OFFICE OF
ACADEMIC AFFAIRS

Office of Assessment & Academic Program
General Education Assessment Results
AY 2019 – 2020

August 2020

Abstract

The Office of Assessment & APR (OA/APR) completed its pilot assessment of the new General Education programming required by the New Mexico Higher Education Department (NMHED). The NMHED established General Education essential skills to be taught and assessed at all New Mexico post-secondary institutions. The University of New Mexico's OA/APR collected completed General Education student artifacts across colleges, schools, and branches in AY 19-20. These General Education artifacts represented four of the five essential skills (the fifth will be assessed next year) and were used to assess information about UNM's new General Education programming. With ~ 790 student artifacts submitted, the OA/APR quantitatively and qualitatively analyzed this robust set of data. In isolation, these ratings are indicative of many variables: (1) student performance, (2) assignment design, (3) alignment of the assessment tool (rubric) and the student work, and (4) rubric dimension selection by each instructor. While student artifacts may have been strong in terms of substantive areas/skills/knowledge required within assignment guidelines, they did not always align with the GE essential skills, which may have resulted in low ratings. Conversely, student artifacts may have been poor in substantive areas/skills/knowledge required within assignment guidelines, but strongly aligned with the GE essential skills, which may have resulted in high ratings. The qualitative analysis brings context, richness, and meaning to the overall GE data results.

INTRODUCTION

Overview

The University of New Mexico's General Education (GE) program is based on new modifications from the NMHED 2019 statewide GE initiative. This initiative includes the adoption of five NMHED GE essential skills, which students develop following the successful completion of the UNM GE Curriculum. Those five skills are;

1. Communication
2. Critical Thinking
3. Personal & Social Responsibility
4. Information & Digital Literacy
5. Quantitative Reasoning

During the statewide GE revision, NMHED aligned each essential skill to a content area. UNM's 3-year General Education assessment cycle allows units to collect, analyze, and report data pertaining to one essential skill per year, mapping to their content area. In this first year, each content area was given a choice between two essential skills to assess that they were aligned to. All essential skills were assessed except for Critical Thinking, which will be assessed by the entire institution in the second year.

Methods

In order to assess the development of these skills, UNM utilized essential skill rubrics developed by the NMHED, and modified them with both AAC&U value rubric language and faculty experts to streamline definitions, promote universal application of skills across disciplines/majors, and ease use.

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

The OA/APR collected equitable and representative samples of student artifact relating to each essential skill. Each college/school/branch was required to select 5-10 sections of GE courses they offer in the designated content area (dependent on the size of the college/school/branch). The instructors of each of the selected course sections provided a minimum of four student artifacts per class and submitted them to the OA/APR.

In preparation for the rating process, OA/APR staff reviewed and normed the rubrics with sample work. Six graduate students were hired by the OA/APR to assist in the analysis of these submitted GE student artifacts. All graduate students underwent FERPA, rubric norming, and GE assessment training. The graduate students were paired and assigned to a rating team where each student rated specific essential skill artifacts individually, and then discussed each rating score with their teammate to increase inter-rater reliability. While rating, the graduate students provided qualitative notes regarding their rating process, the completed artifacts, the instructor assignments, and rubric use. The OA/APR staff met with the graduate student teams weekly (or more) to reconcile any rating challenges and offer assessment coaching as needed.

The OA/APR compiled the ratings and qualitative notes for each essential skill. The quantitative results were visualized in bar graphs (below) and the qualitative narrative was analyzed with Atlas.ti software (also below).

Student Population

When submitting student artifacts for the GE assessment process, the OA/APR requested that instructors associate them with UNM student Banner IDs. Most instructors were able to provide this information. The Office of Institutional Analytics (OIA) assisted the OA/APR in pulling the demographic and academic data of these UNM students. The OA/APR compiled the following information from these associated Banner IDs:

- Total number: **570 students** (some instructors submitted multiple artifacts from the same student)
- Gender: **64% female**, 46% male
- Overall Average GPA: **3.27**
- Student level: **41% Sophomores**, 23% Freshmen, 15% Juniors, 11% other (high school, non- degree seeking), 10% Seniors
- Ethnicity: **48% Hispanic**, 33% White, 6% American Indian, 5% Asian, 2% Two or More races, 2% Non-Res Alien, 2% Race/Ethnicity unknown, 1% Black or Afro American, 0% Native Hawaiian

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

QUANTITATIVE RESULTS

Overall GE Key Takeaways

While reading these overall takeaways, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient. The "skill average" is the overall average rating across all submissions for a particular skill. This is a baseline for this pilot year and should not be considered a benchmark.

Course Level: The **majority** of submitted student artifacts represented **1000** level courses. Across all four essentials skills, artifacts from **2000** level courses **rated higher** on average than those from 1000 level courses.

Course Modality: While artifacts from **face-to-face courses rated higher** across all essential skills, online and hybrid modalities followed closely behind and received similar ratings.

Essential Skills: Across all skills, UNM colleges/schools/branches achieved an artifact rating average of **at least "Emerging."**

Communication: The **majority** of student artifacts **met or were above** skill average at the **"developing"** level. However, artifacts that were written and submitted in a **Second Language** rated lower on average compared to other course code submissions. This Essential Skill had the **most** submissions/student artifacts to be rated, and its highest rated dimension was **Evaluation and Production of Arguments**.

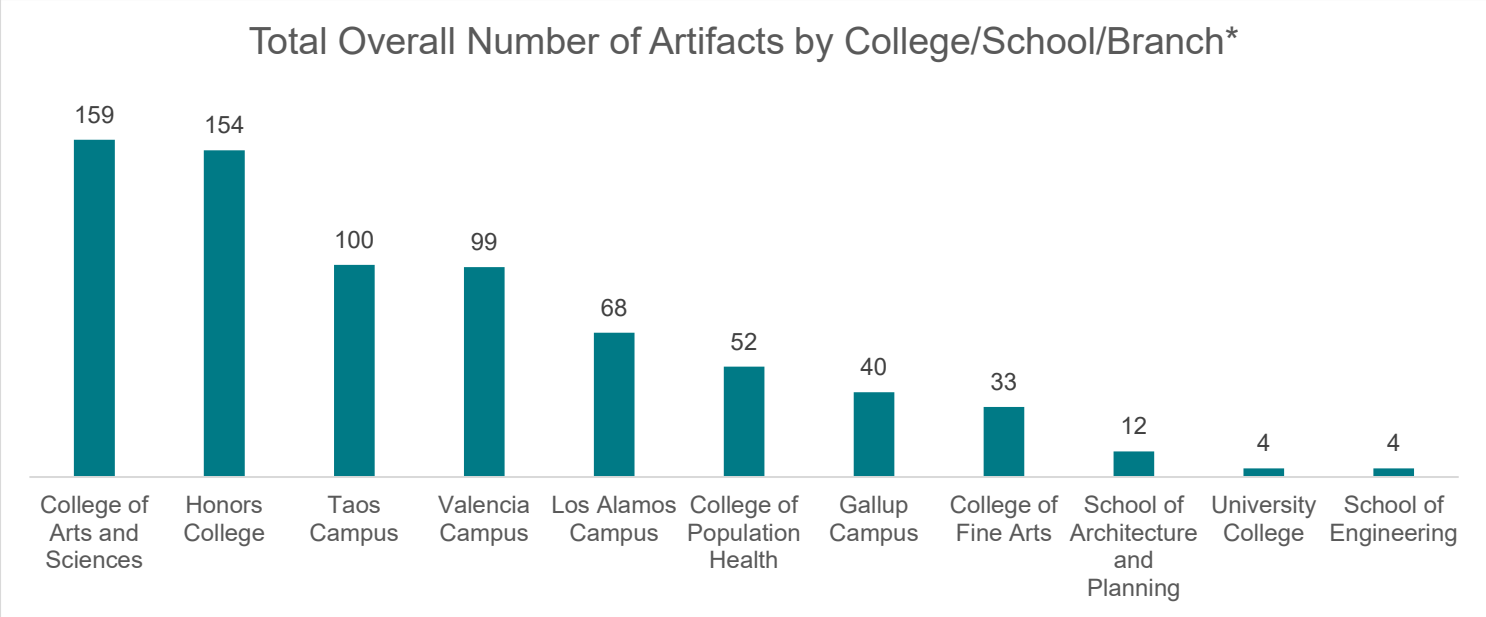
Information & Digital Literacy: The **majority** of student artifacts rated **below** skill average at the "**no evidence" or "emerging"** level across all course codes. This essential skill had the **lowest** overall ratings across all course codes, levels, and modalities. Its highest rated dimension was **Research as Inquiry**.

Personal & Social Responsibility: Approximately **half** of the student artifacts **met** the skill average at the **"emerging"** level. The dimension of "collaboration and teamwork" was shown to be challenging to assess/rate and requires a particular assignment in order to measure this accurately. This skill's highest rated dimension was **Civic Discourse, Civic Knowledge and Engagement**.

Quantitative Reasoning: The **majority** of student artifacts **met** the skill average at the **"developing"** level. Its highest rated dimension was **Communication and/or Representation of Quantitative Information**.

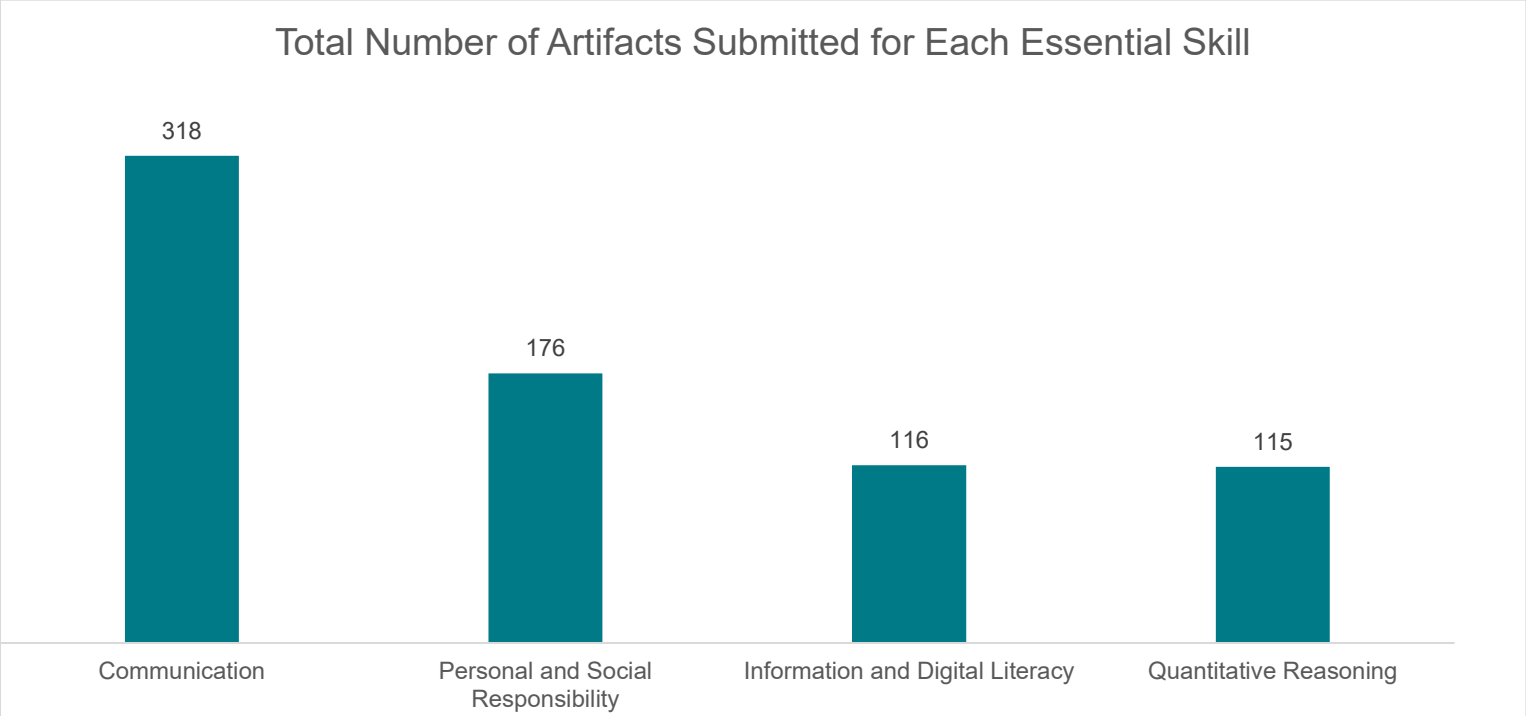
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Total Artifacts Submitted



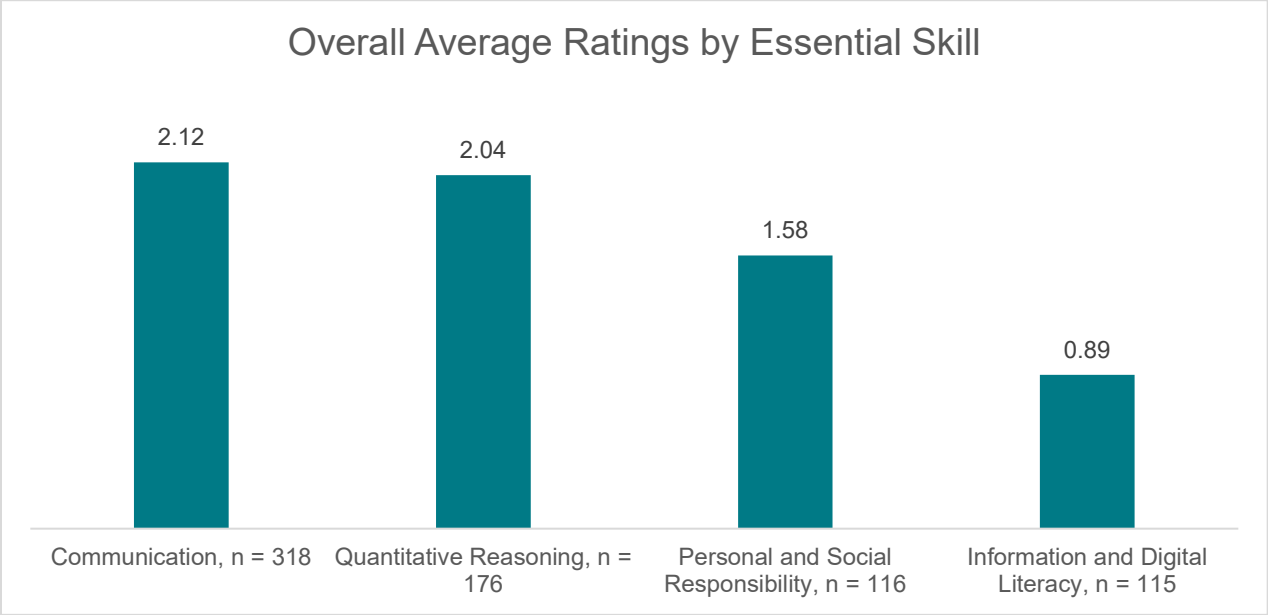
*as of May 27, 2020 (final data pull date)

Total Artifacts by Essential Skill

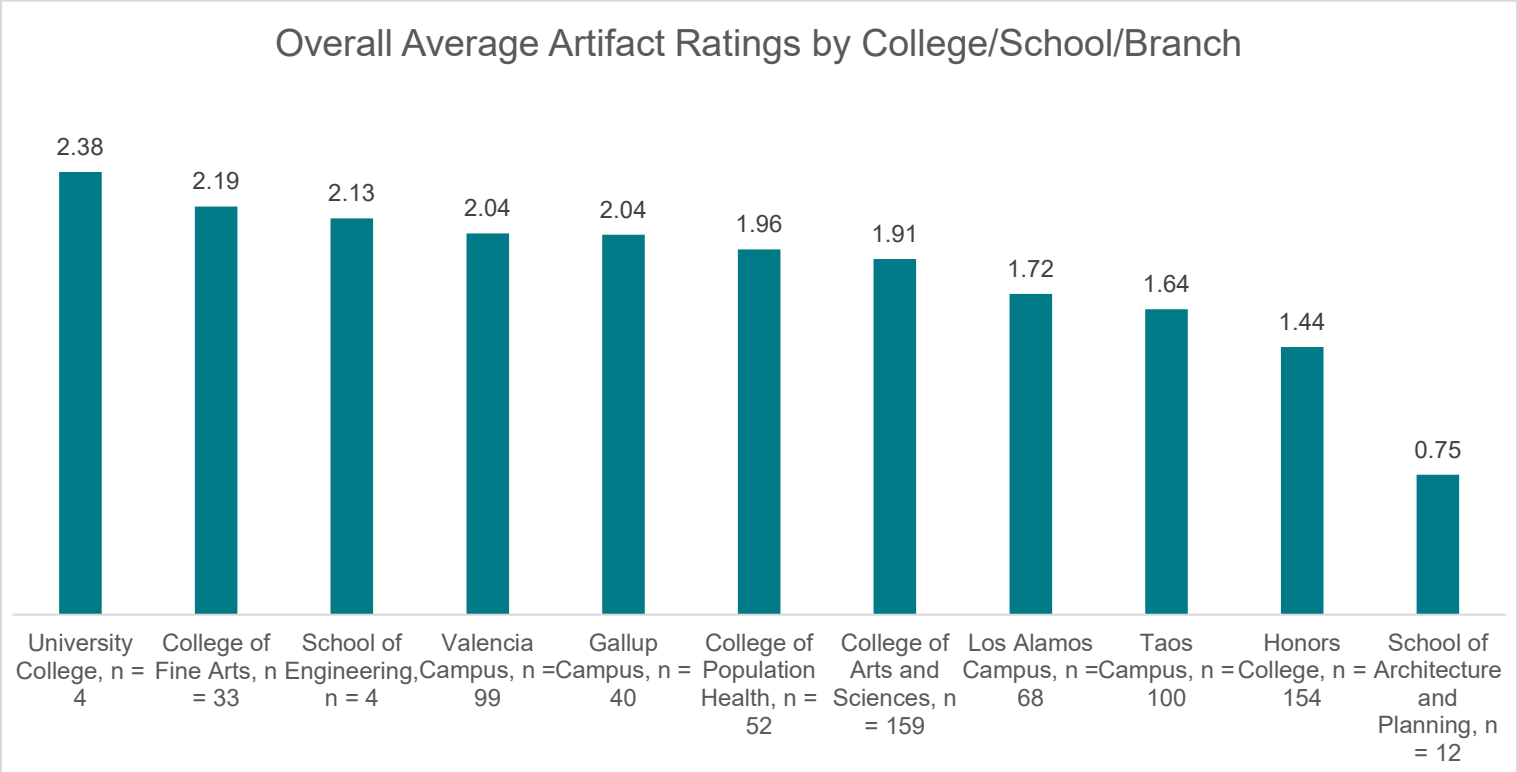


*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Average Artifact Ratings for Each Essential Skill

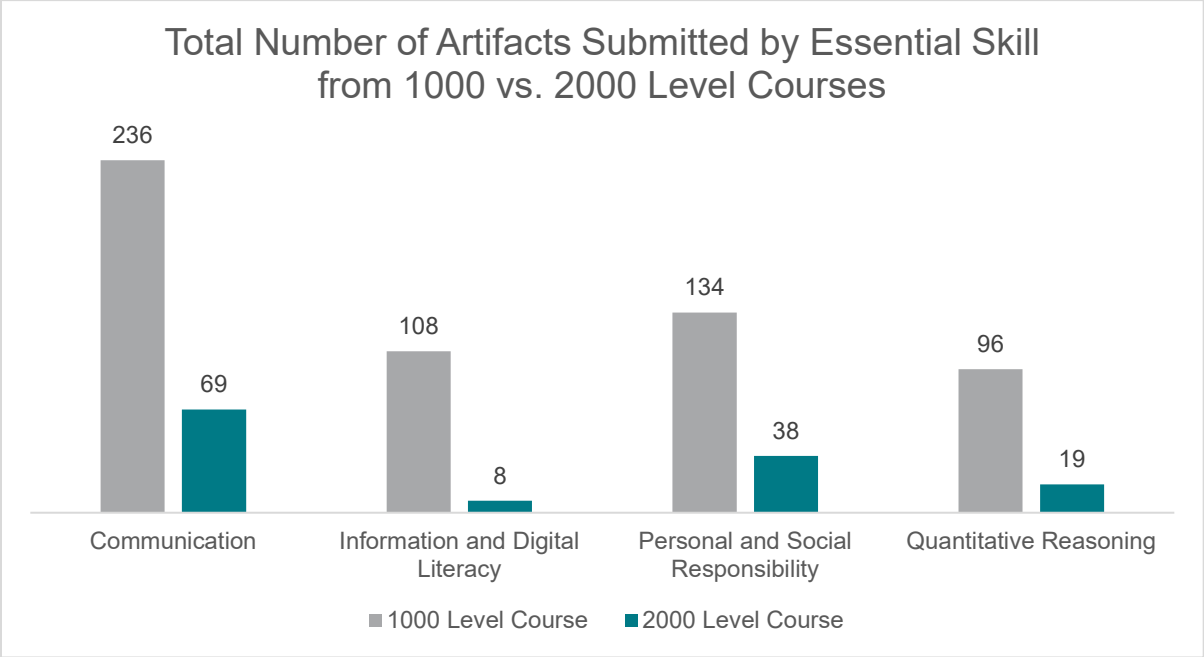


Average Artifact Ratings by College/School/Branch

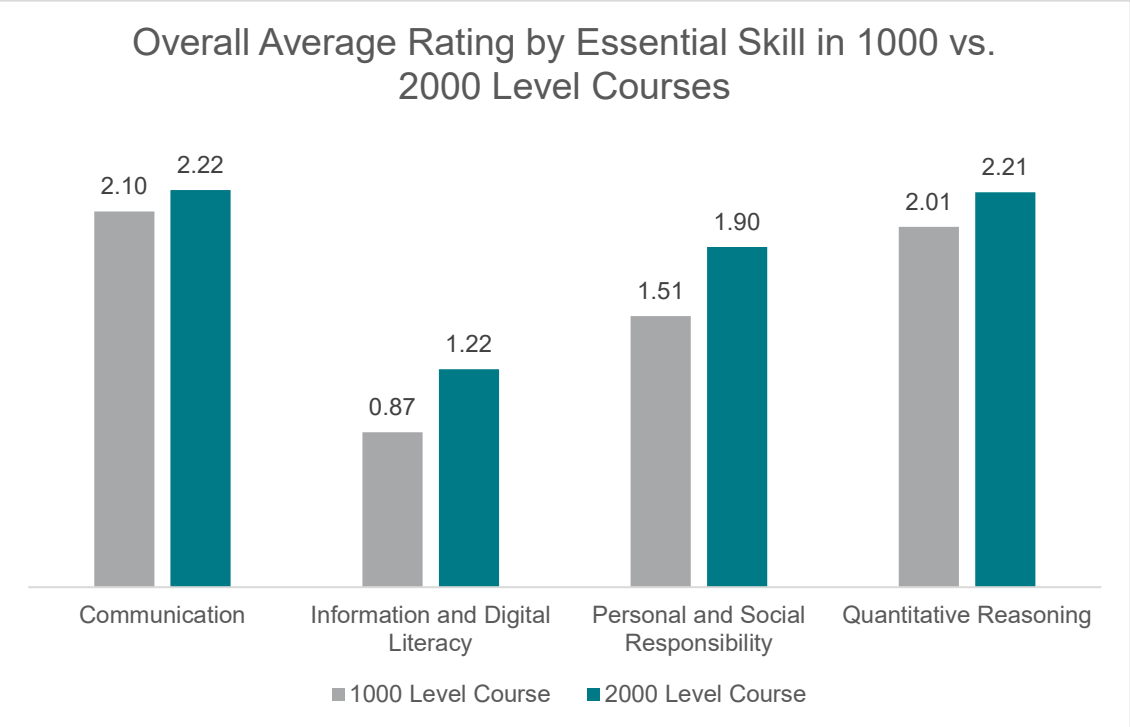


*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Total Number of Artifacts Submitted by Course Level

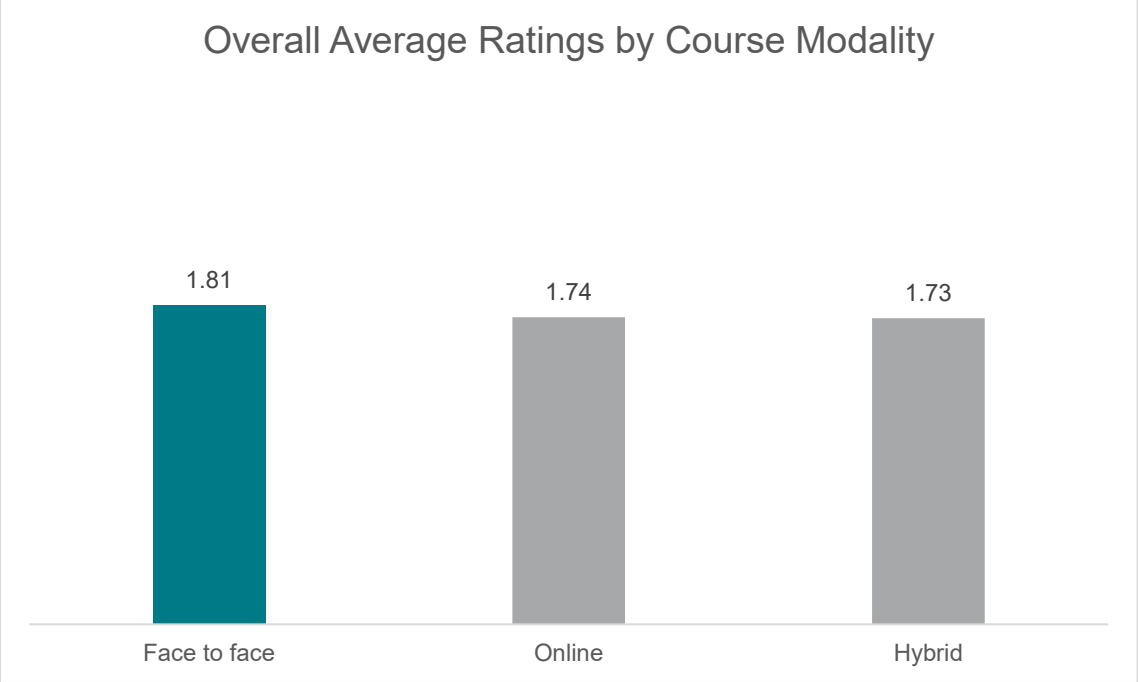


Average Artifact Ratings by Course Level



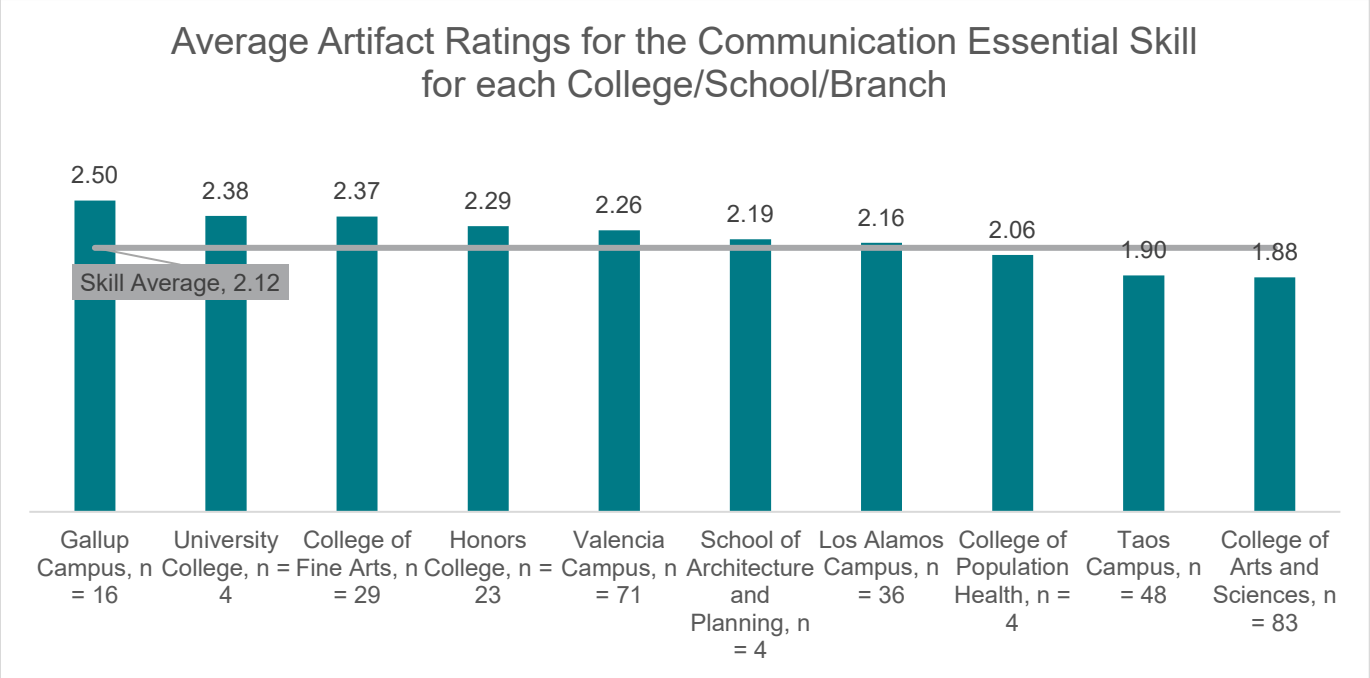
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Average Artifact Ratings per Course Modality



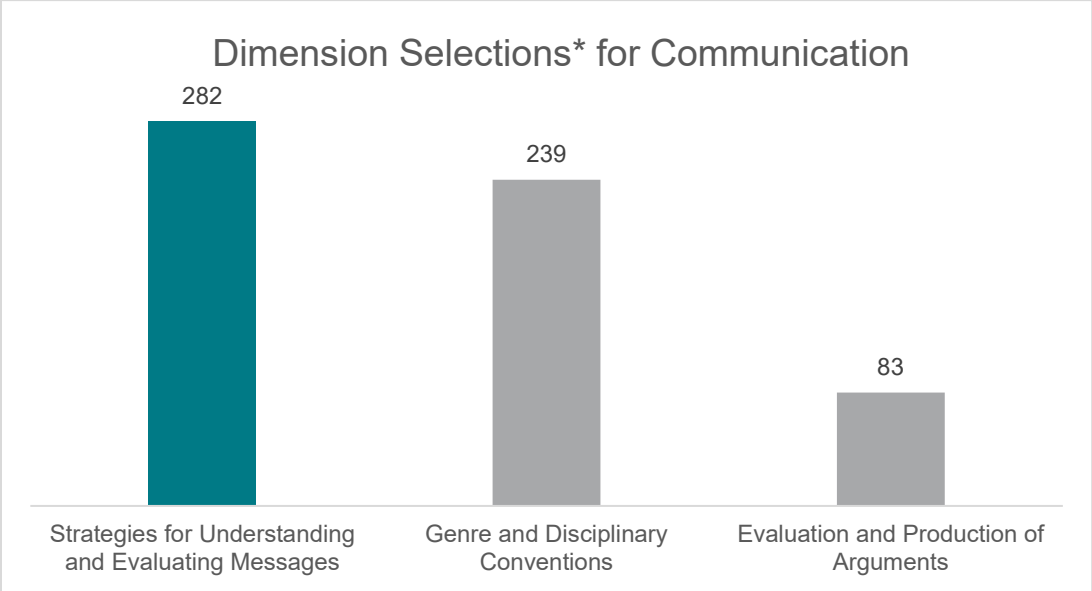
COMMUNICATION ESSENTIAL SKILL

Average Artifact Ratings



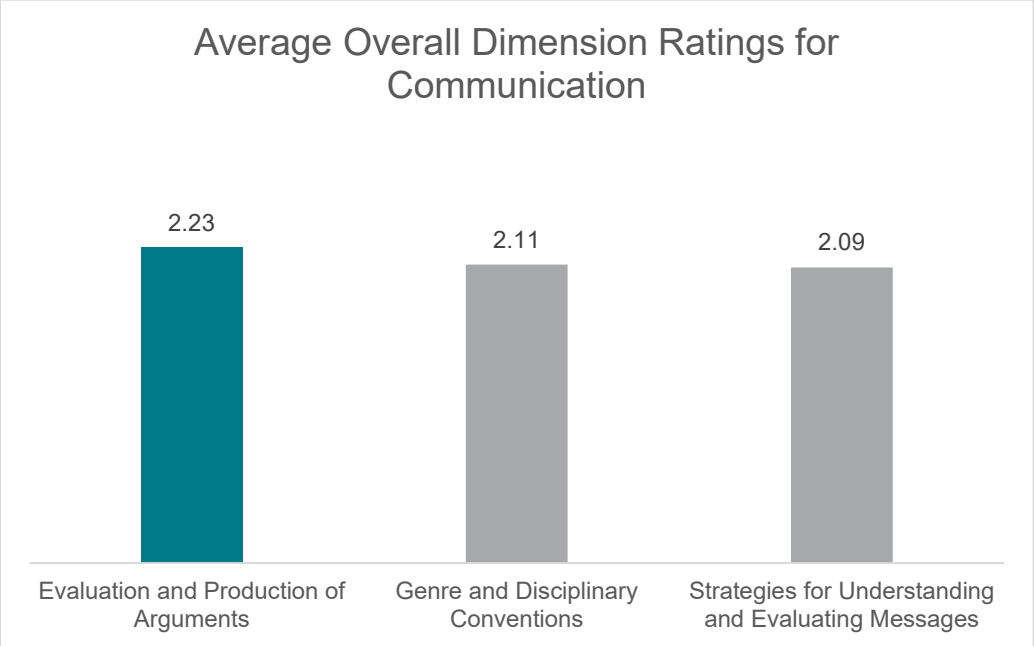
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Total Number of Selections per Dimension



*each artifact was assigned to at least 2 dimensions

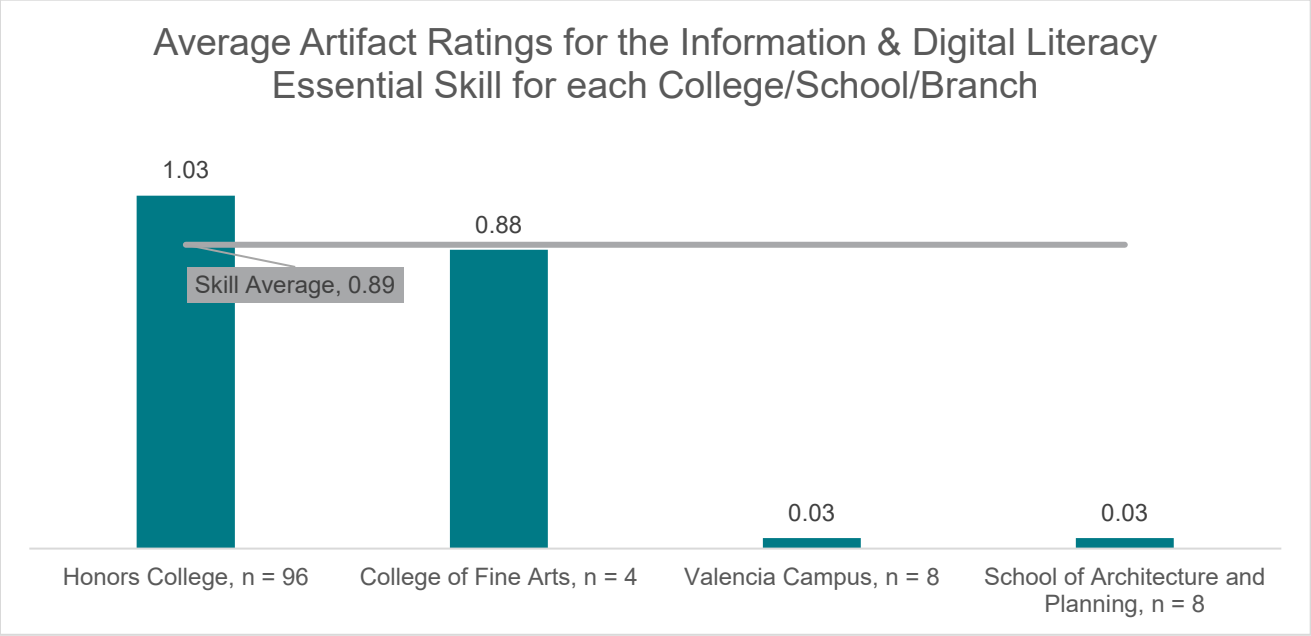
Average Ratings per Each Dimension



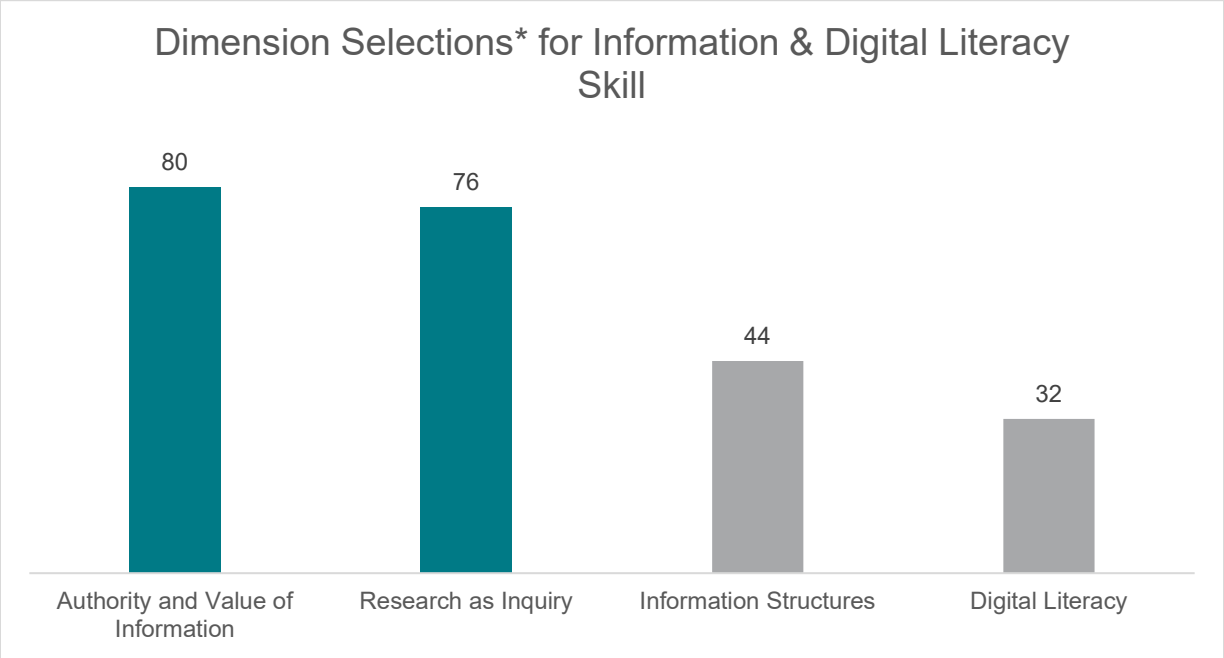
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

INFORMATION & DIGITAL LITERACY ESSENTIAL SKILL

Average Artifact Ratings



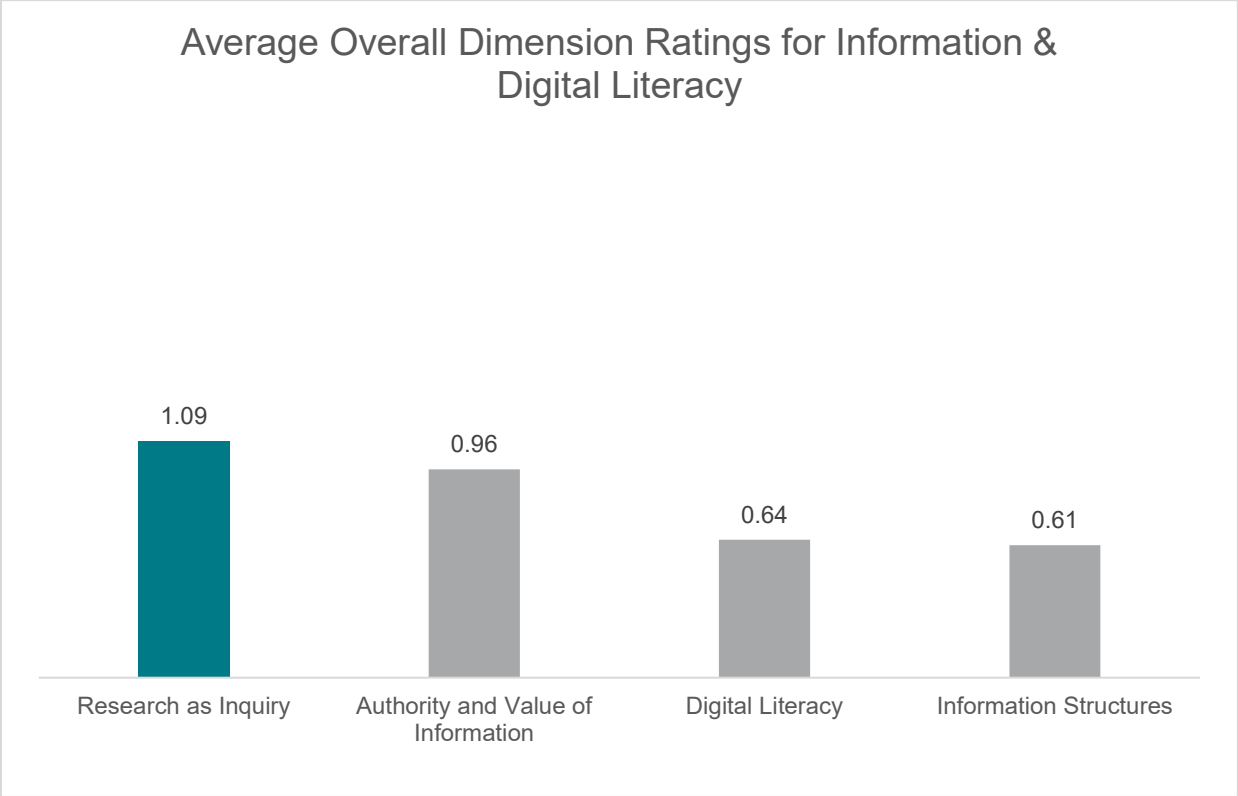
Total Number of Selections per Dimension



*each artifact was assigned to at least 2 dimensions

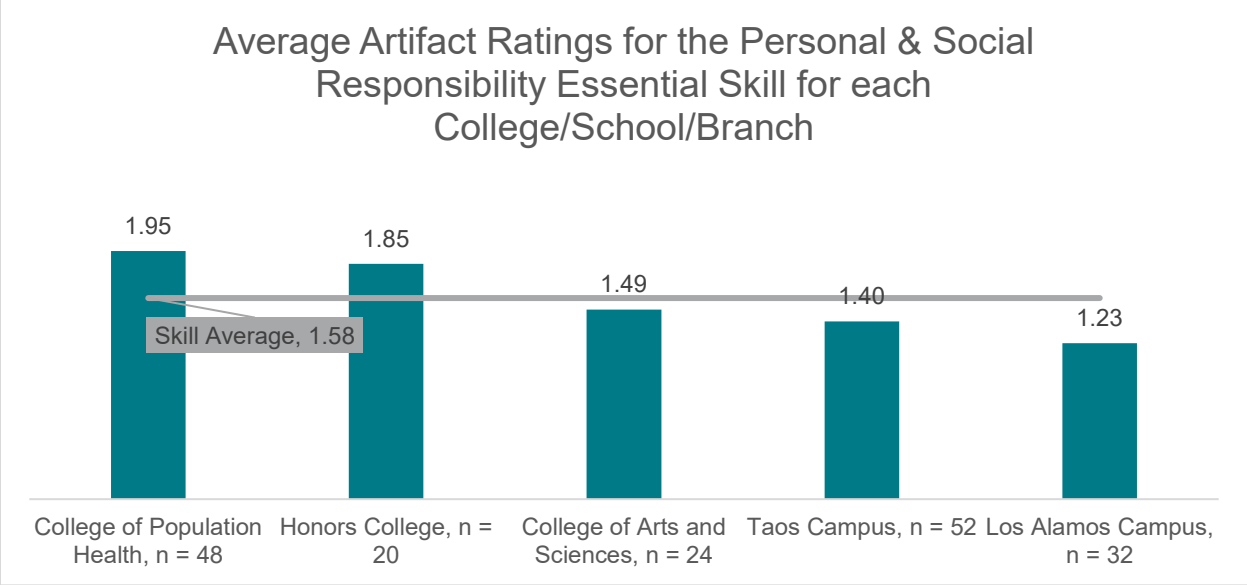
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Average Ratings per Each Dimension



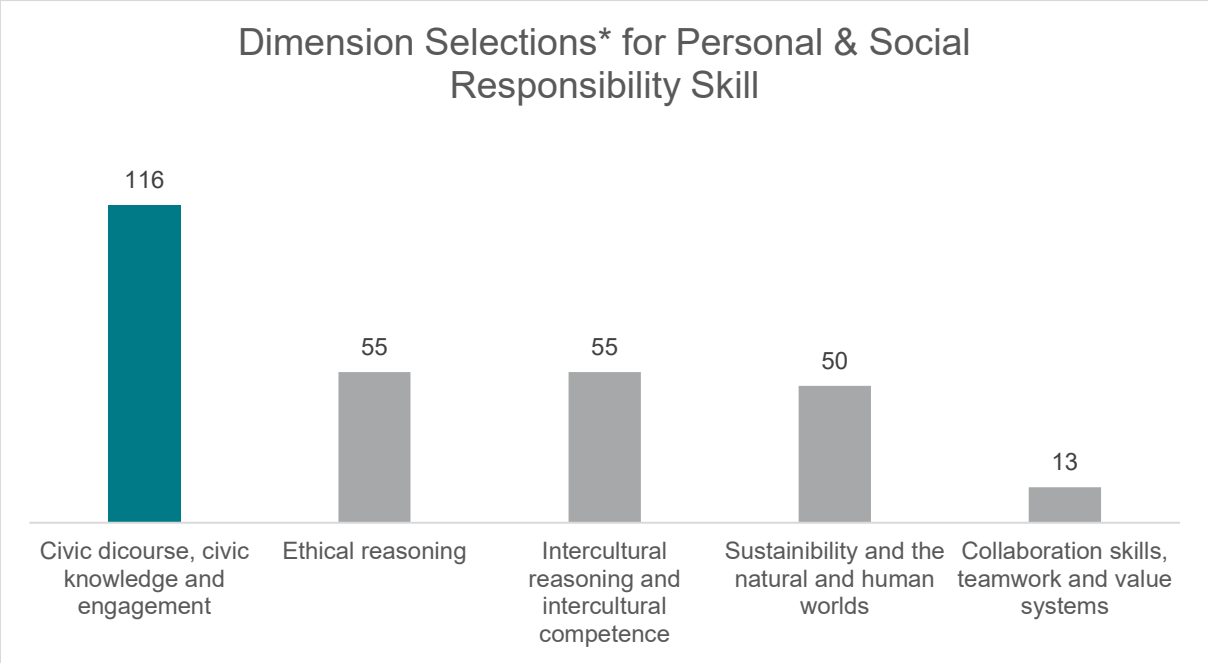
PERSONAL & SOCIAL RESPONSIBILITY ESSENTIAL SKILL

Average Artifact Ratings



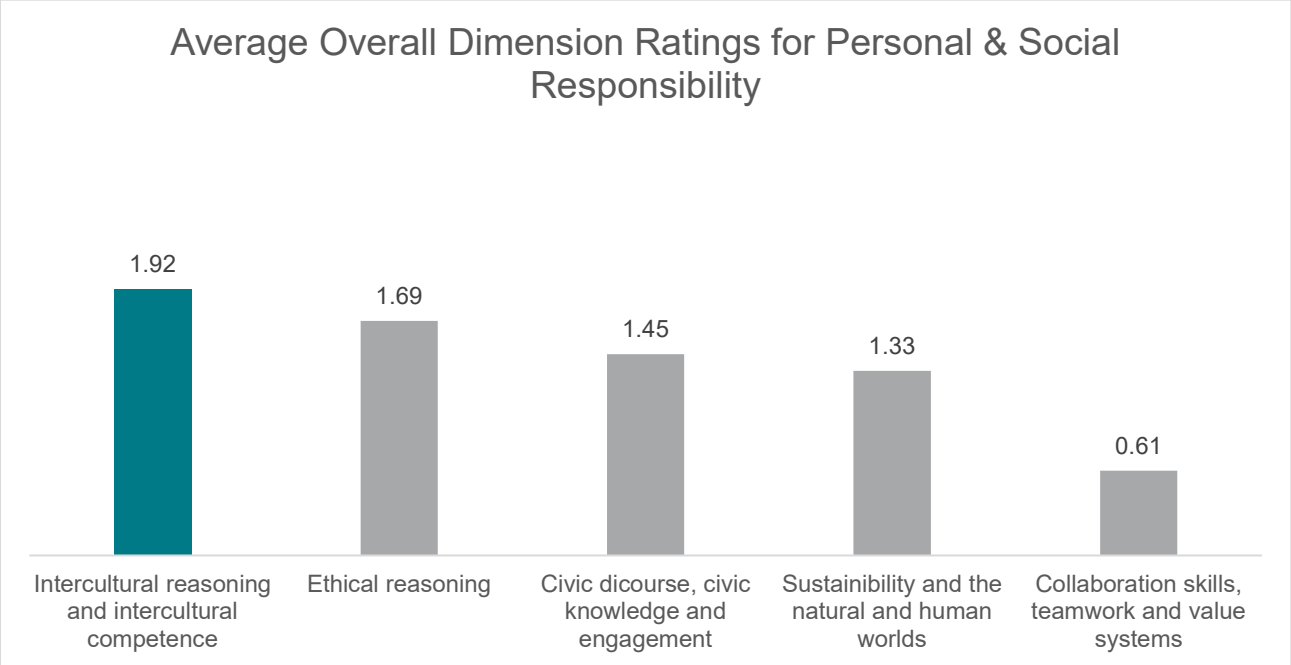
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Total Number of Selections per Dimension



*each artifact was assigned to at least 2 dimensions

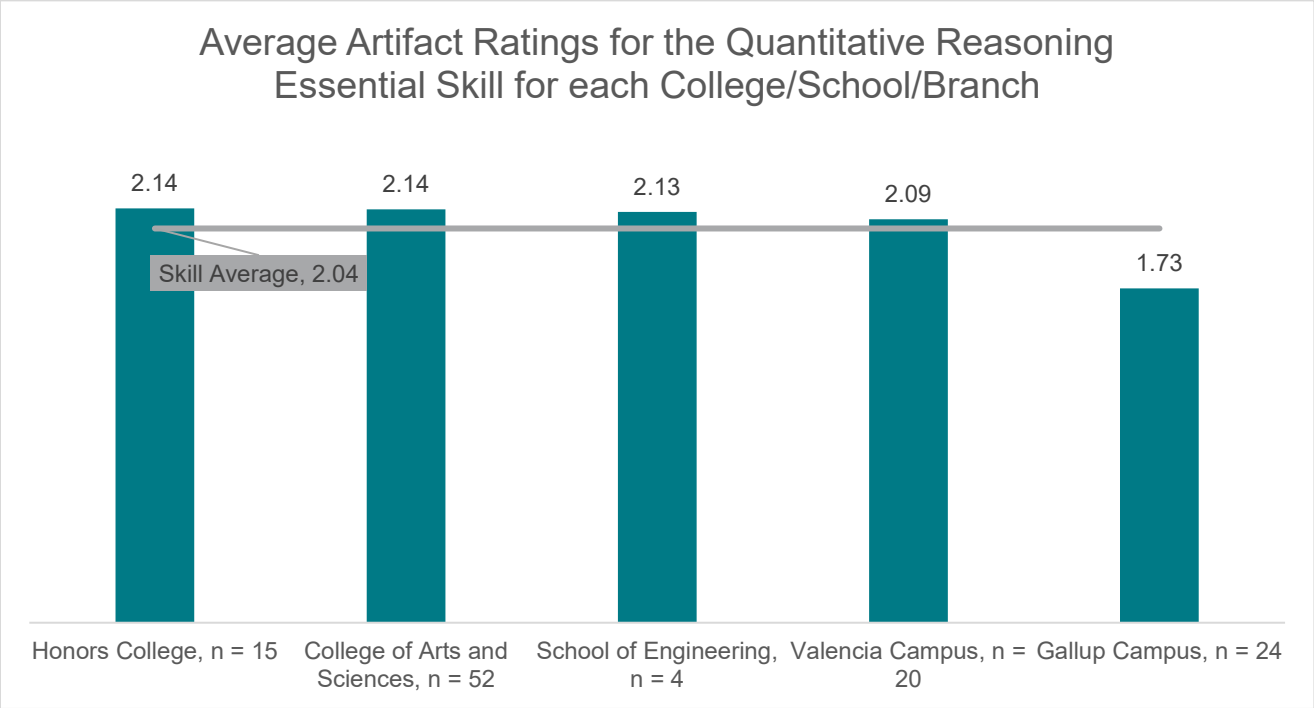
Average Rating per Each Dimension



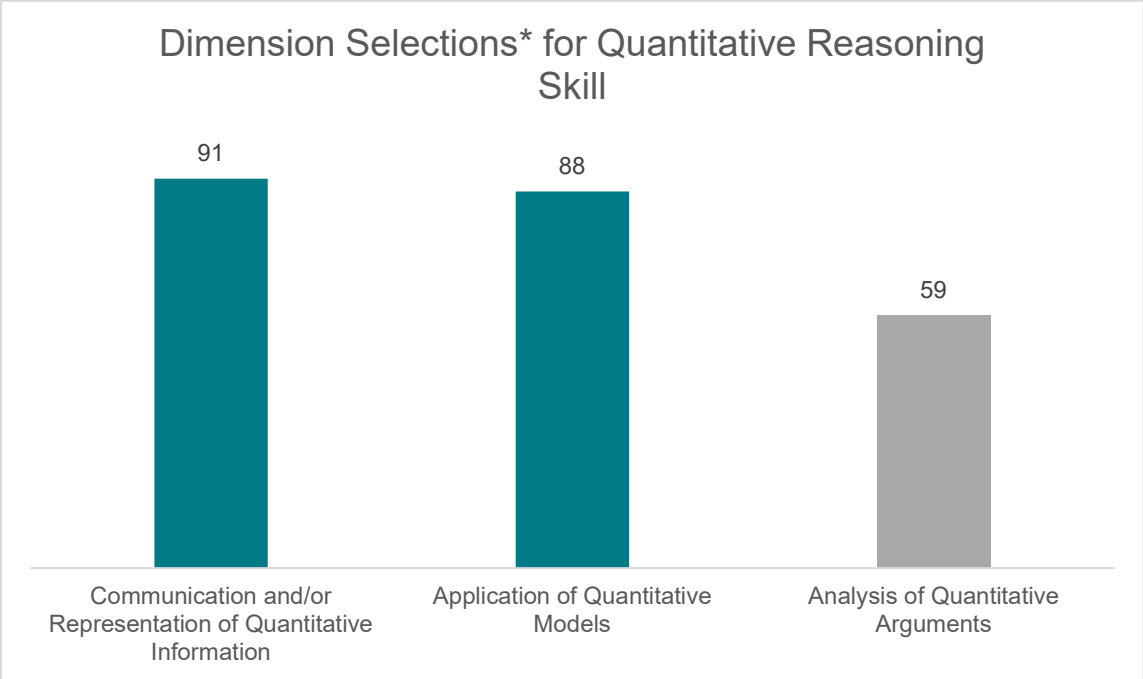
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

QUANTITATIVE REASONING ESSENTIAL SKILL

Average Artifact Ratings



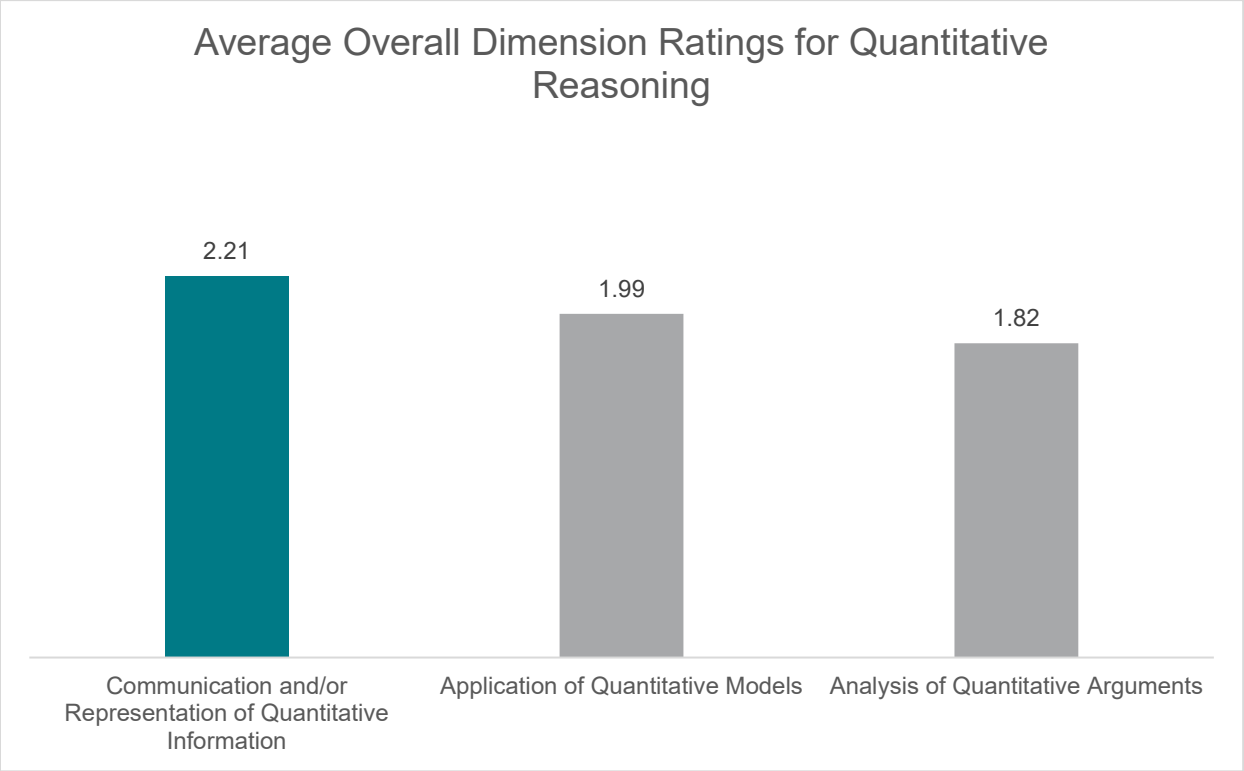
Total Number of Selections per Dimension



*each artifact was assigned to at least 2 dimensions

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Average Ratings per Each Dimension



QUALITATIVE RESULTS

Qualitative Analysis of Graduate Assistant Commentary

Along with rating each submitted artifact, Graduate Assistants (GAs) provided written commentary to support their numerical ratings with context. Each artifact was reviewed by two GAs. This is the qualitative analysis of their 1400 comments. In general, comments fell into one of three thematic groupings:

1. **Student Performance;**
2. **Artifact/Assignment Alignment with the Essential Skill or Dimensions**;
3. **Challenges for Rating**.

Most comments provided by the GAs (approximately half) fell into the **Student Performance** theme. Within this grouping, GAs noted when students performed well (or not) in an assignment and frequently commented on students' use of citations (both good and bad), which were required across all four essential skills. Examples of these comments included:

Performance

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Positive: *Well done! Very good piece of writing. The author addresses all the essential points and evaluates the messages in a great fashion.*

Negative: *It does not answer all of the questions posed by the professor, it does not show an evaluation of the material reviewed in class, and it does not include a conclusion. Doesn't really say anything, has errors.*

Citations

Positive: *Very good references section with quality research. Very well-done, complex solution…thorough.*

Negative: *Inconsistent use of citations, some were within the article…Some claims are not supported by documentation, which makes the rhetoric look weak.*

In some cases, although the artifact was highly rated in the dimensions selected, the GAs found the use of citations problematic. In the case below, while the artifact scored in the developing range (2), the GAs highlighted that the lack of citations was troubling and made it difficult to rate without bias:

This artifact lacks references and external arguments to support claims. There are no citations on the facts mentioned in the article… Good article but it needs a better integration and referencing of sources.

There were also several comments that emphasized the **importance of aligning the assignment/artifact with the essential skill rubric**. Commentary in this theme noted that although the student may have performed well within the assignment, the rating was low because the assignment did not align with the rubric:

Communication: *It's not clear why the instructor selected the 3rd dimension. There does not seem to be any evidence for this dimension.*
It's very difficult to assess Understanding and Evaluation of Messages. There is no clear evidence [of this dimension].

Information & Digital Literacy: *This assignment did not require the use of external sources except for the one in-class required reading [and the dimension requires synthesis of various sources].*

Personal & Social Responsibility: *There is no evidence for any of the dimensions. The assignment doesn't seem to match the rubric.*

Quantitative Reasoning: *This assignment is simply a math solution without detailed information.*

In the cases above, the GAs' notes illustrate that although the artifacts may be well-done, they did not show strong alignment with the two instructor-selected essential skill dimensions. This was especially true when the artifact/assignment required only a quantitative response:

Communication*: This is a math assignment. No argument is presented. Just solved the math.*

Quantitative Reasoning*: No real analysis. Just solved the math problem.*

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Conversely, though less frequently, the GAs also highlighted examples where the assignments were well-aligned with the rubrics:

> **Communication**: *This portfolio project did a great job showing understanding of various rhetorical situations.*
>
> **Information & Digital Literacy:** *Great example of digital literacy! Although the Instagram account was removed, it looked skillfully done from the photos*.
>
> **Personal & Social Responsibility**: *This assignment is an excellent example of a good incorporation of intercultural perspectives to solve issues.*
>
> **Quantitative Reasoning**: *The assignment is about computer science…it requires representation of quantitative data and relation to the real world.*

Finally, GAs noted several factors that made **rating more challenging** for them. Of all 1400 comments, there were only **72 mentions** of difficulties with rating. In some cases, artifacts were simply not valid (either they were duplicates, dead links, or extra assignment details from the instructor, etc.). In others, GAs noted issues with student grammar that was distracting and made rating more difficult:

> *Impossible to follow with so many mistakes. Lots of errors, almost no information.*
>
> *Wonderful genre formatting. However, it lacks citations and references to support argument. No works cited.*

or with assignments that were so short it was difficult to highlight the dimensions from the essential skills:

> *Really brief, topical. Mentions strategies but does not enact them.*

As noted above, GAs also found it difficult to rate assignments that provided little more than a quantitative solution. In a small number of cases, GAs also noted that handwritten artifacts were difficult to read and therefore rate:

> *Handwriting is illegible. It was difficult to read what she/he wrote.*

Artifacts submitted in a second language also proved challenging as some of the languages were not spoken by the GAs (Arabic, Spanish, and French). In most cases, the OA/APR was able to re-assign artifacts to GAs who spoke the languages and reached out to the Arabic department for assistance from a faculty member for artifacts submitted in Arabic.

Pilot Year Assessment Findings

In review and evaluation of the pilot year, the OA/APR examined the new general education assessment process to determine how best to improve data collection and data analysis. The following are areas of strength and areas of improvement that the office identified.

Data collection included the student artifact submission process, the GE workshops/training, communication, the inventory of data collected, and the

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

forms/materials surrounding the student completed work collected. From this pilot assessment cycle, the OA has **updated the GE submission form** to include required fields (rather than optional ones), an item asking what language the student completed the assignment in, group assignment information/descriptions, new course modality options to align with AY2020-21 hybrid teaching options, as well as several other updates that will make the form easier to use..

Communication was instrumental in dispersing information about the new UNM assessment process and will remain so in the subsequent assessment cycles. The OA learned that future communications must **emphasize rubric dimension alignment** and that instructors need to choose dimensions strategically to map the assignment they plan on submitting. Additionally, since regular messaging to all colleges/schools/branches on submissions was beneficial, the **GE submission inventory and its affiliated communications will continue**. Lastly, the OA will **focus communications on the crucial role instructors have** in learning (and teaching) the essential skills as described/portrayed/explained in the UNM GE rubrics.

Training is critical for future GE assessment. The OA believes that specific essential skill training for faculty and TA's is necessary, especially in curriculum building, GE course design and certification, and assessment. This would be beneficial in creating connections to general education programming. Many faculty are new to this general education process - expanding training and providing consistent communications will help embed the NMHED general education reform.

During this pilot year, the OA hired **six graduate students** to assist with the analysis process. The OA office does not have the capacity to analyze this large set of data (~790 student artifacts this year) and discovered that the expertise and education of the graduate students exceeded analysis expectations. The opportunity afforded the graduate students to gain educational assessment and general education knowledge/skills as well as summer income. **The OA plans to mimic this analysis model in years to come.**

Analysis includes the evaluation of the quantitative and qualitative data. The OA learned that **subjectivity and rater bias** played more of an active role in the rubric rating process than initially thought. By assigning rating teams, rater reliability increased, minimizing biases and decreasing the variance in rubric interpretations. Bias awareness was embedded in the analysis training and weekly meetings with raters. Rating issues remain - for example, some raters had difficulty assigning a student artifact with a rating of 0 or "no evidence" and wanted to give credit for student effort. Examples like this will be included in the next set of analysis trainings, as well as **ways to manage these biases in the next assessment cycle.**

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Reconciliation training also needs to be part of the analysis process. The OA planned for this and handled ratings on a case-by-case basis when team members had a rating difference larger than 1-point. In reflection, the OA learned what is effective in reconciling ratings between raters and will incorporate this information into next year's analysis training and process.

Qualitative data was collected by the raters and involved note-taking sections in the analysis forms. The initial idea for the commentary was to provide an opportunity for each rater to share observations on rubric and assignment alignment, excellent and poor student performances, and strong and poor examples of dimension evidence. Even though these specific types of notes were requested, the **note-taking portion was often inadequate or incomplete**, focusing mainly on student performances as noted in the commentary summary above**.** The OA will focus more attention on these notes and the significance of these observations and narrative pieces in future trainings during upcoming assessment years to ensure that they are completed.

Upon analyzing the student artifacts, the graduate students provided feedback to the OA regarding the GE submission process. They suggested the OA create **specific guidelines** for each essential skill submission that would help **aid in alignment between the student artifact, dimension selection and overall essential skill.** They recommended this "checklist" or flowchart be provided as a guide before artifacts are submitted by instructors.

*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No evidence; 1= Emerging; 2= Developing; and 3= Proficient.