



## *Written Communication Artifact Assessment Report 2024-25*

### Executive Summary

In Academic Year 2025, **89 Written Communication artifacts** were scored by two separate raters using a revised version of the AAC&U Written Communication VALUE rubric. The rubric consists of four criteria:

- Content Development
- Genre Conventions
- Sources and Evidence
- Control of Syntax and Mechanics

Each criterion was rated on a five-point scale: Capstone (4), Milestone (3), Milestone (2), Benchmark (1), and Not Present (0); there was also an option to mark an artifact as unscorable.

In the event the average of the scores across all criteria from the 1<sup>st</sup> and 2<sup>nd</sup> raters differed by 1.00 or greater, a 3<sup>rd</sup> rating was generated to normalize the scores. For Academic Year 2025, **28 artifacts (31.5%) required a 3rd score** from an additional rater who did not initially rate the artifacts.

Overall, more than three-quarters of the artifacts were scored on average at the **Milestone (3) level or higher (83.1%)**, with all remaining artifacts being scored at the Milestone (2) level (16.9%). Over one-quarter of the artifacts were scored at the Capstone (4) level (30.3%), and **none of the artifacts were scored at or below the Benchmark (1) level**.

When examining overall normalized ratings by criteria on a five-point scale from 4.00 - 0.00, mean ( $\bar{x}$ ) scores ranged from 2.87 - 2.65, and the **overall mean score was 2.73** (standard deviation ( $\sigma$ ) = 0.79). Criterion-specific highlights include:

- Highest mean score: Control of Syntax and Mechanics ( $\bar{x}$  = 2.87,  $\sigma$  = 0.75)
- Second highest mean score: Sources and Evidence ( $\bar{x}$  = 2.72,  $\sigma$  = 0.79)
- Lowest mean score: Content Development ( $\bar{x}$  = 2.65,  $\sigma$  = 0.80)

When examining the distribution of criteria scores, **Control of Syntax and Mechanics accounted for over one-quarter of criteria scores at the Capstone (4) level (30.4%)**; conversely, Sources and Evidence accounted for most of the criteria scores at the Benchmark (1) level (40.0%).

When compared to previous assessments of Written Communication artifacts, **all four criteria had mean score averages which ranked higher in 2024-25** than any other point throughout Washburn's history of assessing Written Communication. Sources and Evidence had traditionally been the lowest-scored criteria during previous Written Communication assessment cycles, with the **mean score average for 2024-25 ( $\bar{x}$  = 2.72) being a notable increase from the next highest score in 2018-19 ( $\bar{x}$  = 2.47)**.

## Introduction

Written Communication was assessed during the 2024-25 Academic Year 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) uses a modified version (2022) of the AAC&U VALUE rubrics for assessing Written Communication every three years. Artifacts in written format are collected from students in *EN300: Advanced College Writing* and are scored by two or more independent raters using the Outcomes Assessment Projects, formerly Aqua, by the Watermark software platform.

## Review Process

Washburn faculty members were invited to attend the calibration training conducted on May 13, 2025. The 11 faculty who attended the training were assigned 90 artifacts. The artifacts were assigned to be reviewed by two independent raters on four criteria: Content Development, Genre Conventions, Sources and Evidence, and Control of Syntax and Mechanics. These four criteria were scored on a five-point scale: Capstone (4), Milestones (3), Milestone (2), Benchmark (1), and Not Present (0). Reviewers could also assign the status of unscorable to those artifacts that were not appropriate for the purpose of Written communication assessment. One artifact was designated as unscorable, resulting in 89 artifacts being assigned for review.

Of the 11 reviewers who participated in the artifact review process, five scored 16 artifacts, two scored 15 artifacts, one scored 30 artifacts, one scored 17 artifacts, one scored 14 artifacts, and one scored seven (7) artifacts. Overall, a total of 89 artifacts were reviewed two times each, for a total of 178 reviews. When the average difference in scores between the 1<sup>st</sup> and 2<sup>nd</sup> raters differed by 1.00 or greater, a 3<sup>rd</sup> rater was utilized to normalize the ratings. After the initial round of scoring, over two-thirds (68.5%) of averaged ratings did not require a 3<sup>rd</sup> rating; however, 28 artifacts (31.5%) did meet the criteria which required a 3<sup>rd</sup> rating. The process of ascertaining which artifacts required a 3<sup>rd</sup> reviewer is outlined in the next section.

## 3rd Rater Review Process

The differences in averaged ratings ranged from 2.25 – 0.00. The greatest percentage of averaged ratings differed by 0.25 (22.5%). The distribution of scores was positively skewed (right-skewed), meaning the mean of 0.66 was greater than the median of 0.50.

The distribution of averaged rating differences and descriptive statistics are shown in Table 1, Figure 1, and Table 2. A total of 28 artifacts required a 3<sup>rd</sup> rating, highlighted in blue in Table 1. These artifacts had an averaged rating difference range of 2.25 - 1.00.

Table 1. Rating Differences

Rating Difference	Frequency	Percent
0.00	12	13.5%
0.25	20	22.5%
0.50	19	21.3%
0.75	10	11.2%
1.00	10	11.2%
1.25	10	11.2%
1.50	4	4.5%
1.75	1	1.1%
2.00	2	2.2%
2.25	1	1.1%
<b>Total</b>	<b>89</b>	<b>100%</b>

Figure 1. Percent of Rating Differences

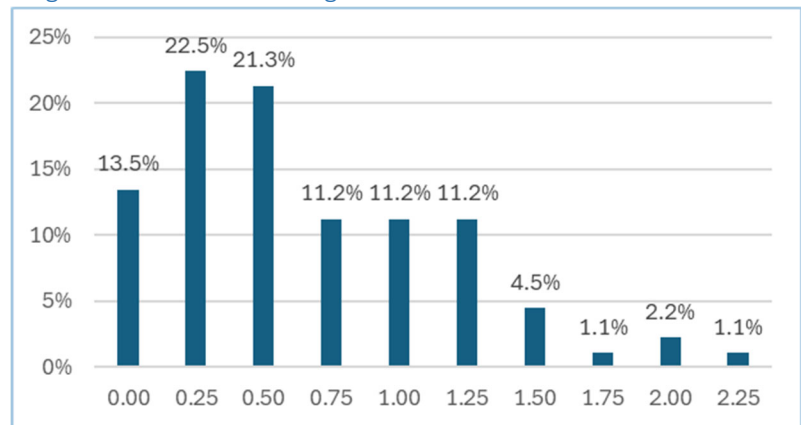


Table 2. Descriptive Statistics for Rating Differences

Mean	Median	Mode	St. Dev
0.66	0.50	0.25	0.52

The 28 artifacts that required a 3<sup>rd</sup> rating were scored by a faculty member who did not initially rate those artifacts. The 3<sup>rd</sup> rater's scores were compared to the first two raters' scores, and the score farthest from the 3<sup>rd</sup> rater's score was removed from further analysis. In two instances, however, the 3<sup>rd</sup> rater's score was either the same as or furthest from the first two raters' scores; the initial scores from the first two raters were kept for analysis in these instances. See Table 3 for means, standard deviations, and both minimum and maximum differences between averaged ratings for artifacts which either needed or did not need a 3<sup>rd</sup> rating.

Table 3. Descriptive Statistics for 3<sup>rd</sup> Rated Artifacts

	Mean	St. Dev	Minimum	Maximum
<b>No 3<sup>rd</sup> Rating Needed</b>	0.36	0.25	0.00	0.75
<b>3<sup>rd</sup> Rating Needed</b>	1.30	0.34	1.00	2.25

The mean of averaged rating differences for the artifacts which did not require a 3<sup>rd</sup> rating was 0.36, compared to 1.30 for the artifacts which did. The standard deviation for the artifacts which did not require a 3<sup>rd</sup> rating was the same at 0.25, compared to 0.34 for the artifacts which did. The minimum difference the artifacts not requiring a 3<sup>rd</sup> rating was 0.00; the maximum difference for artifacts which did require a 3<sup>rd</sup> rating was 2.25. The distribution of normalized averaged rating differences and descriptive statistics after incorporating scores from the 3<sup>rd</sup> rater are shown in Table 4, Figure 2, and Table 5.

Table 4. Normalized Rating Differences

Rating Difference	Frequency	Percent
0.00	19	21.3%
0.25	29	32.6%
0.50	26	29.2%
0.75	12	13.5%
1.00	2	2.2%
1.25	1	1.1%
Total	89	100%

Figure 2. Percent of Normalized Rating Differences

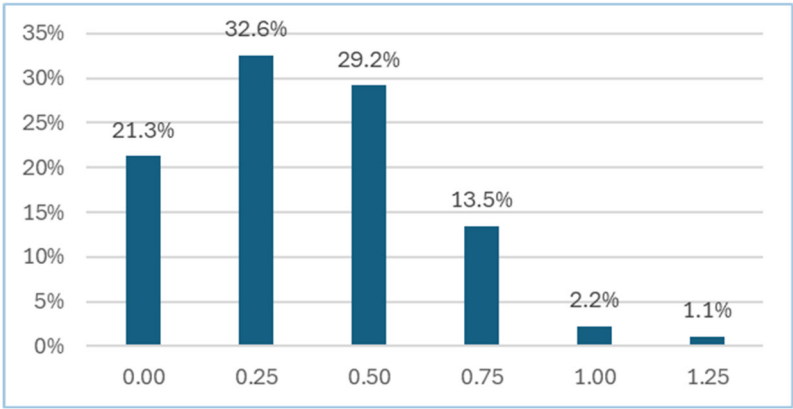


Table 5. Descriptive Stats for Normalized Rating Differences

Mean	Median	Mode	St. Dev
0.37	0.25	0.25	0.28

## Results

### Differences by Criterion

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. Written communication abilities develop through iterative experiences across the curriculum.” The rubric contains four criteria: Content Development, Genre Conventions, Sources and Evidence, and Control of Syntax and Mechanics. Washburn’s Written Communication artifacts had each of these four criteria rated on a five-point scale: Capstone (4), Milestones (3), Milestone (2 (Basic)), Benchmark (1), and Not Present (0).

The 89 Written Communication artifacts were rated by two reviewers on four dimensions for a total of 356 scores. These scores were reviewed to examine differences in averaged ratings per criterion. Table 6 provides descriptive statistics for the differences in averaged ratings by criterion.

Table 6. Descriptive Statistics for Differences in Ratings by Criterion

	Content Development	Genre Conventions	Sources and Evidence	Control of Syntax and Mechanics	Total Difference
Mean	0.47	0.46	0.57	0.49	0.51
St. Dev.	0.55	0.50	0.58	0.57	0.54
Min	0.00	0.00	0.00	0.00	0.00
Max	2.00	1.00	2.00	2.00	2.00

The mean difference for averaged ratings was greatest for Sources and Evidence, meaning the ratings, on average, differed by 0.57 with a standard deviation of 0.58. Conversely, the smallest mean differences were for Genre Conventions, which differed by 0.46 with a standard deviation of 0.50. The minimum averaged rating was 0.00 (or no difference) across all criteria, whereas the maximum averaged rating was 2.00 for all criteria except Genre Conventions, which was 1.00.

### Distribution of Scores

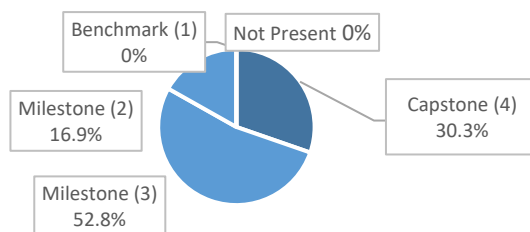
The scores from the two raters were averaged to provide a normalized score for each Written Communication artifact. Table 7 provides the overall distribution of 356 scores for 89 Written Communication artifacts. The ranges of average scores were defined as 4.00 - 3.01 for Capstone (4), 3.00 - 2.01 for Milestone (3), 2.00 - 1.01 for Milestone (2), and 1.00 - 0.01 for Benchmark (1). Those criteria that were not rated were designated as Not Present.

Table 7. Descriptive Data and Statistics for Overall Averaged Ratings

	Capstone (4)	Milestone (3)	Milestone (2)	Benchmark (1)	Not Present	Mean (sd)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	0.00	
<b>Overall (n = 89)</b>	27 (30.3%)	47 (52.8%)	15 (16.9%)	0 (0%)	0 (0%)	2.73 (0.79)

Most artifacts (52.8%) were scored at Milestone (3), followed by Capstone (4) (30.3%). 15 artifacts (16.9%) were rated at Milestone (2), while no artifacts were rated at Benchmark (1) or Not Present (0). See Figure 3 for a visual representation.

Figure 3. Distribution of Scores for Overall Written Communication Artifacts



### Ratings by Criterion

The distribution of average ratings and descriptive statistics for each of the four criteria are in Table 8. The table shows how the artifacts ( $n = 89$ ) were rated on each of the four criteria (total of 356 scores).

Table 8. Descriptive Data and Statistics for Overall Averaged Ratings by Criterion

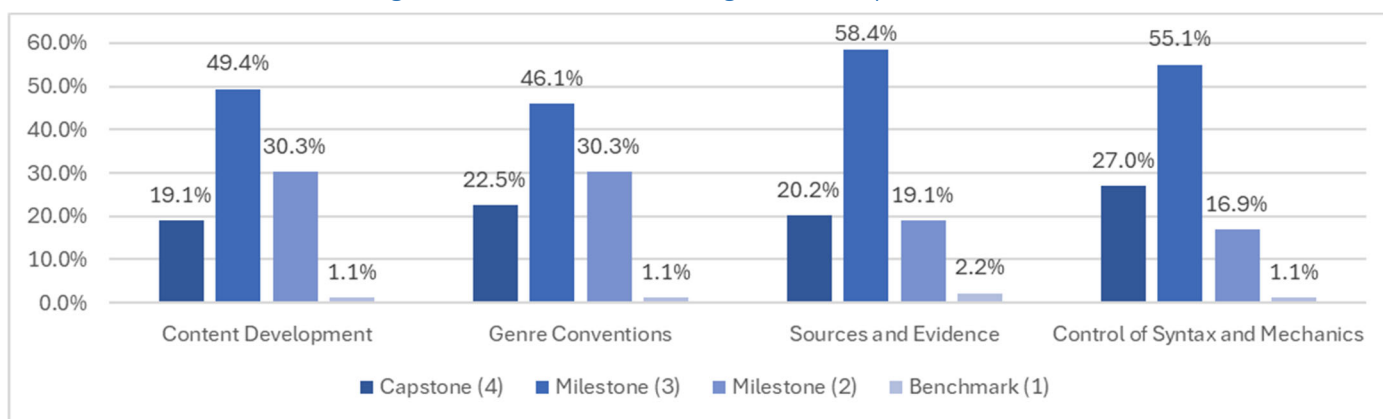
	Capstone (4)	Milestone (3)	Milestone (2)	Benchmark (1)	Not Present (0)	Mean (sd)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	0.00	
<b>Content Development</b>	17 (19.1%)	44 (49.4%)	27 (30.3%)	1 (1.1%)	0 (0%)	2.65 (0.80)
<b>Genre Conventions</b>	20 (22.5%)	41 (46.1%)	27 (30.3%)	1 (1.1%)	0 (0%)	2.67 (0.79)
<b>Sources and Evidence</b>	18 (20.2%)	52 (58.4%)	17 (19.1%)	2 (2.2%)	0 (0%)	2.72 (0.79)
<b>Control of Syntax and Mechanics</b>	24 (27.0%)	49 (55.1%)	15 (16.9%)	1 (1.1%)	0 (0%)	2.87 (0.75)
<b>Overall (n = 356)</b>	79 (22.2%)	186 (52.2%)	86 (24.2%)	5 (1.4%)	0 (0%)	2.73 (0.79)

The averaged ratings of the Sources and Evidence and Control of Syntax and Mechanics scores all fell within the 2.87 - 2.72 range, indicating these criteria scored on the higher end of the Milestone (3) range. The averaged ratings of the Content Development and Genre Conventions scores all fell within

the 2.67 - 2.65 range, indicating these criteria scores were near the mid-point of the Milestone (3) range.

For the Control of Syntax and Mechanics criterion, over 80% of the averaged scores were in the Capstone (4) and Milestone (3) range (82.1%), contributing to the highest overall mean score of the criteria; additionally, Control of Syntax and Mechanics had the lowest standard deviation 0.75, indicating its scores were among the most closely clustered around its mean score. Conversely, Content Development and Genre Conventions had the highest percentage of ratings at or below the Milestone (2) range (31.4%). Refer to Figure 4 for a visual representation of the distribution of ratings by each criterion (excluding Not Present). All four criterion had a nearly normal distribution.

Figure 4. Distribution of Averaged Scores by Criterion



### Comparison to Previous Years Results

Washburn reviewed Written Communication artifacts for the first time 2015-16 ( $n = 199$ ), from which approximately 46% of the artifacts attained an overall averaged score at or above the Milestone (3) performance level. In 2018-19 ( $n = 200$ ) and 2021-22 ( $n = 100$ ), 77.0% and 64.0% of the artifacts were at the same level, respectively. For the current year, 2024-25 ( $n = 89$ ), 83.1% of the artifacts attained an overall averaged score at or above Milestone (3); additionally, the mean of overall normalized, averaged ratings has continued to increase, from 2.53 in 2018-19, to 2.33 in 2021-22, and to 2.73 in 2024-25.

All four criteria had their highest mean scores for averaged ratings by criterion in 2024-25, outpacing the previous highest mean scores in 2018-19. Control of Syntax and Mechanics saw the highest mean score for averaged rankings in 2024-25 at  $\bar{x} = 2.87$ , showing notable growth from 2018-19 at 2.55. Sources and Evidence – which had consistently been one of the lowest scoring criteria during previous assessments of Written Communication artifacts – was the second highest mean score for averaged rankings in 2024-25 at  $\bar{x} = 2.72$ , also showing increases compared to 2018-19 at 2.47. Finally, Content Development increased from  $\bar{x} = 2.56$  in 2018-19 to  $\bar{x} = 2.65$  in 2024-25, and Genre Conventions increased from  $\bar{x} = 2.58$  in 2018-19 to  $\bar{x} = 2.67$  in 2024-25.

### Areas of Consideration and Limitations

Following the 2018-19 Written Communication artifact rating, reviews were adjusted to not be done in pairs, but rather by having each rater receive a random set of artifacts to rate. Removing repeated pairs of raters for artifacts requiring a 3<sup>rd</sup> rater and the overall decreased number of artifacts requiring a 3<sup>rd</sup> rater over time indicates that assigning raters a random set of artifacts has been normalizing scores over time. However, in 2024-25, two artifacts which required a 3<sup>rd</sup> rater had either the same or a greater

difference in ratings when compared to the original two rater's scores, resulting in the original raters' scores being kept for analysis.

The methodology of determining 3<sup>rd</sup> ratings and normalizing scores was modified in 2018-2019, which is important to note when comparing Written Communication assessment results longitudinally. Specifically, the threshold for artifacts requiring a 3<sup>rd</sup> rater was increased from a difference of more than 1.00 or greater ( $\leq 1.01$ ) to a difference of 1.00 or greater ( $\leq 1.00$ ), potentially causing a need for more 3<sup>rd</sup> raters in 2018-19. Regarding normalizing rater's scores: beginning in 2018-19, the 3<sup>rd</sup> rater's score was compared to the first two raters' scores, with the score farthest from the 3<sup>rd</sup> rater's score being removed from further analysis. Prior to this, the methodology was to average the 3<sup>rd</sup> rater's score with the first two raters' scores, resulting in more varied normalized scores due to the normalized ratings for artifacts requiring a 3<sup>rd</sup> rater averaging the scores of three raters instead of two. Consequently, the ranges of the scoring levels for artifacts were modified in 2018-19 to accommodate the normalization of the two ratings as opposed to 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 during the Fall 2024 and Spring 2025 terms. A random sampling was not used to select artifacts for review; the submissions were not required. 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.

# WRITTEN COMMUNICATION VALUE RUBRIC

*for more information, please contact [value@aacu.org](mailto: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 of 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](http://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](http://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](mailto: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 32		Benchmark 1
Content Development	Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's 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.