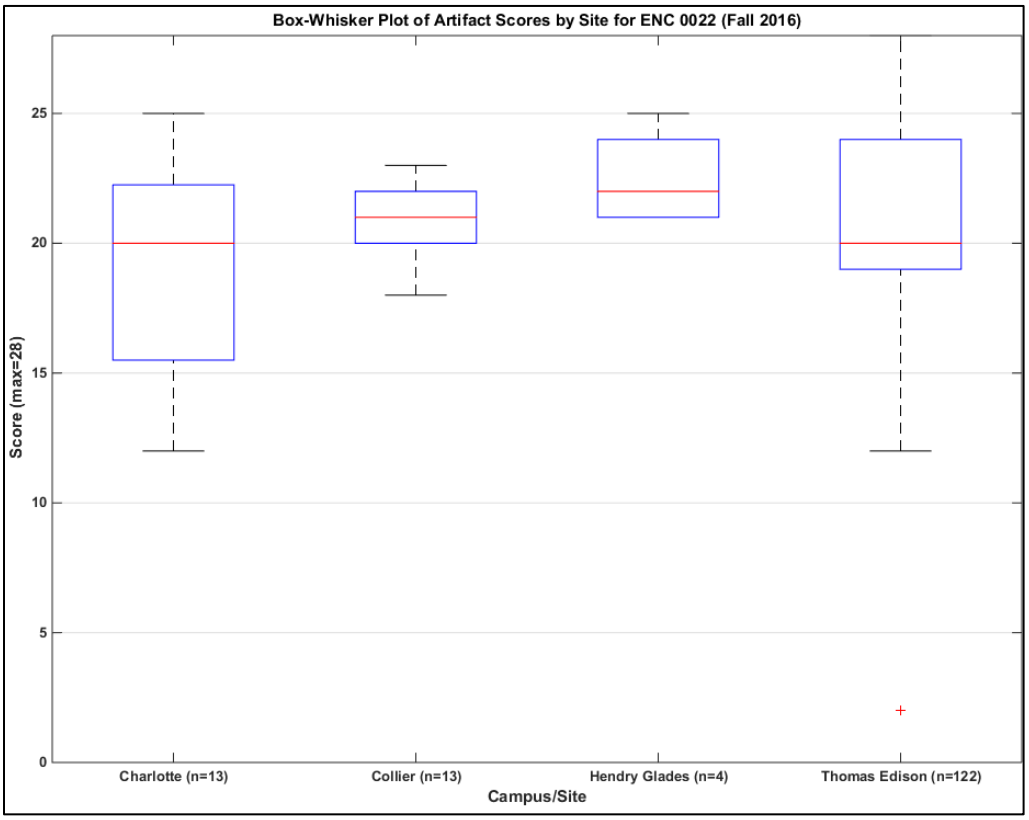

Figure 5. Box-Whisker plot of scores distributed by site for ENC 0022. Red line depicts median score. Upper and lower box boundaries indicate 75% quartile and 25% quartile (box represents central 50% of the scores). Vertical lines represent remaining scores outside central 50% that are not outliers.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

Source of Variation	Sum of squared differences	df	Mean Squares	F_{obs}	p-value	F_{crit}
Between Sites	54.9	3	18.3	1.03	0.383	2.67
Within Sites	2641.0	148	17.8			
Total	2695.9	151				

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 revised based on assessment results of AY 2015-16 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:

- SLO 1: Students must demonstrate the ability to write essays following various rhetorical modes, strategies, and purposes.
- SLO 2 & 3: 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 of measuring the percentage of artifacts scoring a 2 or greater.

For the fall 2016 assessment, 891 artifacts were collected for ENC 1101 from 46 of 59 course sections sampled from 160 course sections offered. The remaining 13 course sections did not report data. The resultant sample represents 23.4% of the population. The lowest scoring rubric dimension by percentage of artifacts scoring a 2 or greater is Documentation at 89% (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%	94%	94%	94%	89%
4	41%	38%	38%	23%	28%
3	40%	38%	41%	52%	40%
2	14%	19%	16%	19%	21%
1	5%	6%	5%	6%	11%
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.

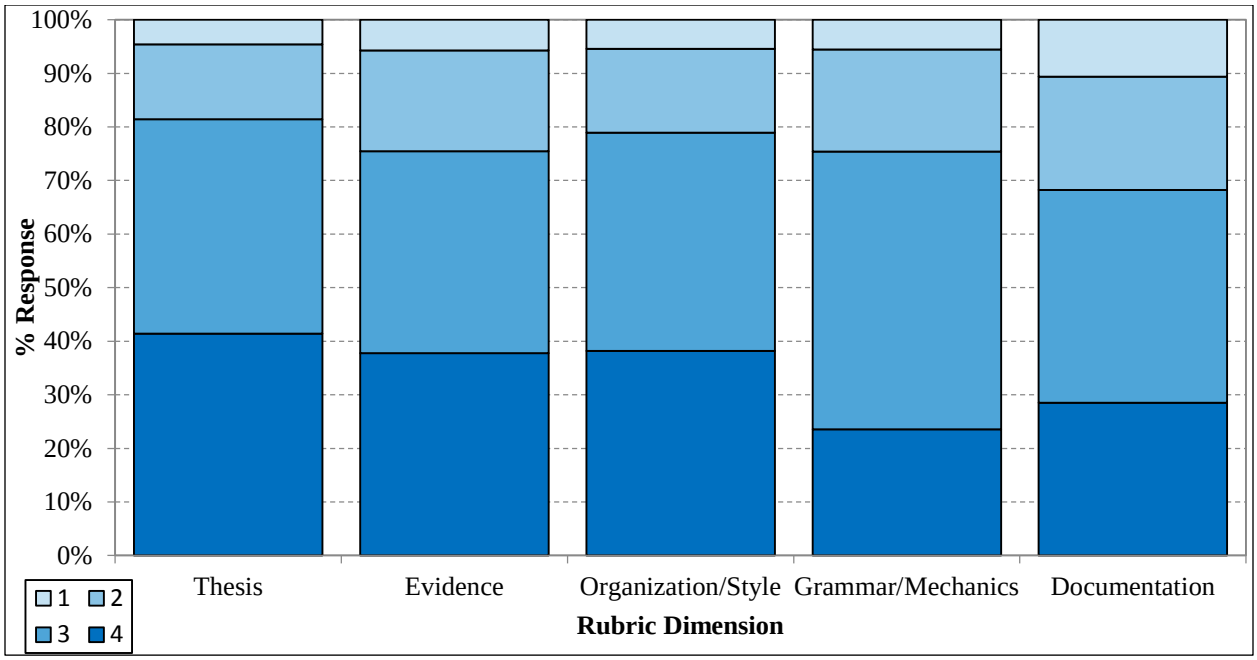

Figure 6. ENC 1101 distribution of rubric scores by dimension.

3.1.2 Descriptive Statistics

Descriptive statistics for ENC 1101 artifacts can be found in Table 6. A histogram of artifact scores for all 891 artifacts is shown in Figure 7. 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 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	891	891	890	890	891	891
Max	4	4	4	4	4	20
Min	0	0	0	0	0	1
Median	3	3	3	3	3	16
Mode	4	4	3	3	3	15
Mean	3.2	3.1	3.1	2.9	2.8	15.1
Standard deviation	0.85	0.90	0.87	0.82	0.97	3.72
Skewness	-0.88	-0.68	-0.78	-0.61	-0.51	-0.90
Kurtosis	0.26	-0.28	0.04	0.25	-0.55	0.56

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

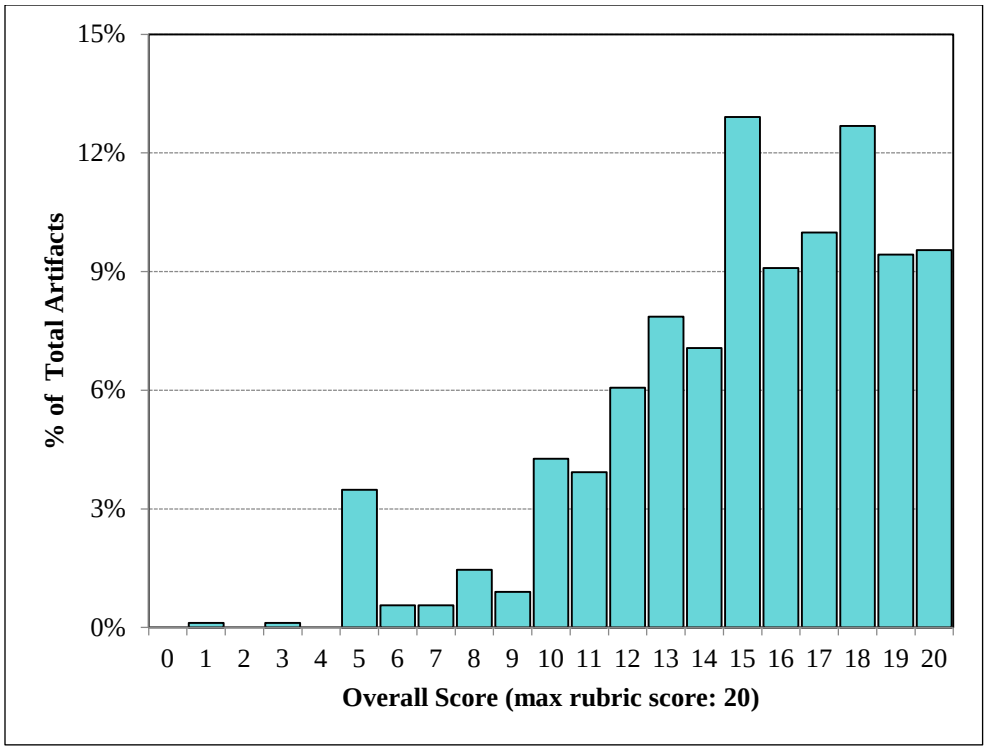

Figure 7. Overall score distribution for ENC 1101 artifacts (fall 2016 term).

A review of the colormap in Figure 8 shows that around 15/20 (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 3.3/4 while the other four dimensions range from 3.5/4 to 3.8/4. From a student performance perspective, average achieving students tend to be equal in all dimensions while over achieving students never extend above average students in the Grammar / Mechanics dimension.

Additionally, at low range scores (12/20 and lower), the Documentation dimension lags behind all others. For example, at an overall score of 10/20, Documentation exhibits average scores of 1.6/4 while the other four dimensions range from 2.0/4 to 2.2/4. From a student performance perspective, under achieving students tend to struggle with Documentation above all other areas.

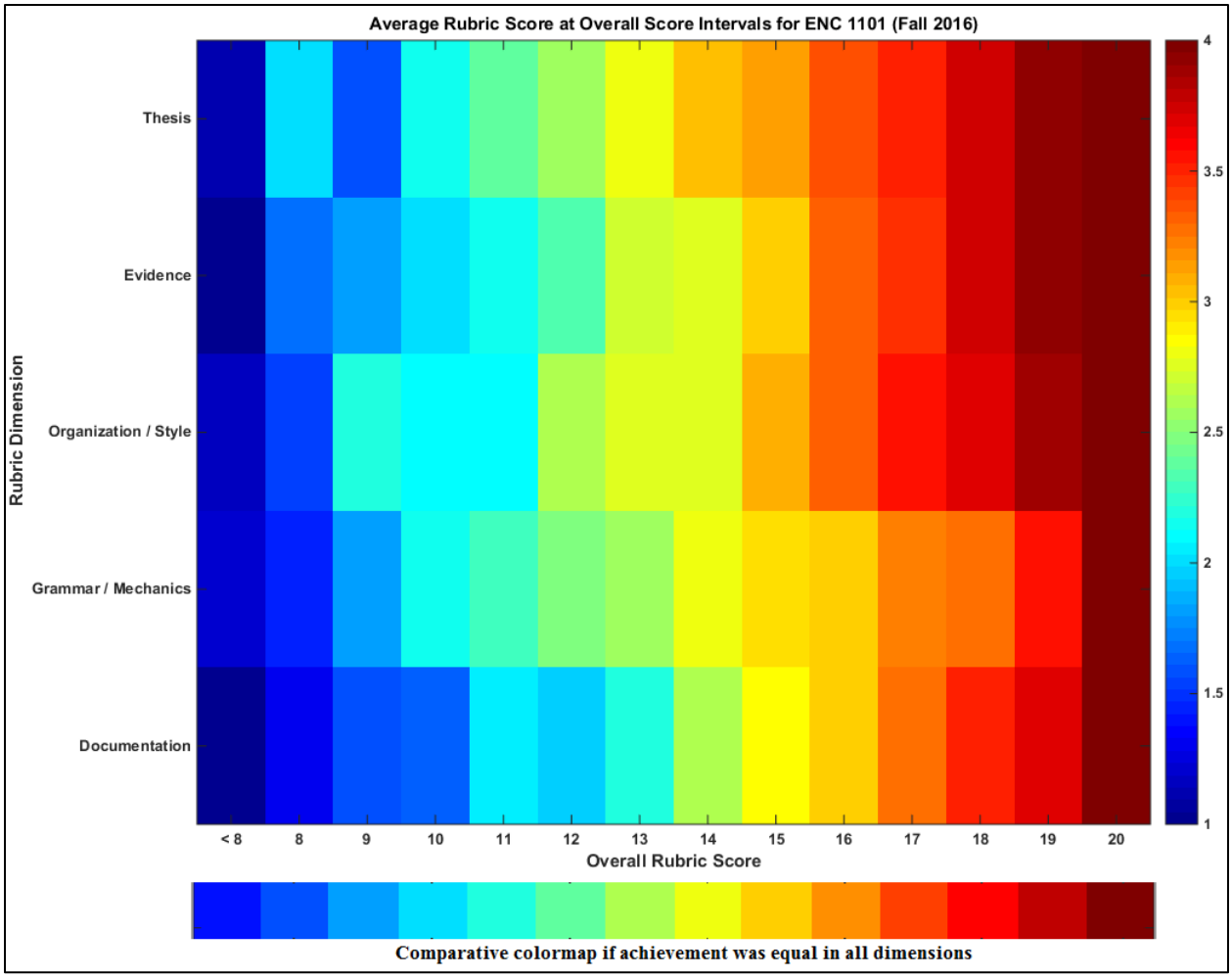

Figure 8. (Top) Colormap of mean scores for each rubric dimension (range: 0-4) based on overall rubric score (combined rubric score of all dimensions, max=20) for ENC 1101. (Bottom) Comparison rubric dimension if dimension score is the same as overall (i.e. artifact overall score is equally distributed 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 comparison of achievement by rubric of fall 2016 results with past results is shown in Figure 9. Results exhibit consistency across all areas over time. The Thesis dimension continues to be the dimension with the highest mean score with a mean score of 3.2/4 in all years. The Grammar/Mechanics and Documentation dimensions continue to be the lowest scoring in all years (2.9/4).

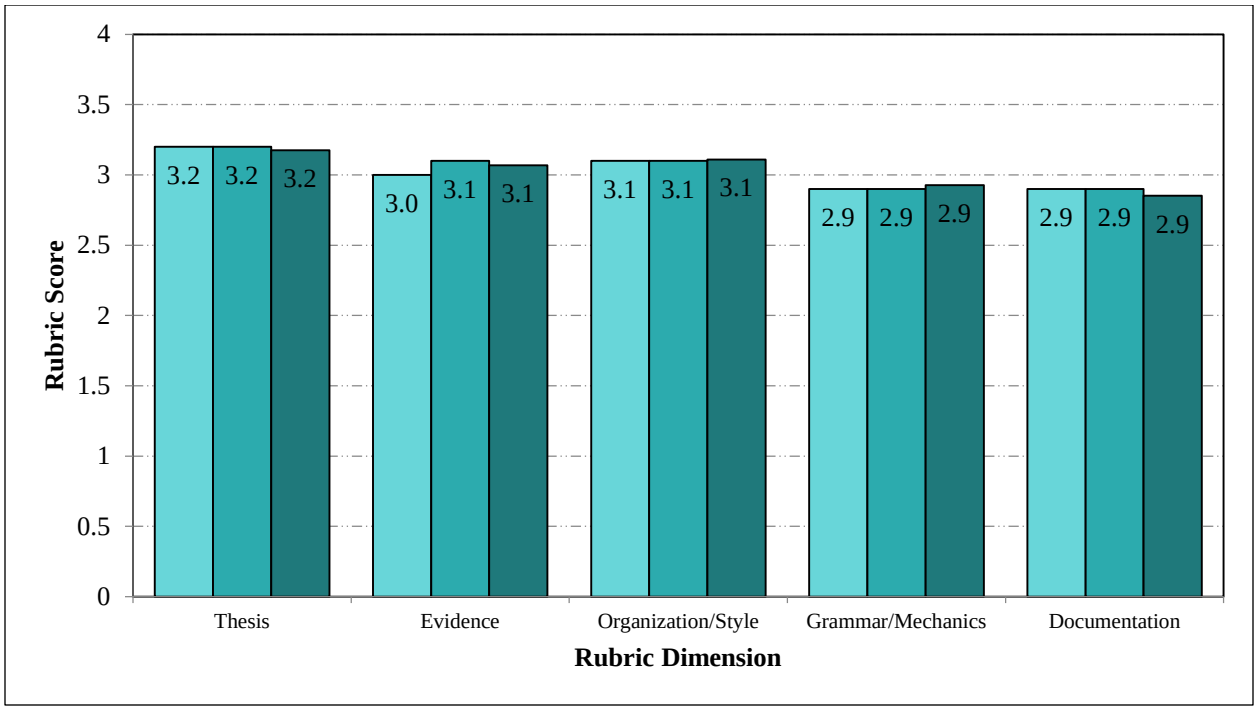

Figure 9. Comparison of mean scores for ENC 1101 through time for fall 2014 (teal),fall 2015 (darker teal), and fall 2016 (darkest teal).

3.2 Comparisons by Site, Format, and Student Type

3.2.1 Dual Enrollment to non-Dual Enrollment Comparison

During the fall 2016 semester, 109 dual enrollment artifacts were collected in ENC 1101 and 782 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 1.0 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 be statistically significantly different. Therefore, we can reject the null hypothesis that the difference in the mean scores of dual enrollment and traditional artifacts can be a result of chance.

df = 889	
Dual enrollment mean	16.0
Dual enrollment standard deviation	3.56
Traditional mean	15.0
Traditional standard deviation	3.73
Effect size	-0.18
p-value	0.008

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 12%. For a graphical representation of this see Figure 10.

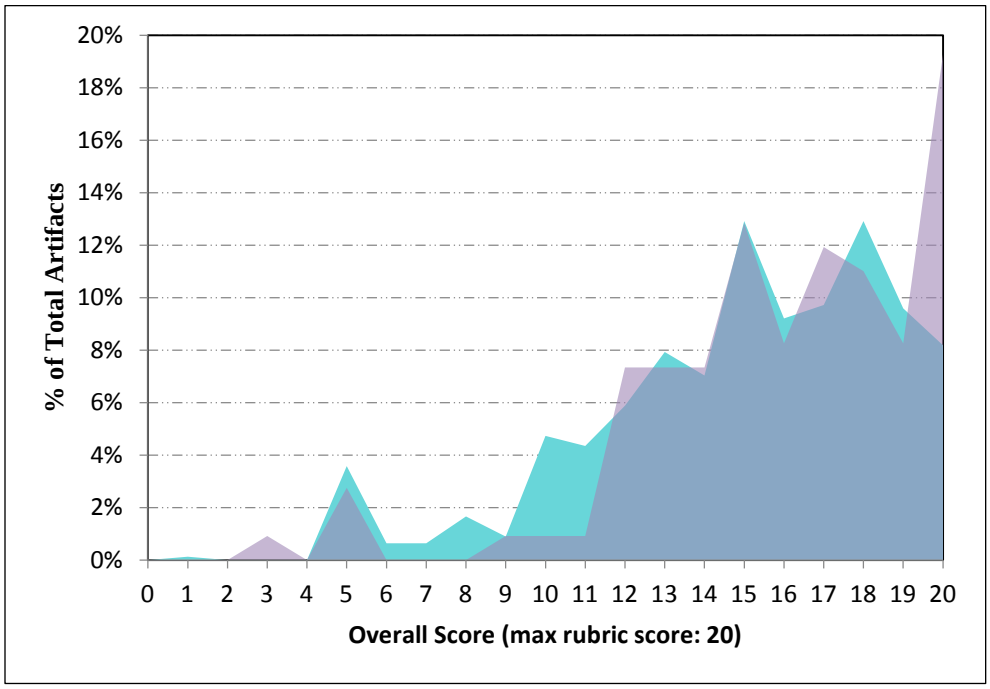

Figure 10. Score distribution for dual enrollment (purple) and traditional (teal) artifacts for ENC 1101.

3.2.2 Online to Traditional Comparison

During the fall 2016 semester, 91 total online artifacts were collected in ENC 1101 and 782 traditional artifacts were collected in ENC 1101. A comparison of mean scores is provided in Table 8. The online artifact mean score is 0.8 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. However, based on the work of Johnson (2013), there is a 17-25% chance that the marginally significant result (p = 0.043) may be false positives (i.e. Type I errors).

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 10%. For a graphical representation of this see Figure 11.

df = 871	
Online mean	15.8
Online standard deviation	3.70
Traditional mean	15.0
Traditional standard deviation	3.73
Effect size	-0.14
p-value	0.043

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

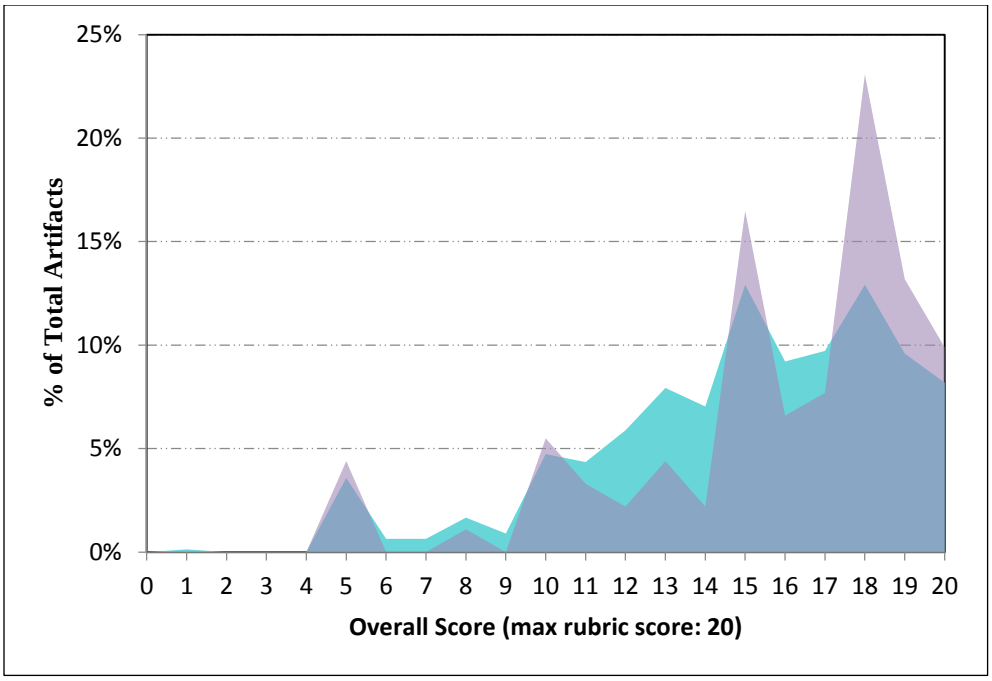

Figure 11. Score distribution for online (purple) and traditional (teal) artifacts of ENC 1101.

3.2.3 Comparison by Site/Campus

Of the 891 artifacts collected from ENC 1101, 31 originated from the Charlotte campus, 219 from the Collier campus, 91 from FSW Online, 24 from the Hendry Glades Center, 417 from the Thomas Edison (Lee) campus, and 109 from offsite (dual enrollment). 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
Charlotte	**3.8**	**3.7**	**3.7**	3.0	**3.2**
Collier	3.3	3.1	3.1	3.0	2.9
FSW Online	3.3	3.2	3.3	3.0	3.1
Hendry-Glades	2.9	2.8	3.1	2.9	2.8
Thomas Edison (Lee)	3.0	2.9	3.0	2.8	2.7
Offsite	3.3	3.4	3.3	**3.1**	2.9

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

Charlotte is consistently the highest exhibiting the highest mean score in four of five dimensions. A plot comparing descriptive statistics of the combined (overall) scores by site is presented in Figure 12. There is extensive overlap across multiple sites although overlap of the central 50% is not shared by all sites. For example, Charlotte exhibits overlap of the central 50% of data with Collier, FSW Online, and offsite, but not with Hendry Glades or Thomas Edison.

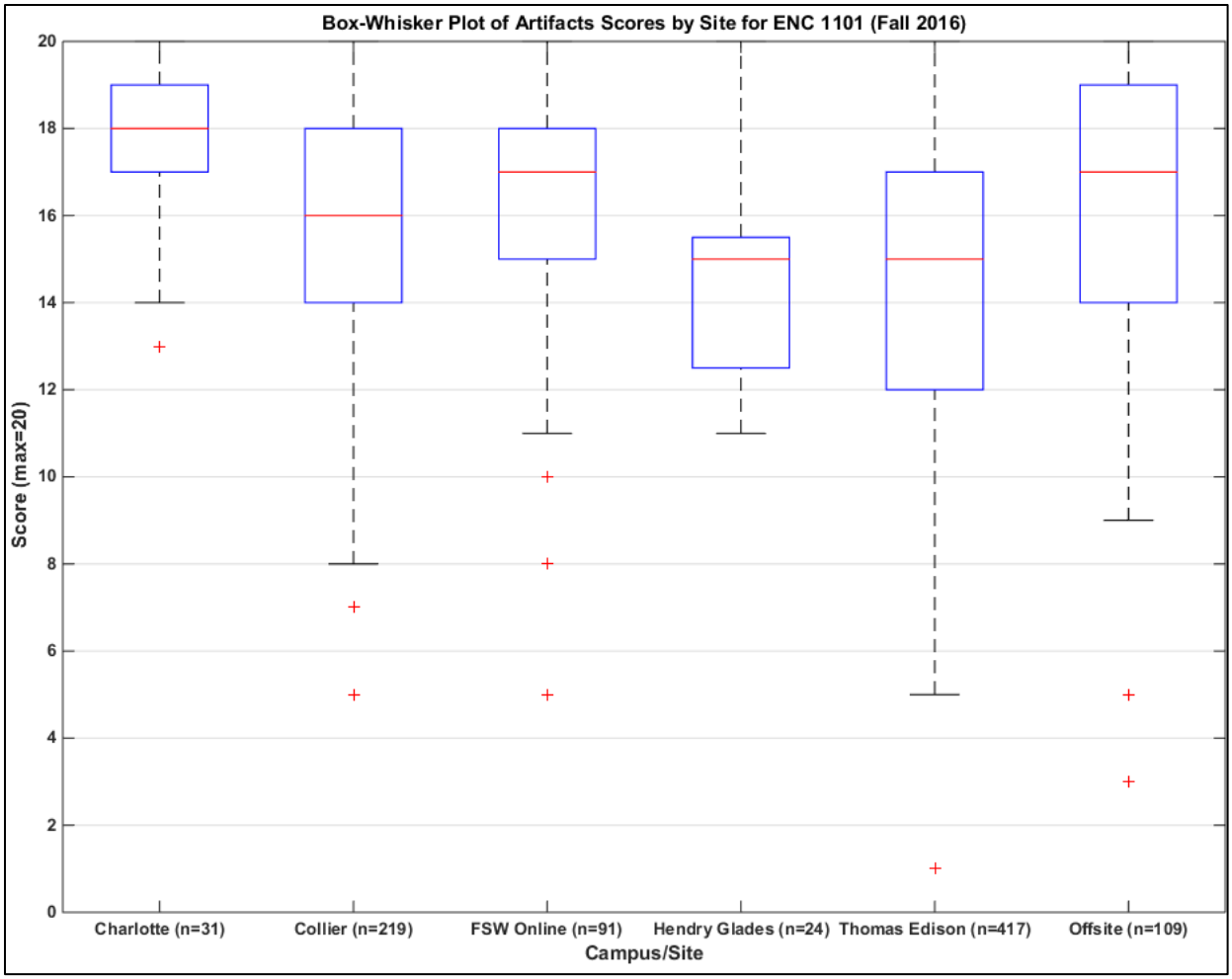

Figure 12. Box-Whisker plot of scores distributed by site for ENC 1101. Red line depicts median score. Upper and lower box boundaries indicate 75% quartile and 25% quartile (box represents central 50% of the scores). Vertical lines represent remaining scores outside central 50% that are not outliers. 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 a statistically significant difference between sites (see Table 10). Therefore, we can reject the null hypothesis that the mean rubric scores at each site are equal to each other and we can 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	511.8	5	102.4	7.66	4.64×10^{-7}	2.22
Within Sites	11,828.1	885	13.4			
Total	12,339.9	890				

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

During the fall 2016 semester, 58 total mini-term artifacts were collected in ENC 1101 and 831 full-term artifacts were collected in ENC 1101. A comparison of mean scores is provided in Table 11. The mini-term artifact mean score is 0.5 higher than full-term 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 not be statistically significantly different. Therefore, we cannot reject the null hypothesis that the difference in the mean scores of mini-term and full-term 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 small effect size. In other words, non-overlap score distribution from mini-term artifacts to full-term artifacts is approximately 5%. For a graphical representation of this see Figure 13.

df = 838	
Mini-term mean	15.6
Mini-term standard deviation	3.76
Full-term mean	15.1
Full-term standard deviation	3.72
Effect size	-0.07
p-value	0.279

Table 11. Comparison of mean scores for mini-term and full-term artifacts. Positive effect sizes indicate a higher mean score for full-term artifacts.

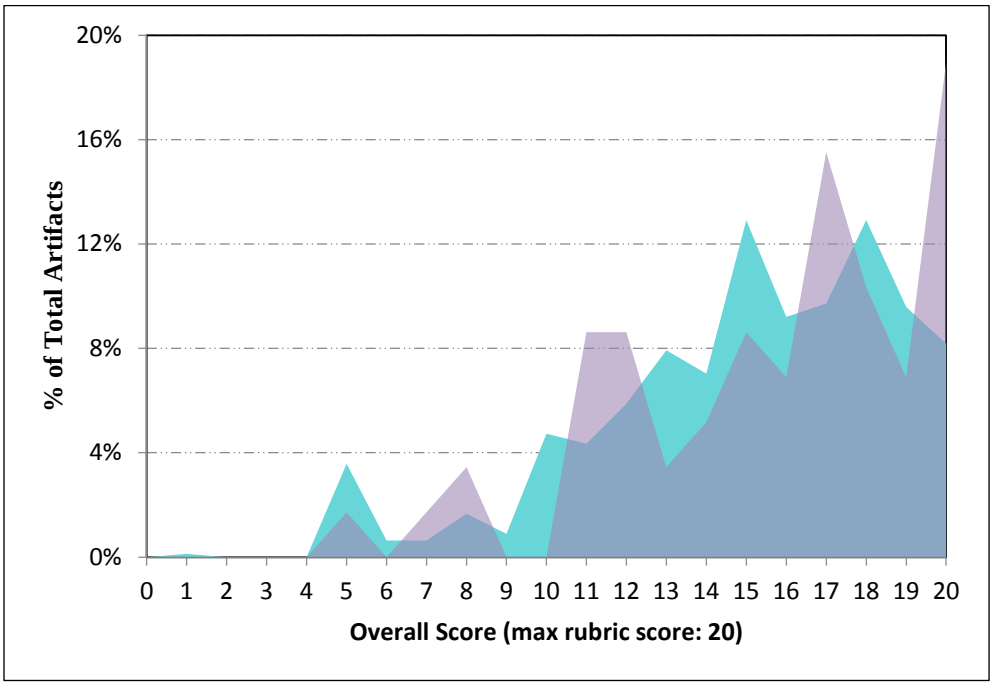

Figure 13. Score distribution for mini-term (purple) and full-term (teal) artifacts of ENC 1101.

4 ENC 1102

4.1 Learning Objectives & Descriptive Statistics

Using common rubric criterion revised based on assessment results of AY 2015-16 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:

- SLO 1: Students must demonstrate the ability to write essays following various rhetorical modes, strategies, and purposes.
- SLO 2 & 3: 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 of measuring the percentage of artifacts scoring a 2 or greater.

For the fall 2016 assessment, 275 artifacts were collected for ENC 1102 from 16 of 19 course sections sampled from 53 course sections offered. One course section did not score all rubric dimensions and so data was excluded as it was unclear if the same rubric was used while the other two sections did not report data. The resultant sample represents 22.5% of the population. The lowest scoring rubric dimension for percentage of artifacts scoring a 2 or greater is Documentation at 91% (Table 12). For a visual comparison of scores by dimension see Figure 14.

Rubric Score	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation
Developing or higher	97%	96%	98%	98%	91%
4	55%	41%	48%	28%	29%
3	33%	37%	35%	55%	45%
2	9%	17%	15%	16%	17%
1	2%	3%	1%	1%	8%
0	1%	1%	1%	1%	1%

Table 12. Percentage of student achievement level by rubric dimension (includes percentage of students scoring in developmental level or higher as per SLO) for ENC 1102.

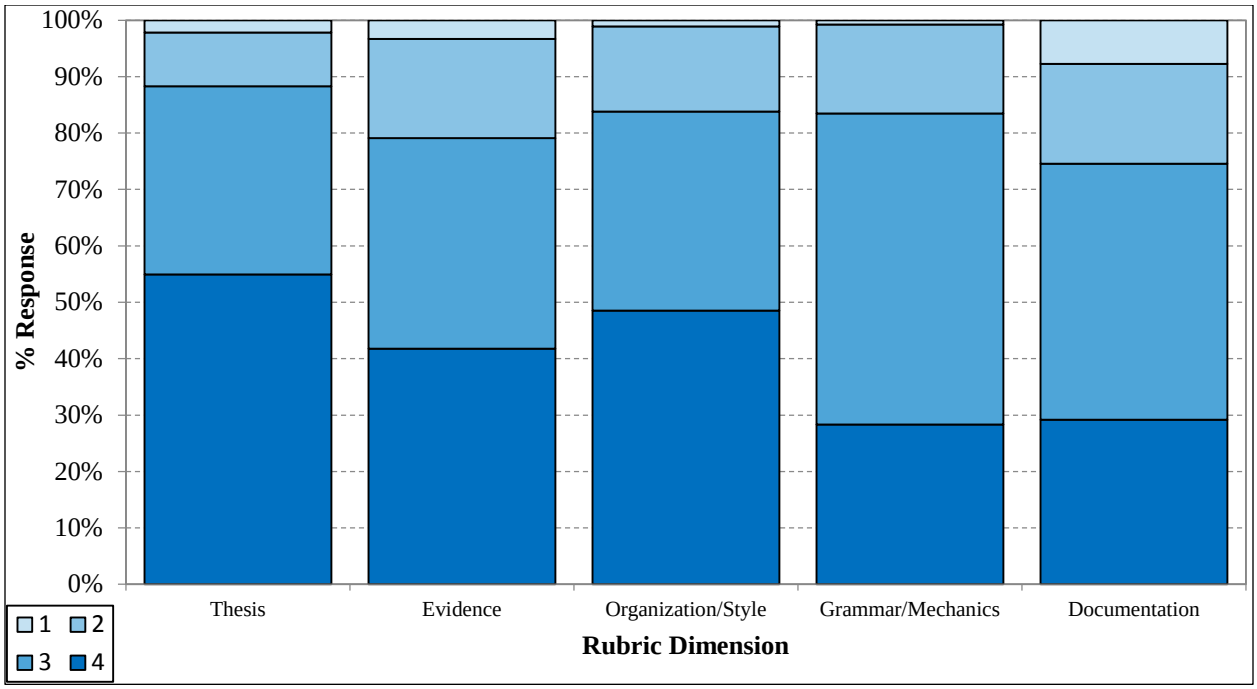

Figure 14. ENC 1102 distribution of rubric scores by dimension.

4.1.2 Descriptive Statistics & Longitudinal Studies

Descriptive statistics for ENC 1102 artifacts can be found in Table 13. A histogram of artifact scores for all 275 artifacts is shown in Figure 15. 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 16). 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	275	275	275	275	275	275
Max	4	4	4	4	4	20
Min	0	0	0	0	0	3
Median	4	3	3	3	3	17
Mode	4	4	4	3	3	17
Mean	3.4	3.2	3.3	3.1	2.9	15.8
Standard deviation	0.80	0.87	0.83	0.75	0.95	3.30
Skewness	-1.40	-0.86	-1.13	-0.86	-0.80	-0.85
Kurtosis	2.11	0.37	1.36	1.96	0.30	0.68

Table 13. Descriptive statistics for ENC 1102 common course assessment.

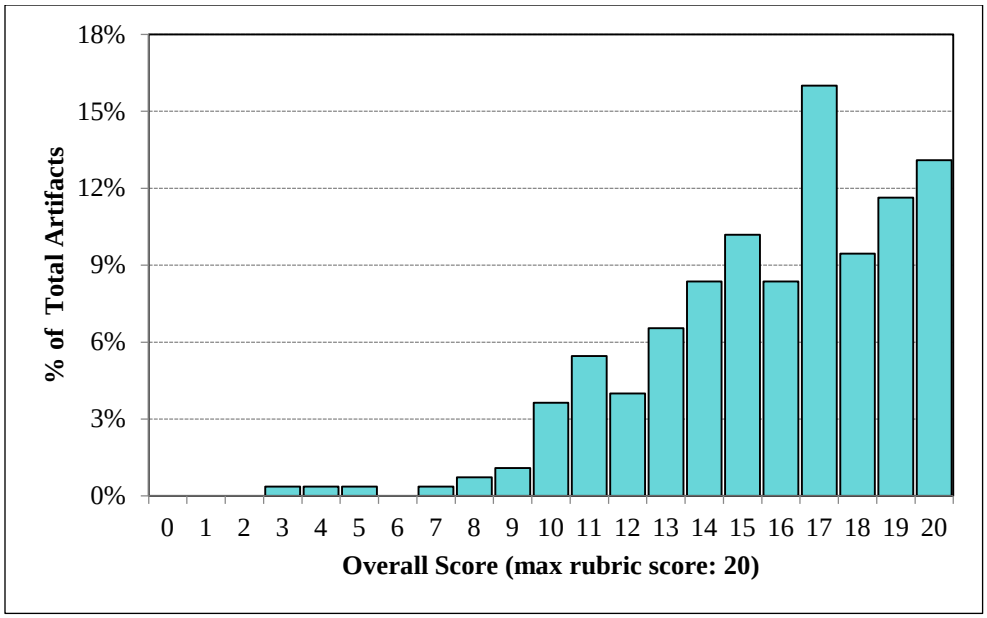

Figure 15. Overall score distribution for ENC 1102 artifacts (fall 2016 term).

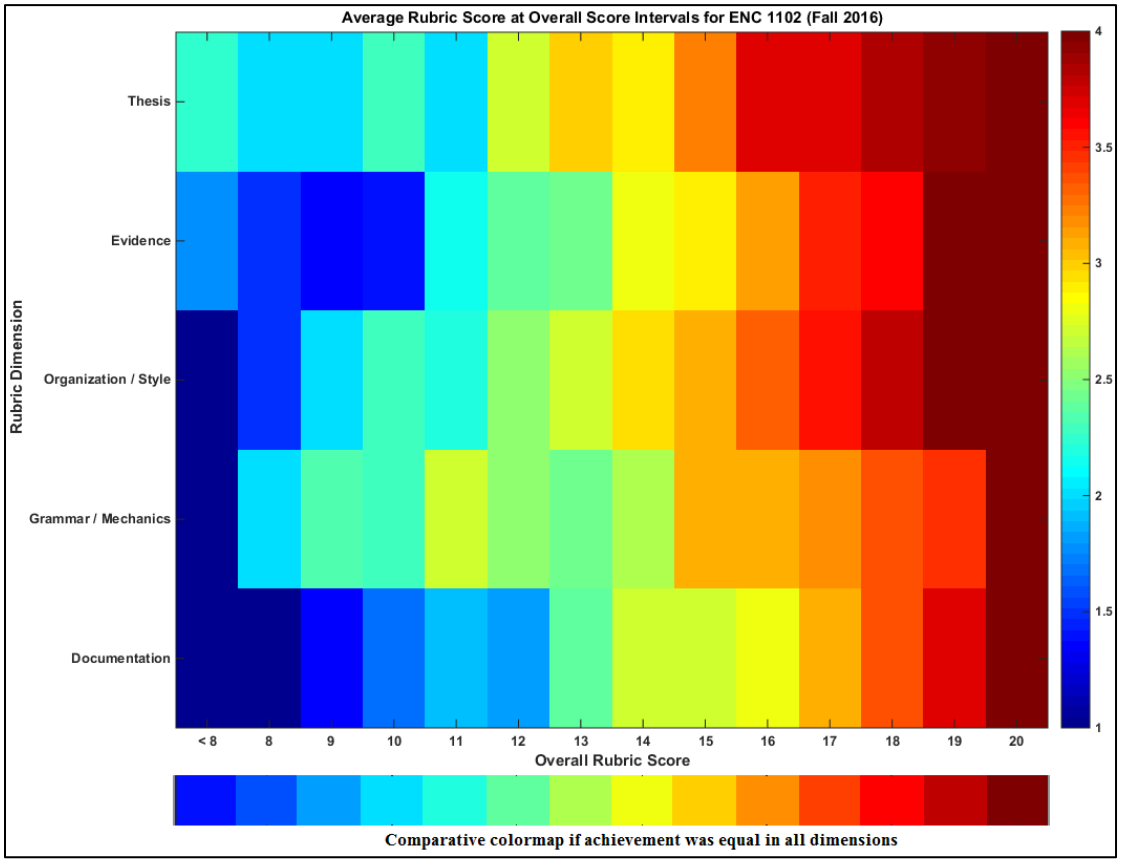

Figure 16. (Top) Colormap of mean scores for each rubric dimension (range: 0-4) based on overall rubric score (combined rubric score of all dimensions, max=20) for ENC 1102. (Bottom) Comparison rubric dimension if dimension score is the same as overall (i.e. artifact overall score is equally distributed 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.

A review of the colormap in Figure 15 shows that the Grammar / Mechanics dimension as the narrowest range of scores. Between 8/20 and 19/20, the Grammar / Mechanics dimension range is 2.0/4 to 3.4/4, a range of 1.4. By comparison, the ranges of other dimensions from 8/20 to 19/20 span from 1.9 to 2.7. At an overall score of 10/20, Thesis dimension is exceptionally strong even at low overall scores. From a student performance perspective, both high achieving and low achieving students exhibit more similar capabilities in Grammar / Mechanics when compared with other dimensions across that range. Whether a student scores a 10/20 or an 18/20, the Grammar / Mechanics dimension would typically be a 2/4 or 3/4, respectively, whereas other dimensions are more likely to be a 1/4 or 4/4, respectively.

The Documentation dimension also exhibits unique characteristics when compared with other dimensions. In mid-to-low range overall scores the Documentation tends to lag behind other dimensions. For example, at an overall score of 12/20, the Documentation dimension exhibits a mean score of 1.8/4. By comparison, at that same overall score other dimensions range from 2.4/4 to 2.7/4. From a student performance perspective, under achieving students tend to struggle most with Documentation.

A comparison of fall 2016 results with past results is shown in Figure 17 below. Results exhibit large increases compared with the last two years of data. Fall 2016 included a substantially larger sample compared to previous years. As a result, sample size includes a more appropriate diversity in courses represented and may be a more appropriate representation of the actual. The Thesis dimension continues to be the dimension with the highest mean score with a mean score in all years. Further, the fall 2016 term is the first in which the Grammar/Mechanics dimension is not the lowest scoring. For fall 2016, the Documentation dimension exhibits the lowest mean score.

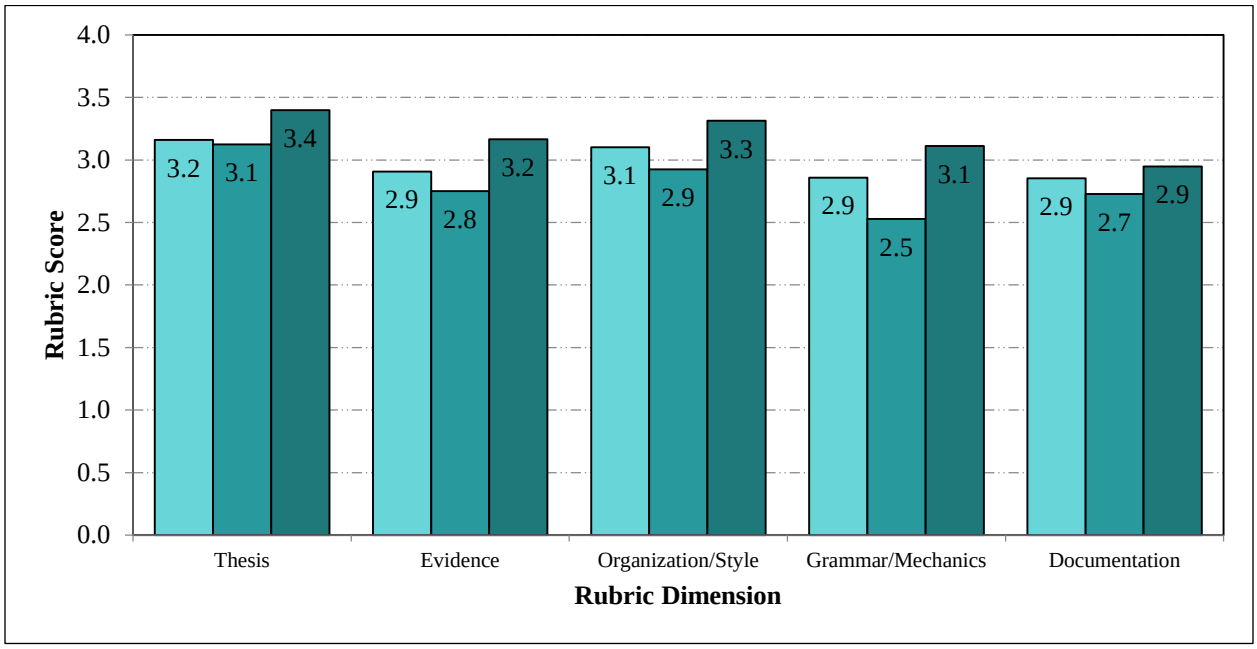

Figure 17. Comparison of mean scores for ENC 1102 through time for fall 2014 (teal), fall 2015 (dark teal), and fall 2016 (darkest teal).

4.2 Comparison by Site, Format, and Student Type

4.2.1 Dual Enrollment to non-Dual Enrollment Comparison

During the fall 2016 semester, 23 dual enrollment artifacts were collected in ENC 1102 and 252 traditional (non-online) artifacts were collected in ENC 1102. A comparison of mean scores is provided in Table 14. The dual enrollment mean score is 2.4 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 be statistically significantly different. Therefore, we can reject the null hypothesis that the difference in the mean scores of dual enrollment and traditional artifacts can be a result of chance.

df = 273	
Dual enrollment mean	18.0
Dual enrollment standard deviation	2.68
Traditional mean	15.6
Traditional standard deviation	3.28
Effect size	-0.49
p-value	3.95×10^{-4}

Table 14. 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 medium effect size. In other words, non-overlap score distribution from online artifacts to traditional artifacts is approximately 32%. For a graphical representation of this see Figure 18.

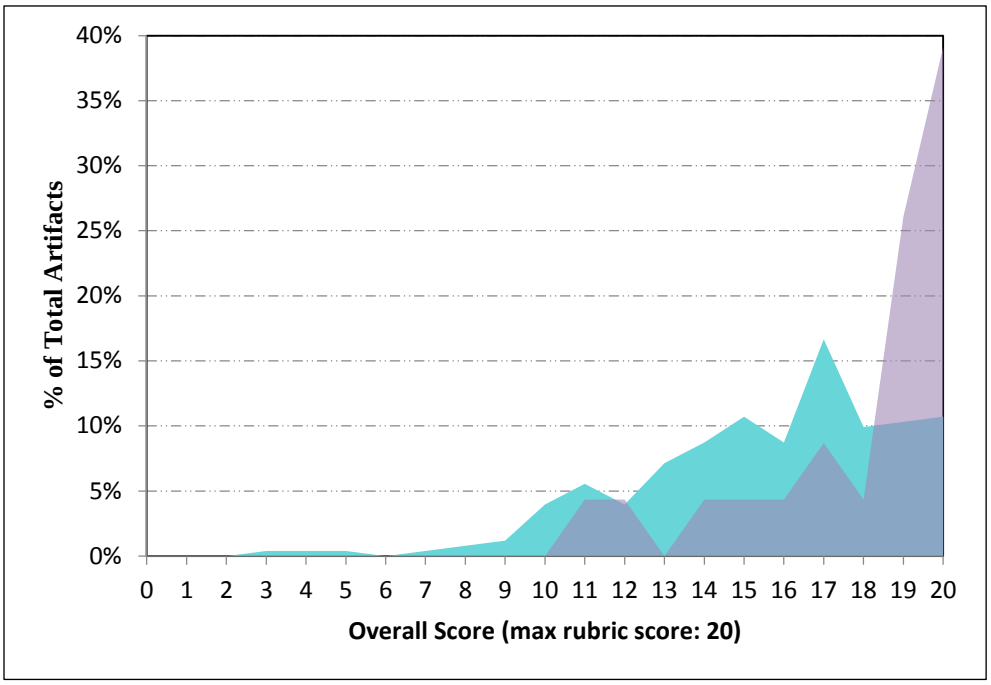

Figure 18. Score distribution for dual enrollment (purple) and traditional (teal) artifacts for ENC 1102.

4.2.2 Online to Traditional Comparison

During the fall 2016 semester, 16 total online artifacts were collected in ENC 1102 and 252 traditional artifacts were collected in ENC 1102. A comparison of mean scores is provided in Table 13. The online artifact mean score is 1.3 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. However, based on the work of Johnson (2013), there is a 17-25% chance that the marginally significant result (p = 0.024) may be false positives (i.e. Type I errors).

df = 266	
Online mean	16.9
Online standard deviation	2.02
Traditional mean	15.6
Traditional standard deviation	3.28
Effect size	-0.30
p-value	0.024

Table 15. 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 small-to-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 19.

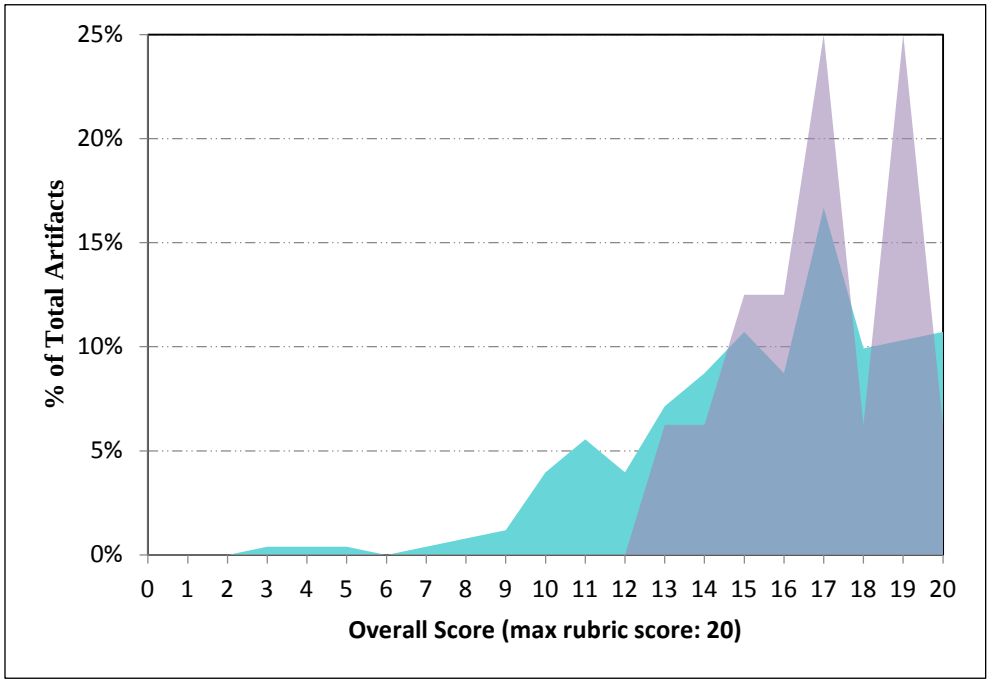

Figure 19. Score distribution for online (purple) and traditional (teal) artifacts of ENC 1102.

4.2.3 Comparison by Site/Campus

Of the 275 artifacts collected from ENC 1102, 42 originated from the Charlotte campus, 44 from the Collier campus, 16 from FSW Online, 11 from the Hendry Glades Center, 139 from the Thomas Edison (Lee) campus, and 23 from offsite (dual enrollment). Mean scores across sites are quite variable. Offsite (dual enrollment) exhibits the highest mean score in two of five dimensions. Collier and FSW Online each exhibit the highest in two others (the sites share the highest for Documentation). A comparison of mean scores by rubric dimension is provided in Table 16. A plot comparing descriptive statistics of the combined (overall) scores by site is presented in Figure 20. There is extensive overlap between sites.

	Thesis	Evidence	Organization / Style	Grammar / Mechanics	Documentation
Charlotte	3.3	3.3	3.1	3.0	3.1
Collier	**3.8**	3.4	3.6	3.4	**3.3**
FSW Online	3.7	3.3	**3.8**	3.0	**3.3**
Hendry Glades	3.2	3.2	3.4	2.9	3.2
Thomas Edison (Lee)	3.2	2.9	3.1	2.9	2.6
Offsite	3.8	**3.7**	3.7	**3.7**	3.2

Table 16. Comparison of mean scores by site for ENC 1102. Bold denotes highest mean score in that dimension among all sites.

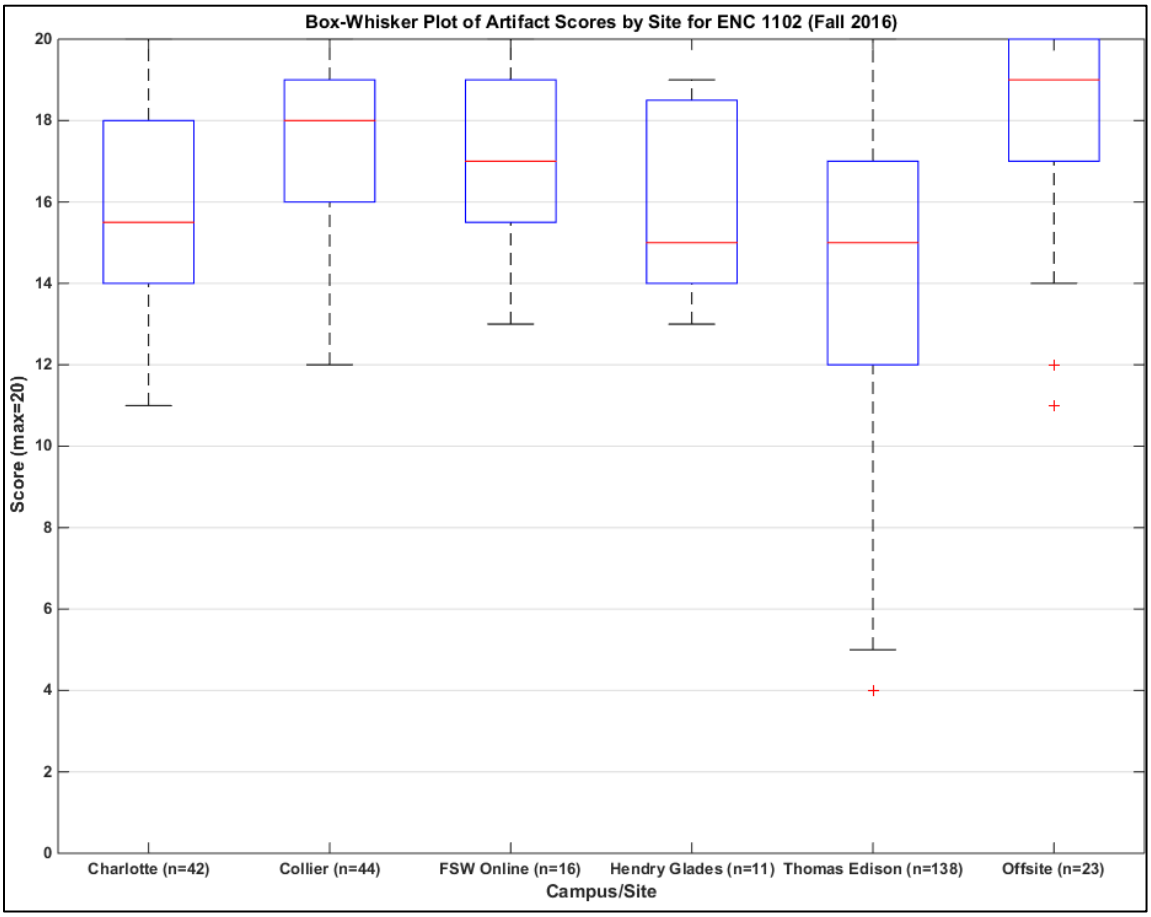

Figure 20. Box-Whisker plot of scores distributed by site for ENC 1102. Red line depicts median score. Upper and lower box boundaries indicate 75% quartile and 25% quartile (box represents central 50% of the scores). Vertical lines represent remaining scores outside central 50% that are not outliers. Red '+'s denote outliers.

A one-way analysis of variance was used to compare means of the combined rubric scores at each site. Results of the ANOVA exhibit a statistically significant difference between sites (see Table 17). Therefore, we can reject the null hypothesis that the mean rubric scores at each site are equal to each other and we can 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	379.4	5	75.9	8.33	$2.45x10^{-7}$	2.25
Within Sites	2440.8	268	9.1			
Total	2820.2	273				

Table 17. Results of one-way ANOVA of combined rubric scores at each site for ENC 1102.

4.2.4 Mini-term to Full-term Comparison

During the fall 2016 semester, 29 total mini-term artifacts were collected in ENC 1102 and 246 full-term artifacts were collected in ENC 1102. A comparison of mean scores is provided in Table 18. The mini-term artifact mean score is 0.6 higher than full-term 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 not be statistically significantly different. Therefore, we cannot reject the null hypothesis that the difference in the mean scores of mini-term and full-term 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 small effect size. In other words, non-overlap score distribution from mini-term artifacts to full-term artifacts is approximately 15%. For a graphical representation of this see Figure 21.

df = 273	
Mini-term mean	15.9
Mini-term standard deviation	4.85
Full-term mean	15.8
Full-term standard deviation	3.08
Effect size	-0.02
p-value	0.882

Table 18. Comparison of mean scores for mini-term and full-term artifacts. Positive effect sizes indicate a higher mean score for full-term artifacts.

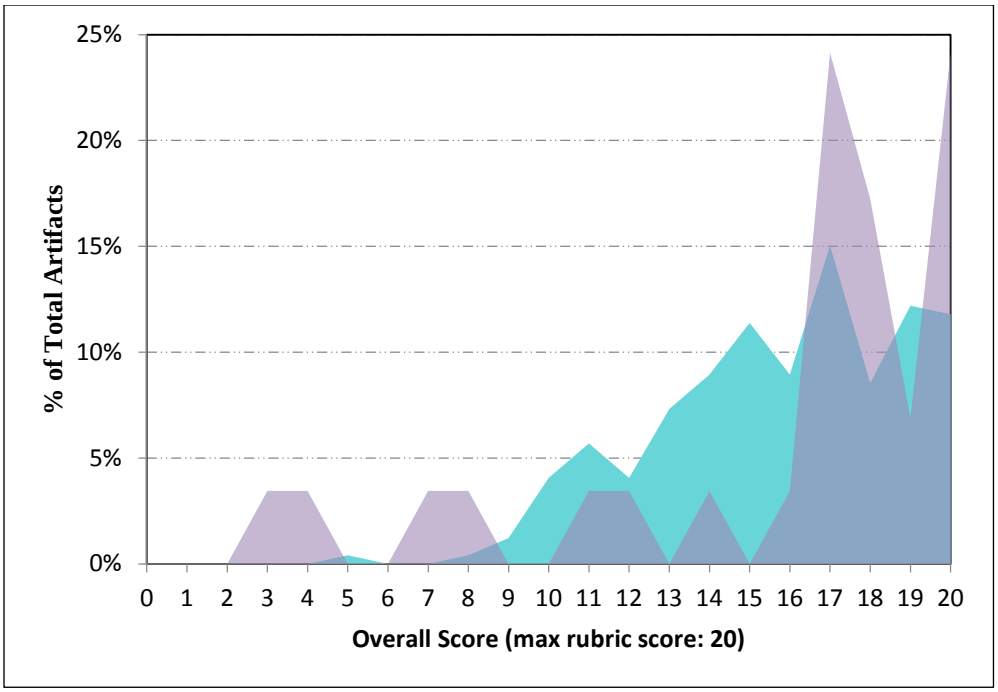

Figure 21. Score distribution for mini-term (purple) and full-term (teal) artifacts of ENC 1102.

5 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 an updated rubric in response to the fall 2015 pilot study with five dimensions for both ENC 1101 and ENC 1102. The assessment plan uses a random sample of 30% 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.

A drilldown of ENC 0022 results are as follows:

1. All seven rubric dimensions had ≥ 80% 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 20/28 and 27/28, and is moderately negatively skewed, meaning scores are shifted towards the higher range.
3. In a study comparing rubric achievement based on overall score, all students are weak in the Research dimension compared with others and high moderate-to-high achieving students are strongest in Organization compared with other dimensions.
4. In a longitudinal study, consistency across all areas except for Research, which exhibits a sharp decline in the most recent term. Fall 2016 data do exhibit an extensive percentage of 0s reported for Research (10%) compared with previous years (0% for fall 2015 and 0% for fall 2014). And while one course section does exhibit 0s universally for all reported scores in Research, 0s are reported in other sections as well, so it does appear to be a real, if less substantial, characteristic of the data.

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 Charlotte campus is the lowest in 6 of 7 dimensions. Collier campus exhibits the highest scores in 4 of 7 dimensions.
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 five rubric dimensions had > 89% achievement at level 2 or higher. The lowest dimension was Documentation.
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 achieving students tend to be equal in all dimensions while over achieving students never extend above average students in the Grammar / Mechanics dimension. Additionally, under achieving students tend to struggle with Documentation above all other areas
4. In a longitudinal study, results exhibit consistency across all areas over time. The Thesis dimension continues to be the dimension with the highest mean score with a mean score of 3.2/4 in all years. The Grammar/Mechanics and Documentation dimensions continue to be the lowest scoring in all years (2.9/4).
5. In a study comparing dual enrollment to traditional (non-online) artifacts, the dual enrollment mean score is 1.0 higher than traditional artifacts and results are statistically significant.
6. In a study comparing online to traditional artifacts, the online artifact mean score is 0.8 higher than traditional artifacts and was statistically significant.
7. In a cross-campus comparison, scores varied greatly across rubric dimensions. Charlotte is consistently the highest exhibiting the highest mean score in four of five dimensions.
8. In a comparison of mini-term courses to full-term courses, the mini-term courses artifact mean score is 0.5 higher than full-term artifacts although results are not statistically significantly different.

A drilldown of ENC 1102 results are as follows:

1. All seven rubric dimensions had > 90% achievement at level 2 or higher. The lowest dimension was Documentation.
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, both high achieving and low achieving students exhibit more similar capabilities in Grammar / Mechanics when compared with other dimensions across that range. Whether a student scores a 10/20 or an 18/20, the Grammar / Mechanics dimension would typically be a 2/4 or 3/4, respectively, whereas other dimensions are more likely to be a 1/4 or 4/4, respectively. Also, under achieving students tend to struggle most with Documentation.

4. In a longitudinal study, results exhibit large increases compared with the last two years of data. Fall 2016 included a substantially larger sample compared to previous years. As a result, sample size includes a more appropriate diversity in courses represented and may be a more appropriate representation of the actual.
5. In a study comparing dual enrollment to traditional (non-online) artifacts, the dual enrollment artifact mean score is 2.4 higher than traditional artifact and was found to be statistically significantly different.
6. In a study comparing online to traditional artifacts, the online artifact mean score is 1.3 higher than traditional artifacts and was found to be statistically significant.
7. In a cross-campus comparison, scores varied greatly across rubric dimensions. Offsite (dual enrollment) exhibits the highest mean score in two of five dimensions. Collier and FSW Online each exhibit the highest in two others (the sites share the highest for Documentation). Results of the ANOVA exhibit a statistically significant difference between sites
8. In a comparison of mini-term courses to full-term courses, the mini-term courses artifact mean score is 0.6 higher than full-term artifacts although results are not statistically significantly different.

6 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.

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.