



Written Communication Artifact Assessment Report 2018-2019

Executive Summary

- In AY 2019, 200 previously written communication artifacts were scored by two separate raters using a revised version of the AAC&U Written Communication VALUE rubric. The rubric consisted of:
 - Four criterion: Content Development, Genre conventions, Sources and Evidence, and Control of Syntax and Mechanics
 - Rated on a four-point scale: Capstone (4), Milestones (3 and 2), and Benchmark (1).
- 3rd ratings were generated if the average score of the 1st rater's scoring and the 2nd rater's scoring differed by 1.00 or greater.
 - 53 (26.5%) artifacts required a 3rd rating by a faculty member who did not initially rate the artifacts.
 - The 3rd rating scores were compared to the first two raters' scores, and score the most distant from the 3rd rater's score was removed from further analysis.
- Overall, almost half of the artifacts were scored at Milestone (3) (57.0%), followed by Milestone (2) at 22.5% of the average ratings. Capstone (4) ratings were 20.0% of written artifacts, and Benchmark ratings were representative of 0.5% of written communication artifacts.
- For overall normalized ratings by criteria, 2.47 to 2.58 was the mean score range, with an overall mean score of 2.54.
 - The criterion with the highest mean score was Genre Conventions with 2.58; 48.5% of written communication artifacts in this criteria fell at the Milestone (3) rating level.
 - Content Development, and Control of Syntax and Mechanics, yielded the second and third highest mean scores (\bar{x} s = 2.56 and 2.55, respectively).
 - Sources of Evidence received the lowest mean score (\bar{x} = 2.47).
- Almost all criteria were rated higher in 2018-19 from 2014-15 and 2015-16. For the 2018-2019 analysis, the highest mean score was Genre Conventions with 2.58, whereas Genre and Disciplinary Conventions was the lowest mean score at 2.32 in 2015-2016 and in the middle range at 2.54 in 2014-2015.

Introduction

Written Communication was assessed during the 2018-19 academic year (AY) using a modified version of the Association of American Colleges and Universities (AAC&U) Written Communication Valid Assessment of Learning in Undergraduate Education (VALUE) rubric (see Appendix A). The AAC&U VALUE rubrics were developed by teams of educational professionals, and include the most frequently identified criteria of learning for different learning outcomes. Washburn University (Washburn) implements performance assessments using modified versions of the AAC&U VALUE rubrics for assessing Written Communication every three years. Artifacts in written format are collected from students in EN 300: Advanced College Writing, and are scored by two or more independent raters using the Aqua by Watermark software platform.

Written Communication Artifact Review Process

Washburn faculty were invited to attend the in-person norming training on May 16, 2019. The 20 faculty who attended the training were assigned 200 previously written papers collected from Fall 2016 to Fall 2018 EN 300: Advanced College Writing courses. The 200 artifacts were reviewed by two independent raters, 20 artifacts each, on four criterion: Content Development, Genre conventions, Sources and Evidence, and Control of Syntax and Mechanics. These four criterion were scored on a four-point scale that consisted of Capstone (4), Milestones (3 and 2), and Benchmark (1). Of the 20 raters who participated in the artifact review process, 75% scored 20 artifacts, 10.0% scored less than 20 artifacts, and 15.0% scored more than 20 artifacts.

When the average difference in scores was equal to or greater than 1.00, a 3rd rater was utilized to normalize the ratings. After the initial round of scoring, the majority (73.5%) of average ratings from the 1st reviewer and the 2nd reviewer did not differ by 1.00 or more, and did not require a 3rd rating. However, there were 53 artifacts (26.5%) that met or exceeded 1.00 and required a 3rd rating.

Written Communication Artifact Review Results

Overall Normalized Ratings

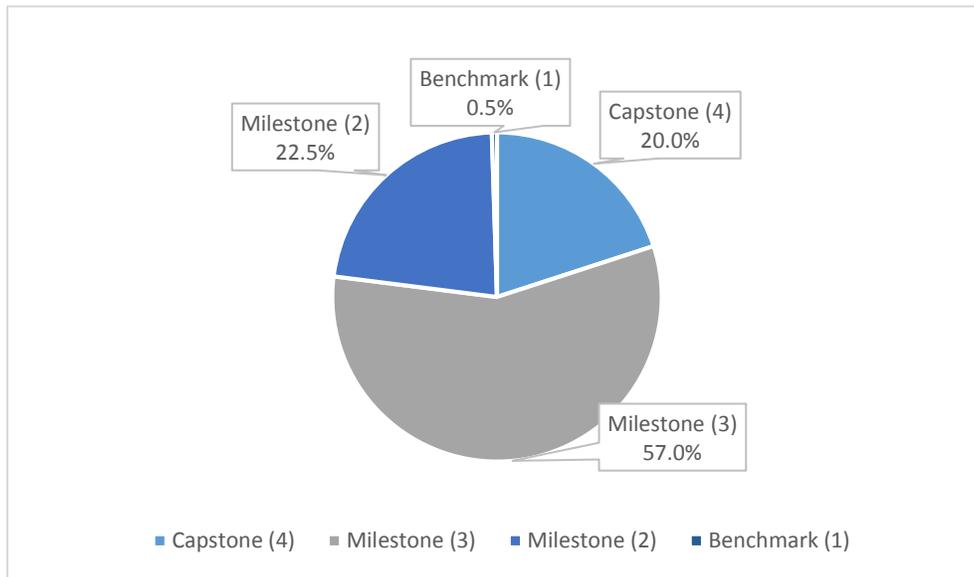
The scores from the two raters were averaged to provide a normalized score for each written communication artifact. Table 1 provides the overall distribution of scores for the 200 written communication artifacts. The range of average scores were categorized as 4.00-3.01 – Capstone, 3.00-2.01 and 2.00-1.01 Milestones, and 1.00-0.01 Benchmark.

Table 1. Descriptive Data and Statistics for Overall Averaged Ratings

	Capstone	Milestones		Benchmark	Mean (<i>sd</i>)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	
Overall (n = 200)	40 (20.0%)	114 (57.0%)	45 (22.5%)	1 (0.5%)	2.53 (0.71)

The majority of artifacts (57.0%) were scored at Milestone (3) on average. The second and third highest percentages of ratings were for Milestone (2) (22.5%) and Capstone (20.0%). Only one artifact was rated as Benchmark on average (0.5%). Figure 1, on the following page, provides a visual representation of the distribution of scores.

Figure 1. Distribution of Scores for Overall Written Communication Artifacts



Overall Normalized Ratings by Criteria

The distribution of average ratings and descriptive statistics for each of the four criteria are in Table 1, below. The table contains the how the artifacts ($n = 200$) were rated on each of the four criteria (total of 800 scores). For example, the row labeled as Content Development contains the scores from all 200 artifacts on that criteria.

Table 2. Descriptive Data and Statistics for Overall Averaged Ratings by Criteria

	Capstone	Milestones		Benchmark	Mean (<i>sd</i>)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	
Content Development	35 (17.5%)	96 (48.0%)	62 (31.0%)	7 (3.5%)	2.56 (0.74)
Genre Conventions	33 (16.5%)	97 (48.5%)	65 (32.5%)	5 (2.5%)	2.58 (0.68)
Sources and Evidence	34 (17.0%)	84 (42.0%)	75 (37.5%)	7 (3.5%)	2.47 (0.70)
Control of Syntax and Mechanics	30 (15.0%)	101 (50.5%)	58 (29.0%)	11 (5.5%)	2.55 (0.72)
Overall ($n = 800$)	132 (16.5%)	378 (47.3%)	260 (32.5%)	30 (3.7%)	2.54 (0.71)

All criterion fell within the 2.47 to 2.58 range, with the overall mean score of 2.54, which indicated that the 200 written artifacts' scores were Milestone (3), on average. The criterion with the highest mean score was Genre Conventions with 2.58; 48.5% of written communication artifacts were scored at the Milestone (3) level and this criteria had the highest percent of artifacts scored in the Capstone (4) level (17.5%). Content Development, and Control of Syntax and Mechanics, yielded the second and third highest mean scores (\bar{x} s = 2.56 and 2.55, respectively); 48.0% and 50.5% of written communication

artifacts were rated Milestone (3) in these areas, respectively. Sources of Evidence received the lowest mean score ($\bar{x} = 2.47$); however, this mean score also indicates that the artifacts were scored at the Milestone (3) level on average. Almost half, 42.0%, of the written communication artifacts for Sources of Evidence were scored with a Milestone (3) rating, yet this criteria had the highest percent of Benchmark (1) scores (37.5%) which likely contributed to lowering the mean score.

See Figures 1-4 below for visual representations of the distributions of ratings by each criteria. All four criteria had distributions of ratings in a normal distribution.

Figure 1. Distribution of Scores for Content Development

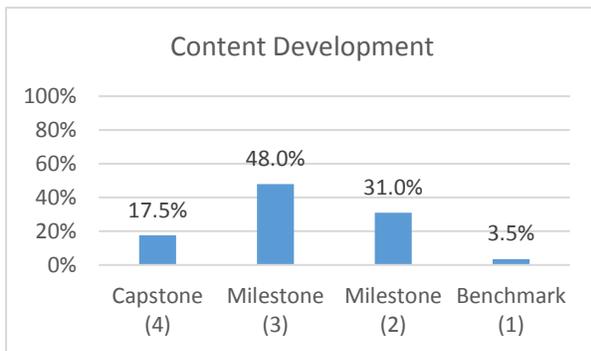


Figure 3. Distribution of Scores for Sources and Evidence

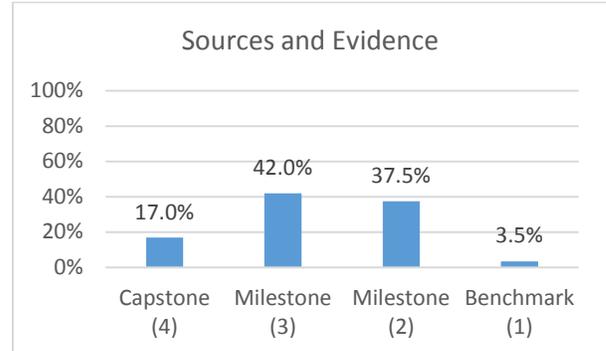


Figure 2. Distribution of Scores for Genre Conventions

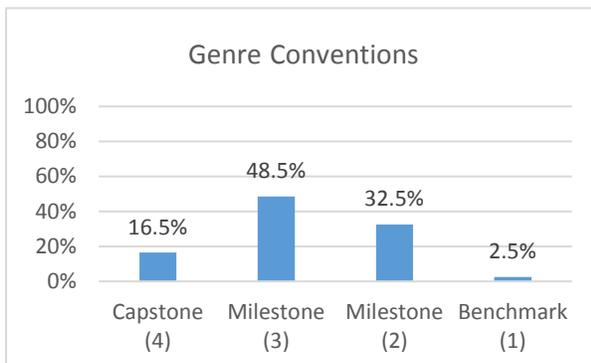
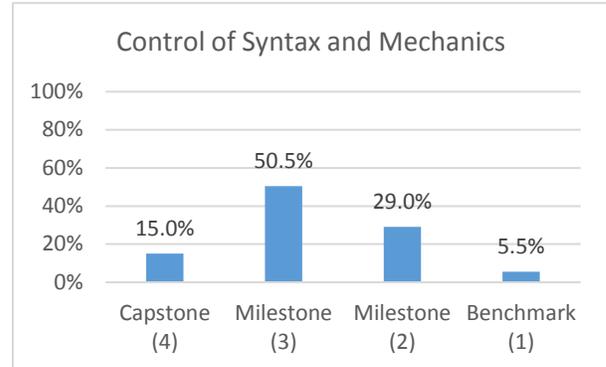


Figure 4. Distribution of Scores for Control of Syntax and Mechanics



Comparison to Previous Years Results

Written communication artifacts were reviewed for the pilot round in 2014-15 ($n = 87$) and in 2015-16 ($n = 199$); these artifacts were rated using a different AAC&U Written Communication VALUE Rubric that contained five traits (Context and Purpose of Writing, Content Development, Genre and Disciplinary Conventions, Sources and Evidence, and Control of Syntax and Mechanics). However, the four criteria of Content Development, Genre and Disciplinary Conventions, Sources and Evidence, and Control of Syntax and Mechanics are similar enough that comparisons can be made. In 2014-15, 58% of the written communication artifacts attained an overall average score Milestone (3) or Capstone (4) levels; in 2015-16, 52% of the average scores were categorized at the Proficient or Advanced performance levels. For

the 2018-19 analysis of scores ($n = 200$), 63.8% of the overall average scores were categorized at the Milestone (3) or Capstone (4) performance levels.

Almost all mean scores were rated higher in 2018-19 from 2014-15 and 2015-16. For the 2018-2019 analysis, the highest mean score was Genre Conventions with 2.58, whereas Genre and Disciplinary Conventions was the lowest mean score at 2.32 in 2015-2016 and in the middle range at 2.54 in 2014-2015. Content Development at 2.56 in 2018-19 was higher than 2014-2015 ($\bar{x} = 2.61$) yet lower in 2015-2016 ($\bar{x} = 2.46$). Control of Syntax and Mechanics retained very similar ratings through the years; this criteria yielded an average rating of 2.55 in 2018-19, whereas in previous years these areas were 2.58 and 2.50. Sources and Evidence was the lowest scored criteria in 2018-19 at 2.47. However, in 2014-15 and 2015-16, this criteria was scored even lower at 2.37 and 2.34, respectively. Context and Purpose was not a scored criteria in 2014-2015 and 2018-19; the 2015-16 mean score was 2.74.

Areas of Consideration and Limitations

All criterion were very similarly rated. Mean scores were within a range of 0.11 (minimum = 2.47, maximum = 2.58), with the overall mean score of 2.54. Sources and Evidence yielded the lowest mean score, which was reflected in the largest amount of Milestone (2) and Milestone (3) ratings (37.5% and 42.0%, respectively). This area could be improved, however, it should be noted that 17.0% Capstone (4) ratings for Sources and Evidence was the second highest percentage of the four criterion.

Written communication artifacts scored during the 2018-19 academic year, were scored in Aqua by Watermark using a modified version of the AAC&U Oral Communication VALUE rubric. The use of the new software and the modified rubric made comparison to previous years difficult to analyze. Caution should be used when examining that section for modifications in processes for the next review cycle.

In addition, when comparing results to previous years, it should be noted that the methodology of determining 3rd ratings and normalized scores was modified in 2018-2019. The cut line for 3rd rater scores was increased from scores with a difference more than 1.00 (≤ 1.01) to include those with differences of 1.00 or greater (≤ 1.00). This may have resulted in more 3rd raters needed for 2018-19. Additionally, when normalizing ratings, the 3rd ratings were compared to the first two raters' scores, and score farthest from the 3rd rater's score was removed from further analysis. In previous years, the 3rd raters' scores had been averaged with the first two raters' scores, resulting in more varied normalized scores for those that required a 3rd rater (i.e., those scores with a 3rd rater were an average of three scores, others were an average of two scores). Consequently, the ranges of the scoring levels of Capstone (4), Milestones (3 and 2), and Benchmark (1) were modified in 2018-19 to accommodate the normalization of the two ratings, not three.

Finally, reasonable efforts were made to collect written communication artifacts from students enrolled in EN 300: Advanced College Writing, a university requirement for Junior level students, from Fall 2016 to Spring 2018. A random sampling was not used to select artifacts for review, however, given that this course is a requirement for all Junior level students, general assumptions could be made about the proficiency level of all students at Washburn in written communication.

Appendix A

WRITTEN COMMUNICATION VALUE RUBRIC

for more information, please contact value@aacu.org



The VALUE rubrics were developed by teams of faculty experts representing colleges and universities across the United States through a process that examined many existing campus rubrics and related documents for each learning outcome and incorporated additional feedback from faculty. The rubrics articulate fundamental criteria for each learning outcome, with performance descriptors demonstrating progressively more sophisticated levels of attainment. The rubrics are intended for institutional-level use in evaluating and discussing student learning, not for grading. The core expectations articulated in all 15 of the VALUE rubrics can and should be translated into the language of individual campuses, disciplines, and even courses. The utility of the VALUE rubrics is to position learning at all undergraduate levels within a basic framework of expectations such that evidence of learning can be shared nationally through a common dialog and understanding of student success.

Definition

Written communication is the development and expression of ideas in writing. Written communication involves learning to work in many genres and styles. It can involve working with many different writing technologies, and mixing texts, data, and images. Written communication abilities develop through iterative experiences across the curriculum.

Framing Language

This writing rubric is designed for use in a wide variety of educational institutions. The most clear finding to emerge from decades of research on writing assessment is that the best writing assessments are locally determined and sensitive to local context and mission. Users of this rubric should, in the end, consider making adaptations and additions that clearly link the language of the rubric to individual campus contexts.

This rubric focuses assessment on how specific written work samples or collections of work respond to specific contexts. The central question guiding the rubric is "How well does writing respond to the needs of audience(s) for the work?" In focusing on this question the rubric does not attend to other aspects of writing that are equally important: issues of writing process, writing strategies, writers' fluency with different modes of textual production or publication, or writer's growing engagement with writing and disciplinarity through the process of writing.

Evaluators using this rubric must have information about the assignments or purposes for writing guiding writers' work. Also recommended is including reflective work samples or collections of work that address such questions as: What decisions did the writer make about audience, purpose, and genre as s/he compiled the work in the portfolio? How are those choices evident in the writing -- in the content, organization and structure, reasoning, evidence, mechanical and surface conventions, and citational systems used in the writing? This will enable evaluators to have a clear sense of how writers understand the assignments and take it into consideration as they evaluate.

The first section of this rubric addresses the context and purpose for writing. A work sample or collections of work can convey the context and purpose for the writing tasks it showcases by including the writing assignments associated with work samples. But writers may also convey the context and purpose for their writing within the texts. It is important for faculty and institutions to include directions for students about how they should represent their writing contexts and purposes.

Faculty interested in the research on writing assessment that has guided our work here can consult the National Council of Teachers of English/Council of Writing Program Administrators' White Paper on Writing Assessment (2008; www.wpacouncil.org/whitepaper) and the Conference on College Composition and Communication's Writing Assessment: A Position Statement (2008; www.ncte.org/cccc/resources/positions/123784.htm)

Glossary

The definitions that follow were developed to clarify terms and concepts used in this rubric only.

- **Content Development:** The ways in which the text explores and represents its topic in relation to its audience and purpose.
- **Evidence:** Source material that is used to extend, in purposeful ways, writers' ideas in a text.
- **Genre conventions:** Formal and informal rules for particular kinds of texts and/or media that guide formatting, organization, and stylistic choices, e.g. lab reports, academic papers, poetry, webpages, or personal essays.
- **Sources:** Texts (written, oral, behavioral, visual, or other) that writers draw on as they work for a variety of purposes -- to extend, argue with, develop, define, or shape their ideas, for example.

WRITTEN COMMUNICATION VALUE RUBRIC

for more information, please contact value@aacu.org
 Revised 2015 for use at Washburn University USLO Assessment



Definition

Written communication is the development and expression of ideas in writing. Written communication involves learning to work in many genres and styles. It can involve working with many different writing technologies, and mixing texts, data, and images. Written communication abilities develop through iterative experiences across the curriculum.

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

	Capstone 4	Milestones		Benchmark 1
		3	2	
Content Development	Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writers understanding, and shaping the whole work.	Uses appropriate, relevant, and compelling content to explore ideas and shape the whole work.	Uses appropriate and relevant content to develop and explore ideas through most of the work.	Uses appropriate and relevant content to develop simple ideas in some parts of the work.
Genre Conventions <i>Formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields (please see glossary).</i>	Demonstrates detailed attention to and successful execution of a wide range of conventions including organization, content, presentation, formatting, and stylistic choices	Demonstrates consistent use of important conventions including organization, content, presentation, and stylistic choices	Includes basic organization, content, and presentation	Attempts to use a consistent system for basic organization and presentation.
Sources and Evidence	Demonstrates skillful use of high-quality, credible, relevant sources to develop ideas	Demonstrates consistent use of credible, relevant sources to support ideas	Demonstrates an attempt to use credible and/or relevant sources to support ideas	Demonstrates an attempt to use sources to support ideas in the writing.
Control of Syntax and Mechanics	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.	Uses straightforward language that generally conveys meaning to readers. The language has few errors.	Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.	Uses language that sometimes impedes meaning because of errors in usage.

Appendix B

3rd Rater Review Methodology

Washburn faculty were invited to attend the in-person norming training on May 16, 2019. The 20 faculty who attended the training were assigned 200 previously written papers collected from Fall 2016 to Fall 2018 EN 300: Advanced College Writing courses. The 200 artifacts were reviewed by two independent raters, 20 artifacts each, on four criterion: Content Development, Genre conventions, Sources and Evidence, and Control of Syntax and Mechanics. These four criterion were scored on a four-point scale that consisted of Capstone (4), Milestones (3 and 2), and Benchmark (1). Of the 20 raters who participated in the artifact review process, 75% scored 20 artifacts, 10.0% scored less than 20 artifacts, and 15.0% scored more than 20 artifacts.

3rd Rater Review Process

The range of differences in ratings was from 0.00 to 2.25. The greatest percentage of average ratings differed by 0.25 (26.5%), followed by 0.50 (20.0%). The distribution of scores were positively skewed (right-skewed) in that the mean (average value) of 0.63 was greater than the median (middle value) of 0.50. The distribution of rating differences and descriptive statistics are in the tables and chart, below.

Table 1B. Rating Differences

Rating Difference	Frequency	Percent
0.00	24	12.0%
0.25	53	26.5%
0.50	40	20.0%
0.75	30	15.0%
1.00	17	8.5%
1.25	15	7.5%
1.50	11	5.5%
1.75	7	3.5%
2.00	2	1.0%
2.25	1	0.5%
Total	200	100%

Figure 1B. Percent of Rating Differences

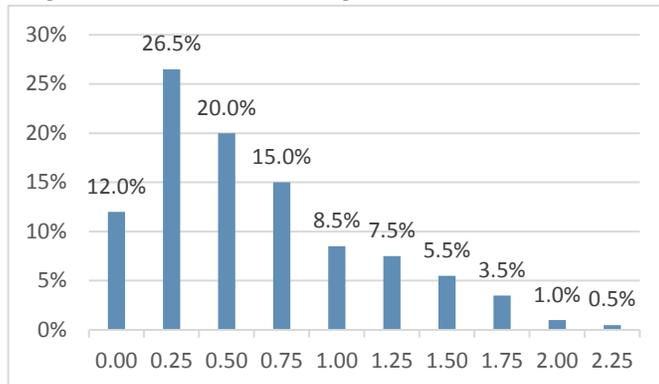


Table 2B. Descriptive Statistics for Rating Differences

Mean	Median	Mode	Standard Deviation
0.63	0.50	0.25	0.50

When the average difference in scores was equal to or greater than 1.00, a 3rd rater was utilized to normalize the ratings. After the initial round of scoring, the majority (73.5%) of average ratings from the 1st reviewer and the 2nd reviewer did not differ by 1.00 or more, and did not require a 3rd rating. However, there were 53 artifacts (26.5%) that met or exceeded 1.00 and required a 3rd rating.

For those requiring a 3rd rating, highlighted in blue in the table on the previous page, there were a total of 53 artifacts. Seventeen (8.5%) artifacts had an average rating difference of 1.00, 15 had rating

differences of 1.25 (7.5%), 11 had rating differences of 1.50 (5.5%), seven had rating differences of 1.75 (3.5%), two had rating differences of 2.00 (1.0%) and one artifact had a 2.25 rating difference (0.5%).

The 53 artifacts that required a 3rd rating were scored by a faculty member who did not initially rate those artifacts. The 3rd rating scores were compared to the first two raters' scores, and score farthest from the 3rd rater's score was removed from further analysis. See Table 3B for means, the minimum difference, and maximum difference between ratings.

Table 3B. Descriptive Statistics for Ratings Selected and Not Selected

	Mean	Standard Deviation	Minimum	Maximum
Rating Selected	0.37	0.23	0.00	1.00
Rating Not Selected	1.17	1.17	0.50	2.75

The mean difference between those ratings that were selected and the 3rd rater's scores was 0.37; the mean difference between those ratings that were not selected and the third rater's scores was 1.17. The minimum difference for those ratings that were selected was 0.00; the maximum difference was 1.00.

Differences by Criteria

The AAC&U Written Communication VALUE rubric defines Written Communication as “the development and expression of ideas in writing. Written communication involves learning to work in many genres and styles. It can involve working with many different writing technologies, and mixing texts, data, and images.” The rubric contained four criterion: Content Development, Genre Conventions, Sources and Evidence, and Control of Syntax and Mechanics. The written artifacts were rated on these four criterion on a four-point scale that consisted of Capstone (4), Milestones (3 and 2), and Benchmark (1).

The 200 written communication artifacts were rated by two different reviewers for a total of 400 scores. These scores were reviewed for each of the four criteria to examine differences in ratings per criteria. Table 4B provides the descriptive statistics for the differences in ratings by criteria.

Table 4B. Descriptive Statistics for Differences in Ratings by Criteria

	Content Development	Genre Conventions	Sources and Evidence	Control of Syntax and Mechanics	Total Difference
Mean	0.53	0.42	0.46	0.54	0.49
Standard Deviation	0.56	0.50	0.53	0.54	0.53
Min	0.00	0.00	0.00	0.00	0.00
Max	2.00	2.00	2.00	2.00	2.00

The mean difference was greatest for the criteria areas of Control of Syntax and Mechanics, and Content Development in that the ratings on average differed by 0.54 and 0.53, respectively. The mean difference was the least for the criteria areas of Sources and Evidence, and Genre Conventions with an average difference in rating of 0.42 and 0.46, respectively. The minimum rating was 0.00 (or no difference) and the maximum was 2.00 across all criteria.