OFFICE OF ACADEMIC AFFAIRS

Office of Assessment & APR

AY 22-23 General Education Assessment Results

Overview:

Introduction:

The goal of this report is to share the results of skills assessment from certified courses in the general education (GE) program at the University of New Mexico (UNM). The Office of Assessment and Academic Program Review (OAAPR) assesses five GE essential skills across three-year cycles based on the New Mexico Higher Education Department's GE requirements. For the 2022-23 academic year, the OAAPR assessed four of the five GE essential skills: Communication (COMM), Information & Digital Literacy (IDL), Personal & Social Responsibility (PSR), Quantitative Reasoning (QR).

Key Takeaways:

- Participation was the highest yet for GE, though not every college or school with GE courses was represented
- Ratings were highly dependent on assignments being aligned to essential skill rubrics
- 22% of artifacts were not aligned with essential skills, and 10% of comments stated that the wrong dimension seemed to have been selected or that another dimension might have worked better
- Math and Art-based artifacts that do not evidence a process or student thinking are difficult to rate for evidence of an essential skill
- Ratings vary somewhat based on the content areas to which essential skills are assigned

Future Improvement Tips:

- Ensure alignment among assignment design, essential skill rubric dimensions, and selected dimensions to enable more accurate ratings
- Provide more random samples of artifacts to give a more representative picture of skill development
- Use the GE assignment collection for exemplary assignments that embed essential skills and promote student skill attainment

Methodology:

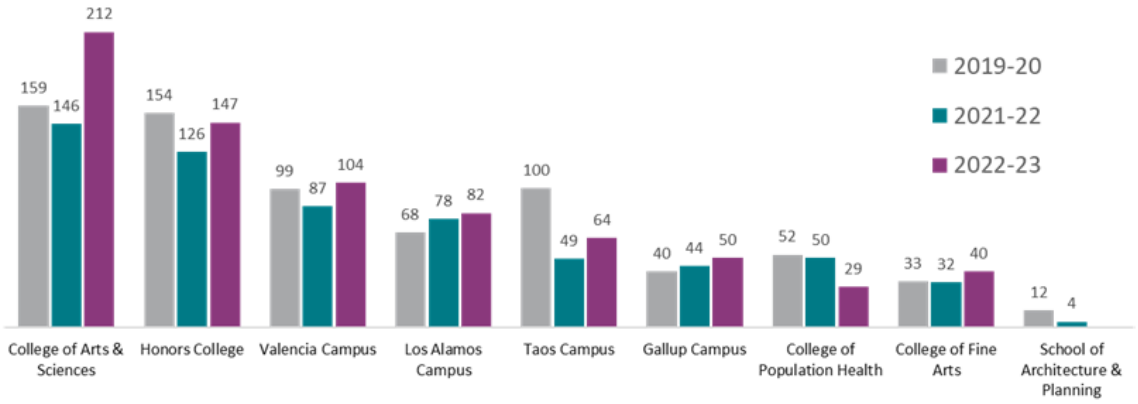
Each year, the OAAPR solicits student artifacts from instructors' assignments relating to the essential skills designated for that year. Each college, school, and branch with GE certified courses is asked to sample artifacts from multiple sections of GE courses in designated content areas (depending on the size of the college, school, or branch).
The OAAPR assesses evidence of GE skills in these artifacts with essential skill rubrics. Graduate student workers trained in FERPA issues, GE assessment, and interrater reliability then use these rubrics to rate submitted artifacts on a scale where 0 = No Evidence, 1 = Emerging, 2 = Developing, and 3 = Proficient. Graduate students also annotate artifacts with qualitative comments to provide additional insight into the rating process. The OAAPR then conducts quantitative analysis of these ratings in Excel and qualitative analysis of comments from graduate student raters.

Individual ratings and comments reflect several variables: student performance, assignment alignment with essential skill rubrics, rubric alignment with student work, and the rubric dimensions selected by each instructor.

Results:

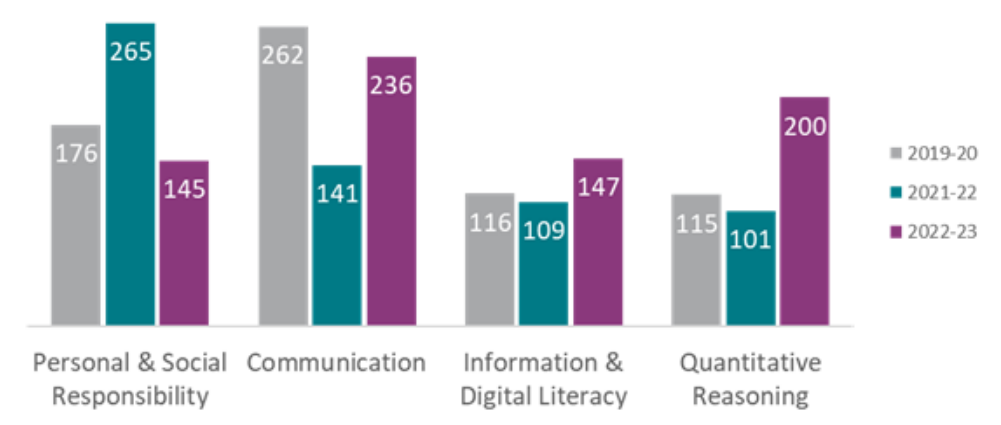
Artifact Submissions:

1. Artifacts Submitted by College/Branch by Academic Year



The OAAPR received 759 artifacts from eight colleges and branches representing 720 identified students (62 artifacts could not be linked to a student). Of the 759 artifacts, 728 were rated. Reasons for not rating submissions included duplicate submissions, inaccessible or unratable files, and submissions from courses outside the GE program. This year saw the highest overall submission counts for years where these four skills were assessed as well as the highest submissions from half of participating colleges and branches since the inception of this GE assessment model.

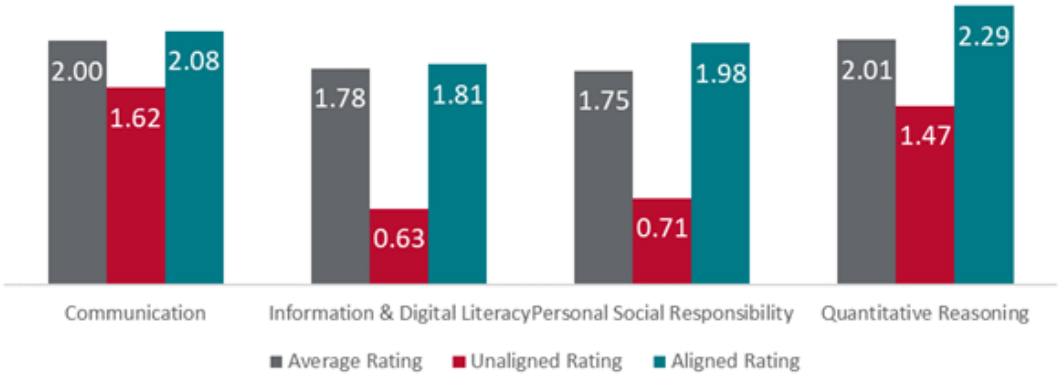
2. Artifacts Submitted by Essential Skill and Academic Year



This year saw the most GE submissions ever for the Information & Digital Literacy and Quantitative Reasoning skills, while Personal & Social Responsibility had its fewest submissions ever.

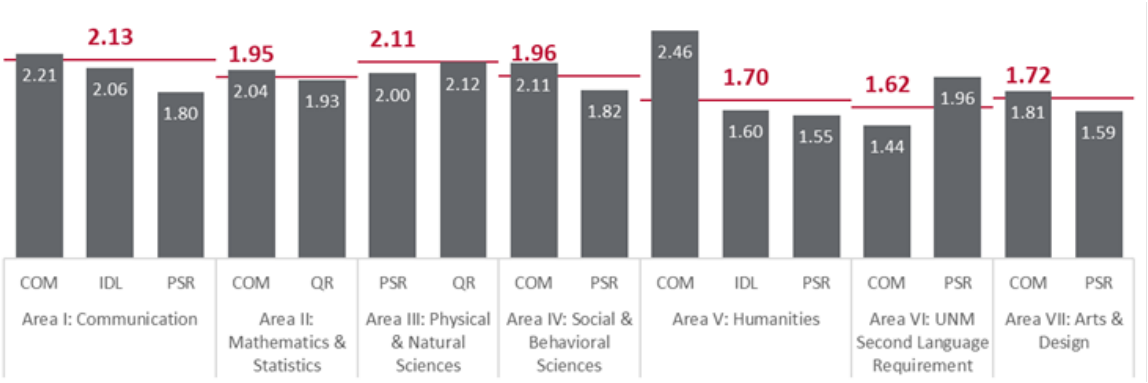
Artifact Ratings:

3. Average Artifact Ratings by Essential Skill for All Artifacts



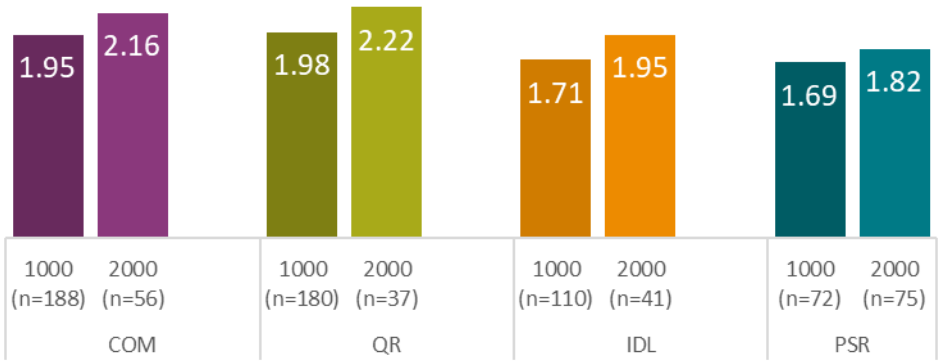
Most artifacts (595, or 82%) were aligned with the chosen essential skill, meaning that the assignments and artifacts directly corresponded to the rubric skill dimensions chosen by instructors. However, those which were not aligned were rated significantly lower, with an average rating difference of 0.93 from aligned artifacts overall. Differences among specific essential skills are shown in the chart above.

4. Average Artifact Ratings by Essential Skill and Content Area



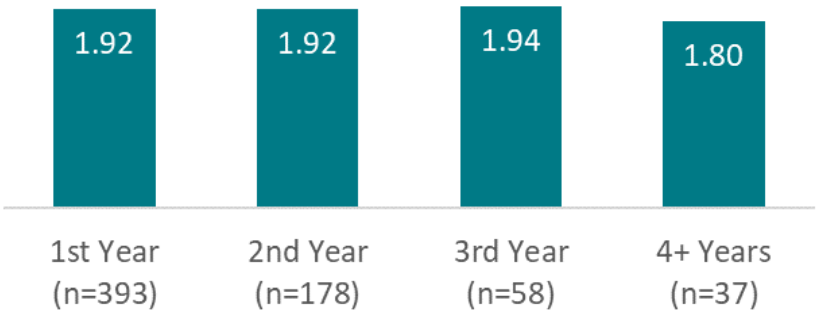
Looking at ratings across content areas, COM had both the highest and lowest ratings by area. Those in humanities were rated almost 2.5, while in second languages the rating was a full point below that. IDL ratings were highest in the communication content area, and both PSR and QR ratings were highest in the physical and natural sciences. There may have been some confusion over course classification or which skill is submitted per area, as evidenced by artifacts that were submitted for PSR in Area 1: Communications and for COM in Area 5: Humanities.

5. Average Artifact Ratings by Essential Skill and Course Level



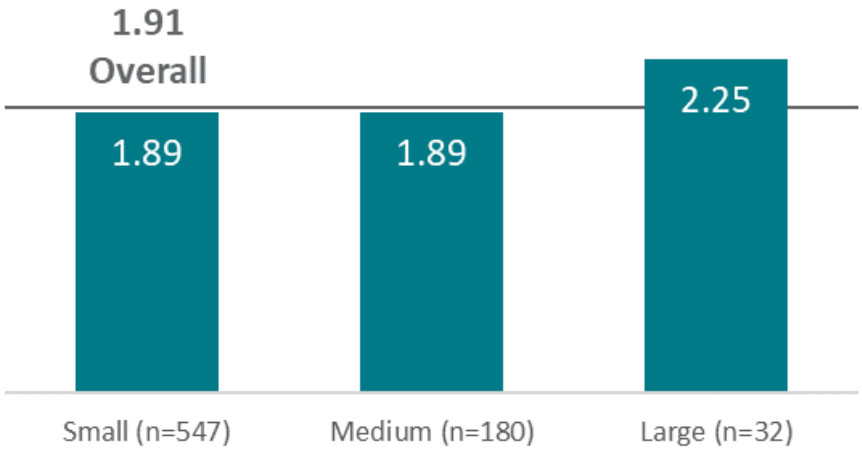
About 76% of submissions were from 1000-level courses, with 24% from 2000-level courses. Overall, 2000-level courses scored higher than 1000-level courses for all essential skill areas, which is to be expected as it suggests skill development between lower- and higher-level courses.

6. Average Artifact Ratings by Student Year



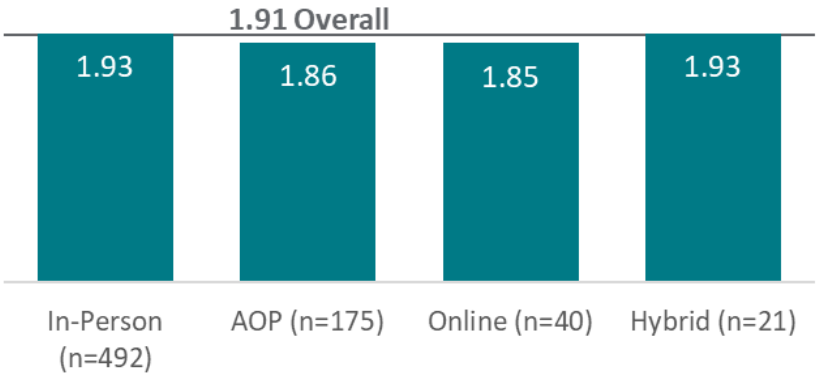
Interestingly, ratings held steady or increased among students in their first three years at UNM before dropping slightly among students who had been at UNM for four or more years.

7. Average Artifact Ratings by Essential Skill and Course Size



Courses were grouped into small (<25), medium (25-100), or large (>100) sizes. Most submissions (72%) came from small courses. Very few (4%) came from courses with more than 100 students, and there was evidence that even these courses may be smaller than was noted on the submission form, as a cursory review of the largest submitted course counts showed that multiple smaller sections had been combined in the final reported numbers. Ratings for small and medium courses were the same, while ratings from large courses were higher. This variation may be due to the small number of samples received for large courses, and other factors such as content area or essential skill sampled could affect this high average rating too.

8. Average Artifact Ratings by Essential Skill and Course Modality

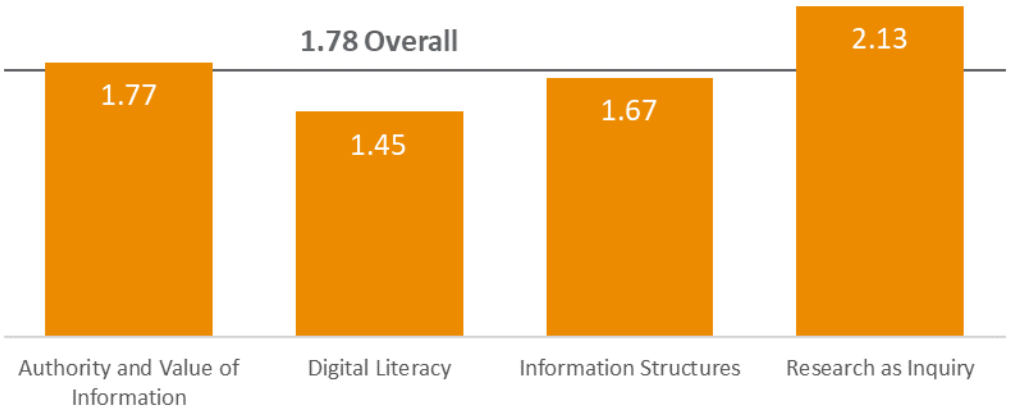


Most artifacts (68%) came from traditional face-to-face courses. Accelerated Online Programs (AOPs) had the next highest amount (24%), with online and hybrid courses having 40 or fewer submissions each. Average ratings were broadly similar across course modalities, suggesting that the method of course delivery does not substantially affect student skill attainment.

9. Average COM Artifact Ratings by Rubric Dimension

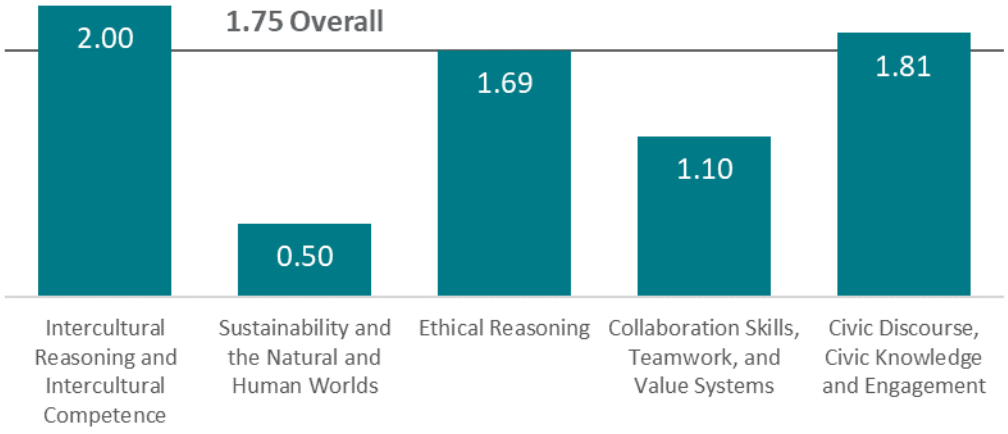
For the Communication skill, students showed the greatest skill development in the 'Genre and Disciplinary Conventions' dimension, and the lowest in 'Evaluation and Production of Arguments.' Overall, this skill was rated as Developing.

10. Average IDL Artifact Ratings by Rubric Dimension



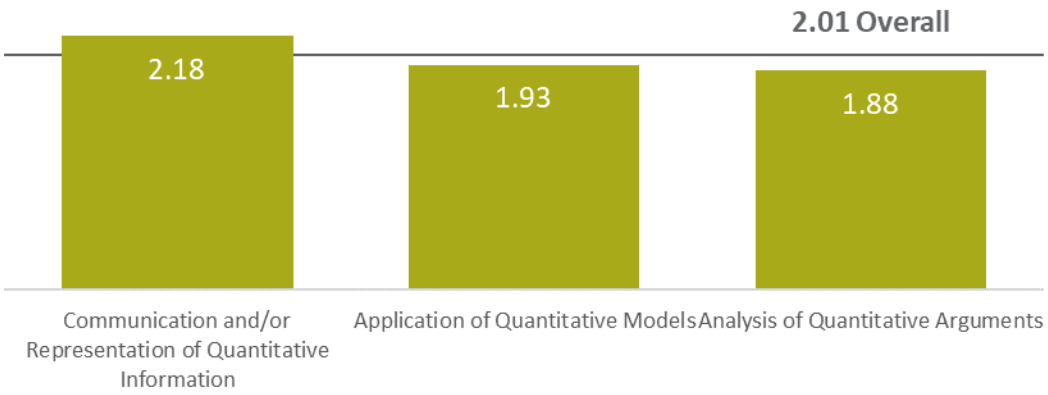
Students performed the best on the 'Research as Inquiry' dimension of Information & Digital Literacy and were rated lowest on the 'Digital Literacy' dimension. Overall, this skill was rated as Emerging.

11. Average PSR Artifact Ratings by Rubric Dimension



When broken down by rubric dimension, 'Sustainability and the Natural and Human Worlds' and 'Collaboration Skills, Teamwork, and Value Systems' had substantially lower ratings than the other dimensions in Personal & Social Responsibility. These ratings may be due to significantly smaller selections for each dimension of around 10 artifacts each. Raters also found many rated artifacts that would have been exemplary for collaboration skills had this dimension been selected. Overall, Personal and Social Responsibility was rated as Emerging.

12. Average QR Artifact Ratings by Rubric Dimension



Ratings of specific QR dimensions were fairly close to each other and to the overall rating, though 'Communication and/or Representation of Quantitative Information' was the highest rated. Overall, Quantitative Reasoning was rated as Developing.

Qualitative Comments:

Notes and comments from graduate student raters provide additional insight into the assessment process and implications for current and future cycles. Analysis of themes showed the emergence of the following categories:

Assignment alignment with essential skill rubrics:

Raters noted both well-aligned and unaligned assignments as they rated and flagged exemplary assignments for inclusion in the OAAPR's sample assignment collection.

Example comments:

- "Great assignment design, clearly aligning the topics of the class with the goals of GE and the specific PSR rubric"
- "Assignment = well-aligned with intercultural reasoning"
- "These might be exemplary assignments, especially for the IS [Information Structures] dimension"

Comments on unaligned assignments often stated that another dimension would have been better selected when submitted or that the assignment did not require students to evidence the essential skill dimension selected for review.

Example comments:

- "I do not see this assignment description or these artifacts as aligned with DL [Digital Literacy]. They may have been posted in an online forum, but that's not evident in the artifacts I'm looking at."
- Civic discourse might be better aligned than ethical reasoning"
- "It seems like none of the topics of the class referred to sustainability and the natural world, so this dimension should not have been selected for evaluation"
- "It is not clear why the second dimension was selected for these artifacts"
- "It seems rather than the third dimension this [artifact] could work better for the second dimension, since students are expected to reflect on the ideas and products of others rather than creating their own."

Difficulty rating assignments:

Raters sometimes noted difficulty rating assignments for reasons besides misalignment, such as duplicate submissions, assignment format, and errors or missing content in the files that were uploaded to the GE submission form.

Example comments:

- "This artifact does not provide enough data about the whole assignment to properly

rate the assignment and student work."

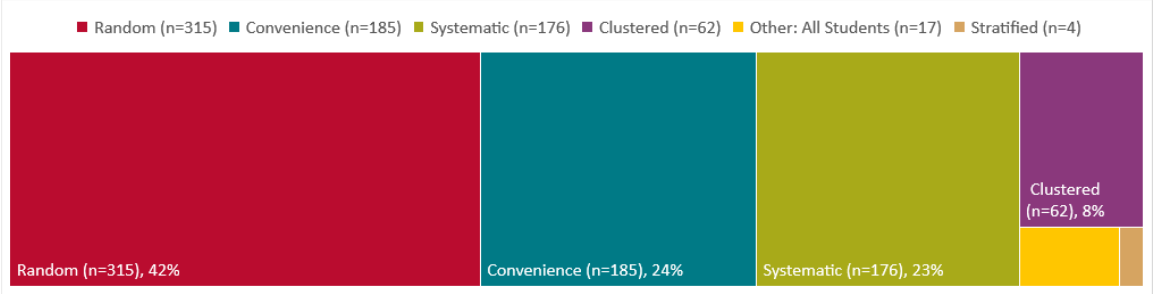
- "The screenshots of the website indicate that there may be more that was generated by the student for the digital literacy component, but based on the artifact alone, there is little evidence regarding the extent of digital literacy."
- "The reading response/discussion board format for artifacts is difficult to assess because there just isn't that much data there"
- "Tough time rating this artifact because it lacks substantive content"
- "Kind of weirdly formatted to the point that some of the info may be cut off and unviewable"

Other Observations:

Additional insights into the GE process came from OAAPR meetings and observations as well as graduate rater comments. The following findings arose from these meetings:

1. Graduate students are requesting GE assessment development to aid in their assignment designs.
2. GE essential skills and their assessment promote interdisciplinarity and unify content areas across UNM.
3. Assessment supports metacognition in students when they self-reflect and reflect on content.
4. Just choosing the right dimension can take an artifact from a 0 to a 3 rating when strongly aligned. Instructors reviewing the rubric dimensions before selecting them for rating is recommended, and reading them before the assignment is created is even better.
5. It can be difficult to have students evidence collaboration in their assignments. When collaboration is selected as an assessment dimension, student reflection and description of the collaborative process (and not solely saying they used a team approach) is solid evidence.
6. There were some cases where instructors submitted assignments that had been designed by TAs to apply to one dimension but were entered under an incorrect dimension in the submission form. This finding reiterates the importance of ensuring alignment at the dimension level, and it also highlights an opportunity for involving TAs in GE assessment.

13. Submissions by Sampling Method



The GE submission form asks instructors to define the sampling method they used for selecting artifacts from their course. Sampling frequencies were calculated using the total number of submission (n=759). Of all sampling methods, random sampling was listed as most common (42%). Next most common was convenience sampling (24%), followed by systematic (23%), and clustered (8%) methods. Two percent submitted all artifacts, and one percent were selected in a stratified method.

Student Demographics:

When submitting student artifacts for the GE assessment process, the OAAPR requests that instructors provide the affiliated UNM student Banner IDs. The OAAPR compiled the following information from these associated Banner IDs.

The number of confirmed students was 720 (some instructors submitted multiple artifacts from the same student, some artifacts came from groups of multiple students, and some artifacts could not be linked to a student). The following data is for all unduplicated students who could be linked to both a submitted artifact and a Banner record.

14. Percent Students Sampled by Registered Campus

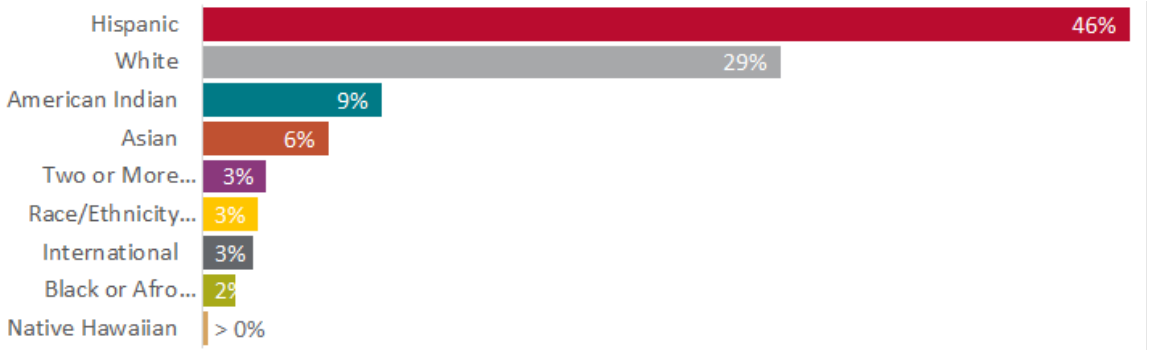
IDs show a good spread of students from across UNM campuses, with a slight oversampling of branch campuses.

- **Gender:** 61% female, 39% male
- **Majors:** Artifacts came from students across 122 different majors/pre-majors, including non-degree and undecided students
- **Student level:** 44% Sophomores, 21% Juniors, 18% Seniors, 11% other (high school, non-degree seeking undergraduate/graduate), 6% Freshman
- **Student date admitted:** First Year 54%, Second Year 24%, Third Year 8%, Four or

more years 6%, and 8% unknown

While the higher proportion of upperclassmen and low proportion of freshmen was initially concerning, looking at admission data shows that most of the sample does come from first-year students. This finding does however raise the question of if artifacts are being selected from the best students, who are more likely to rise quickly through credit hour rankings, than from students who are part-time or moving more slowly. This possible disparity is one of many reasons random sampling is so important.

15. Artifact Submissions by Race and Ethnicity of Students



On a more encouraging note, the race and ethnicity percentages among samples were highly similar to those found across the university in Fall 2022.

Assessment Implications:

It will be important to look for ways to sustain assessment participation in the GE program overall, and among individual colleges, schools, and branches.

Units need to explore ways to raise their overall scores as well as ratings within specific dimensions. One key way of doing so is for instructors to ensure alignment between the essential skill rubric and the assignments on which the submitted artifacts are based. The difference in scores between aligned and unaligned artifacts along with comments from graduate students show the importance of overall alignment and ensuring the correct dimensions are selected for review.

In addition, instructors should ensure that submitted artifacts are indeed the correct assignment, include the complete assignment, and are submitting artifacts in formats that allow raters to accurately determine how they directly connect to the essential skill.

The OAAPR will continue to communicate and provide support for the following areas of improvement:

1. Enhance GE rater training
2. Update submission form to support data collection and analysis
3. Design and provide TA development in GE assessment
4. Expand GE assessment outreach to include chairs and new faculty
5. Maintain the facilitation of GE Success Workshops in collaboration with CTL