Critical Thinking Artifact Assessment Report 2017-2018

Executive Summary

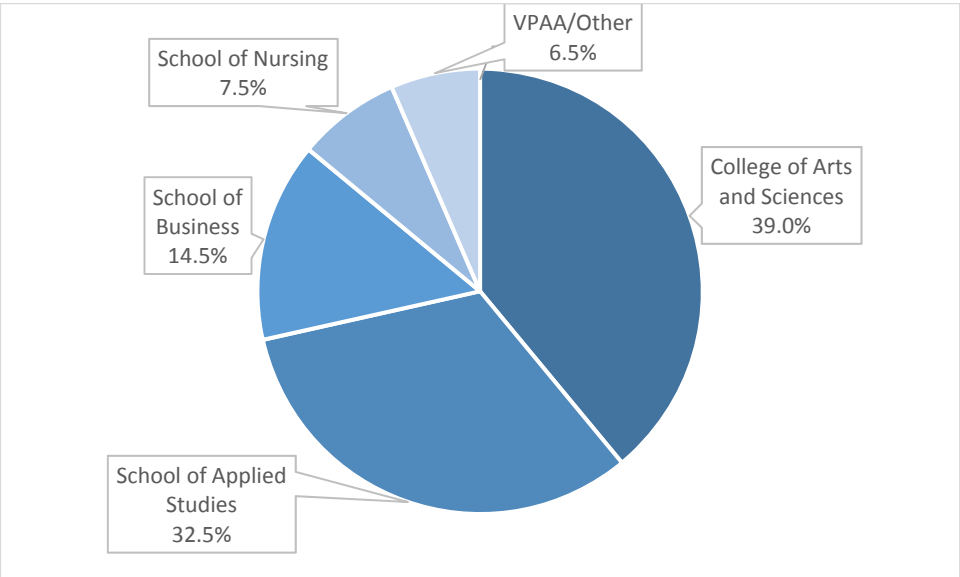
- In AY 2018, 200 previously written critical thinking artifacts were scored by two separate raters using a revised version of the AAC&U Critical Thinking VALUE rubric. The rubric consisted of:
    - Four criterion: Explanation of Issues, Evidence, Student's Position, and Conclusions and Related Outcomes
    - Rated on a five-point scale: Capstone (4), Milestones (3 and 2), Benchmark (1) and Not Present (0).
- Although 54 artifacts needed 3rd ratings, 13 artifacts were not given a 3rd rating and were removed from further analysis, resulting in a total of 187 artifacts. 3rd ratings were generated for 41 (20.5%) of the artifacts, due to the average score of the 1st rater's scoring and the 2nd rater's scoring differing by 1.00 or greater.
    - 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.
- Overall, most of the average ratings were Milestone 3 (47.6%), followed Milestone 2 at 32.6% of the average ratings. The remaining performance levels were 17.6% for Capstone, 1.6% for Benchmark and 0.5% for Not Present.
- For ratings by criteria:
    - The criterion with the highest mean score was Explanation of Issues with 2.59; 46.0% of artifacts resulted in a Milestone 3 rating.
    - Evidence also had average ratings in the Milestone 3 level ($\bar{x}$s = 2.53); 14.4% of artifacts were rated Capstone and 46.5% were rated Milestone 3.
    - Student's Position and Conclusion and Related Outcomes were rated similarly with average ratings of 2.16 and 2.20, respectively, and primarily scored as Milestone 2 (38.0% and 51.9%, respectively).
- All of the mean scores for critical thinking artifacts were higher in 2017-2018 than 2014-2015. Explanation of Issues yielded the highest mean scores in 2014-2015 and 2017-2018, 2.59 and 2.45, respectively.

Introduction

Critical Thinking was assessed during the 2017-18 academic year (AY) using a modified version of the Association of American Colleges and Universities (AAC&U) Critical Thinking 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 Critical Thinking every three years.

Artifacts (*n* = 200) were collected from students in EN 300 Advanced College Writing during the Fall 2016, Spring 2017, and Fall 2017 semesters. EN 300 is designed to be taken in the Junior year and prepares students for advanced academic writing. These 200 artifacts were from students who were enrolled in the majors in the following divisions at Washburn: College of Arts and Sciences (39.0%), School of Applied Studies (32.5%), School of Business (14.5%), School of Nursing (7.5%), and VPAA/Other (6.5%). See Figure 1, below. The College of Arts and Sciences primarily consisted of Psychology (11), Kinesiology (11), and Education (10); the School of Applied Studies was mostly composed of Allied Health (28) and Criminal Justice/Legal Studies (22); School of Business was mostly Pre-Accounting (4); School of Nursing was comprised of Nursing (8) and Pre-Nursing (6); and the VPAA/Other category contained students with majors categorized as Exploratory (13).

Figure 1. Percentage of Artifacts Collected from Student Majors in Divisions at Washburn



Critical Thinking Artifact Review Process

The 200 artifacts were reviewed by two independent raters on four criterion: Explanation of Issues, Evidence, Student's Position, and Conclusions and Related Outcomes. These four criterion were scored on a five-point scale that consisted of Capstone (4), Milestones (3 and 2), Benchmark (1), and Not Present (0).

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, there were 54 artifacts (27.0%) that met or

exceeded 1.00 and required a 3rd rating. However, 13 artifacts were not given a 3rd rating and were removed from further analysis, resulting in a total of 187 artifacts. The 41 artifacts (20.5%) 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 Appendix B for further discussion of the methodology and results of the 3rd rater reviews.

Critical Thinking Artifact Review Results

Overall Normalized Ratings

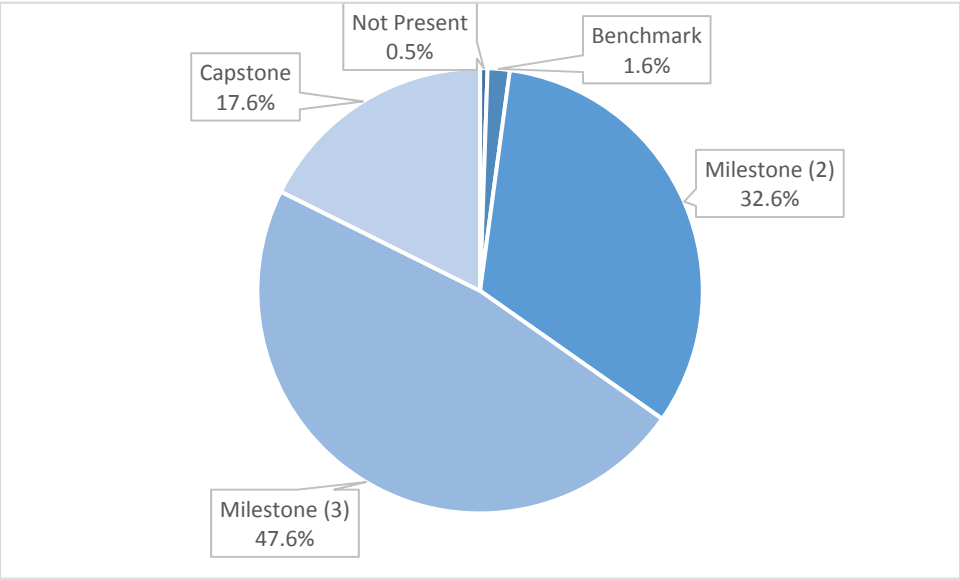
The scores from the two raters were averaged to provide a normalized score for each artifact. Table 1 provides the overall distribution of scores for the 187 Critical Thinking artifacts. The range of average scores were categorized as Capstone (4), Milestones (3 and 2), Benchmark (1), and Not Present (0).

Table 1. Descriptive Data and Statistics for Overall Averaged Ratings

	Capstone	**Milestones**		**Benchmark**	**Not Present**	**Mean (*sd*)**
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	0.00	
Overall (*n* = 187)	33 (17.6%)	89 (47.6%)	61 (32.6%)	3 (1.6%)	1 (0.5%)	2.37 (0.69)

Of the overall normalized ratings, most of the average ratings fell within Milestone 3 (47.6%), followed Milestone 2 at 32.6% of the average ratings. The remaining performance levels were 17.6% for Capstone, 1.6% for Benchmark and 0.5% for Not Present. See Table 1 and the pie chart, Figure 1, for the distribution of scores.

Figure 1. Pie Chart of Distribution of Scores for Overall Critical Thinking Artifacts



The distribution of average ratings and descriptive statistics for the 187 critical thinking artifacts are in Table 2 on the following page. The table contains the how the artifacts (*n* = 187) were rated on each of the four criteria (total of 748 scores). For example, the row labeled as Evidence contains the scores from all 187 artifacts on that particular criteria.

Table 2. Descriptive Data and Statistics for Overall Averaged Ratings

	Capstone	Milestones		Benchmark	Not Present	Mean (*sd*)
	4.00 - 3.01	3.00 - 2.01	2.00 - 1.01	1.00 - 0.01	0.00	
Explanation of Issues	34 (18.2%)	86 (46.0%)	62 (33.2%)	4 (2.1%)	1 (0.5%)	2.59 (0.76)
Evidence	27 (14.4%)	87 (46.5%)	67 (35.8%)	5 (2.7%)	1 (0.5%)	2.53 (0.71)
Student's Position	20 (10.7%)	66 (35.3%)	71 (38.0%)	27 (14.4%)	3 (1.6%)	2.16 (0.86)
Conclusions and Related Outcomes	16 (8.6%)	59 (31.6%)	97 (51.9%)	14 (7.5%)	1 (0.5%)	2.20 (0.72)
Overall (*n* = 748)	97 (13.0%)	298 (39.8%)	297 (39.7%)	50 (6.7%)	6 (0.8%)	2.37 (0.79)

The criterion with the highest mean score was Explanation of Issues with 2.59 or an average score of Milestone 3. Explanation of Issues had 46.0% of artifacts scored at a Milestone 3 level and this criterion yielded the largest percentage of Capstone ratings than the other three criterion. Evidence also had average ratings in the Milestone 3 level ($\bar{x}$s = 2.53); 14.4% of artifacts were rated Capstone and 46.5% were rated Milestone 3. Student's Position and Conclusion and Related Outcomes were rated similarly with average ratings of 2.16 and 2.20, respectively, and primarily scored as Milestone 2 (38.0% and 51.9%, respectively). Overall, mean scores were 2.37 or Milestone 3.

See Figures 2-5 below and on the following page for visual representations of the distributions of ratings. The criteria presented distributions of ratings in a normal curve.

Figure 2. Distribution of Scores for Explanation of Issues

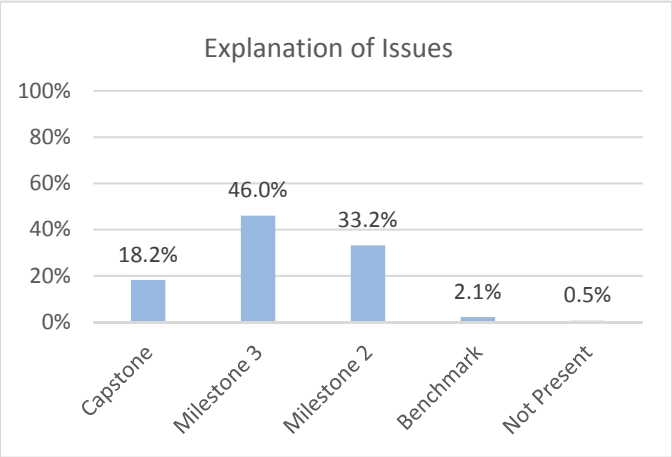


Figure 3. Distribution of Scores for Evidence

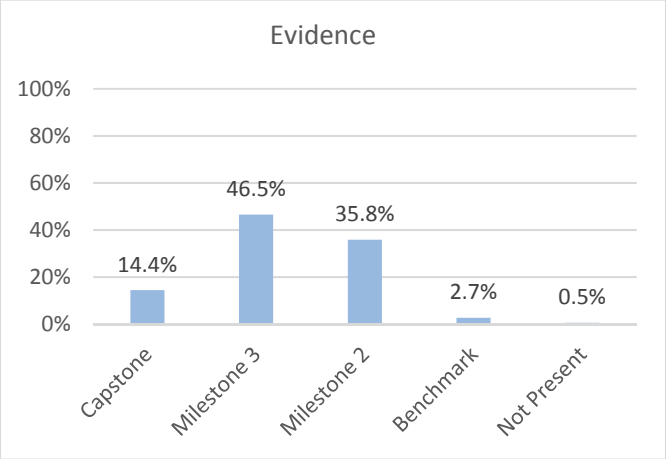


Figure 4. Distribution of Scores for Student's Perspective

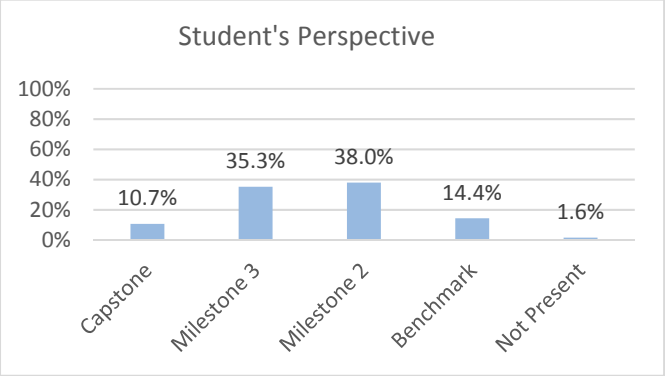
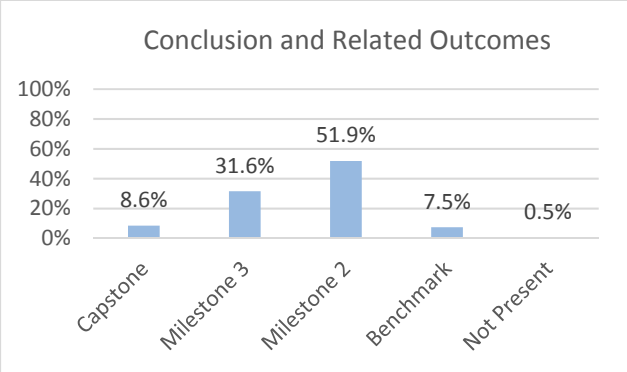


Figure 5. Distribution of Scores for Conclusion and Related Outcomes



Comparison to Previous Years Results

Critical Thinking artifacts were reviewed for the first time in 2014-15 (*n* = 157). In 2014-15, approximately 18% of the artifacts attained an overall average score at the Milestone 3 performance level and Capstone was 1%. In 2017-18, approximately 18% of average scores were categorized at Capstone and 47% were at Milestone 3 performance levels.

All mean scores were rated higher in 2017-18 than in 2014-15. For the 2017-2018 analysis, the highest mean score was Explanation of Issues with 2.59, an increase from 2.45 in 2014-2015. Evidence was the second highest scored criteria in 2014-2015 and 2017-2018, yielding means of 2.31 and 2.53, respectively. Student's Position was scored similarly in 2014-2015 and 2017-2018 with mean scores at 2.11 in 2014-2015 and 2.16 in 2017-2018. Finally, Conclusion and Related Outcomes were scored at 2.20 in 2017-2018, and 2.08 in 2014-2015.

Conclusion, Areas of Consideration and Limitations

Critical Thinking artifacts scored during the 2017-18 academic year, using a modified version of the AAC&U Critical Thinking VALUE rubric, totaled 200. Each artifact was rated by two reviewers. After the initial round of scoring, there were 54 artifacts (27.0%) that met or exceeded 1.00 and required a 3rd rating. However, 13 artifacts were not given a 3rd rating and were removed from further analysis, resulting in a total of 187 artifacts. With the adoption of Aqua by Watermark for artifact assessment projects, monitoring which artifacts that need a third rating should be easier to identify.

The first two and the last two criterion were very similarly rated. Mean scores between the first two criterion, Explanation of Issues and Evidence were within a range of 0.07 (minimum = 2.53, maximum = 2.59). Student's Position and Conclusion and Related Outcomes yielded the lowest mean scores within a range of 0.05 (minimum = 2.16, maximum = 2.20). These latter two areas could be improved.

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, from Fall 2016 to Fall 2017. 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 critical thinking.

Appendix A

CRITICAL THINKING 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 by 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

Association of American Colleges and Universities

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
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.
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.
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.
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.	Conclusion is logically tied to a range of information, including opposing viewpoints; related outcomes (consequences and implications) are identified clearly.	Conclusion is logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes (consequences and implications) are identified clearly.	Conclusion is inconsistently tied to some of the information discussed; related outcomes (consequences and implications) are oversimplified.

Appendix B

3rd Rater Review Methodology

The 187 artifacts were reviewed by two independent raters on four criterion: Explanation of Issues, Evidence, Student's Position, and Conclusions and Related Outcomes. These four criterion were scored on a five-point scale that consisted of Capstone (4), Milestones (3 and 2), Benchmark (1), and Not Present (0). Of the 20 raters who participated in the artifact review process, 19.3% scored 20 artifacts, 48.0% scored more than 20 artifacts, and 32.7% scored more less 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 (31.6%), followed by 0.50 (20.3%), and 0.00 (14.4%). The distribution of scores were positively skewed (right-skewed) in that the mean (average) was 0.56 which was greater than the Median (middle) of 0.50. The distribution of rating differences are shown in the tables and chart, below.

Table 1B. Rating Differences

Rating Difference	Frequency	Percent
0.00	27	14.4%
0.25	59	31.6%
0.50	38	20.3%
0.75	22	11.8%
1.00	19	10.2%
1.25	7	3.7%
1.50	6	3.2%
1.75	5	2.7%
2.00	3	1.6%
2.25	1	0.5%
Total	**187**	**100%**

Figure 1B. Percent of Rating Differences

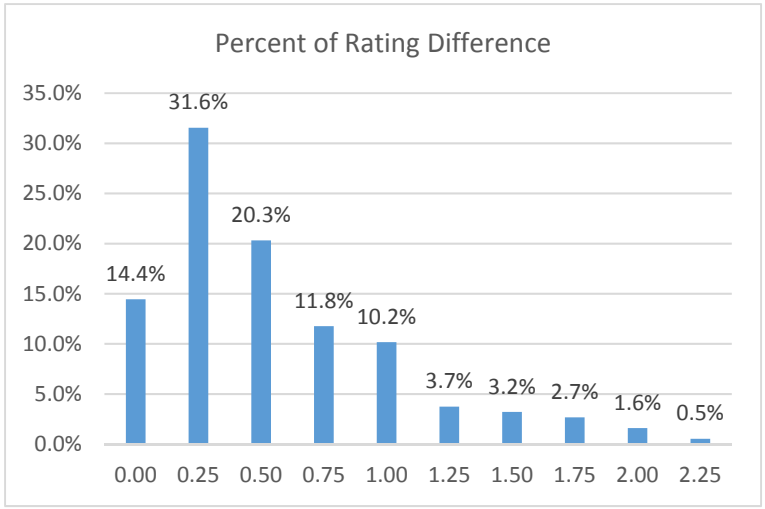


Table 2B. Descriptive Statistics for Rating Differences

Mean	Median	Mode	Standard Deviation
0.56	0.50	0.25	0.48

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 (92.5%) 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 41 artifacts (21.9%) that met or exceeded 1.00 and required a 3rd rating.

For those 41 artifacts requiring a 3rd rating, highlighted in blue in the table above, 19 (10.2%) artifacts had an average rating difference of 1.00, seven had rating differences of 1.25 (3.7%), six had rating differences of 1.50 (3.2%), and five had rating differences of 1.75 (2.7%). Three artifacts had rating differences of 2.00 (1.6%), and one had a rating difference of 2.25 (0.5%).

The 41 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 the 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	0.52	0.55	0.00	2.75
Rating Not Selected	1.22	0.62	0.50	3.00

The mean difference between those ratings that were selected and the 3rd rater's scores was 0.52; the mean difference between those ratings that were not selected and the third rater's scores was 1.22. The minimum difference for those ratings that were selected was 0.00; the maximum difference was 0.50.

Differences by Criteria

The AAC&U Critical Thinking VALUE rubric defines Critical Thinking as "habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion." The rubric contained four criterion: Explanation of Issues, Evidence, Student's Position, and Conclusions and Related Outcomes. The artifacts were rated on these four criterion on a five-point scale that consisted of Capstone (4), Milestones (3 and 2), Benchmark (1), and Not Present (0).

The 187 oral presentations were rated by two different reviewers for a total of 1,496 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

	Explanation of Issues	Evidence	Student's Position	Conclusions and Related Outcomes	Total Difference
Mean	0.85	0.42	0.51	0.52	0.57
SD	0.79	0.54	0.61	0.56	0.65
Min	0.00	0.00	0.00	0.00	0.00
Max	4.00	2.00	3.00	2.00	4.00

The mean difference was greatest for the criteria Explanation of Issues in that the ratings on average differed by 0.85, the minimum rating was 0.00 (or no difference) and the maximum was 4.00. The mean difference was the least for the criteria Evidence with an average difference in rating of only 0.42, a minimum of 0.00 (or no difference) and a maximum of 2.00.