

Analysis of Artifacts from Marshall's Senior Capstone Courses

Academic Year 2017 – 2018

Summer Assessment Workgroup Members: Marie Archambault, Cam Brammer, Kim DeTardo-Bora, Victor Fet, Marty Laubach, Joan St. Germain, Anita Walz, Mary Welch

Summer Assessment Support Staff: Mary Beth Reynolds and Tim Melvin (Office of Assessment), Doug Nichols (Academic Affairs Technical Support)

Executive Summary

Background

In June 2017 the Assessment Workgroup conducted a pilot assessment in which they scored a small sample of capstone project artifacts using the American Association of Colleges and Universities' (AAC&U's) *Critical Thinking* and *Written Communication* Value rubrics. Given the difficulty we have experienced over the years in drawing representative samples of seniors to complete either the *Collegiate Learning Assessment (CLA+)* or Marshall's Senior Assessment, we recommended that staff from the Assessment Office encourage degree programs to use the Blackboard Assignment Module to align their senior capstone assignments with the AAC&U's *Critical Thinking* and *Written Communication* Value rubrics. We recommended that these discussions be incorporated into larger discussions regarding the process of creating assignments in Blackboard and aligning them to appropriate outcomes of Marshall's Baccalaureate Degree Profile (BDP), which we discussed in greater detail in the Blackboard Outcomes Assessment Report. We felt that this has the potential to allow us to evaluate a truly random sample of artifacts from multiple degree programs and apply validated rubrics to assess work that students complete as part of their degree programs. Staff from the Office of Assessment and the Online Design Center met with chairs in all colleges except the College of Information Technology and Engineering during academic year 2017-2018 to ask that they encourage capstone instructors to align capstone projects to Marshall's Capstone Critical Thinking Outcome in Blackboard. Two hundred five (205) artifacts from five academic colleges (Arts and Media, Business, Health Professions, Liberal Arts, and Science) were submitted. The Assessment Workgroup evaluated 200 of these artifacts.

Procedures for 2018 Assessment

General Procedures

Eight faculty representing the Colleges of Arts and Media, Business, Liberal Arts, and Science served as the assessment workgroup for this project. We evaluated capstone artifacts using the AAC&U's *Critical Thinking* and *Written Communication* Value rubrics. These rubrics are included in the supporting documentation. Our sample initially consisted of 200 artifacts. However, during scoring we discovered that artifacts from one of the colleges were group projects (four students per group), with each student in the group uploading the group project. This resulted in our inadvertently assigning the same project paper to several pairs of assessors. A perusal of the 36 artifacts uploaded from these projects revealed ten unique papers; of these, five were scored by four pairs of raters, three by three pairs of raters, and two by two pairs of raters. An additional three uploads from these projects were not able to be opened and scored by assessors. After this discovery, we determined that there were ten unique artifacts within the group of 36 artifacts. However, since these artifacts had been uploaded by each student in the group, each individual upload was scored by a pair of assessors. Eliminating the redundant uploads and the two that could not be assessed resulted in the elimination of 26 of our 200 artifacts from the final analysis. Final scores for the ten group artifacts were determined taking the final scores from each pair of reviewers for each artifact and calculating means. After means were calculated, final scores were rounded to the nearest tenth of a point. In other words, a mean score of 3.18 became 3.0, whereas a mean score of 3.36 became 3.5. In a few cases, where mean scores were equidistant between tenths of points (e.g. 2.25) we rounded up (e.g. to 2.5). An additional two files (from individual project uploads) were not able to be opened or otherwise evaluated and an additional five artifacts were audio files, so were not able to be assessed for written communication. This reduced the number of assessable artifacts to 172 (200 minus 26 from redundant group project uploads; minus an additional two from individual projects uploads not able to be opened for assessment) for critical thinking and 167 for written communication. This project was coordinated by the Office of Assessment.

Scoring Procedures

Evaluators assessed each artifact using the following scale:

Scoring Codes	
These codes were given to artifacts that, in the opinion of the evaluator, were aligned with appropriate outcomes/traits and contained enough information to allow assessment.	
0	The artifact did not demonstrate the minimum level of performance expected at Level 1.
1	The artifact demonstrated Level 1 performance.
2	The artifact demonstrated Level 2 performance.
3	The artifact demonstrated Level 3 performance.
4	The artifact demonstrated Level 4 performance.

Please see the supporting information that follows this summary for a detailed explanation of scoring procedures.

General Information about the Sample

Of the 172 artifacts assessed for critical thinking 33 were from the College of Arts and Media, 16 from the Lewis College of Business, 10 from the College of Health Professions, 54 from the College of Liberal Arts, and 59 from the College of Science. Of the 167 artifacts assess for written communication 28 were from the College of Arts and Media, 16 from the Lewis College of Business, 10 from the College of Health Professions, 54 from the College of Liberal Arts, and 59 from the College of Science.

Results and Analysis

One challenge in reporting results of Blackboard assessment is that, although we assessed 172 artifacts for *Critical Thinking* and 167 for *Written Communication*, each was analyzed by outcome trait. The total number of traits across the two outcomes was 10 (5 each for *Critical Thinking* and for *Written Communication*). Therefore, we tagged a total of 975 traits for *Critical Thinking* and 950 for *Written Communication*.

Outcome	Trait (AAC&U rubric)	Total Traits Aligned	Mean Score	Number of Students Scoring 2.5 – 4	Number of Students Scoring 3.5 – 4
Critical Thinking	Explanation of Issues	172	2.75	135 (78%)	47 (27%)
	Evidence	172	2.48	117 (68%)	16 (9%)
	Influence of Context and Assumptions	172	2.24	100 (58%)	28 (16%)
	Student’s Position	172	2.28	104 (60%)	22 (13%)
	Conclusions and Related Outcomes	172	2.45	107 (62%)	31 (18%)
Total for Critical Thinking		860	2.44	563 (66%)	144 (17%)
Written Communication	Context of and Purpose for Writing	167	2.85	141 (84%)	53 (32%)
	Content Development	167	2.67	126 (75%)	42 (25%)
	Genre and Disciplinary Conventions	167	2.85	140 (84%)	57 (34%)
	Sources and Evidence	167	2.74	131 (78%)	46 (28%)

	Control of Syntax and Mechanics	137	2.70	128 (77%)	43 (26%)
Total for Written Communication		835	2.76	666 (80%)	241 (29%)
Totals		1,695			

A series of paired-samples t-tests, using an adjusted alpha level of .005 to control for Type 1 error, showed that, among the traits of *Critical Thinking*, explanation of issues emerged as a strength, being significantly higher than means for all other traits. The traits influence of context and assumptions and student’s position emerged as weaknesses, with both traits being significantly lower than the other three (explanation of issues, evidence, and conclusions and related outcomes). We also note that, while our paired samples t-tests did not show evidence as a significant weakness, only 9% of seniors scored 3.5 or higher on this trait.

A series of paired samples t-tests, again using an adjusted alpha level of .005 to control for Type I error, showed that, among the traits of *Written Communication*, context and purpose of writing and genre and disciplinary conventions emerged as relative strengths, each being significantly higher than two other traits (content development and control of syntax and mechanics).

These results show that, overall, student performance is stronger in *Written Communication* than in *Critical Thinking* (overall, 29% of this senior sample scored between 3.5 and 4.0 on *Written Communication*, whereas only 17% scored at this level on *Critical Thinking*) when using the AAC&U Value rubrics. Please recall that level 4.0 is “capstone performance” on the AAC&U Value rubrics.

Conclusion

Within *Critical Thinking*, explanation of issues emerged as a relative strength, while student’s position and influence of context and assumptions emerged as relative weaknesses. We suggest a further examination of assignment alignment to tease apart student capabilities from assignment expectations.

Within *Written Communication*, context of and purpose of writing and genre and disciplinary conventions emerged as relative strengths, while content development and control of syntax and mechanics were relative weaknesses. We note that student performance on *Written Communication* was stronger than their performance on *Critical Thinking*.

Overall, 29% of students in this sample scored between 3.5 and 4.0 in *Written Communication*, while only 17% scored at this level in *Critical Thinking*. The Assessment Workgroup will meet in the fall to consider the following issues:

1. Method to increase participation among capstone instructors in having seniors upload signature work in Blackboard.
2. Best method to ensure that duplicate artifacts do not appear in sample.
3. Feedback on the appropriateness (or otherwise) of the two AAC&U rubrics we used for assessment.
4. Best practices to assess students' capstone work.

Supporting Documentation



Capstone
Blackboard Artifact Assessment

Academic Year 2017 – 2018

Outcomes Assessed: AAC&U Rubrics

Outcome	Abbreviation	Traits	Abbreviations
Critical Thinking	CT	Explanation of Issues	Issues
		Evidence	Evidence
		Influence of Context and Assumptions	Context/Assumptions
		Student's Position	Position
		Conclusions and Related Outcomes	Conclusions
Written Communication	WC	Context and Purpose of Writing	Context
		Content Development	Content
		Genre and Disciplinary Conventions	Genre
		Sources and Evidence	Evidence
		Control of Syntax and Mechanics	Syntax/Mechanics

Review Procedures

- Each artifact had two independent raters and usable scores on the 0 – 4 scale were determined in the following manner:
 - If raters assigned the same score, that became the score for the artifact.
 - If raters' scores differed by one point or less, e.g. Rater 1 assigned a score of 1 and Rater 2 a score of 2, the final score was the mean, i.e. 1.5.
 - If raters' scores differed by more than one point, e.g. Rater 1 assigned a score of 1 and Rater 2 a score of 3, the raters met to discuss the rationale for their scores to see if they could agree on a score or, at minimum, scores that differed by no more than one point.
 - If raters' scores differed by more than one point and, after discussion, they were not able to resolve the differences, a third rater was assigned to review the artifact.

Third Readers for this Year's Review

We used two rubrics (*AAC&U Critical Thinking* and *Written Communication*) to assess each artifact, resulting in 400 rubrics (two for each of our 200 artifacts). Of these, we had one trait of one rubric (*Critical Thinking: influence of context and assumptions*) that required a third reader. Original ratings had been a “0” from the first reviewer and a “2” from the second reviewer. The third reviewer independently assigned a score of “1” so, in accordance with our practice, “1” became the final score.

Artifacts Excluded from Analysis of Means Due to Issues that prevented assessors from evaluating the artifacts.

Outcome	Total Artifacts	Total Unduplicated Artifacts	Total Unduplicated Artifacts Eliminated Due to Error	Total Artifacts Eliminated due to Misalignment with Rubric	Total Used for Analysis
Critical Thinking	200	174	2	0	172
Written Communication	200	174	2	5	167

Critical Thinking AAC&U Value Rubric

Critical Thinking QAAC&U Value Rubric

Traits	Level 0	Level 1	Level 2	Level 3	Level 4
Explanation of issues	Does not meet Level 1	Issue/problem to be considered critically is stated without clarification or description.	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, described, and clarified so that understanding is not seriously impeded by omissions.	Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.
Evidence <i>Selecting and using information to investigate a point of view or conclusion</i>	Does not meet Level 1	Information is taken from source(s) without any interpretation/evaluation. Viewpoints of experts are taken as fact, without question.	Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis. Viewpoints of experts are taken as mostly fact, with little questioning.	Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis. Viewpoints of experts are subject to questioning.	Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis. Viewpoints of experts are questioned thoroughly.
Influence of context and assumptions	Does not meet Level 1	Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.	Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others' assumptions than one's own (or vice versa).	Identifies own and others' assumptions and several relevant contexts when presenting a position.	Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.
Student's position (perspective, thesis/hypothesis)	Does not meet Level 1	Specific position (perspective, thesis/hypothesis) is stated, but is simplistic and obvious.	Specific position (perspective, thesis/hypothesis) acknowledges different sides of an issue.	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) is imaginative, 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).
Conclusions and related outcomes (implications and consequences)	Does not meet Level 1	Conclusion is inconsistently tied to some of the information discussed; related outcomes (consequences and implications) are oversimplified.	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 logically tied to a range of information, including opposing viewpoints; related outcomes (consequences and implications) are identified clearly.	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.

Written Communication AAC&U Value Rubric

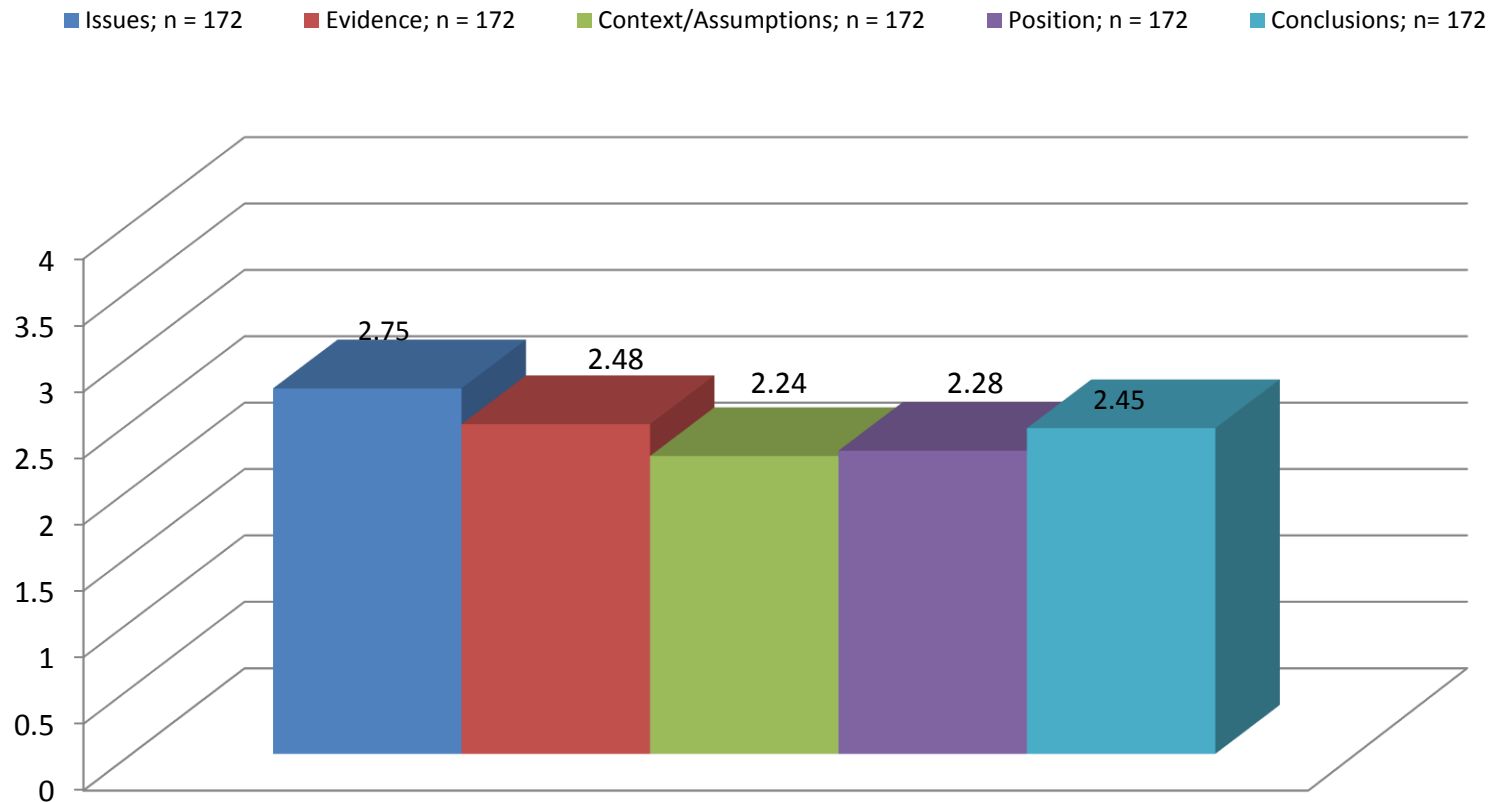
Written Communication AAC&U Value Rubric

Traits	Level 0	Level 1	Level 2	Level 3	Level 4
Context of and Purpose for Writing <i>Includes considerations of audience, purpose, and the circumstances surrounding the writing task(s).</i>	Does not meet Level 1	Demonstrates minimal attention to context, audience, purpose, and to the assigned tasks(s) (e.g., expectation of instructor or self as audience).	Demonstrates awareness of context, audience, purpose, and to the assigned tasks(s) (e.g., begins to show awareness of audience's perceptions and assumptions).	Demonstrates adequate consideration of context, audience, and purpose and a clear focus on the assigned task(s) (e.g., the task aligns with audience, purpose, and context).	Demonstrates a thorough understanding of context, audience, and purpose that is responsive to the assigned task(s) and focuses all elements of the work.
Content Development	Does not meet Level 1	Uses appropriate and relevant content to develop simple ideas in some parts of the work.	Uses appropriate and relevant content to develop and explore ideas through most of the work.	Uses appropriate, relevant, and compelling content to explore ideas within the context of the discipline and shape the whole work.	Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's understanding, and shaping the whole work.
Genre and Disciplinary Conventions <i>Formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields (please see glossary).</i>	Does not meet Level 1	Attempts to use a consistent system for basic organization and presentation.	Follows expectations appropriate to a specific discipline and/or writing task(s) for basic organization, content, and presentation	Demonstrates consistent use of important conventions particular to a specific discipline and/or writing task(s), including organization, content, presentation, and stylistic choices	Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific discipline and/or writing task (s) including organization, content, presentation, formatting, and stylistic choices
Sources and Evidence	Does not meet Level 1	Demonstrates an attempt to use sources to support ideas in the writing.	Demonstrates an attempt to use credible and/or relevant sources to support ideas that are appropriate for the discipline and genre of the writing.	Demonstrates consistent use of credible, relevant sources to support ideas that are situated within the discipline and genre of the writing.	Demonstrates skillful use of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline and genre of the writing
Control of Syntax and Mechanics	Does not meet Level 1	Uses language that sometimes impedes meaning because of errors in usage.	Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.	Uses straightforward language that generally conveys meaning to readers. The language in the portfolio has few errors.	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.

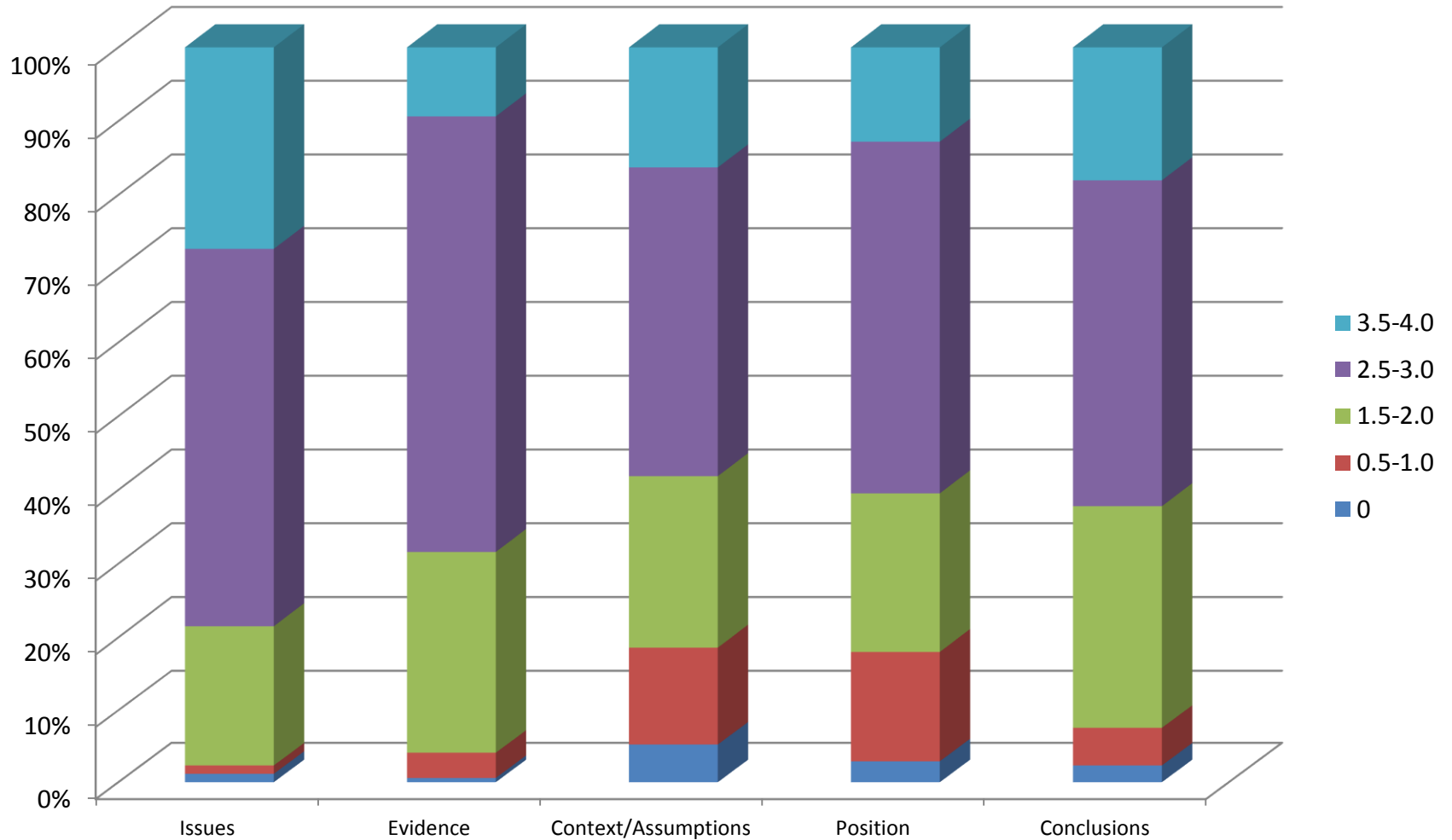
Critical Thinking: Overall Analysis

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.

AAC&U Rubric



Critical Thinking



Critical Thinking

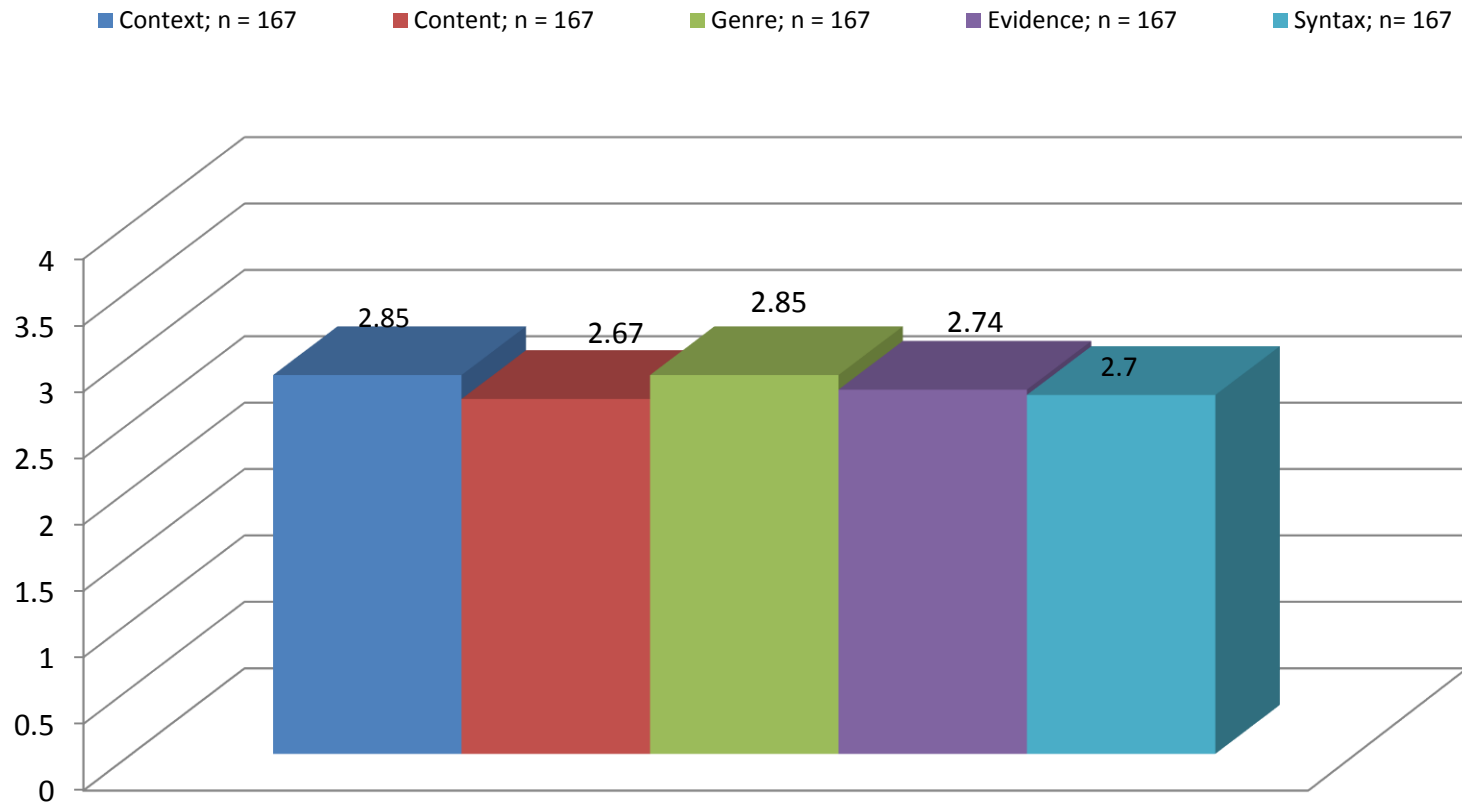
Inter-Rater Agreement Results (Based on all 200 artifacts assessed)

Trait/ Performance Level	Issues; Kappa = .086; Kappa Liberal = .774	Evidence; Kappa = .143; Kappa Liberal = .812	Context/Assumptions; Kappa = .144; Kappa Liberal = .739	Position; Kappa = .106; Kappa Liberal = .679	Conclusions; Kappa = .151; Kappa Liberal = .758
Agree on score	69 (35%)	82 (41%)	66 (33%)	63 (32%)	72 (36%)
Difference = 1 point or less	89 (45%)	83 (42%)	84 (42%)	78 (39%)	82 (41%)
Difference = 1.5 to 2 points	29 (15%)	24 (12%)	20 (10%)	35 (18%)	31 (16%)
Difference > 2 points	3 (2%)	2 (1%)	15 (8%)	14 (7%)	3 (2%)
Agree on Unable to Score due to error	5 (3%)	5 (3%)	5 (3%)	5 (3%)	5 (3%)
Score + Missing Second Rater Score	5 (3%)	4 (2%)	10 (5%)	5 (3%)	7 (4%)
Total	200 (100%)	200 (100%)	200 (100%)	200 (100%)	200 (100%)

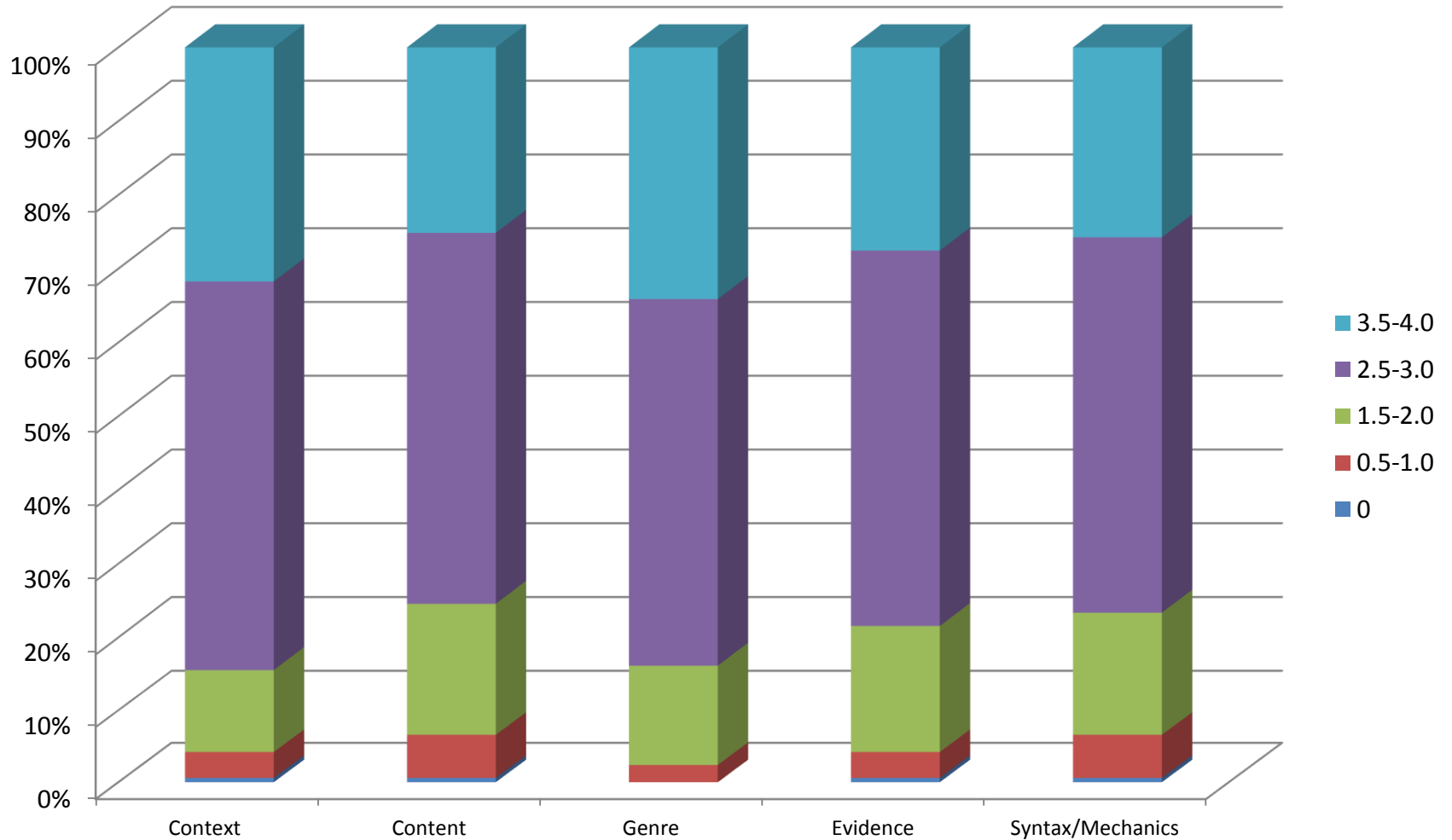
Written Communication: Overall Analysis

Mean Scores on a scale of 0 – 4, with 4 being the highest possible score.

AAC&U Rubric



Written Communication



Written Communication

Inter-Rater Agreement Results (Based on all 200 artifacts assessed)

Trait/ Performance Level	Context; Kappa = .161; Kappa Liberal = .835	Content; Kappa = .193; Kappa Liberal = .804	Genre; Kappa = .067; Kappa Liberal = .775	Evidence; Kappa = .126; Kappa Liberal = .794	Syntax/Mechanics; Kappa = .188; Kappa Liberal = .826
Agree on score	75 (38%)	74 (37%)	55 (28%)	68 (34%)	74 (37%)
Difference = 1 point or less	88 (44%)	83 (42%)	97 (49%)	88 (44%)	86 (43%)
Difference = 1.5 to 2 points	19 (10%)	27 (14%)	31 (16%)	23 (12%)	24 (12%)
Difference > 2 points	3 (2%)	3 (2%)	2 (1%)	3 (2%)	1 (1%)
Agree on Unable to Score due to error	5 (3%)	5 (3%)	5 (3%)	5 (3%)	5 (3%)
Agree on unable to score due to no written artifact	5 (3%)	5 (3%)	5 (3%)	5 (3%)	5 (3%)
Score + Missing Second Rater Score	5 (3%)	3 (2%)	5 (3%)	8 (4%)	4 (2%)
Two missing rates scores	0	0	0	0	1 (1%)
Total	200 (100%)	200 (100%)	200 (100%)	200 (100%)	200 (100%)