Critical Thinking Artifact Assessment Report 2020-2021

Executive Summary

- In AY 2021, 114 Critical Thinking artifacts were scored by two separate raters using a revised version of the AAC&U Critical Thinking VALUE rubric. The rubric consists of four criteria:
    - Explanation of Issues
    - Evidence
    - Student's Position
    - Conclusions and Related Outcomes

  Each criterion was rated on a five-point scale: Capstone (4), Milestones (3 and 2), Benchmark (1), and Not Present (0). It also included the option to mark an artifact as unscorable.
- A 3rd rating was generated if the average of the 1st rater's score and the 2nd rater's score differed by 1.00 or greater.
    - 32 artifacts (28.3%) required a 3rd score from a rater who did not initially rate the artifacts.
    - The 3rd scores were compared to the first two raters' scores, and the score most distant from the 3rd rater's score was removed from further analysis.
- Overall, almost half of the artifacts were scored on average at Milestone (2) (41.2%), followed by another third of the artifacts at Milestone (3) (35.4%). Capstone (4) ratings were representative of 11.9% of artifacts, Benchmark (1) ratings were representative of 10.2% of artifacts, and Not Present (0) ratings were representative of 1.3% of artifacts.
- For overall normalized ratings by criteria, 2.47 to 2.02 was the mean score range, with an overall mean score of 2.24 on the five-point scale from 4.00 to 0.00.
    - The criterion with the highest mean score was Explanation of Issues (2.47); 33.3% of Critical Thinking artifacts were rated at the Capstone (4) level.
    - The criterion with the second highest mean score was Evidence (2.38).
    - The criteria with the lowest mean scores were Student's Position (2.09) and Conclusions and Related Outcomes (2.02).
- Almost all criteria were rated higher in 2017-18 than they were in 2014-15 or in 2020-21. But for the 2020-21 analysis, Explanation of Issues had the highest mean score at 2.47, whereas in 2017-18 its mean was 2.59, and in 2014-15 its mean was 2.45.

Introduction

Critical Thinking was assessed during the 2020-21 academic year (AY) using a modified version of the Association of American Colleges and Universities (AAC&U) 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 a modified version (2015) of the AAC&U VALUE rubrics for assessing Critical Thinking 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.

Review Process

Washburn faculty were invited to attend the calibration training conducted via Zoom on May 25, 2021. The 14 faculty who attended the training were assigned 114 artifacts collected from Fall 2020 and Spring 2021 EN 300: Advanced College Writing courses. The artifacts were assigned to be reviewed by two independent raters on four criteria: Explanation of Issues, Evidence, Student's Position, and Conclusions and Related Outcomes. These four criteria were scored on the five-point scale of Capstone (4), Milestones (3 and 2), and Benchmark (1), with the additional Not Present (0) for scoring criteria that did not meet Benchmark (1) level performance. Reviewers could also assign the status of unscorable to those artifacts that were not appropriate for the purpose of critical thinking assessment.

Of the 14 reviewers who participated in the artifact review process, one scored 20 artifacts, 11 scored 17 artifacts, one scored 10 artifacts, and one scored 9 artifacts. One artifact was designated as unscorable, which resulted in a total of 113 artifacts reviewed two times each, for a total of 226 reviews.

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 (71.7%) 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 32 artifacts (28.3%) that met or exceeded 1.00 average difference in scores and required a 3rd rating. The process of determining which artifacts needed a 3rd reviewer is outlined in the next section.

3rd Rater Review Process

The differences in ratings ranged from 0.00 to 2.50. The greatest percentage of average ratings differed by 0.25 (22.3%), followed by 0.75 (20.5%). The distribution of scores was positively skewed (right-skewed) in that the mean (average value) of 0.68 was greater than the median (middle value) of 0.50. The distribution of rating differences and descriptive statistics are shown in the tables and chart on the following page.

A total of 32 artifacts required a 3rd rating, highlighted in blue in Table 1. Twelve artifacts (10.7%) had an average rating difference of 1.00, seven had rating differences of 1.25 (6.3%), six had rating differences of 1.50 (5.4%), four had rating differences of 1.75 (3.6%), two had rating differences of 2.00 (1.8%) and one artifact had a 2.50 rating difference (0.9%).

Table 1. Rating Differences

Rating Difference	Frequency	Percent
0.00	14	12.4%
0.25	25	22.3%
0.50	18	16.1%
0.75	23	20.5%
1.00	12	10.7%
1.25	7	6.3%
1.50	6	5.4%
1.75	4	3.6%
2.00	2	1.8%
2.50	1	0.9%
Total	**113**	**100%**

Figure 1. Percent of Rating Differences

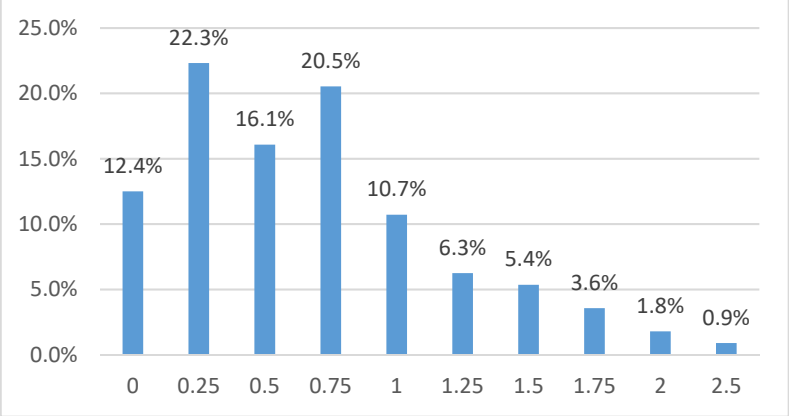


Table 2. Descriptive Statistics for Rating Differences

Mean	Median	Mode	St. Dev.
0.68	0.50	0.25	0.53

The 32 artifacts that required a 3rd rating were scored by a faculty member who did not initially rate those artifacts. The 3rd rater's scores were compared to the first two raters' scores, and the score farthest from the 3rd rater's score was removed from further analysis. See Table 3 for means, minimum difference, and maximum difference between ratings.

Table 3. Descriptive Statistics for Ratings Selected and Not Selected

	Mean	St. Dev	Minimum	Maximum
Rating Selected	0.41	0.32	0.00	1.00
Rating Not Selected	1.32	0.48	0.25	2.25

The mean difference between those ratings that were selected and the 3rd rater's scores was 0.41; the mean difference between those ratings that were not selected and the third rater's scores was 1.32. The standard deviation was greater for artifacts for which ratings were not selected (*sd* = 0.48) than for artifacts for which ratings were selected (*sd* = 0.32). The minimum difference for those ratings that were selected was 0.00; the maximum difference was 1.00.

Results

Differences by Criterion

The AAC&U Critical Thinking VALUE rubric defines Critical Thinking as "a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion." The rubric contains four criteria: Explanation of Issues, Evidence, Student's Position (perspective, thesis/hypothesis), and Conclusions and Related Outcomes (implications and consequences). The written artifacts were rated on these four criteria on a five-point scale consisting of Capstone (4), Milestones (3 and 2), and Benchmark (1).

The 113 Critical Thinking artifacts were rated by two reviewers on four dimensions for a total of 452 scores. These scores were reviewed for each of the four criteria to examine differences in ratings per criterion. Table 4 provides the descriptive statistics for the differences in ratings by criterion.

Table 4. Descriptive Statistics for Differences in Ratings by Criterion

	Explanation of Issues	Evidence	Student's Position	Conclusions and Related Outcomes	Total Difference
Mean	0.50	0.58	0.53	0.51	0.53
St. Dev.	0.58	0.65	0.60	0.54	0.59
Min	0.00	0.00	0.00	0.00	0.00
Max	2.00	3.00	3.00	2.00	3.00

The mean difference was greatest for Evidence and Student's Position in that the ratings on average differed by 0.58 and 0.53, respectively. The mean difference was smallest for Conclusions and Related Outcomes and Explanation of Issues with an average difference in rating of 0.51 and 0.50, respectively. The minimum rating was 0.00 (or no difference) across all criteria, whereas the maximum rating was 2.00 for Explanation of Issues and Conclusion and Related Outcomes and 3.00 for Evidence and Student's Position.

Distribution of Scores

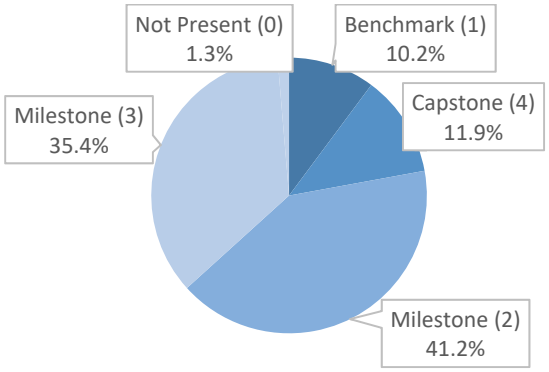
The scores from the two raters were averaged to provide a normalized score for each Critical Thinking artifact. Table 1 provides the overall distribution of 452 scores for 113 Critical Thinking artifacts. The ranges of average scores were defined as 4.00-3.01 Capstone, 3.00-2.01 and 2.00-1.01 Milestones, and 1.00-0.01 Benchmark. Those criteria that were not rated are designated Not Present.

Table 5. 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	
Overall (*n* = 452)	54 (11.9%)	160 (35.4%)	186 (41.2%)	46 (10.2%)	6 (1.3%)	2.24 (0.84)

The majority of artifacts (41.2%) were scored at Milestone (2), followed by Milestone (3) (35.4%). Fifty-four (54) artifacts (11.9%) were rated at Capstone (4), 46 (10.2%) were rated at Benchmark (1), and six (1.3%) were rated at Not Present (0). See Figure 2 below for a visual representation.

Figure 2. Distribution of Scores for Overall Critical Thinking Artifacts



Ratings by Criterion

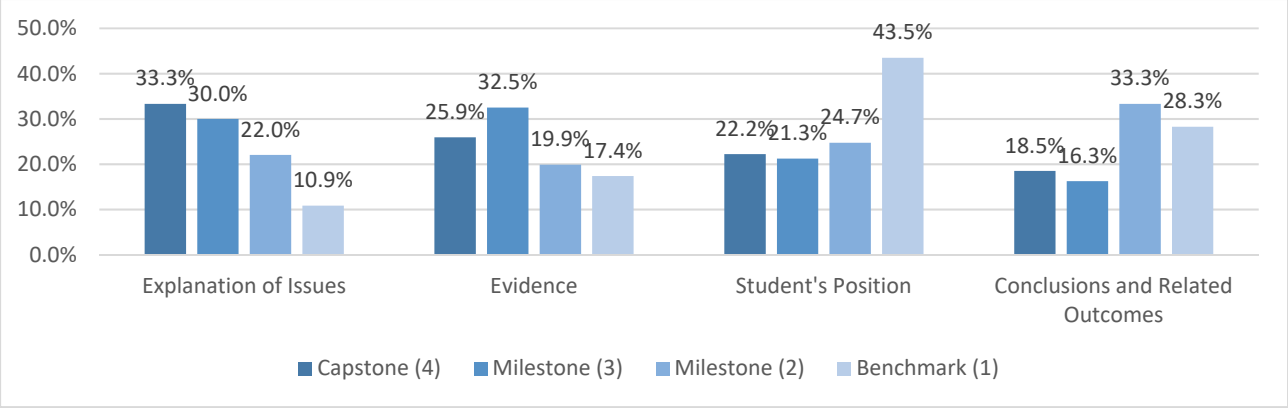
The distribution of average ratings and descriptive statistics for each of the four criteria are in Table 6, below. The table shows how the artifacts (*n* = 113) were rated on each of the four criteria (total of 452 scores).

Table 6. 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	
Explanation of Issues	18 (33.3%)	48 (30.0%)	41 (22.0%)	5 (10.9%)	1 (16.7%)	2.47 (0.79)
Evidence	14 (25.9%)	52 (32.5%)	37 (19.9%)	8 (17.4%)	2 (33.3%)	2.38 (0.82)
Student's Position	12 (22.2%)	34 (21.3%)	46 (24.7%)	20 (43.5%)	1 (16.7%)	2.09 (0.88)
Conclusions and Related Outcomes	10 (18.5%)	26 (16.3%)	62 (33.3%)	13 (28.3%)	2 (33.3%)	2.02 (0.79)
Overall (*n* = 452)	54 (100%)	160 (100%)	186 (100%)	46 (100%)	6 (100%)	2.24 (0.84)

For Explanation of Issues and Evidence, the ratings fell within the 2.47 to 2.38 mean range, indicating that these artifacts' scores were in the mid-point of the Milestone (3) range, on average. Further, one third of scores for Explanation of Issues (33.3%) were in the Capstone (4) range, and Evidence received around a quarter of scores (25.9%) in the Capstone (4) range. However, the highest percent of scores for Evidence fell in the Milestone (3) range (32.5%), which likely contributed to its lower mean score (2.38) compared with Explanation of Issues (2.47). Student's Position and Conclusions and Related Outcomes were near the Milestone (2) range, but these two criteria actually fell toward the lower end of the Milestone (3) range. Most scores for Student's Position fell within the Benchmark (1) range (43.5%). Finally, a third of the Conclusions and Related Outcomes scores (33.3%) fell within the Milestone (2) range. See the bar chart below for a visual representation of the distribution of ratings by each criterion (excluding Not Present). Evidence was the only criterion with a normal distribution of ratings. Explanation of Issues was positively skewed while Student's Position and Conclusions and Related Outcomes were negatively skewed.

Figure 3. Distribution of Scores by Criterion



Comparison to Previous Years Results

Critical Thinking artifacts were reviewed for the first time in 2014-15 (*n* = 157). In 2014-15, approximately 52% of the artifacts attained an overall average score at the Milestone (3) performance level. In 2017-18 (*n* = 200), 47.6% were at Milestone (3) performance levels. For the current year, 2020-21, 35.4% were categorized as Milestone (3); however, 41.2% were scored at the Milestone (2) performance level in 2020-2021.

All criteria were rated higher in 2017-18 than in 2014-15 or 2020-21. For the 2017-18 analysis, the highest mean score was Explanation of Issues with 2.59, higher than 2.45 in 2014-15, but slightly lower than the 2020-21 score of 2.47. Evidence was the second highest scored criterion in 2014-15, 2017-18, and 2020-21, yielding means of 2.31, 2.53, and 2.38, respectively. Student's Position was scored similarly in all three years with mean scores of 2.11 in 2014-15, 2.16 in 2017-18 and 2.09 in 2020-21. Finally, Conclusions and Related Outcomes was scored the lowest all three years with 2.20 in 2017-18, 2.08 in 2014-15, and 2.02 in 2020-21.

Areas of Consideration and Limitations

Critical Thinking artifacts scored during the 2017-18 and 2020-21 academic years were scored using the Watermark Outcomes Assessment Projects (formerly known as Aqua) 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 the 2014-15 year difficult to analyze.

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 2017-18. The cut line for 3rd rater scores was increased from scores with a difference of more than 1.00 (≤ 1.01) to include those with a difference of 1.00 or greater (≤ 1.00). This may have resulted in more 3rd raters needed for 2020-21. Additionally, when normalizing ratings, the 3rd ratings were compared to the first two raters' scores, and the 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 2017-18 to accommodate the normalization of the two ratings.

Finally, reasonable efforts were made to collect Critical Thinking artifacts from students enrolled in EN 300: Advanced College Writing, a university requirement for Junior level students during the Fall 2020 and Spring 2021 terms. A random sampling was not used to select artifacts for review; however, the submissions were voluntary. 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 Critical Thinking.

Appendix A

CRITICAL THINKING VALUE RUBRIC

for more information, please contact value@aacu.org

Association of American Colleges and Universities

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

Critical thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.

Framing Language

This rubric is designed to be transdisciplinary, reflecting the recognition that success in all disciplines requires habits of inquiry and analysis that share common attributes. Further, research suggests that successful critical thinkers from all disciplines increasingly need to be able to apply those habits in various and changing situations encountered in all walks of life.

This rubric is designed for use with many different types of assignments and the suggestions here are not an exhaustive list of possibilities. Critical thinking can be demonstrated in assignments that require students to complete analyses of text, data, or issues. Assignments that cut across presentation mode might be especially useful in some fields. If insight into the process components of critical thinking (e.g., how information sources were evaluated regardless of whether they were included in the product) is important, assignments focused on student reflection might be especially illuminating.

Glossary

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

- Ambiguity: Information that may be interpreted in more than one way.
- Assumptions: Ideas, conditions, or beliefs (often implicit or unstated) that are "taken for granted or accepted as true without proof." (quoted from www.dictionary.reference.com/browse/assumptions)
- Context: The historical, ethical. political, cultural, environmental, or circumstantial settings or conditions that influence and complicate the consideration of any issues, ideas, artifacts, and events.
- Literal meaning: Interpretation of information exactly as stated. For example, "she was green with envy" would be interpreted to mean that her skin was green.
- Metaphor: Information that is (intended to be) interpreted in a non-literal way. For example, "she was green with envy" is intended to convey an intensity of emotion, not a skin color.

CRITICAL THINKING VALUE RUBRIC

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

Definition

Critical thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.

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** 3	 2	**Benchmark** 1	**Not Present** 0
Explanation of issues	Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.	Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.	Issue/problem to be considered critically is stated but description leaves some terms undefined, ambiguities unexplored, boundaries undetermined, and/or backgrounds unknown.	Issue/problem to be considered critically is stated without clarification or description.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Evidence *Selecting and using information to investigate a point of view or conclusion*	Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis.	Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis.	Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis.	Information is taken from source(s) without any interpretation /evaluation.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Student's position (perspective, thesis/hypothesis)	Specific position (perspective, thesis/hypothesis) is sophisticated, taking into account the complexities of an issue. Limits of position (perspective, thesis/hypothesis) are acknowledged. Others' points of view are synthesized within position (perspective, thesis/hypothesis).	Specific position (perspective, thesis/hypothesis) takes into account the complexities of an issue. Others' points of view are acknowledged within position (perspective, thesis/hypothesis).	Specific position (perspective, thesis/hypothesis) acknowledges different sides of an issue.	Specific position (perspective, thesis/hypothesis) is stated, but is simplistic and obvious.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.
Conclusions and related outcomes (implications and consequences)	Conclusions and related outcomes (consequences and implications) are logical and reflect student's informed evaluation and ability to place evidence and perspectives discussed in priority order.	Conclusions are logically tied to a range of information, including opposing viewpoints; related outcomes (consequences and implications) are identified clearly.	Conclusions are logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes (consequences and implications) are identified clearly.	Conclusions are inconsistently tied to some of the information discussed; related outcomes (consequences and implications) are oversimplified.	Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.