OFFICE OF
ACADEMIC AFFAIRS

Office of Assessment & Academic Program Review

General Education Assessment Report

Fall 2022

INTRODUCTION

The University of New Mexico's Office of Assessment & APR (OA/APR) collected General Education student artifacts across colleges, schools, and branches in AY 21-22. These General Education artifacts represented quantitative reasoning, communication, information & digital literacy and personal & social responsibility (four of the five essential skills) and were used to assess UNM's General Education programming.

The OA/APR quantitatively and qualitatively analyzed 616 student artifacts. In isolation, these ratings are indicative of many variables: (1) student performance, (2) assignment alignment with the essential skill, (3) alignment of the assessment tool (rubric) and the student work, and (4) rubric dimension selection by each instructor.

Overview

The University of New Mexico's General Education (GE) program is based on 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. During the GE revision, a statewide task force aligned general education content with three essential skills per 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. The five skills are:

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

The statewide task force further defined each "essential skill" through identification of multiple "component skills." For more information on the organization of UNM's General Education Program and alignment of content areas with essential skills and component skills of essential skills, see gened.unm.edu.

Methods

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 this skill across disciplines/majors, and ease use.

The OA/APR collected samples of student artifacts from instructors relating to the essential skills. Each college/school/branch was required to select 5-10 sections of GE courses per designated content area (dependent on the size of the college/school/ branch).

Six graduate students underwent FERPA, rubric norming, and GE assessment training to rate the artifacts. 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 as well.

ratings and qualitative notes for the assigned essential skills. The quantitative results are visualized in bar graphs (below) and the qualitative narrative were analyzed with Atlas.ti software (also below).

Student Population

When submitting student artifacts for the GE assessment process, the OA/APR requests that instructors provide the affiliated UNM student Banner IDs. 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: **665 students** (some instructors submitted multiple artifacts from the same student, while some artifacts came from groups of multiple students)
- Gender: **60.6% female**, 38.8% male, 0.6% not reported
- Student level: **39.7% Sophomores**, 23.0% Juniors, 20.3% Seniors, 8% other (high school, non-degree seeking undergraduate/graduate, nursing levels), 7.7% First-Years, 1.4% not reported
- Ethnicity: **44.5% Hispanic**, 34.4% White, 7.7% American Indian, 5.1% Asian, 3.5% Two or More Races, 1.5% International,1.5% Race/Ethnicity unknown, 1.2% Black or Afro American, 0.6% not reported.
- Majors: Artifacts came from students across **122 different majors/pre-majors**, including non-degree and undecided students (new data this year)

QUANTITATIVE RESULTS

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.

Essential Skills: UNM colleges/schools/branches achieved a **rating average of 1.85** across all artifacts, at the "proficient" to "developing" range.

Communication The majority of student artifacts met or were above skill average at the "developing" level. Unlike Year 1 where the highest rated dimension was Evaluation and Production of Arguments, this year's results were **highest in the Genre/Disciplinary Conventions dimensions**.

Information & Digital Literacy: The majority of student artifacts were just shy of the "developing" level, and was markedly higher than Year 1's overall rating. This year's highest rated IDL dimension was **Research as Inquiry**. All IDL artifacts were found to align with the skill rubric.

Personal & Social Responsibility: The majority of student artifacts were in the "proficient" to "developing" range. This year's highest rated PSR dimensions were the **Intercultural Reasoning** and the **Ethical Reasoning** dimensions, the same as in Year 1.

*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.

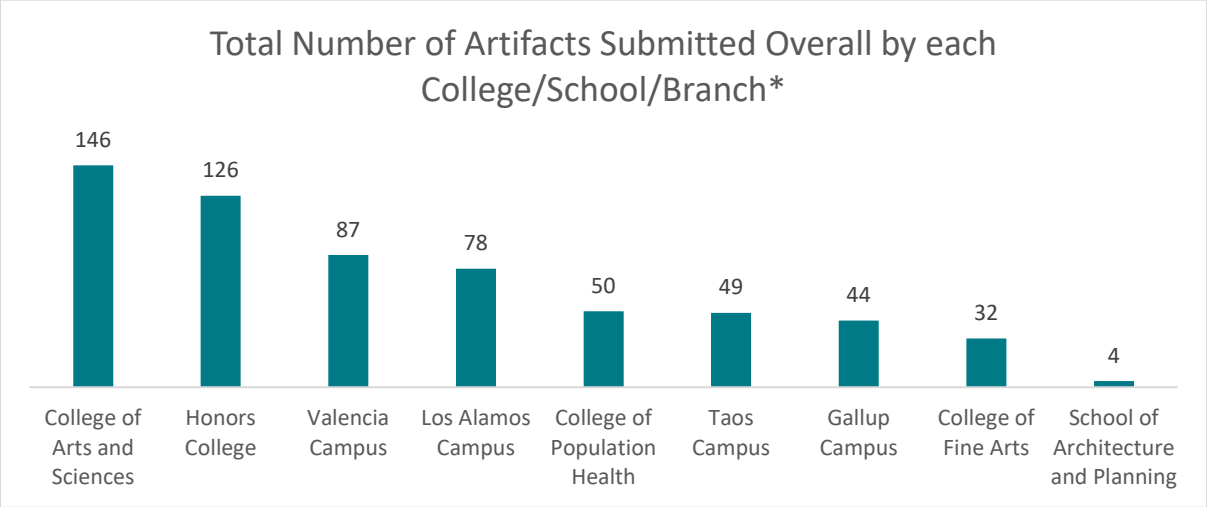
Quantitative Reasoning: The majority of student artifacts were in the "proficient" to "developing" range. This year's highest rated QR dimension was the **Communication and/or Representation of Quantitative Information**, the same as in Year 1.

Course Level: The **majority** of submitted student artifacts represented **1000** level courses. Communication and IDL Artifacts from **2000** level courses **rated higher on** average than those from 1000 level courses.

Course Modality: Artifacts from **hybrid courses rated highest**, at a Developing level.

Class Size: Artifacts from **large-sized classes rated highest,** though only 20 artifacts were submitted from large courses. **Small-sized** classes were the majority of submitted artifacts, and rated "highly emerging" overall.

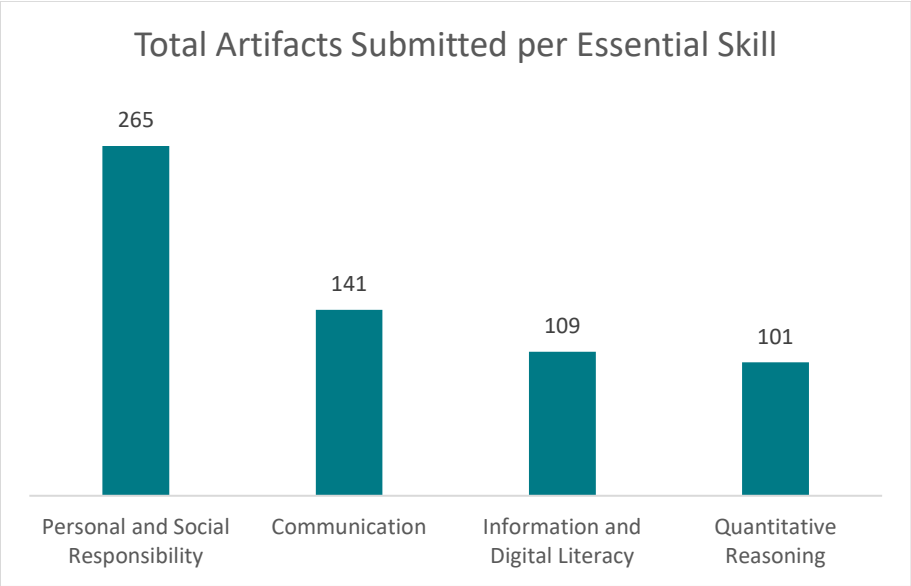
Total Artifacts Submitted

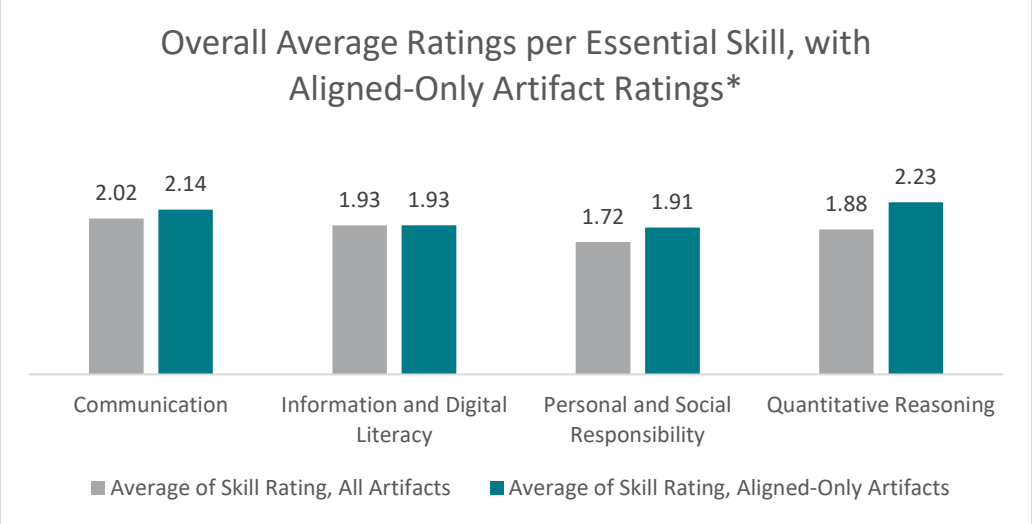


*As of May 16, 2022 (final data pull date)

*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.

Overall Average Essential Skill Submissions & Ratings



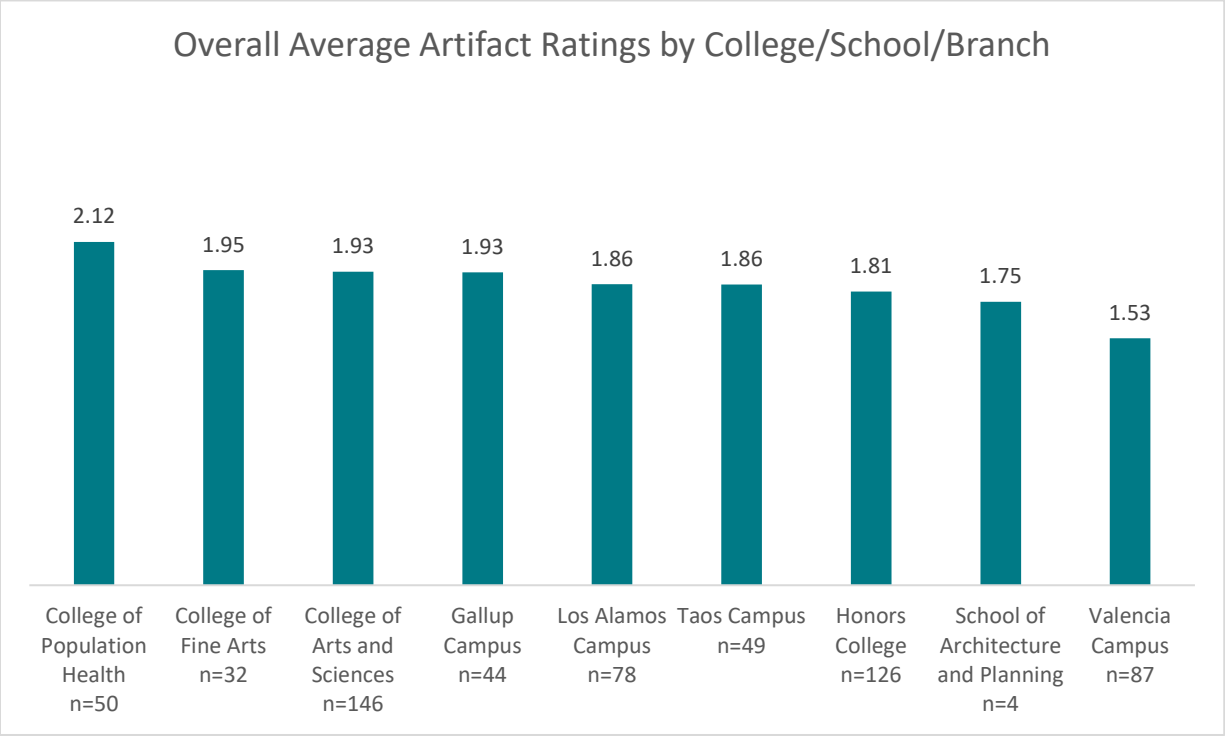


*All submitted artifact/assignment descriptions were assessed for alignment to the skill rubrics. These graphs illustrate submissions and ratings for artifacts that were found to be aligned, compared to the overall submissions/ratings of all artifact submissions, regardless of alignment.

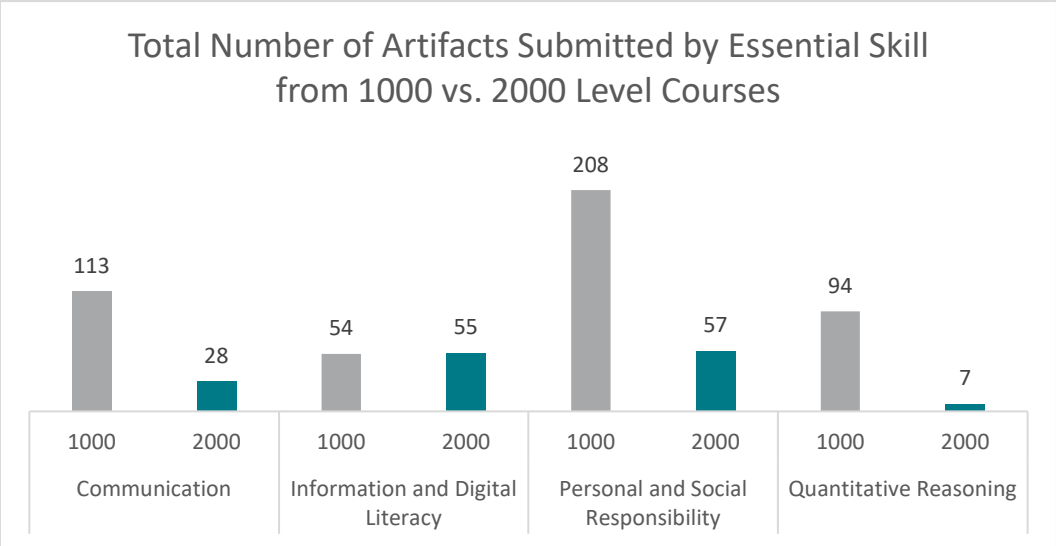
All further graphs below are for **ALL ARTIFACTS**, and do not include this alignment distinction.

*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 by College/School/Branch, across all skills

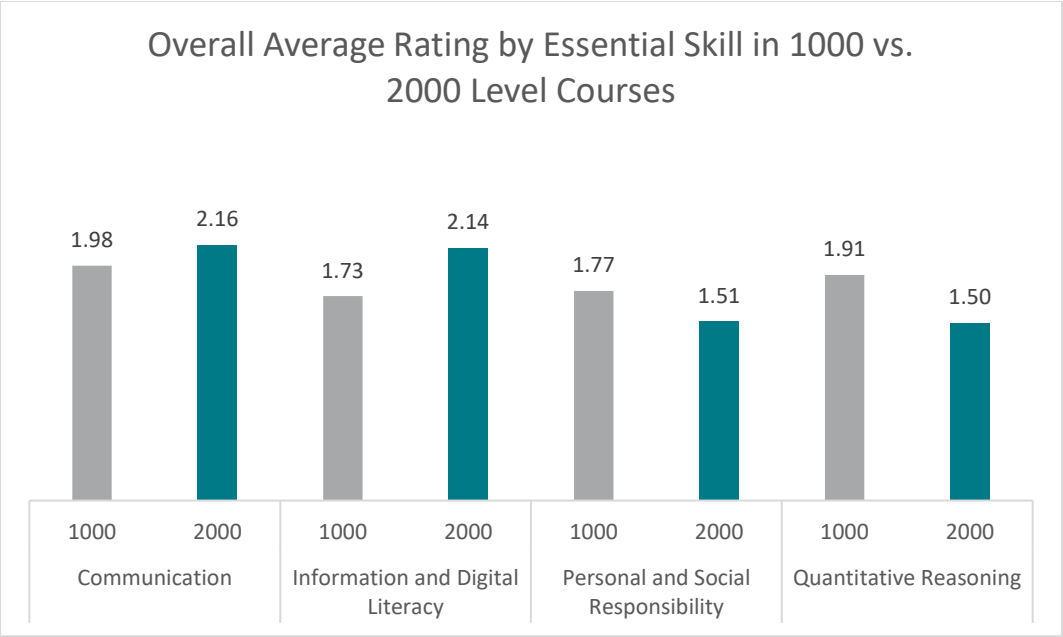


Total Number of Artifacts Submitted by Course Level



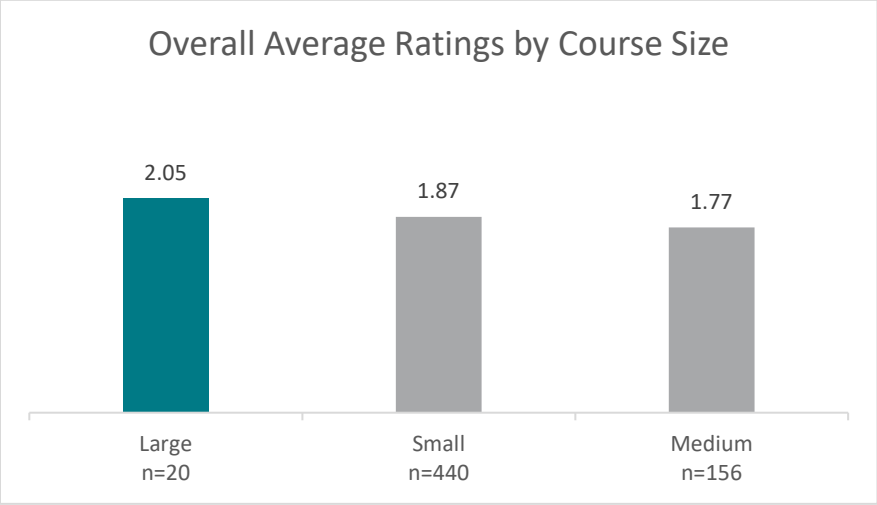
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No Evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Average Artifact Ratings by Course Level



*With only 7 2000-level artifacts submitted for QR, caution should be taken in comparing the decrease from 1000 to 2000 level ratings

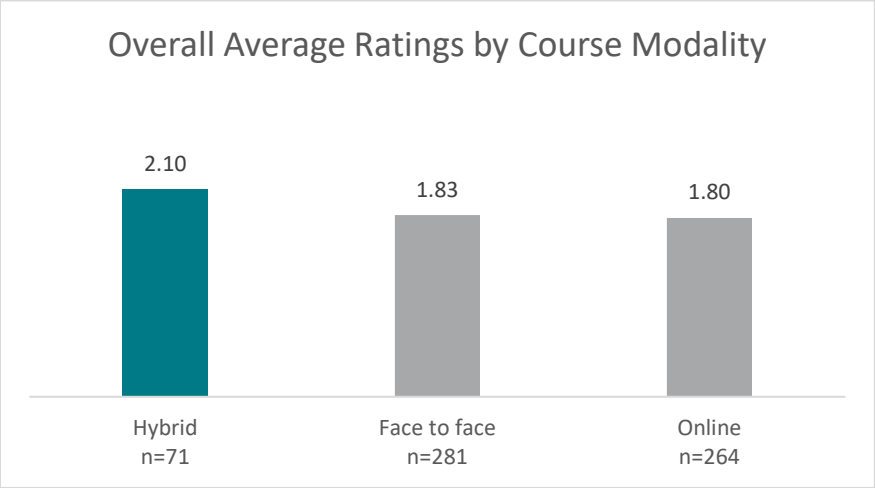
Average Artifact Ratings per Course Size



*Small equates to courses with 25 or fewer students; larger courses are those with 100+ students.

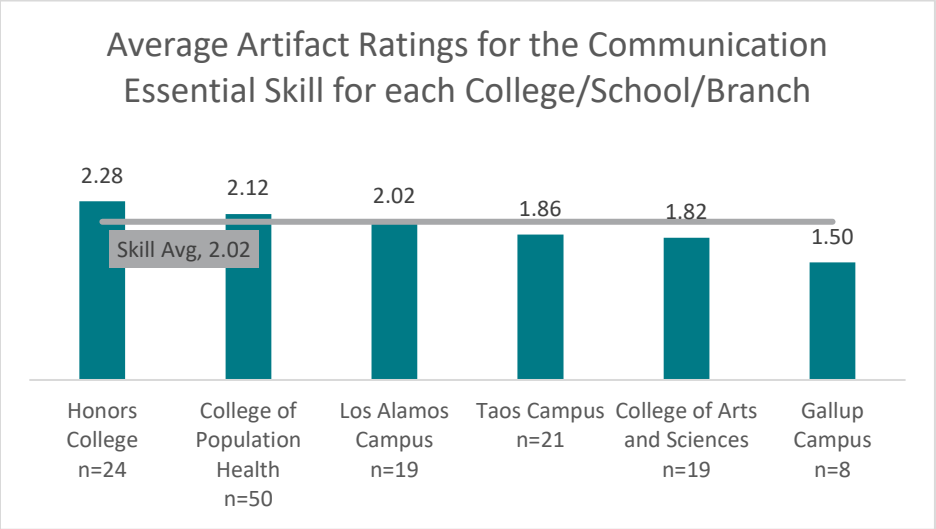
*When reviewing quantitative results, please keep in mind the rating scale range of 0-3 used to assess the submitted GE student artifacts: 0 = No Evidence; 1= Emerging; 2= Developing; and 3= Proficient.

Average Artifact Ratings per Course Modality



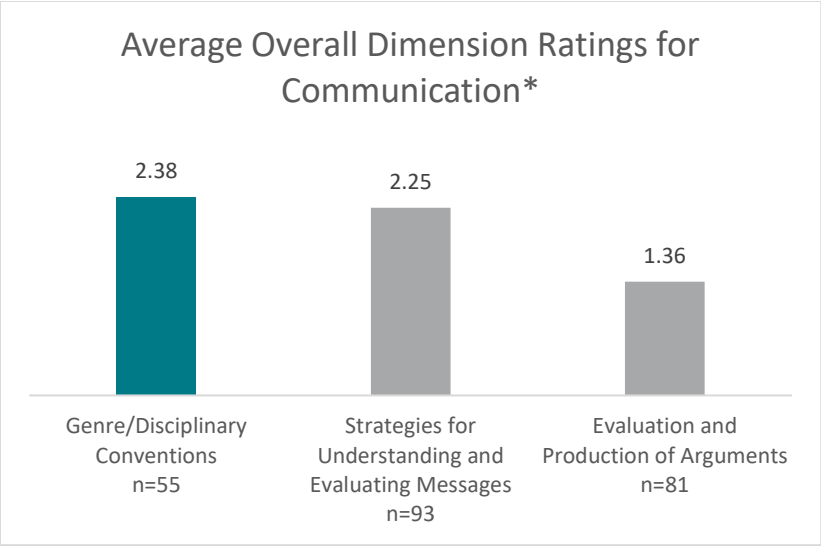
COMMUNICATION ESSENTIAL SKILL

Average Communication Artifact Ratings, by C/S/B



*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.

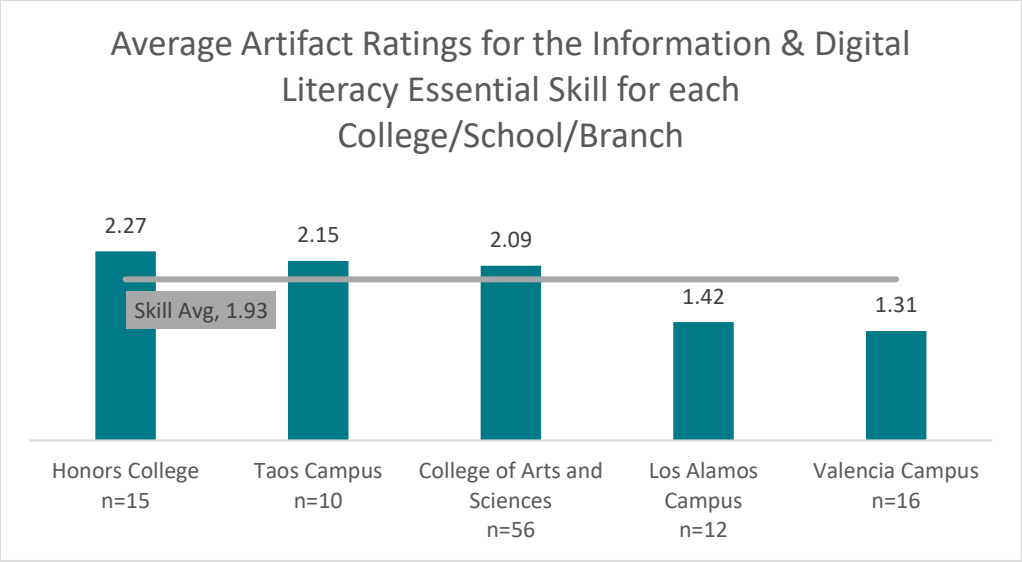
Average Rating for Each Communication Dimension



*N-Value surpasses total Communication Artifacts of 141 as artifacts were associated with more than 1 dimension.

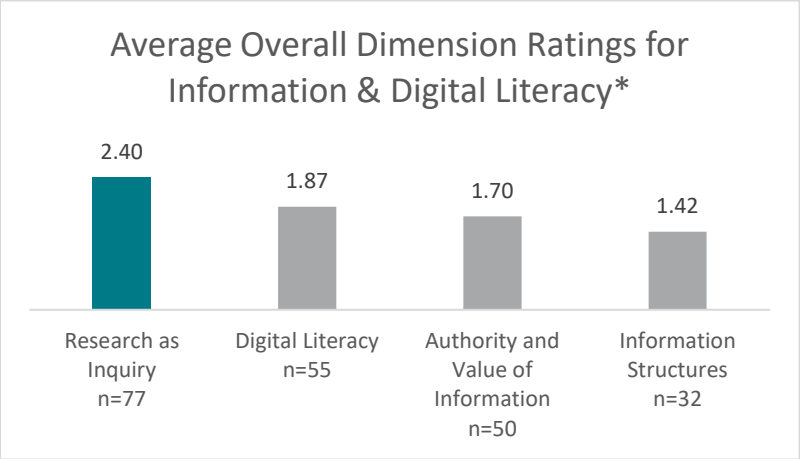
INFORMATION & DIGITAL LITERACY ESSENTIAL SKILL

Average Information & Digital Literacy Artifact Ratings, by C/S/B



*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.

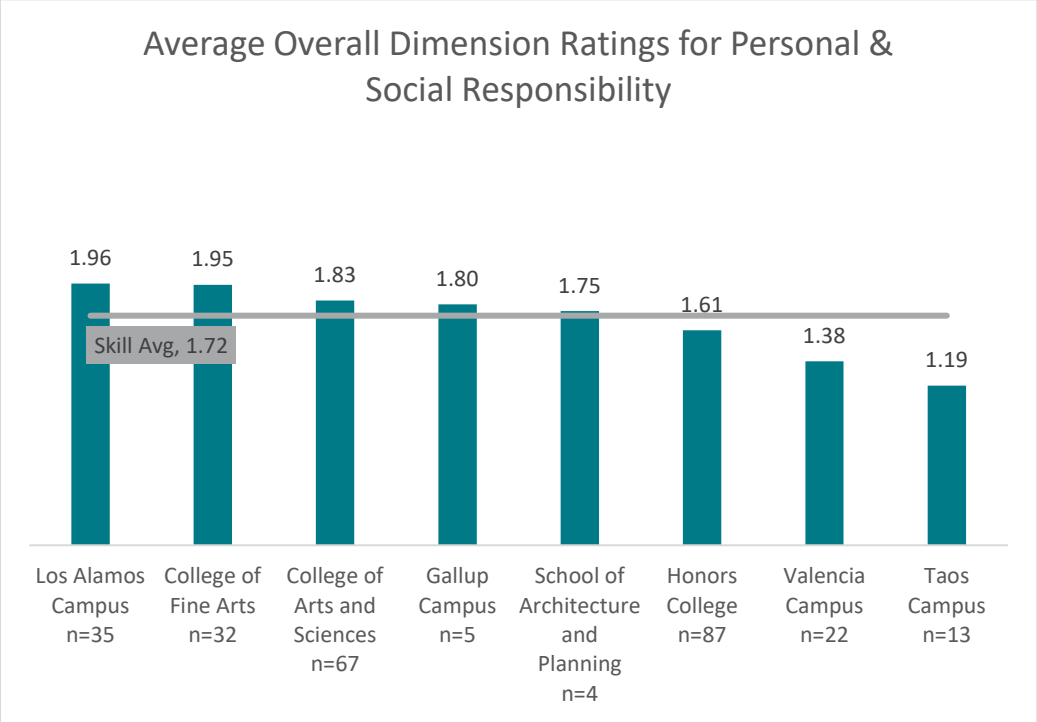
Average Rating for Each Information & Digital Literacy Dimension



*N-Value surpasses total IDL Artifacts of 109 as artifacts were associated with more than 1 dimension.

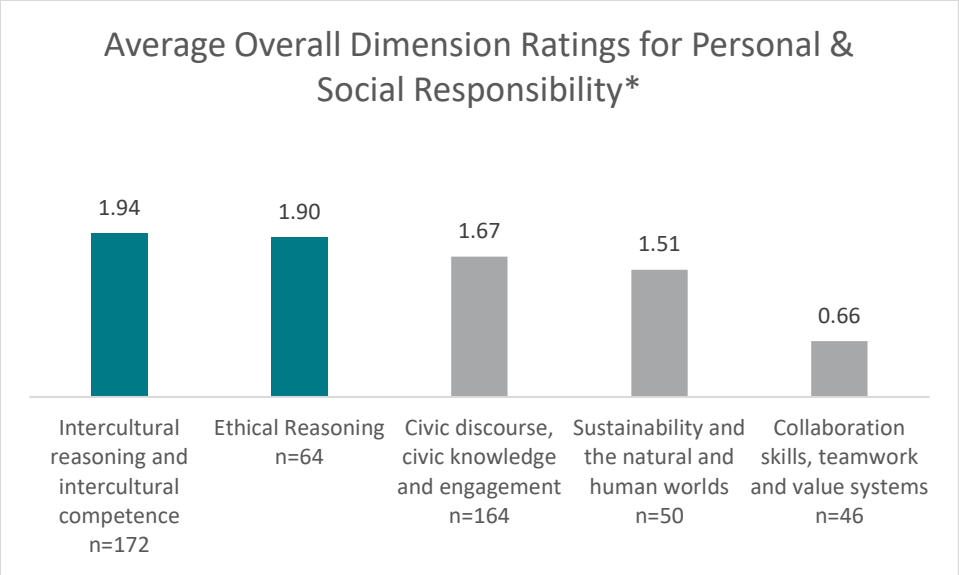
PERSONAL & SOCIAL RESPONSIBILITY ESSENTIAL SKILL

Average Personal & Social Responsibility Artifact Ratings, by C/S/B



*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.

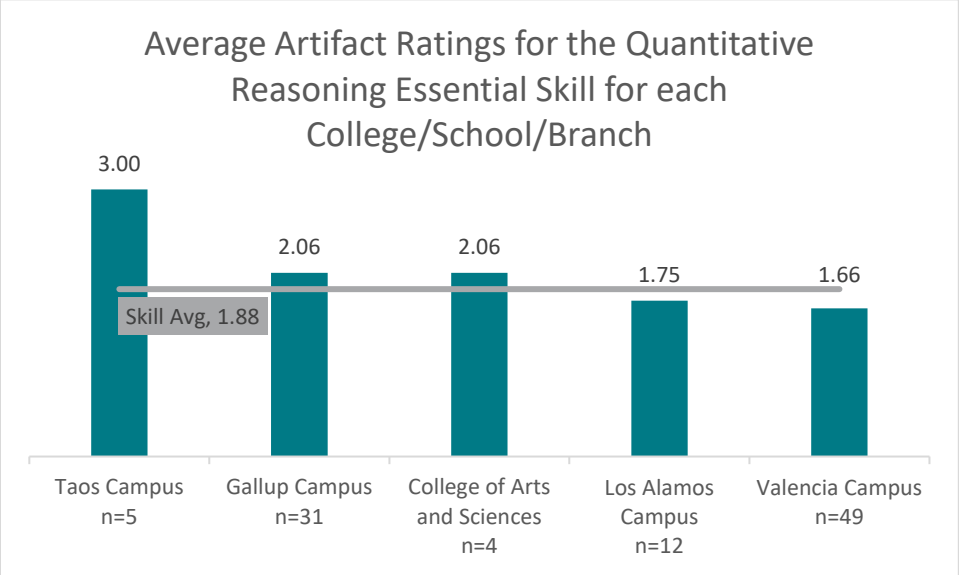
Average Rating for Each Personal & Social Responsibility Dimension



*N-Value surpasses total PSR Artifacts of 265 as artifacts were associated with more than 1 dimension.

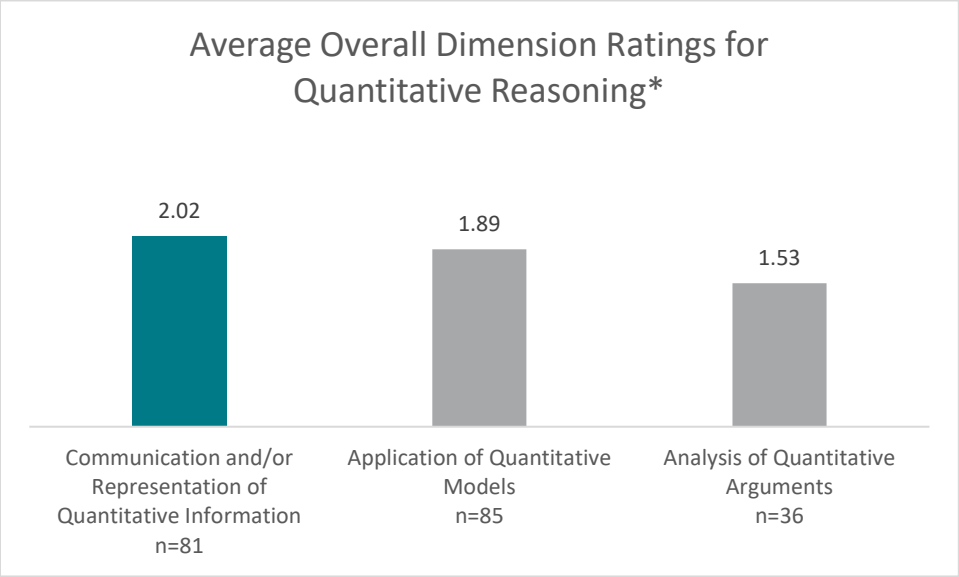
QUANTITATIVE REASONING ESSENTIAL SKILL

Average Quantitative Reasoning Artifact Ratings, by C/S/B



*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 Rating for Each Quantitative Reasoning Dimension



*N-Value surpasses total QR Artifacts of 101 as artifacts were associated with more than 1 dimension.

QUALITATIVE RESULTS

Qualitative Analysis of Graduate Assistant Commentary

Along with rating each submitted artifact, Graduate Assistants (GAs) provided written commentary to give context. Each artifact was reviewed by two GAs, so each artifact has two sets of qualitative notes from each of the students. Below is the analysis for those comments related solely to the rating experience.

There were 334 comments on the assessment process; many were linked to several different qualitative codes for a grand total of 574 "code occurrences." Percentages listed below divide the frequency of the codes within a thematic grouping by the grand total of "code occurrences." In general, rater comments fell into one of four thematic groupings:

1. **Assignments: Design, Instructions, or Artifact Type (44%)**
2. **Dimension/Skill Alignment (38%)**
3. **Justification for Rating / General Notes on Student's Work (14%)**
4. **Rubric Application & Form Suggestions (4%)**

Assignment Design, Instructions, or Artifact Type

44.% (253) codes emerged within this theme, though codes fell into three sub-categories:

1. The assignment itself did not require students to evidence the essential skill/dimension – 135 code occurrences where the raters noted that the assignment itself did not require students to provide evidence or "do work" that

*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.

aligned with the provisions of the skill/dimension rubric. This is an example of how assignment design affects the overall rating. One rater wrote:

…there is very little prompting in the assignment guidelines that would encourage students to craft an artifact that fulfills the "Proficient" dimension; for example, there are not a lot of solutions being compared/contrasted, there are not many strategies, ethical solutions, or solutions to personal and global problems being demonstrated.

Another wrote:
This assignment has the potential to take advantage of all the levels of the selected dimensions, yet the way it is formulated seems to restrict the students into mostly the emergent level, where they can only describe the issues at hand with no need to explore perspectives and offer solutions.

2. Assignment instructions/design posed issues for rating – 95 code occurrences pertained to the assignment information provided by instructors. For example, one rater wrote:

   This assignment lends itself for multiple outcomes, from an analysis of an individual comic, to a historical comparison of social issues. Because of this, the rubric may or may not fit the artifacts

3. Artifact type is challenging to rate – 23 code occurrences discussed the difficulty in rating specific types of artifacts. Notably, assignments that would inherently include an oral component but was not submitted (i.e., PowerPoints without the presentation narrative). One rater wrote:

   I feel uncertain that these assignments can be said to "align" with AVI or DL …was there a verbal delivery that accompanied these artifacts? I feel less and less convinced that PP artifacts should be accepted if they're not accompanied by audio files. It's not a complete use of the PP genre -- if there is no audio -- what is a PP without accompanying presentation?

Dimension/Skill Alignment

Raters noted strong and/or poor dimension(s)/skill alignment for more than a third (38%) of all coding occurrences. Particularly, raters stating that an assignment was not well-aligned to one or both chosen dimensions, or that another dimension would have been a better choice for the artifacts. For example, one rater wrote:

Rating Evaluation and Production of Arguments dimension would fit better than Rating Strategies for Understanding and Evaluating Messages.

Codes also included raters stating that an artifact was well-aligned with the selected dimensions. One rater wrote:

*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 instructor in this assignment provides the students with the component of a function and expects them to come up with the function described in the problem. Students don't need to think a lot to get the function definition correct. But how they communicate their answers demonstrate if they are "versatile communicators who can respond to a diverse range of audiences, purposes, and contexts" … All the students in this assignment demonstrated their ability to communicate mathematically.

Justification for Rating / General Notes on Student's Work

While this qualitative section of the rater form was designed to create a space for notes about the assessment process itself, there was still a substantial amount of code occurrences justifying ratings (14%). For example, one rater wrote:

…writing style of this particular artifact makes it difficult to comprehend.

Rubric Application & Form Suggestions

Lastly, 4% of the code occurrences (21) included internal rater discussions on the topics of the GE Skill rubrics themselves, how raters should apply certain rubric elements to different works, and what might be done in the future to aid raters. For the discussion on the rubrics themselves, one rater wrote:

The information structures dimension is… weird to rate -- the two components (especially of the emerging and developing columns) seem almost unrelated and definitely not apparent in all the artifacts that have been designated to be rated for IS in these artifacts.

For the Rater Form itself, one rater wrote:

Something that would be helpful would be for there to be an identifying marker on each assignment as to which artifact it corresponds to; the instructor ID is there, but not necessarily the artifact ID.

ASSESSMENT IMPLICATIONS

While reviewing and evaluating this year's General Education assessment data, the OA/APR developed several possible solutions to improve data collection and data analysis. The following are areas of improvement that the office identified.

Data Collection/Form Adjustments

Upon evaluating the previous year's data, the OA/APR **updated the GE submission form** to include required fields (rather than optional ones) and added a sampling question to gauge how student artifacts are being selected by instructors. The OA/APR recommends random selection to give an accurate representation of our student's skills (i.e., otherwise purposefully selecting the best performing students may artificially inflate ratings, etc.). From this year's analysis, more than 75% of all artifacts were chosen through random sampling methods (37%), through clustered sampling (6%), or through

*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.

systematic sampling methods (32%). However, more than 15% of all artifacts were also chosen through more selective sampling methods (convenience and stratified); artifacts from these more purposefully selected students had higher overall ratings than artifacts from randomly chosen students, showing just how impactful an instructor's sampling method may be to the assessed student pool and results. Sampling methodology will continue to be required of instructors during the submission process.

To gain more meaningful information on the assessment process itself and to allow for more dynamic qualitative analysis, the OAAPR provided GAs with two qualitative comment sections for each artifact; one for general notetaking and one for any notes on the assessment process itself. The use of the qualitative data note-taking sections by raters is still in question; despite changes to the notetaking sections that delineated between "general notes on why you rated an artifact the way you did" and "notes on the assessment process itself", notes on the assessment process still had a noticeable percentage of code occurrences related to the justifications of ratings. As with the previous two years, these comments do not provide the insight desired on improvements, recommendations, etc. for the GE assessment process. The OA/APR will focus **more attention on the training of GAs** on the different qualitative note sections.

The graduate students also provided feedback to the OA/APR regarding the GE submission process. They suggested **a more detailed "what elements of the assignment align to the skill/dimensions chosen"**, to aid future raters in understanding what parts of an assignment they should truly hone in on when rating certain dimensions.

Raters noted that having the artifact IDs on the artifacts themselves would be helpful. Instructors could superimpose the banner ID of the student on the artifact before submitting (but after selecting the particular artifact) so that raters can make sure they are looking at the correct artifact when rating – some instructors currently already do this, but it is certainly a minority of submitted artifacts.

Communication Improvements

The OA/APR learned that future communications must **emphasize assignment and rubric dimension alignment** to instructors so that essential skill dimensions strategically map to the assignment being submitted. This has proven to be especially true after analysis revealed that **artifacts well-aligned to the four skill rubrics rated higher on average overall** than those the raters noted were not well-aligned to the rubrics.

In comparison to Year 1, which analyzed 725 artifacts submitted across these four essential skills, the office notes a decrease in student artifact submissions this year, as artifacts totaled 616. Since the OA/APR has determined that regular messaging to all colleges/schools/branches on submissions was beneficial, the **GE submission inventory update and its affiliated communications will continue**.

*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.

Expanded Training Opportunities

The OA/APR believes that specific essential skill training for instructors and teaching assistants is necessary, especially in GE course design, certification, and assessment. **Co-facilitated workshops with CTL** are in progress, including an "assignment collection" to serve as good examples for instructors that wish to know more about what type of assignments align better with the various skills.

Rating and Analysis Process

Analysis includes evaluation of the quantitative and qualitative data. The OA/APR previously learned that **subjectivity and rater bias** played more of an active role in the rubric rating process than initially thought. By assigning rating teams in previous years, rater reliability increased, minimizing biases, and decreasing the variance in rubric interpretations. Bias awareness was similarly embedded in the analysis training and weekly meetings with raters. Rating challenges remain - for example, **some raters had difficulty applying the rubrics to different kinds of assignments**. Particularly challenging this year were artifacts attributed to the Sustainability dimension within Personal & Social Responsibility, and the various dimensions within Information & Digital Literacy. These issues will be included in the next set of rater trainings, as well as **ways to manage biases in the next assessment cycle.**

Importance of Rubric Alignment

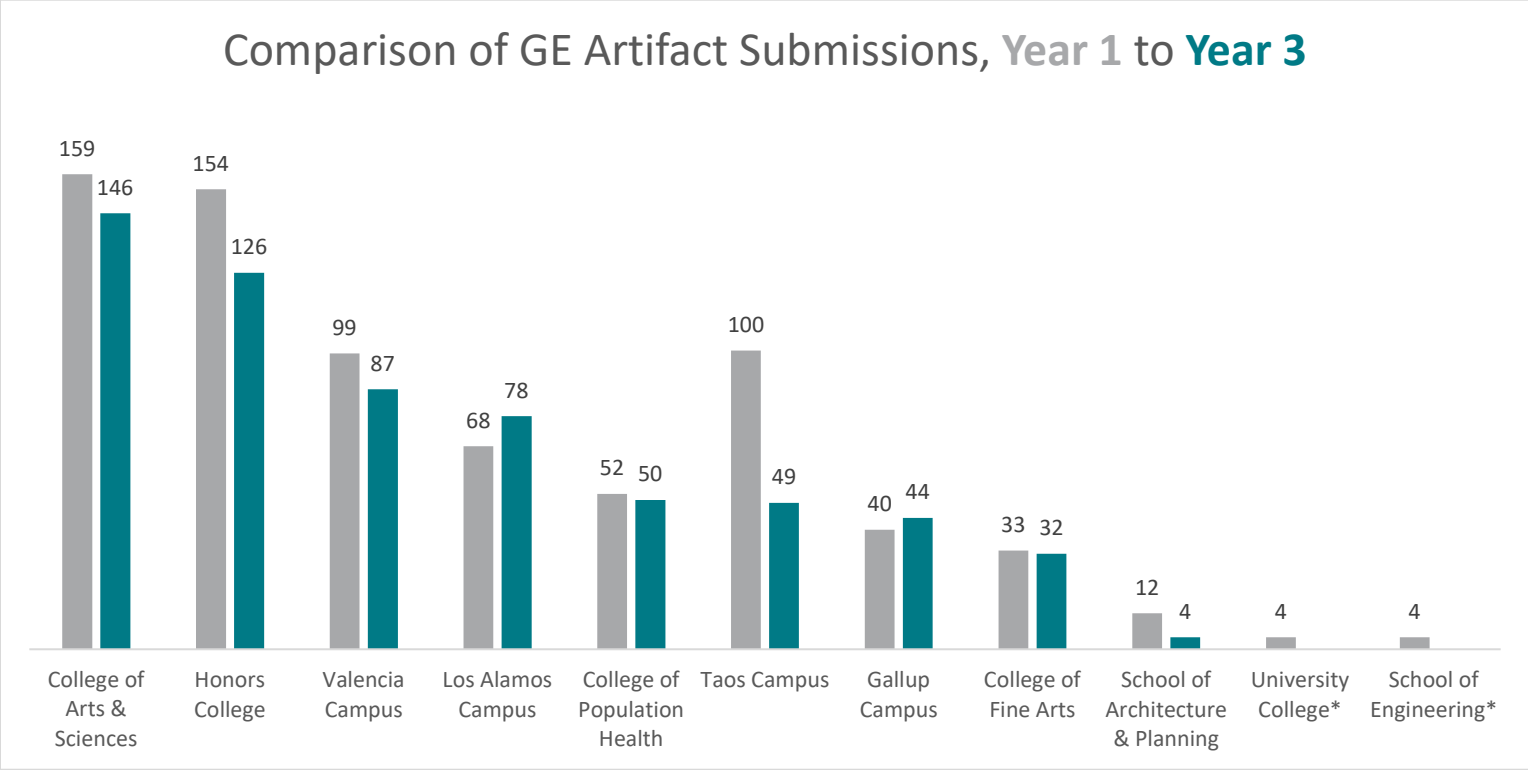
Raters noted that the **most difficult assignments to rate were the ones that were not well-aligned to the rubric**. Assignments that were short (less than 300 words); provided only a computational product (without written explanation of the work); provided multiple choice answers (exams or quizzes); provided only a PowerPoint without the recording of the presentation; or asked students to reflect on/provide solely an opinion were especially difficult to rate. Interestingly, raters found portfolios less challenging this year compared to last year when they were assigned to Critical Thinking; most portfolios this year were assigned to Information & Digital Literacy.

Raters found assignments that were well aligned to the rubric easier to rate. The **most important factor seemed to be whether the assignment had been created with the rubric in mind**. Raters even identified some examples where questions on exams worked well because the instructor had aligned the narrative response to the rubric. Some of these were assignments that required written explanation (750 words or more), oral explanation (recorded/video presentations), and/or that required students to critically interpret or evaluate information from several sources (essays, presentations, performance analyses). It was noted in the analysis debrief meeting that it is important that instructors **must read the various rating definitions within a dimension to truly understand what the dimension relates to.**

*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.

Appendix A: Comparison of Results to Year 1

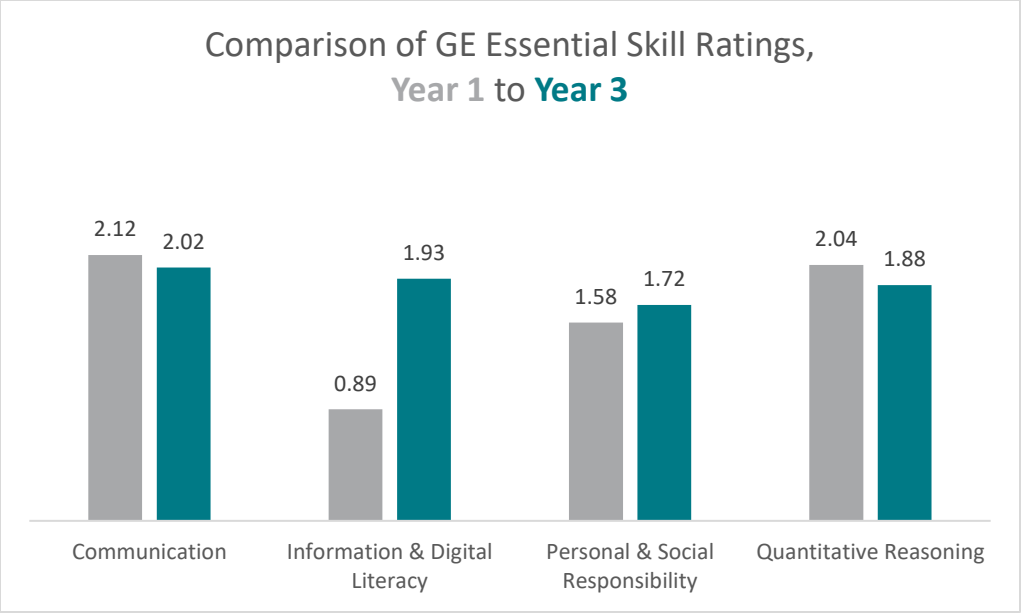
By design, Year 1 and Year 3 of the GE assessment cycle analyze the same four essential skills. This allows for a comparison of results.



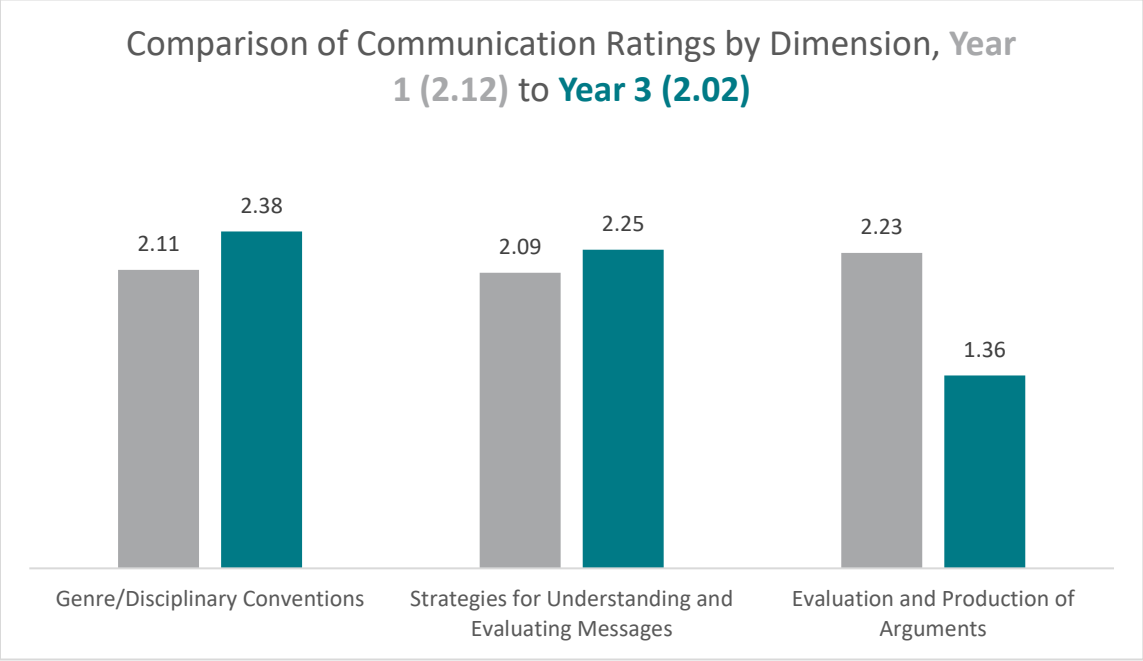
*UC and SOE were not tasked with submitting to the GE Assessment process following Year 1

From Year 1 to Year 3, artifacts submitted to the OAAPR for analysis decreased from 725 to 616. 7 out of 9 colleges/schools/branches had lower submissions in Year 3 compared to Year 1.

*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.

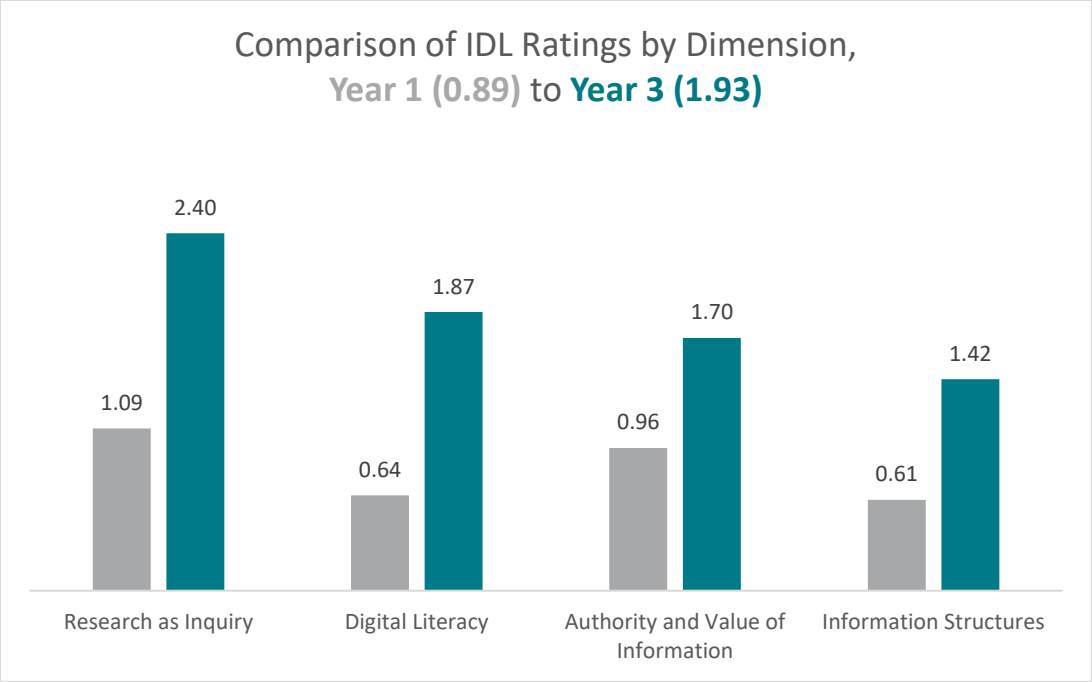


Overall skill ratings shifted since Year 1: the Communication rating slightly decreased from 2.12 to 2.02; the IDL rating increased significantly from 0.89 to 1.98; the PSR rating slightly increased from 1.58 to 1.72; and the QR rating slightly decreased from 2.04 to 1.88.

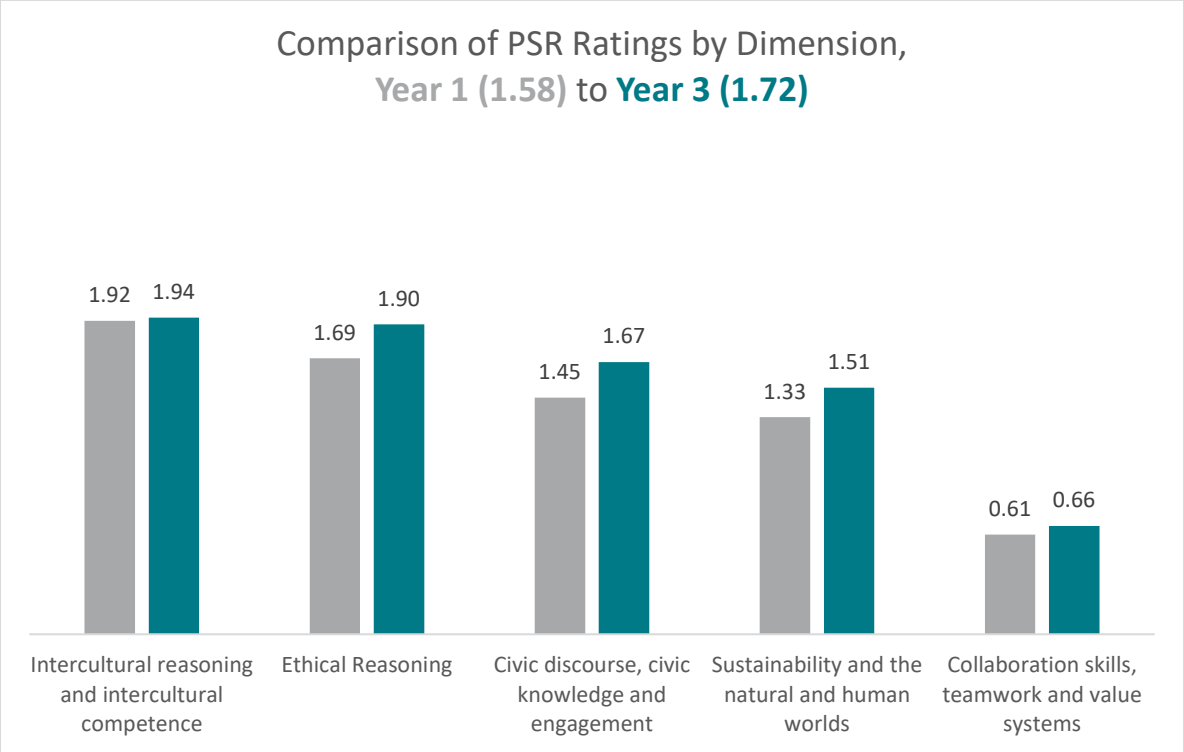


Since Year 1, two of the three Communication dimension ratings slightly increased. However, the rating for dimension *Evaluation and Production of Arguments* decreased significantly from 2.23 to 1.36.

*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.

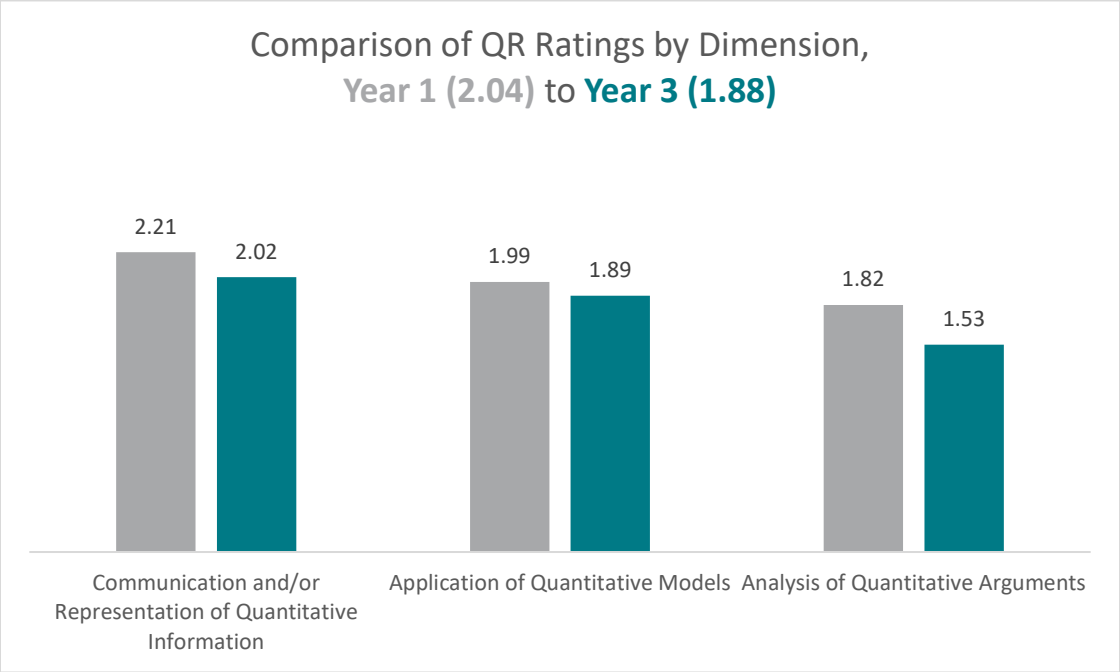


All four IDL dimension ratings increased significantly compared to Year 1. However, *Information Structures* was the lowest rated dimension in both years.



While all PSR dimension ratings improved since Year 1, the most challenging dimension to evidence in student work tends to be the dimension *Collaboration, Skills, Teamwork and Value Systems*.

*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.



Across Year 1 & 3 results, the lowest rated Quantitative Reasoning dimension is *Analysis of Quantitative Arguments*.

*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.