Office of Assessment & APR:
AY 20-21 General Education Assessment
Results

Abstract

The University of New Mexico's Office of Assessment & APR (OA/APR) collected completed General Education student artifacts across colleges, schools, and branches in AY 20-21. These General Education artifacts represented one of the five essential skills and were used to assess information about UNM's new General Education programming.

With 841 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 alignment with the essential skill, (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 skill of Critical Thinking, 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 skill, 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 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. The Critical Thinking essential skill was assessed during this annual cycle.

Methods

To assess the development of Critical Thinking, UNM utilized an essential skill rubric 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.

*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 artifacts relating to Critical Thinking. 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 the Critical Thinking 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 associatedBanner IDs:
- Total number: **781 students** (some students had artifacts submitted from more than one instructor)
- Gender: **62.10% female**, 37.64% male, 0.26% not reported
- Student level: **38.54% Sophomores**, 25.22% Juniors, 17.16% Seniors, 6.66% Graduate Students, 6.40% Freshmen, 6.02% other (high school, non-degree seeking undergraduate, nursing levels, and not reported),
- Ethnicity: **48.14% Hispanic**, 33.55% White, 5.38% American Indian, 4.35% Asian, 3.33% Two or More races, 1.92% International, 1.66% Black or Afro American, 1.41% Race/Ethnicity unknown, 0.26% not reported

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 Critical Thinking and should not be

*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.

considered a benchmark.

Course Level: The **majority** of submitted student artifacts represented **1000** level courses. Artifacts from **2000** level courses **rated higher** on average than those from 1000 level courses.
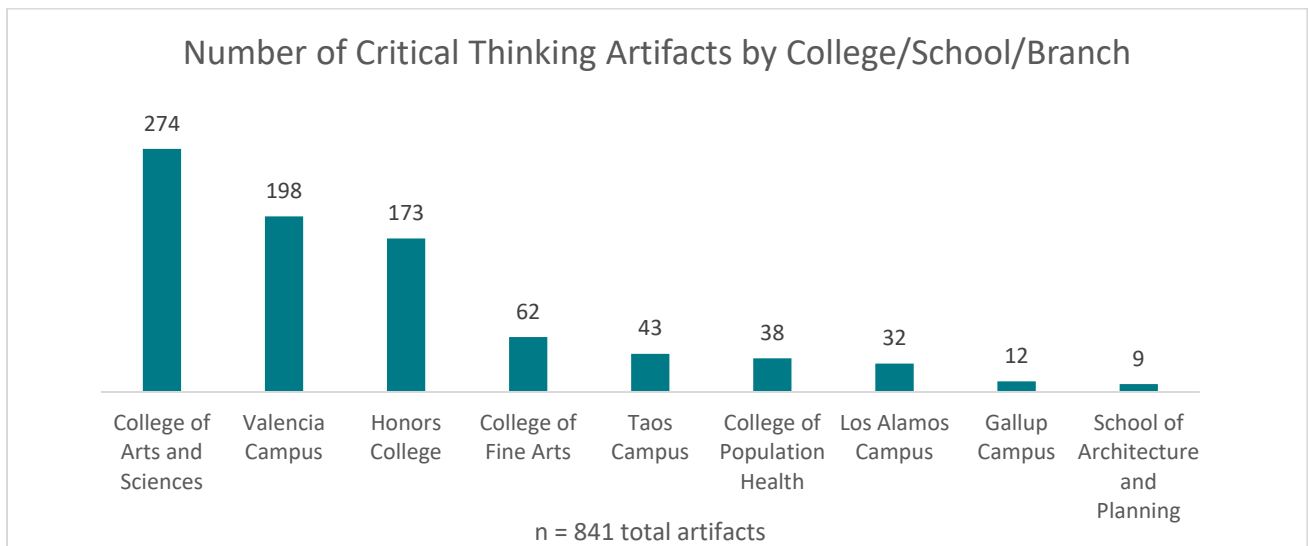
Course Modality: While artifacts from **face-to-face courses rated higher**, this modality had only 20 artifacts submitted. Online course submissions made up the majority of artifacts and rated "highly emerging" overall.

Class Size: Artifacts from **medium-sized** classes **rated highest** among the participants who submitted this information (the amount was much lower than the overall n).

Critical Thinking Skill: UNM colleges/schools/branches achieved an artifact rating average of 1.99 across all artifacts. Four colleges/schools/branches surpassed this average rating.

Critical Thinking Dimensions: Problem Setting is the highest rated dimension while **Evidence Evaluation** is the lowest rated.
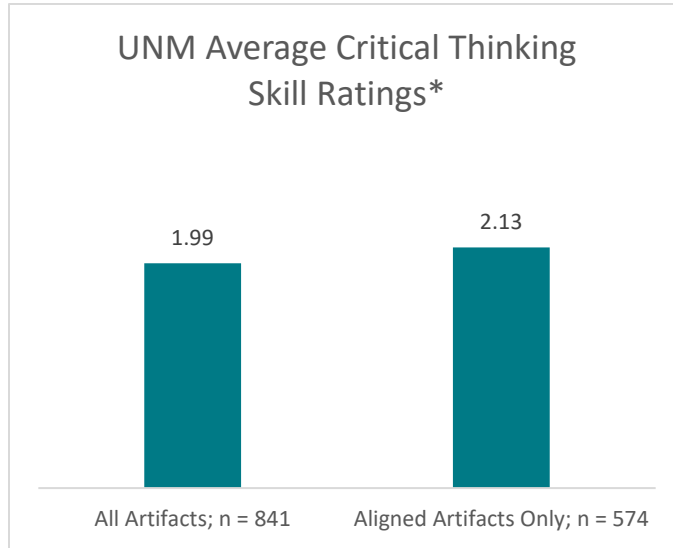
Total Artifacts Submitted

Number of Critical Thinking Artifacts by College/School/Branch

College/School/Branch	Number
College of Arts and Sciences	274
Valencia Campus	198
Honors College	173
College of Fine Arts	62
Taos Campus	43
College of Population Health	38
Los Alamos Campus	32
Gallup Campus	12
School of Architecture and Planning	9

n = 841 total artifacts
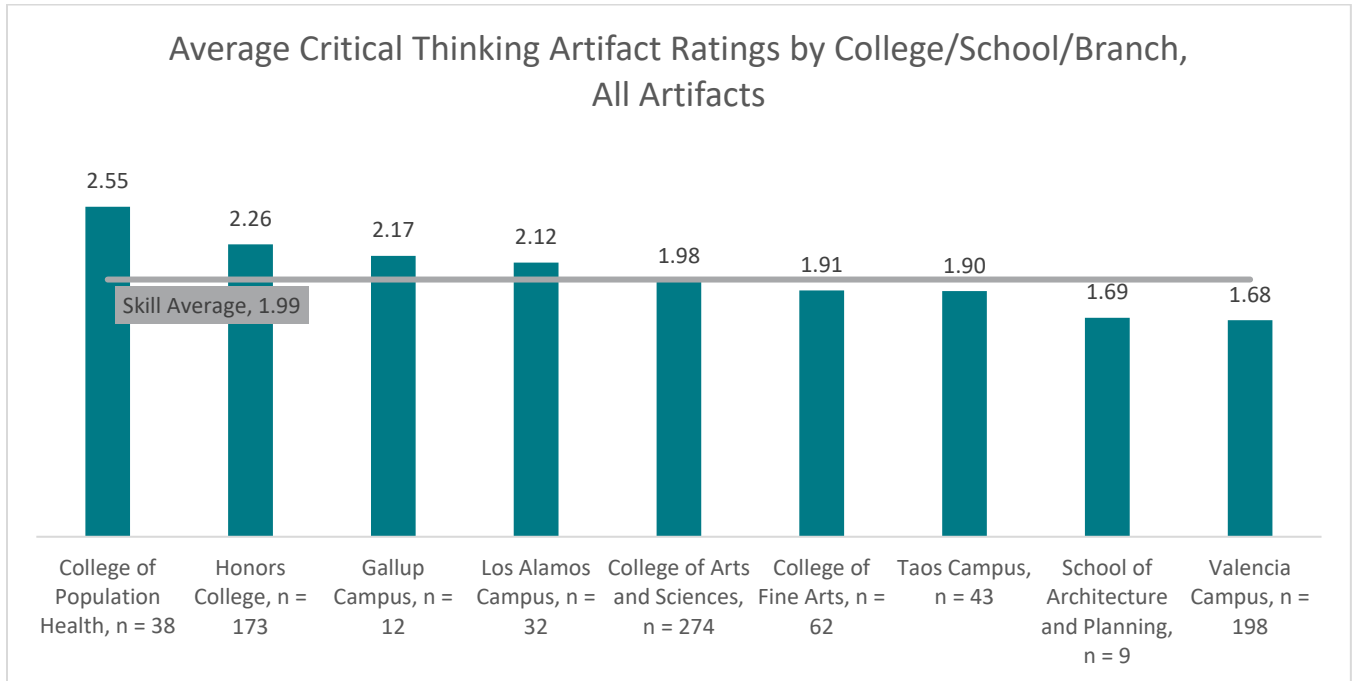
*as of May 27, 2021 (final data pull date)

*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 Critical Thinking Skill Ratings, All Artifacts vs Aligned Artifacts

UNM Average Critical Thinking Skill Ratings*

All Artifacts; n = 841 — 1.99

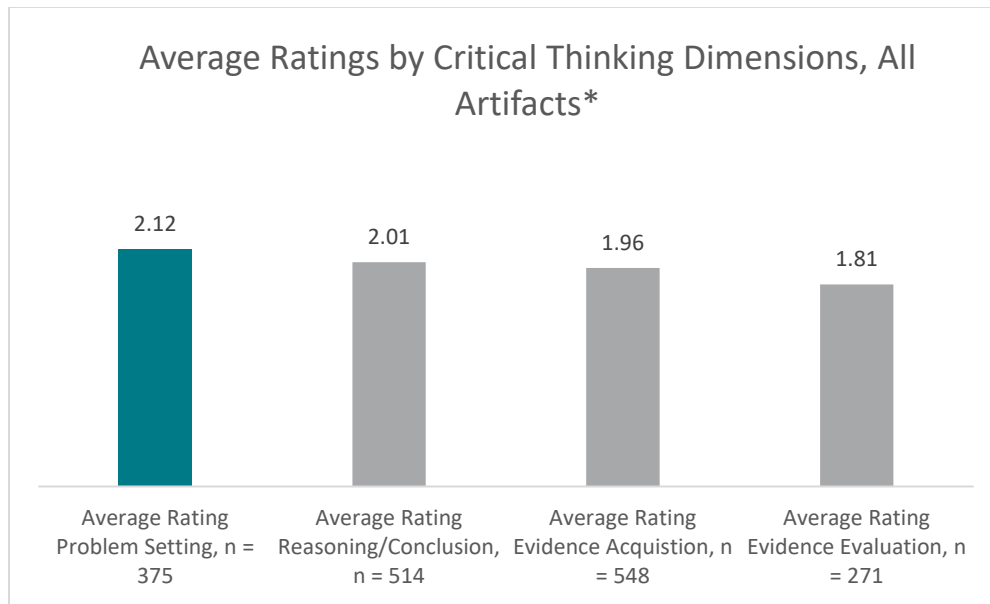Aligned Artifacts Only; n = 574 — 2.13

*All submitted artifact/assignment descriptions were assessed for alignment to the Critical Thinking rubric. This table illustrates ratings for artifacts that were found to be aligned, compared to the overall ratings of all artifact submissions, regardless of alignment.

Average Critical Thinking Skill Ratings, by C/S/B

Average Critical Thinking Artifact Ratings by College/School/Branch, All Artifacts

Skill Average, 1.99

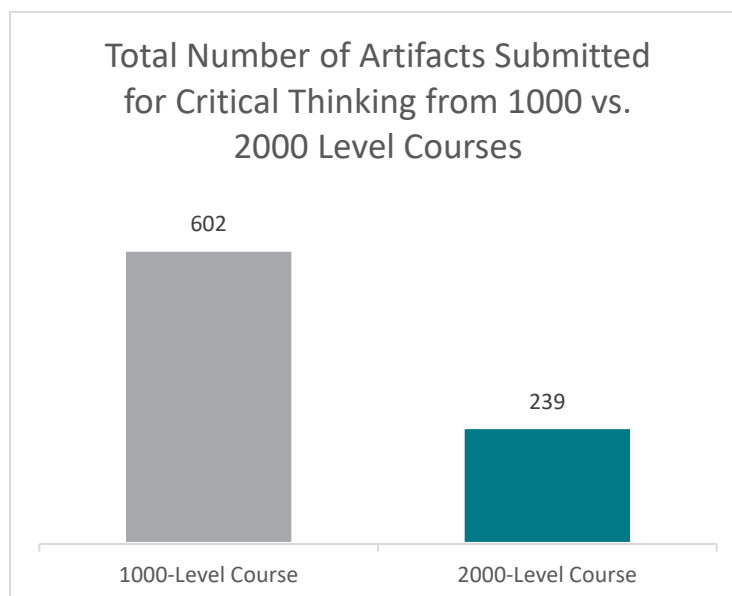| College/School/Branch	Rating
College of Population Health, n = 38	2.55
Honors College, n = 173	2.26
Gallup Campus, n = 12	2.17
Los Alamos Campus, n = 32	2.12
College of Arts and Sciences, n = 274	1.98
College of Fine Arts, n = 62	1.91
Taos Campus, n = 43	1.90
School of Architecture and Planning, n = 9	1.69
Valencia Campus, n = 198	1.68

*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 Critical Thinking Dimension

Average Ratings by Critical Thinking Dimensions, All Artifacts*

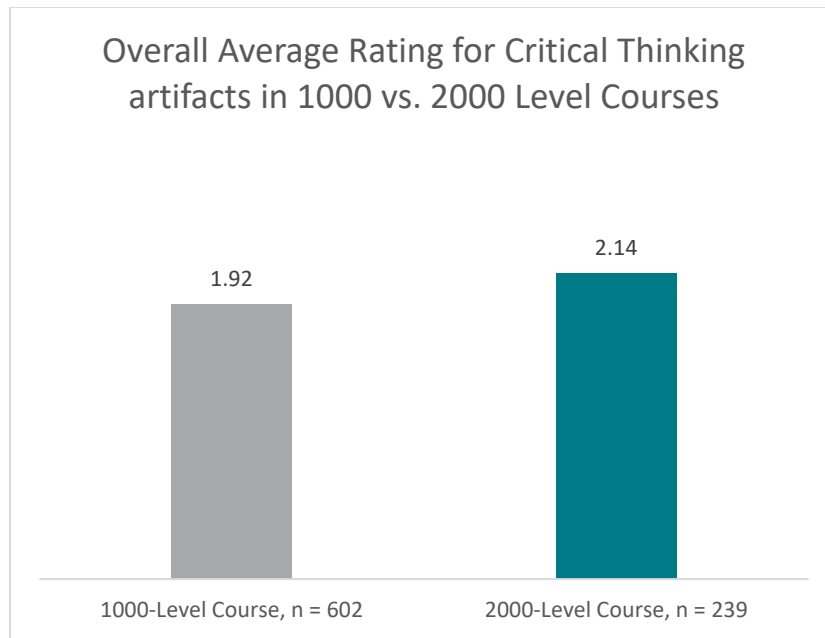|			
2.12	2.01	1.96	1.81
Average Rating Problem Setting, n = 375	Average Rating Reasoning/Conclusion, n = 514	Average Rating Evidence Acquistion, n = 548	Average Rating Evidence Evaluation, n = 271

*N-Value exceeds 841 artifacts as each submission was allowed to correspond to multiple dimensions

Total Number of Critical Thinking Artifacts Submitted, by Course Level

Total Number of Artifacts Submitted for Critical Thinking from 1000 vs. 2000 Level Courses

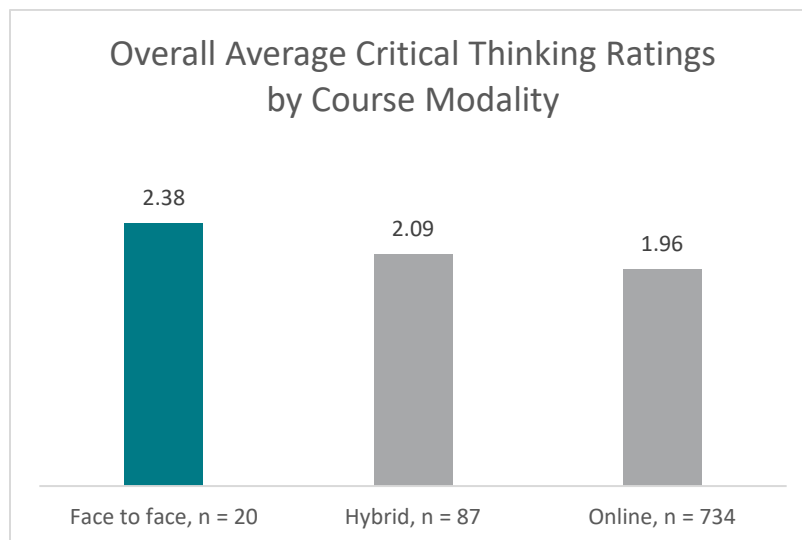|	
602	239
1000-Level Course	2000-Level Course

*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 Critical Thinking Artifact Ratings, by Course Level

Overall Average Rating for Critical Thinking artifacts in 1000 vs. 2000 Level Courses

1.92 — 1000-Level Course, n = 602
2.14 — 2000-Level Course, n = 239

Average Critical Thinking Artifact Ratings, by Course Modality

Overall Average Critical Thinking Ratings by Course Modality

2.38 — Face to face, n = 20
2.09 — Hybrid, n = 87
1.96 — Online, n = 734

*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.
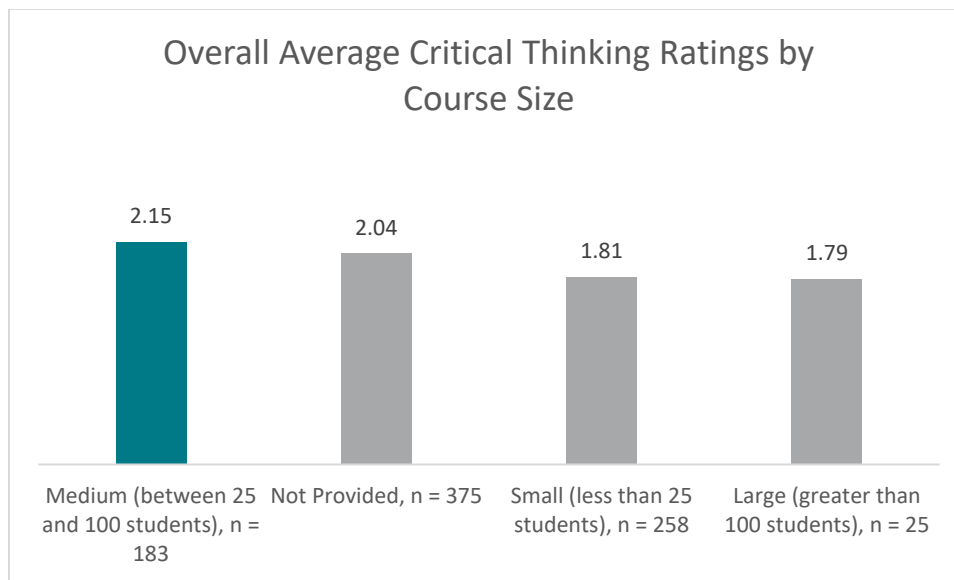
Average Critical Thinking Artifact Ratings, by Course Size

Overall Average Critical Thinking Ratings by Course Size

Course Size	Rating
Medium (between 25 and 100 students), n = 183	2.15
Not Provided, n = 375	2.04
Small (less than 25 students), n = 258	1.81
Large (greater than 100 students), n = 25	1.79

QUALITATIVE RESULTS

Qualitative Analysis of Graduate Assistant Commentary

Along with rating each submitted artifact, Graduate Assistants (GAs) provided written commentary to support their numerical rating and give context. Each artifact was reviewed by two GAs. The following is a qualitative analysis of their 1,312 comments, which resulted in 654 coded quotations; many were linked to several different codes for a grand total of 1,178 "code occurrences." Percentages listed below divide the frequency of the codes within a thematic grouping by this grand total of "code occurrences." In general, GA comments fell into one of four thematic groupings:

1. **Justification for rating (52%)**
2. **Student Performance (19%)**
3. **Artifact/Assignment Alignment with the Essential Skill or Dimensions (18%)**
4. **Challenges for Rating (11%)**

Justification for Rating

Because GAs worked in pairs and utilized the comments section to spark conversation around norming their ratings, many comments were coded into a "justification for rating"

*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.

theme. Of the 1,178 code occurrences, 52% (615) were coded as **justification for rating**. For example, one rater wrote:

> *For this set of artifacts 254-57, I have given all 3s; these students are masterfully demonstrating the acquisition of evidence while also thoroughly considering this evidence; because of the nature of the rhetorical analysis they're employing in these artifacts, they demonstrate proficient evidence evaluation as they consider the various viewpoints expressed in the speeches they're looking at - they each take a balanced role in considering the credibility of a variety of claims in these speeches.*

Student Performance
Additionally, many comments provided by the GAs fell into the **Student Performance** theme. Of the 1,178 code occurrences, 19% (221) were coded as either **Strong**, **Mixed**, or **Poor** Student Performance. Examples of these comments include:

> **Strong:** *WOW! What a great read! The problem is clearly outlined from the beginning, and the student provides all necessary information for the reader to track with their direction. Evidence is provided to sufficiently address the problem at hand, and there is thorough consideration of this evidence. What a great take on gender and witchcraft in Oz!*

> **Mixed:** *Student establish[ed] their question/problem from the beginning, investigating the creation of a COVID play, but the parameters are ambiguous in this problem set. The student gestures toward a conclusion in what they hope the audience would get out of a play of this nature, but the conclusion is oversimplified and tenuously related to the rest of the information provided.*

> **Poor:** *Student showed poor ability to evaluate evidence, and also poor ability to reach logical conclusions.*

While comments such as these serve the purpose of informing inter-rater reliability and ultimately improve confidence in the ratings overall, these comments do not provide the insight desired on improvements, recommendations, etc. for the GE assessment process. GA comments followed a similar pattern in the pilot year. Therefore, as detailed in the **Assessment Findings** section below, the Office of Assessment & APR (OA/APR) will change the GA rater form to alleviate this going forward.

Artifact/Assignment Alignment
18% of the code occurrences (217) emphasized the **importance of aligning the assignment/artifact with the Critical Thinking 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. One rater stated:

> *These are some of the few assignments where the students actually needed to acquire sources and yet we don't have to rate evidence acquisition!*

10

*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.

Another wrote:

> *The submitted artifact is not designed to correctly measure Problem Setting or Reasoning/Conclusion.*

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 Critical Thinking dimensions. This was especially true when the artifact/assignment required only a computational or short text response. One rater commented:

> *Way too short of an assignment to gauge critical thinking…this is a reading comprehension assignment*

Another GA pair wrote:

> *Since the assignment description (graph and equation) is not attached, we don't know if the student used the correct interval… [the student] just record[ed] the answers.*

Conversely (though less frequently), the GAs also highlighted examples where the assignments were well-aligned with the rubrics. Of a video artifact submitted, one GA wrote:

> *Great example of critical thinking through the oral channel. Good citations and analysis of the citations.*

Another wrote:

> *This assignment specifically focuses on critical thinking, which is helpful.*

Challenges

Finally, GAs noted several factors that made **rating more challenging** for them. 11% of code occurrences (125) were attributable to problems or difficulties with rating assignments/artifacts. 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 incomplete assignments or assignments that were either too long or too short. One rater noted:

> *This is a blank assignment, reads as "no response recorded" so assuming it was left blank by student.*

Another 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.

Super short assignment, did not demonstrate evidence acquisition from the video, ideas were not fully developed.

Assessment Findings

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

Data collection included the student artifact submission process, the GE workshops/training, communication, the inventory of data collected, and the forms/materials surrounding the student completed work collected. After evaluating last year's cycle, the OA/APR **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. This year, the form will also require a response for course size and teaching modality.

Qualitative data was provided by the raters through the use of note-taking sections in the analysis forms. The initial idea for the commentary section was to provide an opportunity for each rater to share observations on rubric and assignment alignment, in addition to providing examples of strong and poor dimension evidence. Even though these specific types of notes were requested, the **note-taking portion was often inadequate**, focusing mainly on justifying the rating selected or the student performances as noted in the commentary summary above**.** The OA/APR will focus more attention on the note section and the significance of the observations and narrative pieces in future trainings to ensure that they are completed. Next year's analysis forms will include **two separate columns** for GA notes/commentary. One column will request justification for their rating that they can use to discuss and norm with their partner, and a second column requesting specific examples of alignment of assignments to the essential skill rubrics and dimensions.

Communication Improvements

Communication was instrumental in dispersing information about the new UNM assessment process and will remain so in the subsequent assessment cycles. The OA/APR 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. This has proven to be especially true after analysis revealed that artifacts well aligned to the Critical Thinking rubric rated higher on average overall than those the GAs noted were not well aligned to the Critical Thinking rubric.

*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.

Additionally, 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**. Lastly, the OA/APR 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.

Expanded Training Opportunities

Training is critical for future GE assessments to be successful. The OA/APR believes that specific essential skill training for faculty and teacher assistants is necessary, especially in curriculum building, GE course design and certification, and assessment. The training 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 New Mexico Higher Education general education reform.

Rating and Analysis Process

Again this year, the OA/APR hired **six graduate students** to assist with the analysis process. The OA/APR staff does not have the capacity to analyze this large set of data (841 student artifacts this year) and discovered that the expertise and education of the graduate students was incredibly valuable. This opportunity is beneficial to the GAs as well - they gained educational assessment and general education knowledge/skills. **The OA/APR plans to replicate this analysis model in the future.**

Analysis includes the 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. This was replicated for this year's Critical Thinking artifact analysis. Bias awareness was similarly embedded in the analysis training and weekly meetings with raters. Rating issues remain - for example, some raters had difficulty creatively applying the Critical Thinking rubric to different kinds of assignments. 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.**

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

*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.

Importance of Rubric Alignment

Though they were not the majority, there were comments provided by the GAs where they identified assignment examples for further review. GAs noted that the **most difficult assignments to rate were the ones that had not been well-aligned to the rubric**. Assignments where students provided several pieces of work (portfolios); 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 an opinion were especially difficult to rate.

GAs 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**. GAs 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).

The graduate students also provided feedback to the OA/APR regarding the GE submission process. They suggested that the OA/APR should 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 that this "checklist" or flowchart (which could also include the above-mentioned **suggestions for preferred assignment types and formats)**, 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.