

2015-2016 Written Communication Assessment Report: Results Summary

Executive Summary

One hundred and ninety-nine student papers were scored by two separate raters using AAC&U Written Communication VALUE Rubric. Five traits (Context and Purpose of Writing, Content Development, Genre and Disciplinary Conventions, Sources and Evidence, and Control of Syntax and Mechanics) were scored on a four-point scale with 1=Benchmark, 2=Milestone A, 3=Milestone B, and 4=Capstone. Analyzed scores are averages of individual raters' scores. A total of 10%, or 20 of the 199 papers, had average scores differing more than 1.00 point requiring a 3rd rater which was an improvement from the pilot assessment in which 36%, or 12 of the 33 papers, required an additional rater. The mean differences between raters' scores for the written communication traits decreased from the pilot in 2014-2015 to the 2015-2016 assessment, as did the standard deviations of those means- showing that individual raters' scores became closer together, and the validity and reliability of the scores given by the raters have improved due to the enhancements in the rater training. For 2015-2016, raters were most consistent on the scoring of Control of Syntax and Mechanics and least consistent in the scoring of Content Development.

Fifty-two percent, or 104 of the 200 student papers achieved an Overall Total Average score for Written Communication classified at the Capstone or Milestone B performance level; this was a slight decrease from the pilot assessment in which 58%, or 50 of the 87 papers were classified at the Capstone or Milestone B performance levels. The Written Communication traits of Context and Purpose of Writing and Control of Syntax and Mechanics were where Washburn students were strongest in 2015-2016, and the lowest traits included Genre and Disciplinary Conventions and Sources and Evidence. The percent of student papers classified at the Capstone or Milestone B performance levels remained consistent from 2014-2015 for Sources and Evidence and Control of Syntax and Mechanics, and decreased slightly for Genre and Disciplinary Conventions and Content Development. The number of student papers scored increased dramatically from 87 in 2014-2015 to 200 in 2015-2016 which increases in the amount of variation among the student assessments- normally accompanied by a decrease in mean scores and an increased standard deviations of those mean scores. However, the score means only declined a bit and the standard deviations decreased showing that validity/reliability of the raters' scores improved.

Key Findings for 2015-2016 Rater Differences

- 10% of the papers had average total scores differing more than 1 point- requiring a 3rd rater; 18% of the scores from the two raters did not differ, 63% differed less than 1 point, and 9% by 1 point
- Raters were most consistent for Control of Syntax and Mechanics with 97% of ratings differing by 0-1 points; followed by Sources and Evidence and Context and Purpose of Writing traits, at 95% and 93%
- Content Development seemed to be the trait in which raters were least consistent with 88% of the ratings differing by 0 or 1, and almost 12% of the trait ratings differing by 2-3 points

Key Findings for 2015-2016 Performance

- 52% of student papers achieved Overall Total Average scores categorized at the Milestone B or Capstone performance levels
- 6% of the papers attained an overall total average score at the Capstone performance level, 47% scored at Milestone B, 43% at Milestone A, 6% at Benchmark, and 0% at the Unobserved performance level

Key Findings for 2015-2016 Performance (continued)

- 73% attained average scores classified at the Milestone B or Capstone performance levels for the Context and Purpose trait, 64% for Syntax and Mechanics, 53% for Content Development, and 50% for the Sources and Evidence and for the Disciplinary Conventions traits
- The Context and Purpose trait had the highest mean average score at 2.74, followed by Syntax and Mechanics at 2.50 and Content Development at 2.46; Sources and Evidence rendered a mean score of 2.34, and the Disciplinary Conventions trait had the lowest mean at 2.32
- Washburn students performed best on the Context and Purpose trait, and performed the lowest on the Disciplinary Conventions and Sources and Evidence written communication traits

Key Findings for Pilot 2014-2015 and 2015-2016 Rater Differences

- 87 papers were scored (33 by two raters) for the 2014-2015 pilot, which increased to 199 papers scored by two individual raters for 2015-2016
- 18% of the Overall Average Total scores from two raters did not differ in 2015-2016, an increase compared to 0% in 2014-2015
- 10% had a difference of greater than 1 point and required additional rater in 2015-2016 which was a decrease from the 36% of student papers that required additional rater in 2014-2015
- Mean differences between raters for the overall total average scores in 2014-2015 was 0.89 with a standard deviation of 0.472, and the mean difference between raters for 2015-2016 decreased by .042 to 0.47 with the standard deviation decreasing by 0.026 to 0.446
- Mean differences between raters' scores for the written communication traits decreased from the pilot in 2014-2015 to the 2015-2016 assessment, as did the standard deviations of those means- showing that individual raters' scores became closer together, and the validity and reliability of the scores given by the raters have improved

Key Findings for Pilot 2014-2015 and 2015-2016 Performance

- 58% of the 87 student presentations in 2014-2015 attained a rounded overall average score classified at the Milestone B or Capstone performance levels; this decreased by 5.5 percentage points in 2015-2016 where 52% of the 200 overall total average scores were categorized at the Milestone B or Capstone performance levels
- The percent of rounded average scores classified at the Milestone B or Capstone performance levels decreased from 2014-2015 to 2015-2016 for all four of the written communication traits assessed in both years: Disciplinary Conventions (57.5% to 49.5%), Content Development (58.6% to 53.0%), Syntax and Mechanics (64.4% to 63.5%), and Sources and Evidence (50.6% to 50.0%)
- The Context and Purpose written communication trait was not assessed during the pilot, but almost 73% of student papers scored at the Milestone B or Capstone performance levels in 2015-2016
- An increase in the number of students assessed (87-200) leads to an increase in the amount of variation among scores; resulting in decreased mean scores and increased standard deviations.
- Overall total average and trait mean scores did decrease slightly from 2014-2015 to 2015-2016 (tenths of a point); but the standard deviations decreased instead of increased, except for the Syntax and Mechanics trait which increased slightly.

2015-2016 Written Communication Assessment Report: Results Summary

Introduction: A total of 199 senior student papers were scored by two separate raters using AAC&U Written Communication VALUE Rubric. Five traits (Context and Purpose of Writing, Content Development, Genre and Disciplinary Conventions, Sources and Evidence, and Control of Syntax and Mechanics) were scored on a four-point scale with 1=Benchmark, 2=Milestone A, 3=Milestone B, and 4=Capstone. When there were only two raters (their overall mean scores did not differ by more than 1.0) the mean for each trait and the overall total is the average of the two rater's scores. If the two raters differed by more than 1.0 for the overall total scores, then a third rater also scored the paper. In cases of three raters, the mean for each trait and the overall total is the average of the three raters' scores.

Differences between Individual Raters: Approximately 18% of the Overall Total Average scores from the two raters did not differ, 63% differed less than 1 point, and almost 9% differed by 1 full point. Therefore, 90% of the student papers did not have scores that differed enough to require a 3rd rater, but 10% had a difference of greater than 1 point (see Table 1) and an additional rater was necessary.

Table 1. Frequencies and Percentages of Overall Average Score Differences between Raters

Average Score Difference	Frequency	Percent	Cumulative Percent
.00	36	18.1%	18.1%
.20	63	31.7%	49.8%
.40	28	14.1%	63.9%
.60	21	10.6%	74.5%
.80	14	7.0%	81.5%
1.00	17	8.5%	90.0%
1.20	10	5.0%	95.0%
1.40	3	1.5%	96.5%
1.60	4	2.0%	98.5%
1.80	1	0.5%	99.0%
2.00	1	0.5%	99.5%
2.20	1	0.5%	100.0%
Total	199	100.0%	10.0%

The mean difference between raters for the overall average scores was 0.468 with a standard deviation of 0.446, and differences ranged from 0 to 2.20 (see Table 2). The differences between raters for the individual traits ranged from 0 to 3 points for the Purpose, Content, and Conventions traits, and ranged from 0 to 2 points for the Evidence and Mechanics traits. The mean rater differences for Content Development had the highest mean difference at 0.62, followed by the Genre and Disciplinary Conventions trait at 0.60. The mean difference between rater scores for Sources and Evidence was 0.59, Context and Purpose was 0.57, and the trait with the smallest mean difference between raters was Control of Syntax and Mechanics at 0.55.

Table 2. Descriptive Statistics for Rater Score Differences

Statistic (N=199)	Purpose Difference	Content Difference	Conventions Difference	Evidence Difference	Mechanics Difference	Overall Average Score Difference
Mean	0.57	0.62	0.60	0.59	0.55	0.468
Std. Dev.	.639	.699	.666	.595	.556	0.446
Mode	0	0	0	1	1	0.00
Minimum	0	0	0	0	0	0.00
Maximum	3	3	3	2	2	2.20

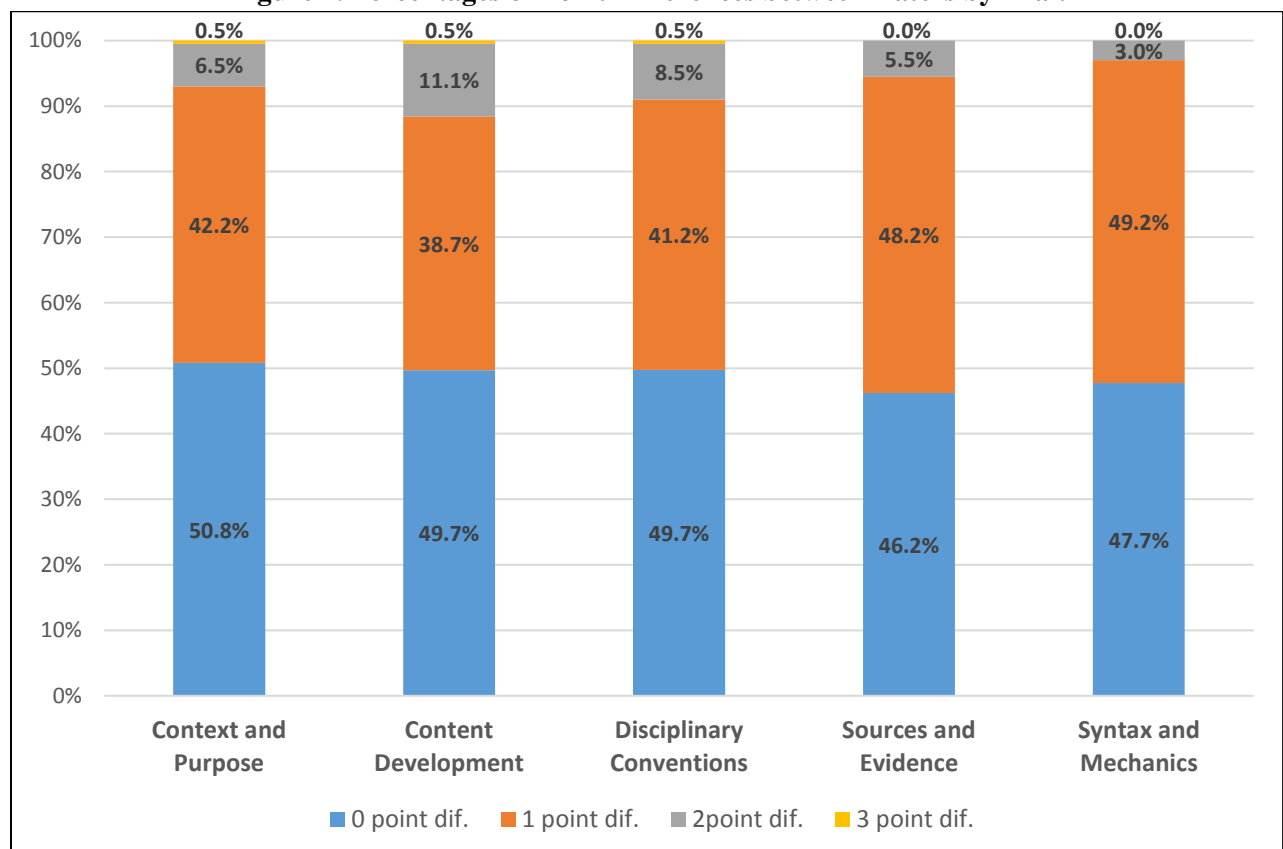
Frequencies and percentages of the difference scores for each trait on the Written Communication rubric were computed; and the 0- and 1-point score differences, as well as the 2- and 3-point score differences, were combined to provide frequencies and percentages for the closer and more widely disparate scores by trait (see table below). Raters seemed to be most consistent for the Mechanics trait scores with 97% of the ratings for that trait differing by 0-1 points. Raters were fairly consistent for the Evidence and Purpose traits, with 95% and 93% of the ratings having score differences equal to 0 or 1, respectively; and the Conventions trait had 91% of the ratings differing by 0-1 points. Content Development seemed to be the trait in which raters were least consistent with 88% of the ratings differing by 0 or 1, and almost 12% of the trait ratings differing by 2-3 points.

Table 3. Frequencies and Percentages of Rater Score Differences by Trait

Rater Score Differences	Purpose		Content		Conventions		Evidence		Mechanics	
	Freq.	Perc.	Freq.	Perc.	Freq.	Perc.	Freq.	Perc.	Freq.	Perc.
0	101	50.8%	99	49.7%	99	49.7%	92	46.2%	95	47.7%
1	84	42.2%	77	38.7%	82	41.2%	96	48.2%	98	49.2%
0 or 1	185	93.0%	176	88.4%	181	91.0%	188	94.5%	193	97.0%
2	13	6.5%	22	11.1%	17	8.5%	11	5.5%	6	3.0%
3	1	0.5%	1	0.5%	1	0.5%	0	0.0%	0	0.0%
2 or 3	14	7.0%	23	11.6%	18	9.0%	11	5.5%	6	3.0%

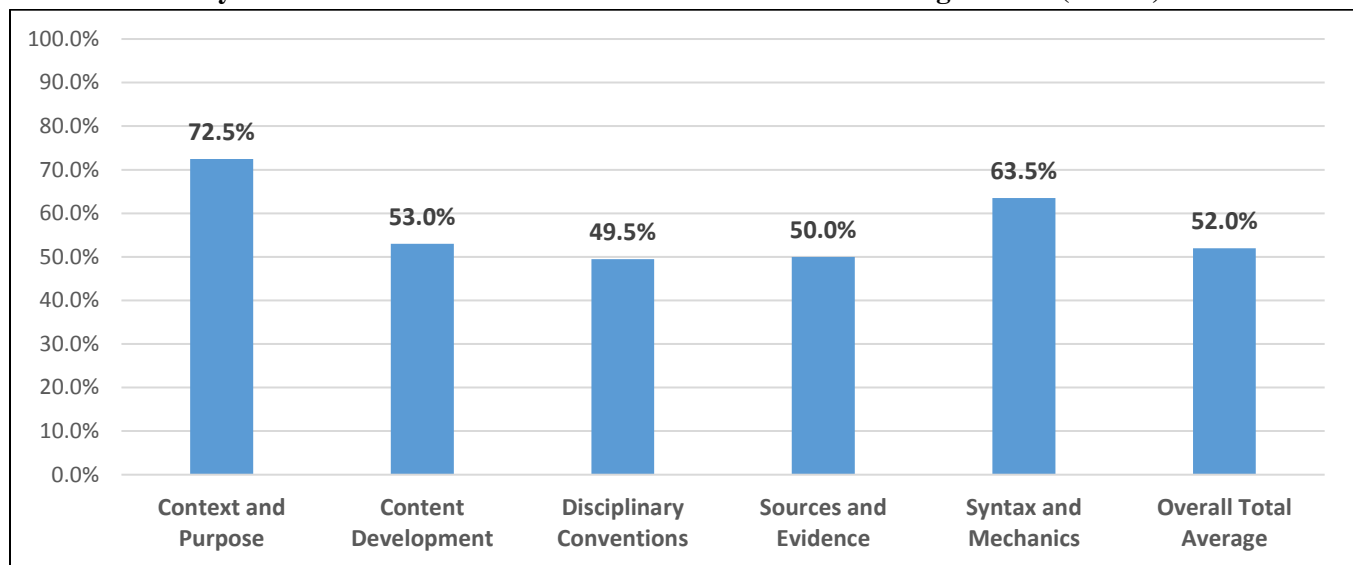
Figure 1 visually presents the percentages of ratings for each trait on the AAC&U Written Communication VALUE Rubric by the point differences between rater scores.

Figure 1. Percentages of Point Differences between Raters by Trait



Percent Scoring at Milestone B or Capstone Performance Levels for Overall Total Average and Traits: Of the 200 student papers, 52% achieved an overall total average score at the Milestone B (score=3) or Capstone (score=4) performance levels (see Figure 2). Almost 73% of the student papers attained average scores at the Milestone B or Capstone levels for the Context and Purpose written communication trait, 64% for Syntax and Mechanics, and 53% achieved scores of 3 or 4 (the Milestone B or Capstone performance levels) for the Content Development trait. Half (50%) of the student papers reached the Milestone B or Capstone performance levels with scores of 3 or 4 for the Sources and Evidence trait and for the Disciplinary Conventions trait assessed by the AAC&U Written Communication Rubric.

Figure 2. Percent of Students with Scores of 3 (Milestone B) or 4 (Capstone) by Written Communication Trait and Overall Total Average Scores (N=200)



Descriptive Statistics for Overall Total Average and Average Trait Scores: The overall total scores for the Written Communication papers ranged from 0.80 to 4.00 with a mean or “average score” of 2.47, and a median or “middle score” of 2.50. Four different modes or “most frequent scores” were revealed in the overall total average scores with 13 students each scoring 1.80, 2.20, 2.30, and 2.70 (see table 4). The consistency of the mean and median, accompanied by the standard deviation, or “average score distance” from the mean, of 0.638 indicate that the overall total average scores mostly cluster around the mean. Also, the score frequency sharply decreases towards the minimum value of 0.80 and maximum value of 4.00 creating a more pointed/steep shape for the score distribution than that of the normal bell curve.

A larger standard deviation value signifies the scores are more spread out on the distribution, while a smaller standard deviation signifies that more of the scores are clustered at or near the mean, and the frequency of scores above or below the mean taper off towards the ends of the score distribution. If the mean is equal to the median, then the distribution is symmetric and has zero skewness. A positive skew occurs when the mean is greater than the median so the bulk of the score distribution is concentrated on the left of the mean (lower scores). A negative skew occurs when the mean is less than the median so the bulk of the score distribution is concentrated on the right of the mean (higher scores).

Table 4. Descriptive Statistics for Written Communication Assessment Average Scores (N=200)

Statistic	Context and Purpose	Content Development	Disciplinary Conventions	Sources and Evidence	Syntax and Mechanics	Overall Total Average
Mean	2.74	2.46	2.32	2.34	2.50	2.47
Median	3.00	2.50	2.33	2.42	2.50	2.50
Mode	3.00	2.00	2.00	2.00	3.00	1.80, 2.20, 2.30, 2.70
Std. Dev.	0.670	0.758	0.756	0.774	0.687	0.638
Min.	1.00	0.50	0.50	0.00	0.50	0.80
Max.	4.00	4.00	4.00	4.00	4.00	4.00

The AAC&U Written Communication Rubric trait with the highest mean average score was Context and Purpose of Writing at 2.74, followed by Syntax and Mechanics at 2.50 and Content Development at 2.46 (see Table 4). Both the median and mode for the average Context and Purpose scores were 3.00 which is a little higher than the mean, and the standard deviation was 0.670- meaning the Context and Purpose scores were just barely negatively skewed but clustered around the mean. The median for Syntax and Mechanics at 2.50 was equal to the mean and lower than the mode at 3.00; thus, the Syntax and Conventions scores with a standard deviation of 0.687 were fairly normally distributed around the mean. The standard deviation for the Content Development scores was 0.758, and the median at 2.50 was slightly higher than the mean but the mode at 2.00 was slightly lower than the mean- indicating that the scores were just barely negatively skewed with a moderate spread surrounding the mean.

The Sources and Evidence trait rendered the next highest average score mean at 2.34, followed by written communication trait with the lowest mean, Disciplinary Conventions at 2.32 (see Table 4). The median for the Sources and Evidence average scores at 2.42 was slightly higher than the mean and the mode at 2.00 was slightly lower than the mean- showing a faint positive skew in their distribution. The Sources and Evidence scores are more spread out surrounding the mean with a standard deviation of 0.774. The median for the Disciplinary Conventions trait average scores at 2.33 was practically equal to the mean and the mode at 2.00 was only barely lower than the mean, indicating an almost normal distribution of scores. The Disciplinary Conventions trait scores revealed a moderate standard deviation of 0.756- suggesting the scores are fairly spread out around the mean.

The Sources and Evidence trait of written communication revealed average score values that ranged from a minimum of 0.00 (Unobserved) to a maximum of 4.00 (Advanced). The Content Development, Disciplinary Conventions, and Syntax and Mechanics traits showed average score ranges from .50 (below the Benchmark performance level) to 4.00. The Context and Purpose trait revealed a range of average scores from 1.00 (Benchmark) to 4.00.

Percent Scoring at Each Performance Level for Overall Total Average and Trait Averages: After averaging the 2 or 3 rater scores for the overall total and rounding those scores to whole numbers that correspond to the performance levels, 6% of the 200 student papers attained a rounded overall total average score at the Capstone performance level (see Figure 3/Figure 5). Almost 47% of the student papers achieved overall total average scores at the Milestone B level, 43% at Milestone A, 6% at Benchmark, and 0% of the overall total average scores for the Written Communication Assessment scored at the Unobserved performance level.

Figure 3. Percent of Student Papers by Overall Total Average Score and Performance Level (N=200)

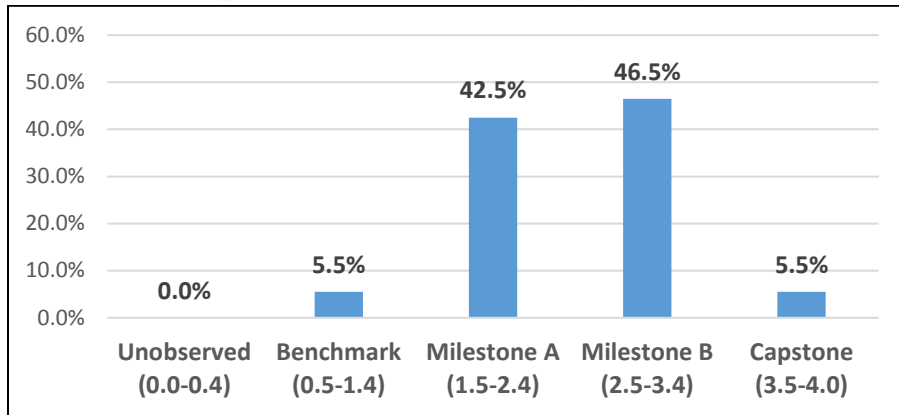
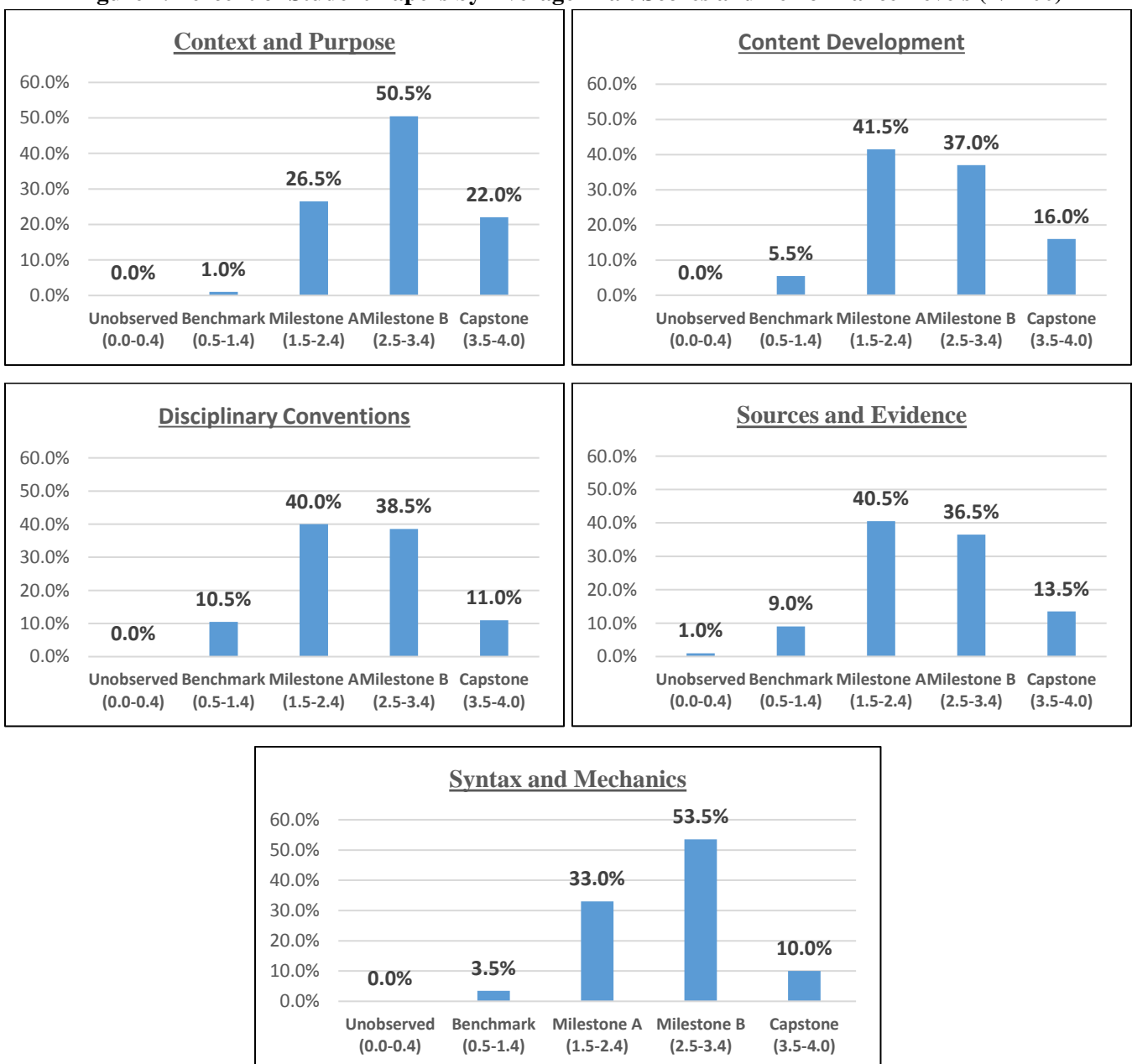


Figure 4. Percent of Student Papers by Average Trait Scores and Performance Levels (N=200)

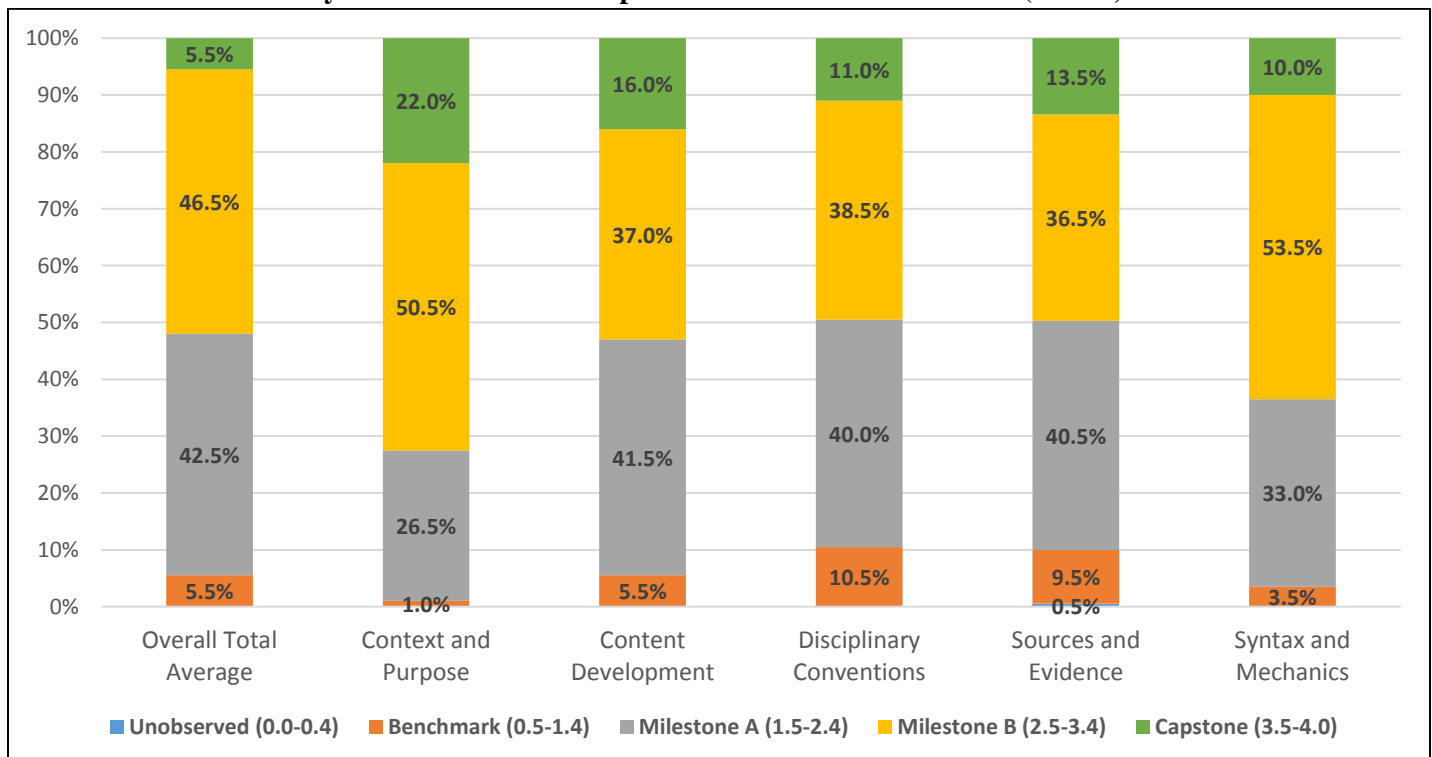


Twenty-two percent of the written communication assessments achieved an average Context and Purpose trait score at the Capstone performance level, 51% scored at the Milestone B level, 27% at the Milestone A level, 1% scored at the Benchmark level, and 0% scored at the Unobserved performance level for Context and Purpose (see Figure 4/Figure 5). Sixteen percent of the 200 papers achieved an average Content Development trait score at the Capstone performance level, 37% scored at the Milestone B level, 42% at the Milestone A level, 6% scored at the Benchmark level, and 0% scored at the Unobserved performance level for the Content Development trait.

Eleven percent of the student papers achieved an average Disciplinary Conventions trait score at the Capstone performance level, 39% scored at the Milestone B level, 40% at the Milestone A level, 11% scored at the Benchmark level, and 0% scored at the Unobserved performance level for Disciplinary Conventions (see Figure 4/Figure 5). Fourteen percent of the written communication assessments achieved an average Sources and Evidence trait score at the Capstone performance level, 37% scored at the Milestone B level, 41% at the Milestone A level, 10% scored at the Benchmark level, and 0% scored at the Unobserved performance level for the Sources and Evidence trait.

Ten percent of the written communication assessments achieved an average Syntax and Mechanics trait score at the Capstone performance level, 54% scored at the Milestone B level, 33% at the Milestone A level, 4% scored at the Benchmark level, and 0% scored at the Unobserved performance level for the Syntax and Mechanics trait (see Figure 4/Figure 5). The written communication trait that seems to render the best performance by students is Context and Purpose, followed by Syntax and Mechanics, Content Development. The written communication traits with the lowest student performance are Disciplinary Conventions and Sources and Evidence.

Figure 5. Comparison of Average Trait and Overall Total Scores by Percent of Student Papers at each Performance Level (N=200)



Changes from Pilot 2014-2015 to 2015-2016 Written Communication Results

Introduction: Considering that the Written Communications assessment during 2014-2015 was a pilot project- the procedures were new for faculty, and it was the first time that the training for raters scoring using the AAC&U rubric was given- much had been learned about the process. The pilot included 87 student papers collected and scored by at least one rater and only 33 had two individual rater scores, but this increased to 199 presentations for 2015-2016 and all papers were scored by two raters. The score differences between individual raters decreased and student scores for the average overall total and each written communication trait increased.

Differences between Individual Raters: Approximately 18% of the overall average scores from the two raters did not differ in 2015-2016, compared to 0% in 2014-2015 (see Table 5). For 2015-2016 over 82% differed less than 1 point and almost 9% differed by 1 full point, compared to 48% differing less than 1 point and 15% differing by 1 point in 2014-2015. In 2015-2016, 90% of the presentations did not have overall average scores that were disparate enough to require a 3rd reader, and 10% had a difference of 1.20 to 2.20 points and an additional rater to score the presentations was necessary. Of the 33 papers with two raters in 2014-2015, 63% of the presentations did not have overall average scores more than 1 point disparate, and 36% had a difference of 1.25 to 2.00 points.

Table 5. Percent of Overall Total Average Differences between Raters by Difference Scores

Score Difference	2014-2015 (N=33)			2015-2016 (N=199)		
	Frequency	Percent	Cum. Percent	Frequency	Percent	Cum. Percent
0.00	0	0.00%	0.00%	36	18.10%	18.10%
0.20				63	31.70%	49.80%
0.25	5	15.15%	15.15%			
0.38	1	3.03%	18.18%			
0.40				28	14.10%	63.90%
0.50	6	18.18%	36.36%			
0.60				21	10.60%	74.50%
0.75	4	12.12%	48.48%			
0.80				14	7.00%	81.50%
1.00	5	15.15%	63.64%	17	8.50%	90.00%
1.20				10	5.00%	95.00%
1.25	8	24.24%	87.88%			
1.40				3	1.50%	96.50%
1.50	2	6.06%	93.94%			
1.60				4	2.00%	98.50%
1.75	1	3.03%	96.97%			
1.80				1	0.50%	99.00%
2.00	1	3.03%	100.00%	1	0.50%	99.50%
2.20				1	0.50%	100.00%
Total	33	100.00%	36.36%	199	100.00%	100.00%

The mean difference between raters for the overall total average scores in 2014-2015 was 0.89 with a standard deviation of 0.472, and differences ranged from 0.25 to 2.00 (see Table 6). The mean difference between raters for the overall average scores for 2015-2016 decreased by .042 to 0.47, the standard deviation decreased by 0.026 to 0.446, and for the minimum score difference decreased from 0.25 to 0.00, but the maximum difference increased by 0.20 to 2.20.

Table 6. 2014-2015 and 2015-2016 Descriptive Statistics for Rater Score Differences

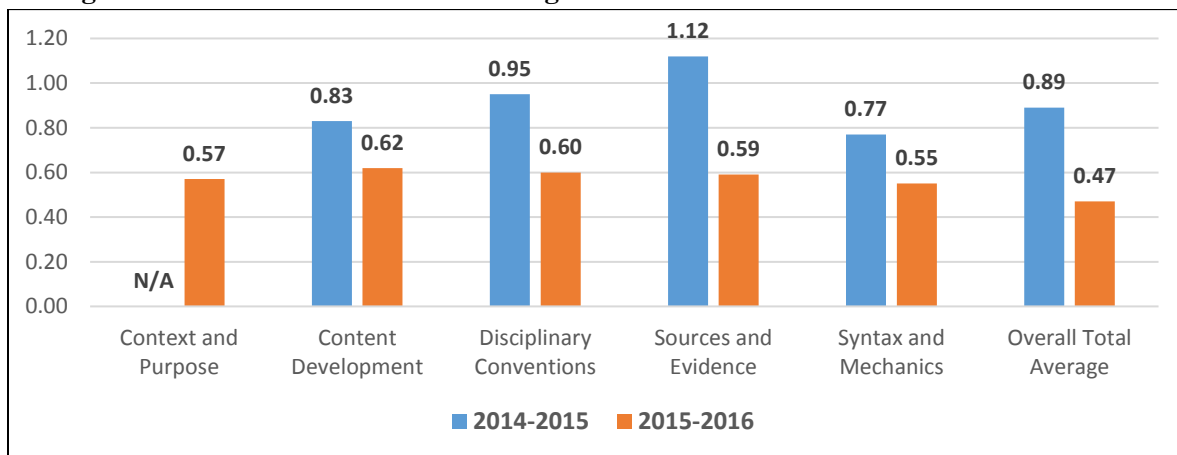
Statistic	Context and Purpose		Content Development		Disciplinary Conventions		Sources and Evidence		Syntax and Mechanics		Overall Total Average	
	14-15	15-16	14-15	15-16	14-15	15-16	14-15	15-16	14-15	15-16	14-15	15-16
Mean	N/A	0.57	0.83	0.62	0.95	0.60	1.12	0.59	0.77	0.55	0.89	0.47
Std. Dev.	N/A	0.639	1.000	0.699	1.000	0.666	1.000	0.595	1.000	0.556	0.472	0.446
Mode	N/A	0.00	0.61	0.00	0.69	0.00	0.83	1.00	0.72	1.00	1.25	0.00
Minimum	N/A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00
Maximum	N/A	3.00	2.00	3.00	2.00	3.00	3.00	2.00	2.00	2.00	2.00	2.20

2014-2015 N=33 / 2015-2016 N=199

For the written communication rubric traits scored in 2014-2015 (which did not include Context and Purpose) the differences between raters ranged from 0 to 2 points for Content Development, Disciplinary Conventions, and Syntax and Mechanics; differences ranged from 0 to 3 points for Sources and Evidence. For 2015-2016 the differences between raters ranged from 0 to 2 points for Sources and Evidence and Syntax and Mechanics, and ranged from 0 to 3 points for Context and Purpose, Content Development, and Disciplinary Conventions (see Table 6). The range of rater score differences between years was consistent for the Syntax and Mechanics trait; increased for Content Development and Disciplinary Conventions, but decreased for the Sources and Evidence trait.

For the pilot assessment in 2014-2015, the mean differences between raters' scores for the AAC&U written communication rubric traits ranged from 0.77-1.12, but decreased in 2015-2016 for a range of 0.55-0.62 (see Table 6 / Figure 5). In 2014-2015 the Sources and Evidence trait had the highest mean difference at 1.12 but decreased to 0.59 in 2015-2016 (see Table 6/Figure 5). The Disciplinary Conventions trait rater score mean difference for the pilot was 0.95 but decreased the next year to 0.60, the Content Development mean difference decreased from 0.83 to 0.62, the Syntax and Mechanics trait decreased from 0.77 to 0.55, and the mean score difference between raters for the Context and Purpose trait in 2015-2016 was 0.57. The mean difference between rater scores for all written communication traits decreased from the pilot in 2014-2015 to the assessment administered in 2015-2016, as did the standard deviations of those means- showing that individual raters' scores became closer together, and the validity and reliability of the scores given by the individual raters have improved.

Figure 5. 2014-2015 to 2015-2016 Changes in Mean Differences between Raters' Scores



Frequencies and percentages of the difference scores for each trait on the Written Communication rubric were computed; and the 0- and 1-point score differences, as well as the 2- and 3-point score differences, were combined to provide frequencies and percentages for the closer and more widely disparate scores by trait (see Table 7). The scores for the Content Development trait maintained the same percentage of 0-1 point differences (88%) and 2-3 point differences (12%) between raters from 2014-2015 to 2015-2016, whilst the rater scores for the traits of Disciplinary Conventions, Sources and Evidence, Syntax and Mechanics, and the Overall Average Total scores became closer or more consistent.

Table 7. Percent of Differences between Raters by Trait

Difference between Raters (Rounded)	Context and Purpose		Content Development		Disciplinary Conventions		Sources and Evidence		Syntax and Mechanics		Overall Total Average	
	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016
0	N/A	51%	24%	50%	24%	50%	21%	46%	39%	48%	18%	64%
1	N/A	42%	64%	39%	55%	41%	48%	48%	42%	49%	70%	33%
0 or 1	N/A	93%	88%	88%	78%	91%	70%	94%	82%	97%	88%	97%
2	N/A	7%	12%	11%	21%	9%	24%	6%	18%	3%	12%	3%
3	N/A	1%	0%	1%	0%	1%	6%	0%	0%	0%	0%	0%
2 or 3	N/A	7%	12%	12%	21%	9%	30%	6%	18%	3%	12%	3%

From 2014-2015 to 2015-2016, the percent of individual raters' scores differing by 0-1 points increased substantially from 70% to 94% for the Sources and Evidence trait of Written Communication (see Table 7), and a considerable increase from 82% to 97% was revealed for the Syntax and Mechanics trait. The consistency between raters for the Disciplinary Conventions trait, as defined by a small score difference of 0-1 points, increased moderately from 78% to 91%, as did the rater consistency for the Overall Total Average scores increasing from 88% in 2014-2015 to 97% in 2015-2016.

Descriptive Statistics for Overall Total Average and Trait Scores: For the pilot in 2014-2015, the overall total average scores for the 87 written communication presentations ranged from 0.50 to 4.00 with a mean of 2.52 and standard deviation of 0.640. In 2015-2016, the overall total average scores for the 200 written communication presentations ranged from 0.80 to 4.00 with a mean of 2.47 and a standard deviation of 0.638 (see Table 8). The mean overall total average score decreased slightly from the pilot to 2015-2016, but the standard deviation of scores from the mean also decreased slightly. The range of the overall total average scores decreased as the minimum overall total average score increased slightly.

Table 8. Descriptive Statistics Changes for Written Communication Average Scores

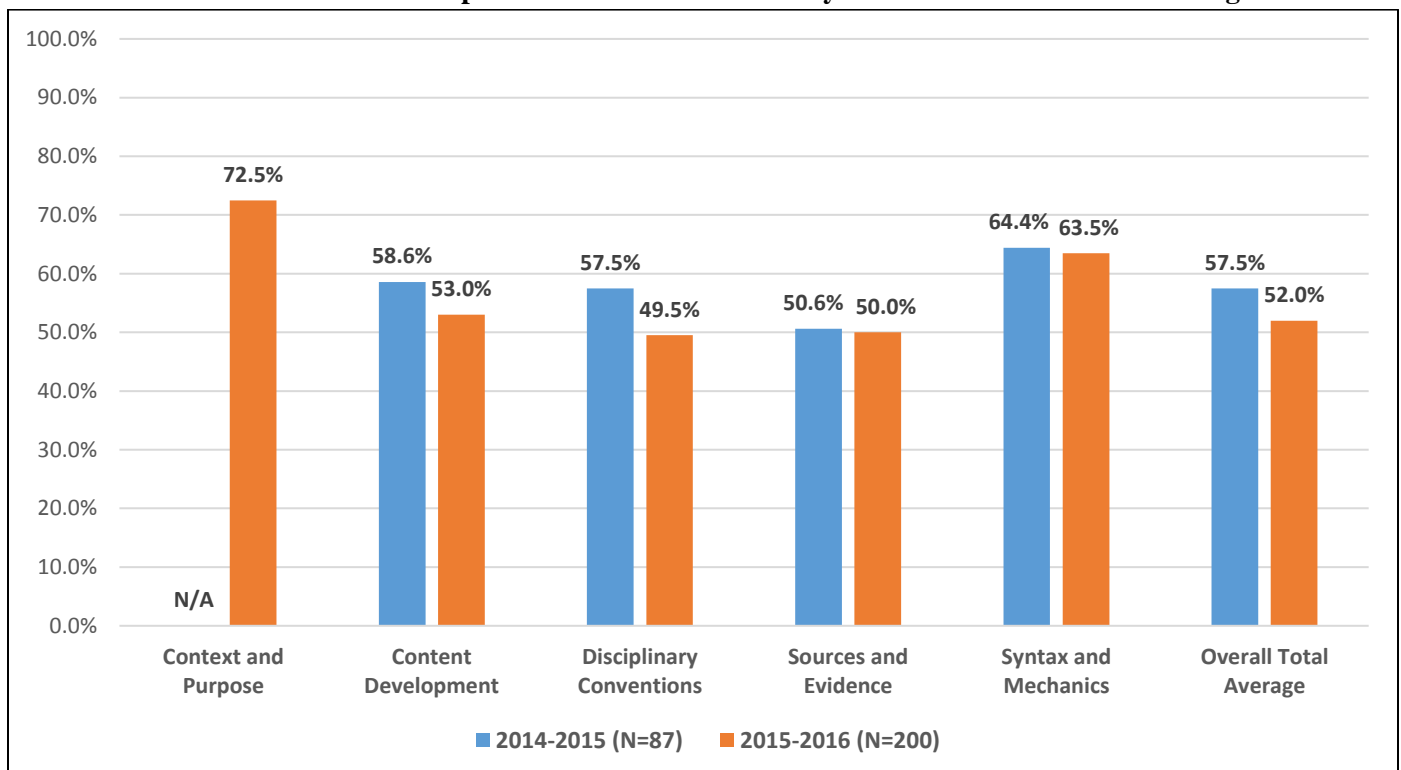
Statistic	Context and Purpose		Content Development		Disciplinary Conventions		Sources and Evidence		Syntax and Mechanics		Overall Total Average	
	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016	2014-2015	2015-2016
2014-2015 N=87 2015-2016 N=200												
Mean	N/A	2.74	2.61	2.46	2.54	2.32	2.37	2.34	2.58	2.50	2.52	2.47
Median	N/A	3.00	2.50	2.50	2.50	2.33	2.50	2.42	2.50	2.50	2.50	2.50
Mode	N/A	3.00	2.00	2.00	3.00	2.00	2.00	2.00	3.00	3.00	2.50	1.80, 2.20, 2.30, 2.70
Std. Dev.	N/A	0.670	0.846	0.758	0.705	0.756	0.838	0.774	0.652	0.687	0.640	0.638
Min.	N/A	1.00	0.00	0.50	1.00	0.50	0.00	0.00	1.00	0.50	0.50	0.80
Max.	N/A	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00

The descriptive statistics for the five AAC&U written communication rubric traits assessed revealed that the Context and Purpose trait attained the highest mean score for 2015-2016 at 2.74, but was not part of the rubric used for the assessment during the pilot in 2014-2015. The Syntax and Mechanics trait obtained the second highest mean in 2015-2016, but it decreased slightly by 0.08 from 2.58 to 2.50 as the standard deviation increased slightly by 0.035 from 0.652 to 0.687. The Content Development trait achieved the third highest mean in 2015-2016 at 2.46, but this was a decrease of 0.15 from 2.61 in 2014-2015 (see Table 8). The standard deviation for Content Development decreased by 0.088 from 0.846 in the pilot to 0.758.

The Sources and Evidence written communication trait attained a mean of 2.34 in 2015-2016, which was a slight decrease of 0.03 from the mean of 2.37 in 2014-2015 (see Table 8). However, the standard deviation for the Sources and Evidence trait also decreased by 0.064 from 0.838 during the pilot to 0.774 in 2015-2016. The Disciplinary Conventions trait rendered the lowest mean in 2015-2016 at 2.32, and this was a decrease of 0.22 from a mean of 2.54 in 2014-2015. The standard deviation for the Disciplinary Conventions trait increased slightly from 0.705 during the pilot to 0.756 in 2015-2016.

Percent Scoring at Milestone B or Capstone Performance Levels for Overall Total Average and Traits: In 2014-2015, 57.5% of the 87 student papers assessed using the AAC&U Written Communication rubric achieved a rounded overall total average score of 3.0 or above, and were classified as reaching the Milestone B or Capstone performance levels (see Figure 6). The percent achieving a rounded overall total average score of 3.0 or higher in 2015-2016 decreased by 5.5 percentage points, and 52% of the 200 student papers were categorized at the Milestone B or Capstone performance levels.

Figure 6. Comparison of the Percent of Student Papers Scoring at the Milestone B or Capstone Performance Levels by Trait and Overall Total Average

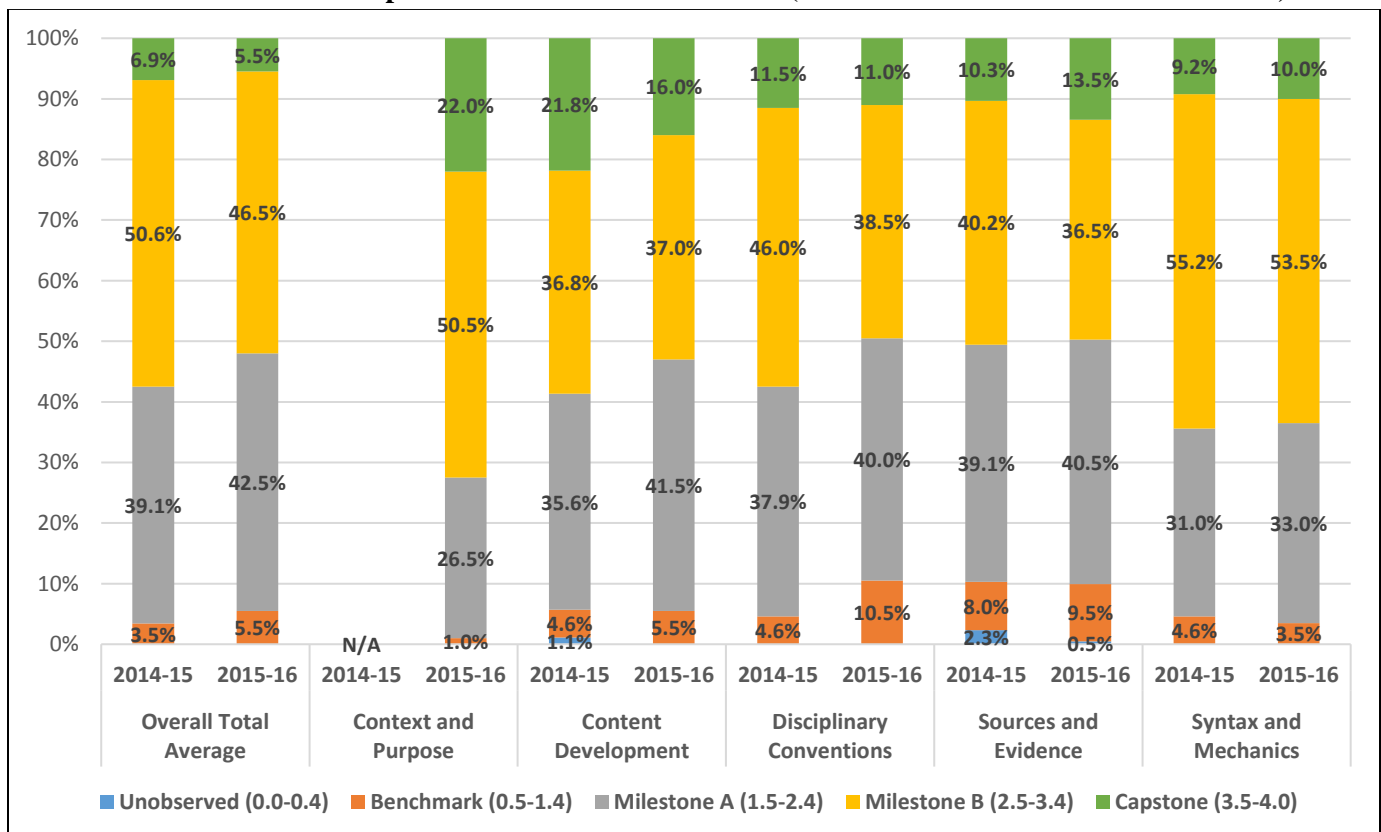


The written communication traits that rendered the highest percent of scores classified at the Milestone B or Capstone performance levels in 2015-2016 were Context and Purpose, Syntax and Mechanics, and Content Development (see Figure 6). For the Context and Purpose trait 72.5% of students papers scored at the Milestone B or Capstone performance levels, but Context and Purpose was not assessed using the pilot rubric in 2014-2015. The Syntax and Mechanics trait rendered 63.5% of scores classified at the Milestone B or Capstone performance levels in 2015-2016, which was a decrease of almost 1 percentage point from 64.4% of the Syntax and Mechanics trait scores at 3.0 or higher in 2014-2015. Fifty-three percent of the Content Development trait scores in 2015-2016 were at the Milestone B or Capstone levels with rounded scores of 3.0 or higher, which was a 5.6 percentage point decrease from 2014-2015 when 58.6% of the Content Development trait scores were classified as Milestone B or Capstone.

The Sources and Evidence trait attained 50.0% of scores at the Milestone B or Capstone performance levels in 2015-2016, and this was approximately the same percentage of assessments (50.6%) that had attained rounded Sources and Evidence trait scores of 3.0 or higher for the pilot in 2014-2015 (see Figure 6). The Disciplinary Conventions trait rendered 49.5% of scores categorized as Milestone B or Capstone in 2015-2016, which was a decrease of 8 percentage points from 57.5% of Disciplinary Conventions trait scores reaching Milestone B or Capstone during the pilot.

The 2014-2015 and 2015-2016 percentages of student written communication assessments scoring at the Unobserved, Benchmark, Milestone A, Milestone B, and Capstone performance levels for the overall total average and the five separate written communication trait scores are shown in Figure 7.

Figure 7. 2014-15 and 2015-16 Comparison of Trait and Overall Total Average Scores (Rounded) by Percent of Student Papers at each Performance Level (2014-2015 N=87 and 2015-2016 N=200)



Conclusion: In 2015-2016, 199 student papers were scored by two separate raters using a revised version of the AAC&U Written Communication Rubric. Five separate traits (Context and Purpose, Content Development, Disciplinary Conventions, Sources and Evidence, and Syntax and Mechanics) were rated on a five-point scale of 0=Unobserved, 1=Benchmark, 2=Milestone A, 3=Milestone B, and 4=Capstone. The trait scores were averaged to calculate an overall total average score for each student paper. Eighteen percent of the overall average scores from the two raters did not differ, 63% differed less than 1 point, and 9% differed by 1 full point. A total of 10% of the 199 overall total average rater scores had a difference of 1.20 to 2.20 points, and an additional rater was required to score the student paper. The individual raters were very consistent in their scores for the Syntax and Mechanics trait with 97% of the ratings having score differences equal to 0 or 1; followed by Sources and Evidence with 94.5%, Context and Purpose with 93%, and Disciplinary Conventions with 91% of raters scores having differences of 0 or 1. Content Development seemed to be the written communication trait in which raters were least consistent with 88.4% of the scores differing by 0 or 1.

Fifty-two percent of the 200 written communication assessments achieved an overall total average score that after rounding would be classified at the Milestone B or Capstone performance levels. Almost 6% of the student papers attained an overall total average score at the Capstone performance level, 46.5% scored at the Milestone B level, 42.5% at Milestone A, 5.5% at Benchmark, and 0% scored at the Unobserved performance level. Approximately 73% percent of student papers attained average scores classified as Milestone B or Capstone performance levels for the Context and Purpose trait, 63.5% for Syntax and Mechanics, 53% for Content Development, 50% for Sources and Evidence, and 49.5% for the Disciplinary Conventions trait. The Context and Purpose trait had the highest mean average score at 2.74, followed by Syntax and Mechanics at 2.50 and Content Development at 2.21. The Sources and Evidence trait rendered a mean score of 2.34, followed by Disciplinary Conventions at 2.32. The Disciplinary Conventions and Sources and Evidence traits seem to be where Washburn students performed the lowest; and students performed best on the Context and Purpose trait compared to the other written communication traits assessed.

The changes in rater differences and student scores between the pilot in 2014-2015 and the assessment administered in 2015-2016 were examined. The number of student papers scored by two independent raters increased from 33 to 199, the percentage of presentations in which the overall average scores from the two raters did not differ increased from 0% in 2014-2015 to over 18% in 2015-2016, and the percent of presentations in which the overall average scores from the two raters differed more than one point decreased from 36% to 10%. The mean difference between raters for the overall average scores in 2014-2015 was 0.89 with a standard deviation of 0.472, and in 2015-2016 the mean difference decreased to 0.47 with the standard deviation decreasing to 0.446. The mean differences between raters' scores for the written communication traits decreased from the pilot in 2014-2015 to the 2015-2016 assessment, as did the standard deviations of those means- showing that individual raters' scores became closer together, and the validity and reliability of the scores given by the individual raters have improved.

Almost 58% of the 87 total student papers in 2014-2015 attained a rounded overall average score classified at the Milestone B or Capstone performance levels, but this decreased by almost 6 percentage points in 2015-2016, and 52% of the 200 overall average scores were categorized at the Milestone B or

Capstone performance levels. The percent of rounded scores classified at the Milestone B or Capstone performance levels decreased from 2014-2015 to 2015-2016 for all four of the previously assessed written communication traits. The Syntax and Mechanics trait went from 64.4% of the average scores being classified at the Milestone B or Capstone performance levels in 2014-2015 to 63.5%, Content Development decreased from 58.6% to 53%, Disciplinary Conventions from 57.5% to 49.5%, and Sources and Evidence from 50.6% to 50%. The Context and Purpose written communication trait was not assessed during the 2014-2015 pilot, but 72.5% of average trait scores were categorized at the Milestone B or Capstone performance levels.

The number of student papers scored for the Written Communication assessment increased radically from 87 in 2014-2015 to 200 in 2015-2016. In research the mere increase in the number of students assessed would be expected to lead to an increase in the amount of variation among the students and their assessments. The increased variation would usually manifest itself in decreased mean scores and increased standard deviations of those mean scores. The overall total average mean score and the written communication trait mean scores did decrease from the pilot in 2014-2015 to 2015-2016, but this decrease was very small (tenths of a point), but the standard deviations of the overall total average score mean and the trait means decreased instead of increased, except for the Syntax and Mechanics trait which increased slightly.

In terms of what the assessment data suggests for written communication instruction in the future, perhaps more concentration should be placed upon the traits of Disciplinary Conventions, Sources and Evidence, and Content Development. The Disciplinary Conventions trait rendered only 49.5% of student papers reaching the Milestone B or Capstone performance levels, Sources and Evidence reached 50%, and the Content Development trait rendered 53% of the student papers scoring in the Milestone B or Capstone performance levels. Improving these individual written communication traits will improve the overall total average scores for the written communication assessment.

The training held for the individual raters assessing the written communication assessments seemed to be more effective in 2015-2016 than for the pilot assessment in 2014-2015. Ninety percent of the overall total average scores for student papers provided by pairs of individual raters were not disparate enough to warrant an additional rater, but 10% (20 papers) did have overall total average scores that differed more than 1 point between rater pairs and an additional rater was necessary. The interrater reliability for three of the five written communication traits assessed by the revised AAC&U rubric was 93% or higher. However, the Content Development trait showed 88% of rater pairs with 0-1 point difference, but almost 12% with differences between raters of 2-3 points; and this could be improved by stressing the scoring of that trait and providing extra examples during the rater training. Disciplinary Conventions is another trait that might benefit from extra attention during the rater training; 91% of rater pairs showed a 0-1 point difference, but 9% of rater pairs exhibited larger 2-3 point differences between their scores of the Disciplinary Conventions trait.