



University of Maine

General Education Assessment Report
May 2019

*General Education Committee of the Faculty Senate & Office of Institutional Research
and Assessment*

Table of Contents

Executive Summary	3
Overview	3
Artifact Collection	5
AAC&U VALUE Rubric Design	5
Scoring Session	7
Results	7
Key Findings	12
Recommendations	13
Discussion of Results and Future Plans	13

Executive Summary

At the University of Maine (UMaine), the general education curriculum makes up one third of a student's academic experience. The purpose of general education assessment is to evaluate how well our students are meeting the general education learning objectives and to better understand areas needing improvement. There are five overarching general education areas and nine discrete general education areas:

1. Western cultural tradition
 - a. Social context and institutions
 - b. Cultural diversity and international perspectives
 - c. Population and the environment
 - d. Artistic and creative expression
2. Ethics
3. Quantitative literacy
4. Writing
5. Science foundations

In spring 2019, the Social Context and Institutions block was assessed. Faculty used modified AAC&U VALUE rubrics to score hard copies of student artifacts and submit the scores in Google Forms. A group of 33 participants scored 170 artifacts. The information from this report will be used to inform the general education curriculum and future assessment activities.

Key Findings:

- Inter-rater reliability percentages indicate a low to moderate agreement among raters.
- Results indicate that scorers did not think many of the student artifacts were aligned to the rubrics.
- From the artifacts with evidence of meeting the rubric categories, the majority of scores were at a level 2.

Recommendations:

- The rubric should be reevaluated by the General Education subcommittee of the Academic Affairs Committee of the Faculty Senate.
- More rubric training is needed for faculty and scorers.
- The artifact collection process should be improved to ensure artifacts are more closely aligned with the rubric.

Overview

Introduction

The general education curriculum intends to help develop “broadly educated persons who can appreciate the achievements of civilization, understand the tensions within it, and contribute to resolving them” (UMaine Catalog,

2019).¹ There are many ways for a student to fulfill the requirements of the general education curriculum, and it makes up one third of a student's academic experience. Five broad categories comprise the general education curriculum:

1. Human Values and Social Contexts
 - a. Western Cultural Tradition
 - b. Social Context and Institutions
 - c. Cultural Diversity and International Perspectives
 - d. Population and the Environment
 - e. Artistic and Creative Expression
2. Ethics
3. Quantitative Literacy
4. Writing
5. Science

Additionally, a capstone experience specific to the student's major is required. More details can be found at <https://umaine.edu/facultysenate/committees/general-education-committee/>.

In spring 2019, the Social Context and Institutions block of the general education curriculum was assessed. The Social Context and Institutions block focuses “upon ways in which social contexts shape and limit human institutions (defined broadly to include customs and relationships as well as organizations). The specific focus may be upon ways in which social contacts and institutions interact with human values, the role of institutions in expressing cultural values, or the social and ethical dimensions attendant upon particular academic disciplines.” Students completing the general education area of Social Context and Institutions will be able to:

1. Identify, describe and analyze social contexts and human institutions
2. Recognize and critically evaluate the interaction between social contexts and human institutions

Organization and Reporting

The general education assessment scoring session was organized by the General Educational Committee of the Faculty Senate and the Office of Institutional Research and Assessment. The rubrics used to assess student artifacts were adapted from the AAC&U VALUE rubrics. Faculty were organized into teams based on the general education classes they taught. They designed rubrics aligned to the nine general education areas using criteria from the VALUE rubrics. While only two general education areas have been assessed so far, the rubrics to assess each of the areas have been developed and will be used in future scoring sessions.

A statistically representative sample of artifacts from lower division courses and upper division courses was randomly selected by the Office of Institutional Research and Assessment to be evaluated by faculty reviewers. Individual faculty members whose courses were randomly selected were contacted via email to submit student artifacts that aligned to the general education area being assessed. Ten students were then randomly selected (by the Office of Institutional Research and Assessment) from each group of artifact submissions. Any identifying information of the student or the course was redacted by the Office of Institutional Research and Assessment.

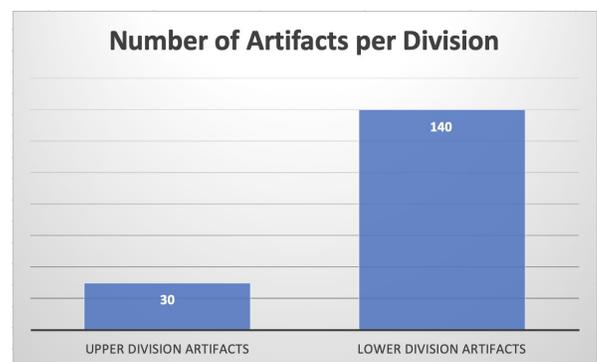
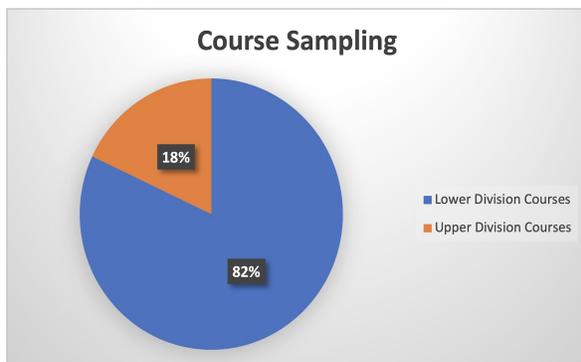
After completing the scoring sessions, the results were collected and analyzed by the Office of Institutional Research and Assessment and reported back to the General Education Committee of the Faculty Senate.

¹ University of Maine Catalog, <http://catalog.umaine.edu/index.php>.

Artifact Collection

Student artifacts were collected from 17 course sections from a stratified random sample from the Registrar's list of course sections. Representative of lower division and upper division proportions of general education courses, 14 lower division courses (82%) and three upper division courses (18%) were randomly sampled. Ten students were randomly selected from each course; however, due to small class size and tardy submissions, less than ten artifacts were collected from five of the courses. In total, 170 artifacts were collected for the scoring session. There were 140 artifacts from lower division courses and 30 from upper division courses.

The artifacts were embedded course assignments collected from courses that met the general education designation as determined by the Undergraduate Program Curriculum Committee (UPCC). All artifacts were redacted so the student and course would be unidentifiable.



Rubric Design

The rubric used for the scoring event was developed by a team of faculty using the AAC&U VALUE rubrics. For the Social Context and Institutions general education block, five rubric categories were assessed. The categories (represented in the rubric below) are as follows:

1. Knowledge of Cultural Worldview Frameworks
2. Interpretation: Making sense with sources as blueprints for meaning
3. Global Self-Awareness
4. Analysis of Knowledge
5. Perspective Taking

Human Values and Social Contexts: Social Contexts and Institutions

Preamble

Courses included in the Social Contexts and Institutions category focus upon the ways in which social contexts shape and limit human institutions (defined broadly to include customs and relationships as well as organizations). The specific focus may be upon ways in which social contexts and institutions interact with human values, the role of institutions in expressing cultural values, or the social and ethical dimensions attendant upon particular academic disciplines.

Student Learning Outcomes

Students completing the general education area of Social Context and Institutions will be able to:

1. Identify, describe and analyze social contexts and human institutions
2. Recognize and critically evaluate the interaction between social contexts and human institutions

Description	Level 4 - Capstone	Level 3 - Milestone	Level 2 - Milestone	Level 1 - Benchmark
Knowledge of cultural worldview frameworks	Demonstrates sophisticated understanding of the complexity of elements important to members of another culture in relation to its history, values, politics, communication styles, economy, or beliefs and practices.	Demonstrates adequate understanding of the complexity of elements important to members of another culture in relation to its history, values, politics, communication styles, economy, or beliefs and practices.	Demonstrates partial understanding of the complexity of elements important to members of another culture in relation to its history, values, politics, communication styles, economy, or beliefs and practices.	Demonstrates surface understanding of the complexity of elements important to members of another culture in relation to its history, values, politics, communication styles, economy, or beliefs and practices.
Interpretation Making sense with sources as blueprints for meaning	Provides evidence that s/he can use an appropriate epistemological lens and that s/he can also engage with sources as part of a continuing dialogue within and beyond a discipline or community.	Articulates an understanding of the multiple ways of engaging with sources and the range of interpretive strategies particular to one's discipline(s) or community	Demonstrates that s/he can engage sources purposefully, choosing among interpretive strategies depending on the context	Can identify purpose(s) for using sources, relying on an external authority such as an instructor for clarification of the task.
Global Self-Awareness	Effectively addresses significant issues in the natural and human world based on articulating one's identity in a global context.	Evaluates the global impact of one's own and others' specific local actions on the natural and human world.	Analyzes ways that human actions influence the natural and human world.	Identifies some connections between an individual's personal decision-making and certain local and global issues.
Analysis of Knowledge	Connects and extends knowledge (facts, theories, etc.) from one's own academic study/field/discipline to civic engagement and to one's own role in civic life, politics, and government.	Analyzes knowledge (facts, theories, etc.) from one's own academic study/field/discipline making relevant connections to civic engagement and to one's own role in civic life, politics, and government.	Begins to connect knowledge (facts, theories, etc.) from one's own academic study/field/discipline to civic engagement and to one's own role in civic life, politics, and government.	Begins to identify knowledge (facts, theories, etc.) from one's own academic study/field/discipline that is relevant to civic engagement and to one's own role in civic life, politics, and government.
Perspective Taking	Evaluates and applies diverse perspectives to complex subjects within natural and human systems in the face of multiple and even conflicting positions (i.e. cultural, disciplinary, and ethical.)	Synthesizes other perspectives (such as cultural, disciplinary, and ethical) when investigating subjects within natural and human systems.	Identifies and explains multiple perspectives (such as cultural, disciplinary, and ethical) when exploring subjects within natural and human systems.	Identifies multiple perspectives while maintaining a value preference for own positioning (such as cultural, disciplinary, and ethical).

Scoring Session

The scoring session took place on May 17, 2019. Thirty-three UMaine faculty and staff gathered for the half-day scoring event. Participants were asked to independently view a training video and score one calibration artifact prior to the event. After a few introductory remarks on the day of the event, raters participated in a discussion about the independent calibration session. Following the discussion, all participants scored one additional artifact, interrater reliability was calculated, and follow-up training was offered.

The rubric used during the scoring event was comprised of four levels: 4 (capstone), 3 (milestone), 2 (milestone), and 1 (benchmark). If an artifact did not meet a level 1 on a rubric category, the scorers were instructed to indicate a score of “0.” However, if the artifact as a whole did not seem to address a particular rubric category, then the scorers were instructed to indicate “NA.” For example, a level 1 in the rubric category of Global Self-Awareness calls for evidence that the student “*identifies some connections between an individual’s personal decision-making and certain local and global issues.*” If the paper clearly contained elements of wrestling with global and local issues, but the student did not make connections to personal decision making, the artifact would be given a score of “0” for that rubric category. However, if there was no evidence that any part of the artifact was expected to address global or local issues, the artifact would be given a score of “NA.”

Once the group was appropriately calibrated, participants began scoring student artifacts. Each participant had an individual folder with 10-11 artifacts to score. All scores were collected via Google Forms. Most artifacts were scored twice. After the live scoring session, the scorers participated in facilitated discussions using the questions below. A comprehensive summary of the discussion feedback can be requested through the Office of Institutional Research and Assessment.

Table Questions:

1. When our general education system started in 1996 we had 550 courses. Now we have 924. Some of our regional peer institutions have around 450.
 - a. What are your thoughts on the number of gen ed classes at UMaine?
 - b. Is this system flexible or random?
2. Give us your thoughts on this scoring process.
 - a. How could it be improved?
 - b. Would you prefer the scoring session be online (but still have a forum where we discuss results)?
 - c. How do we attract participants without a stipend?
3. Faculty participation is an important aspect of this process.
 - a. Do you have suggestions on how to get more faculty involved?
 - b. How do you generate interest in incorporating more best practices into gen ed classes?

Results

Our general education assessment scoring process is modeled after AAC&U’s VALUE Institute’s rubrics and methodology. While rubrics more commonly ascend upward in level, the VALUE rubrics descend from level 4 to level 1. The purpose of this is to emphasize an “assets-based---versus deficit-focused---approach to assessment of

student learning.”² Scorers are trained to assume students are capable of attaining the highest level and working down from there.

After the scoring event was completed, the data were downloaded from the Google Forms spreadsheet and cleaned for analysis. The results were analyzed by rubric category and student level. Although the rubric categories are numeric (4,3,2,1), the data are categorical and do not reflect a true scale. For example, the distance between a level 1 and a level 2 may not be equal to the distance between a level 3 and a level 4. The data are therefore presented as percentages of scores per rubric category. For the purposes of analyzing these data, all NAs were changed to zeroes. To determine a final score for each rubric category among artifacts that had two scorers, two general rules were applied (as adopted from AAC&U’s method). If an artifact was given a score of 0 by at least one scorer, the 0 was retained and the other score was deleted. Secondly, if an artifact was given a score of two different whole numbers (4,3,2,1), the score was averaged and rounded up to the nearest whole number. For example, if an artifact was given scores of 3 and 4 on a rubric category, the final score for that category would be a 4.

Inter-rater Reliability

Several measures of inter-rater reliability were calculated, a process modeled after that used by AAC&U to prepare their yearly VALUE Reports (Drezek McConnell & Rhodes, 2017). Each artifact was scored by two raters. For the purposes of determining inter-rater reliability we have treated all NA’s as 0 and consider a 0 as meaning that there wasn’t enough evidence of either alignment or student ability. In the table below, exact agreement reports the percentage of artifacts scored exactly the same by the two raters while adjacent agreement reports the percentage of agreement within one rubric level of each other (for example if one rater scored 1 while the other scored 2). From these results we see that exact agreement was generally low: approximately 34% across rubric categories. However, the relationship was notably stronger when we looked at the percentage in relatively close agreement: approximately 74% across rubric categories. The remainder of the statistics in the table adjust for chance agreement in different ways. The weighted Cohen’s Kappa takes into account the degree of disagreement between raters. The weighted Brennan-Prediger Kappa coefficient controls for the number of scoring levels (0-4) and the Gwet’s AC coefficient considers how uniform the rating distribution was for each rubric category. Categories with uniform distributions are harder to rate (less consensus) than ones with more peaked distributions (more consensus). While the Cohen’s Kappa results indicate relatively low agreement the Brennan-Prediger and Gwet’AC results indicate moderate levels of agreement. In general, absolute agreement was rare, but raters tended to be in the same ballpark with their application of the rubric criteria.

	1. Knowledge of Cultural Worldview Frameworks	2. Interpretation: Making sense with sources as blueprints for meaning	3. Global Self-Awareness	4. Analysis of Knowledge	5. Perspective Taking
Exact Agreement (%)	35.88%	38.24%	31.18%	28.82%	35.29%
Adjacent Agreement (%)	72.94%	78.82%	73.53%	73.53%	73.53%
Weighted Cohen's Kappa	0.13	0.22	0.09	0.12	0.1
Weighted Brennan-Prediger	0.48	0.6	0.52	0.49	0.54
Weighted Gwet's AC	0.69	0.73	0.67	0.62	0.71

² Value Institute Report 2017-2018.

Scoring Results

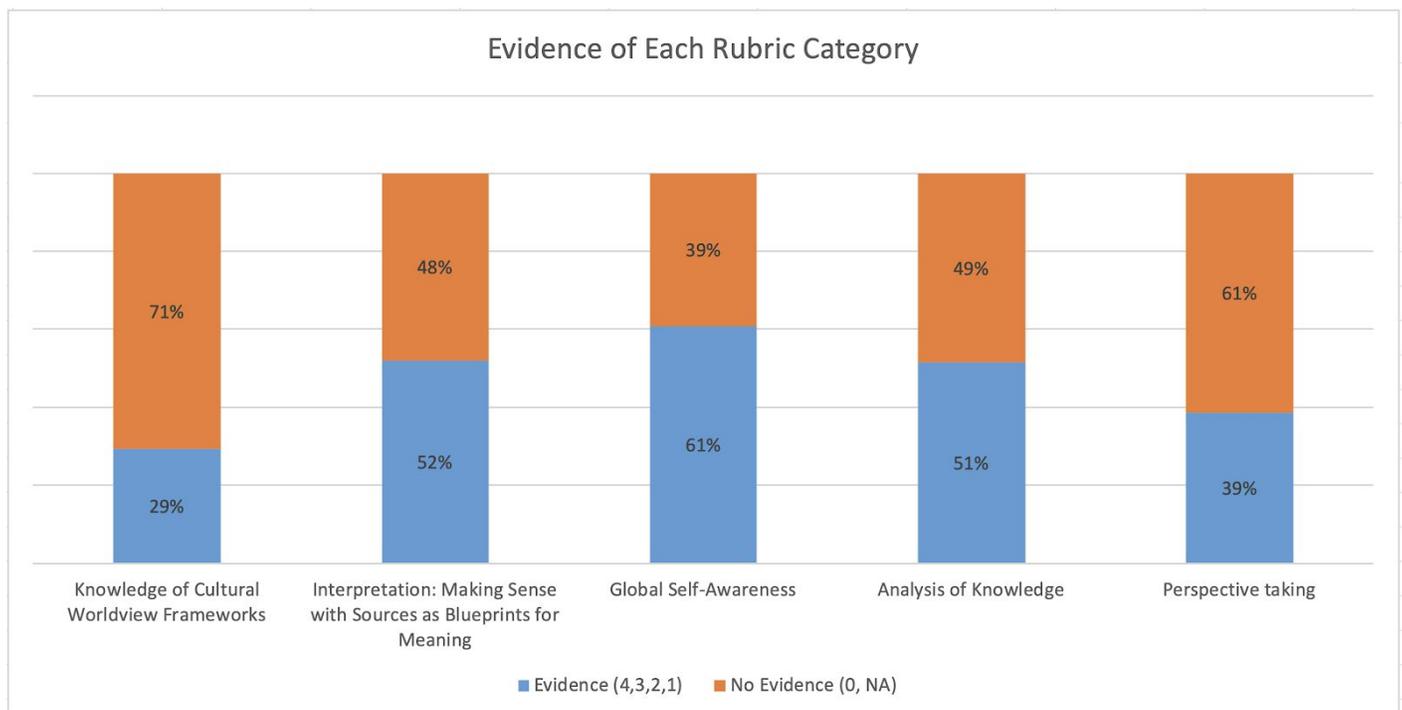
The results of the scoring session were analyzed three different ways: a.) evidence of skills emphasized in rubric vs. no evidence, b.) score distributions per rubric category, and c.) score distributions per student level.

The Social Context and Institutions rubric contains five categories:

1. Knowledge of Cultural Worldview Frameworks
2. Interpretation: Making Sense with Sources as Blueprints for Meaning
3. Global Self-Awareness
4. Analysis of Knowledge
5. Perspective Taking

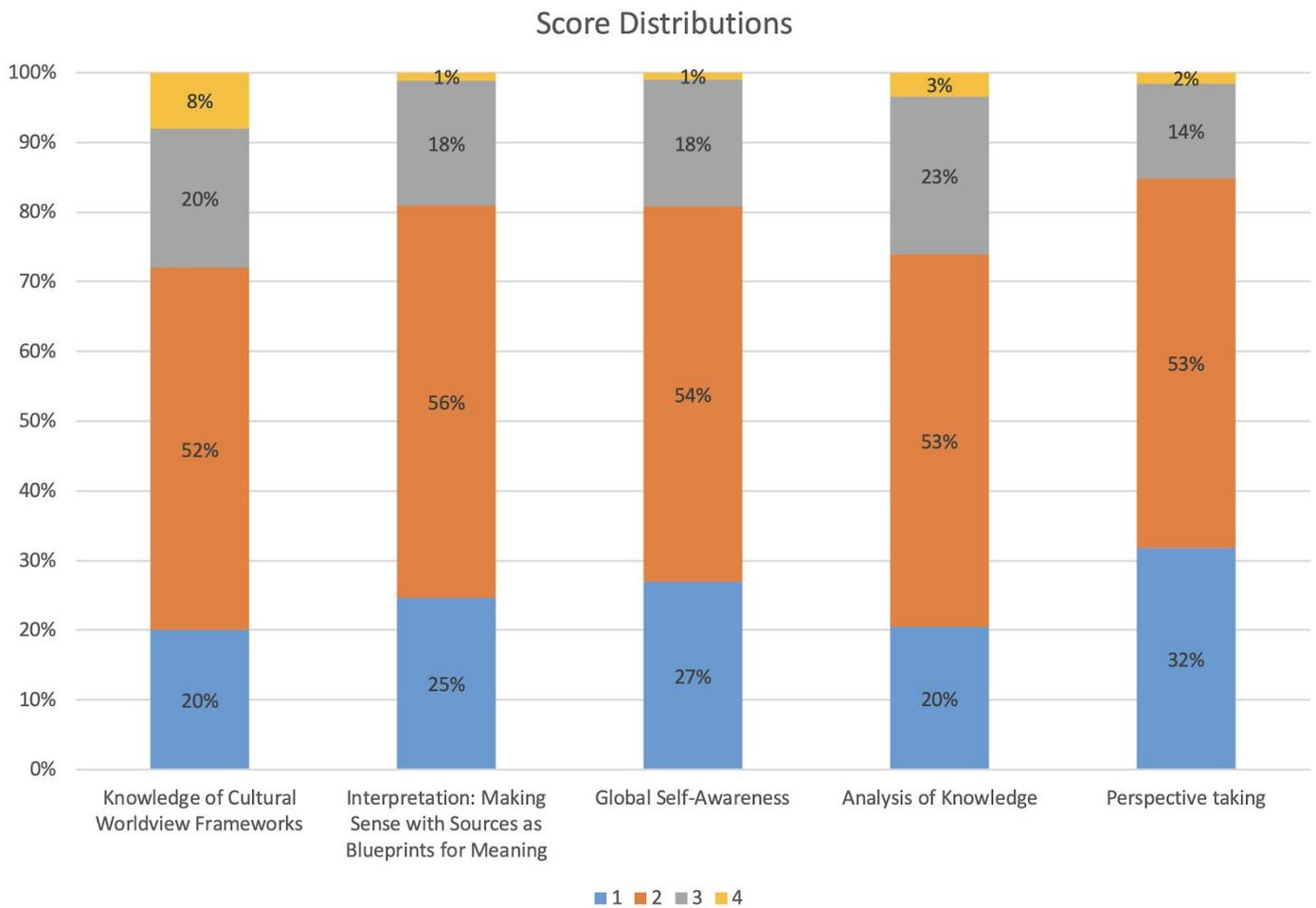
Scorers determined to what extent an artifact met the criteria in each rubric category and gave it a corresponding score (4,3,2,1,0, NA). Many artifacts were given a score of “0” or “NA,” indicating that there was no evidence of reaching at least a benchmark level (1) or there was no evidence of alignment between the artifact and the rubric category. The chart below depicts the percentage of artifacts that displayed evidence of the skills or knowledge required by at least a benchmark level in each rubric category (bottom percentage-blue) versus artifacts that did not show evidence of achieving benchmark or alignment (top percentage-orange). The majority of artifacts in two rubric categories (“Knowledge of Cultural Worldview Frameworks” and “Perspective Taking”) did not show evidence of meeting at least a benchmark level. There also is a sizable percentage of artifacts with no evidence of achieving a benchmark level in the other three categories.

Evidence vs. No evidence



The stacked bar charts below depict the score distribution across rubric categories when all course types and student levels are considered together. Percentages of students scoring at each level (4,3,2,1) per rubric category are shown descending from a level 4 to a level 1. Represented in the chart are the percentages of artifacts showing evidence of achieving at least a benchmark level-- scores of 0 were not included in this analysis. The table below the chart gives a more nuanced description of the counts and percentages for each rubric category based on the total of artifacts with evidence versus those with no evidence. From the artifacts with evidence of meeting the rubric categories, the majority of scores were at a level 2.

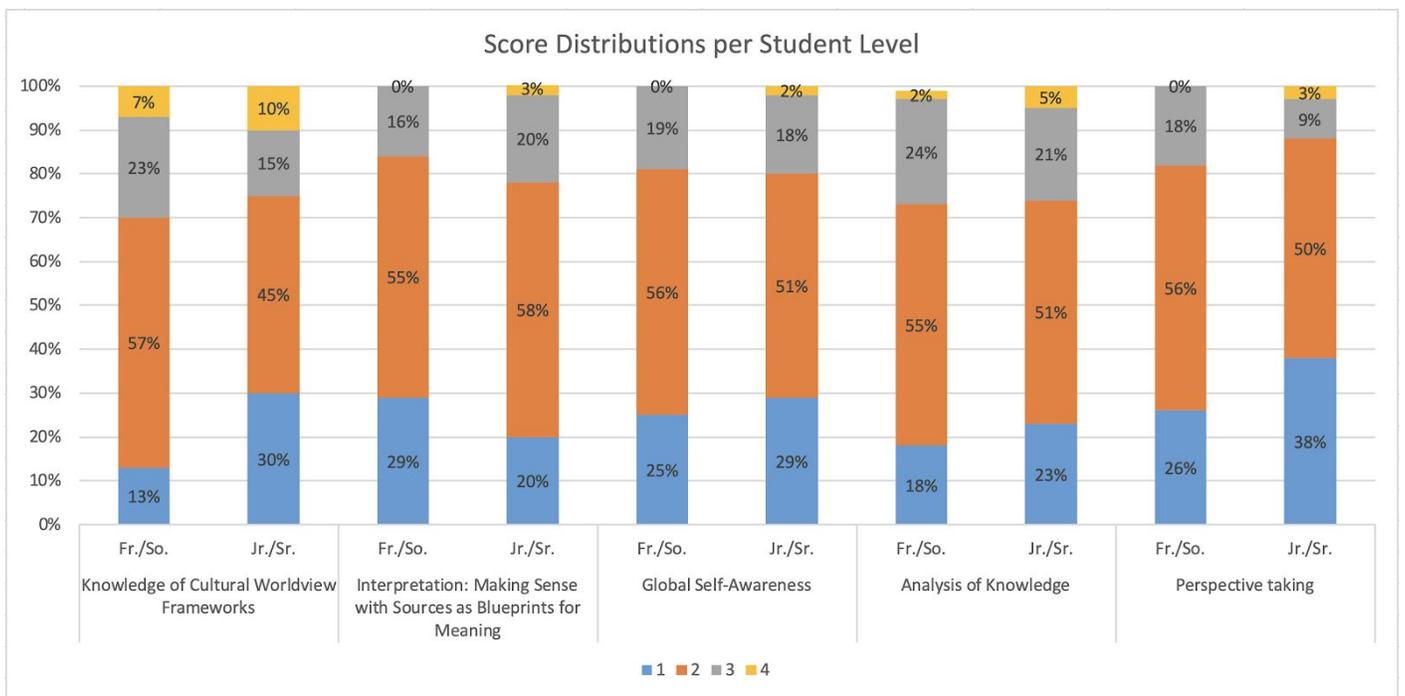
Score Distributions per Rubric Category



	Capstone		Milestones				Benchmark		Total with Evidence (4,3,2,1)		No Evidence (0, NA)	
	4		3		2		1		Count	%	Count	%
	Count	%	Count	%	Count	%	Count	%				
1. Knowledge of Cultural Worldview Frameworks	4	8%	10	20%	26	52%	10	20%	50	29%	121	71%
2. Interpretation: Making Sense with Sources as Blueprints for Meaning	1	1%	16	18%	50	56%	22	25%	89	52%	82	48%
3. Global Self-Awareness	1	1%	19	18%	56	54%	28	27%	104%	61%	67	39%
4. Analysis of Knowledge	3	3%	20	23%	47	53%	18	20%	88%	51%	83	49%
5. Perspective taking	1	2%	9	14%	35	53%	21	32%	66%	39%	105	61%

Score Distributions per Student Level

The following chart and table give a breakdown of scores per rubric category by class standing (freshman & sophomore, junior & senior). Regardless of class standing, the majority of artifacts were scored at a level 2 or below. A more detailed table of the breakdown of percentages per category follows this chart.



	Capstone 4		Milestones 3		2		Benchmark 1		Total with Evidence (4,3,2,1)		No Evidence (0, NA)	
	n	%	n	%	n	%	n	%	n	%	n	%
First-Years and Sophomores												
1. Knowledge of Cultural Worldview Frameworks	2	7%	7	23%	17	57%	4	13%	30	33%	62	67%
2. Interpretation: Making Sense with Sources as Blueprints for Meaning	0	0%	8	16%	27	55%	14	29%	49	53%	43	47%
3. Global Self-Awareness	0	0%	11	19%	33	56%	15	25%	59	64%	33	36%
4. Analysis of Knowledge	1	2%	12	24%	27	55%	9	18%	49	53%	43	47%
5. Perspective taking	0	0%	6	18%	19	56%	9	26%	34	37%	58	63%
Juniors and Seniors												
1. Knowledge of Cultural Worldview Frameworks	2	10%	3	15%	9	45%	6	30%	20	26%	58	74%
2. Interpretation: Making Sense with Sources as Blueprints for Meaning	1	3%	8	20%	23	58%	8	20%	40	51%	38	49%
3. Global Self-Awareness	1	2%	8	18%	23	51%	13	29%	45	58%	33	42%
4. Analysis of Knowledge	2	5%	8	21%	20	51%	9	23%	39	50%	39	50%
5. Perspective taking	1	3%	3	9%	16	50%	12	38%	32	41%	46	59%

Key Findings

1. Inter-rater reliability percentages indicate a low to moderate agreement among raters.

Although calibration sessions took place before the official scoring began, inter-rater reliability remained low to moderate. Consequently, we cannot say there was strong agreement on how artifacts scored against the rubric. Given the low reliability and feedback from the scorers regarding the confusing language in components of the rubric, we conclude that the Social Contexts and Institutions rubric should be evaluated by the Academic Affairs and General Education Committee of the Faculty Senate. Further training on the rubric, whether it is revised or stays in its current form, is necessary.

2. Results indicate that scorers did not think many of the student artifacts were aligned to the rubrics.

The “Evidence of Each Rubric Category” chart shows a startling reality (one which was also echoed during the scoring event): many of the artifacts seemed to have no relation to this set of general education learning outcomes or corresponding rubric. Without decent alignment of the artifacts to the rubric, the scoring session can seem more random than helpful for understanding the general education curriculum.

3. From the artifacts with evidence of meeting the rubric categories, the majority of scores were at a level 2.

When removing all scores of “0” from the analysis, the majority of artifacts were scored at a level 2. Results from the Western Cultural Tradition 2018 scoring session yielded a similar finding.

Recommendations

1. More rubric training is needed.

The Social Contexts and Institutions rubric clearly posed challenges during the scoring event. Discussion before the official scoring began focused on intended meaning of rubric categories and semantics. It is possible this was a particularly difficult general education area and rubric to define. However, in future scoring sessions, more attention will be paid to preparing scorers and reaching a shared agreement on the meaning of the rubric categories and descriptions. Additionally, future calibration sessions will contain “anchor artifacts” that act as examples of a low scoring and high scoring student artifacts.

2. Gather artifacts more closely aligned to the rubric.

To ensure stronger alignment between the submitted assignments and rubrics in future scoring sessions, a small group of individuals will consider the assignment prompt before the artifacts are collected. This will help the organizers of the scoring session detect any poorly aligned assignments and permit time to request another assignment. Additionally, doing this will hopefully have the tangential impact of spreading awareness about the general education rubrics to increase alignment to the general education learning outcomes, as evidenced by course assignments.

Discussion of Results and Future Plans

Assessment data collected from the general education assessment session will be discussed within the Academic Affairs and General Education Committee of the Faculty Senate and then shared internally and publicly as appropriate to encourage discussion about programmatic changes.

The results from this year’s scoring session yielded some surprising findings. We did not anticipate that so many artifacts would be scored “0” or “NA.” We hypothesize this occurred because of two reasons: a.) many of the artifacts were not strongly aligned to the rubric and b.) there was confusion surrounding the rubric and the process of differentiating how and when to use a score of “0” versus “NA.” We plan to alter future scoring sessions and not give an option for scoring “0” or “NA.” Only one option will be given to communicate that an artifact does not meet the benchmark level. Additionally, when analyzing the scores by class standing, juniors and seniors did not score markedly different than first-years and sophomores. It is expected that as students progress in their academic journey, they will leverage their experiences and knowledge into new courses. The general education rubrics are designed to measure skills and habits, not a defined knowledge set. Therefore, one would reasonably expect that artifacts from juniors and seniors would tend to score higher.

It is clear from this report and from the Feedback Report (based on attendees’ feedback) that there is a need for greater awareness about the general education learning outcomes and corresponding rubrics. Plans are underway to disseminate the rubrics more widely, in hopes that they will eventually be used in assignment design or else altered to be more usable.

Assessment of the general education curriculum is ongoing. The next general education assessment session will take place in January 2020, and the general education area “Cultural Diversity and International Perspectives” will be assessed. Further, a team of five individuals who attended the 2019 Institute on General Education and Assessment at the University of Vermont, hosted by the Association of American Colleges and Universities (AAC&U) developed an action plan for fostering a dialogue about the general education curriculum at UMaine. This group represented individuals from the Academic Affairs and General Education Committee of the Faculty Senate as well as Institutional

Research and Assessment. The dialogue regarding the future of general education at UMaine will continue throughout 2019-2020. The outcomes of these conversations, the assessment data collected, and other data to be collected are expected to inform serious consideration of targeted programmatic changes that will improve the curriculum and enhance the student's overall learning experience at UMaine.