



**Office of  
Assessment & APR**

**AY23-24 GE  
Assessment Results**

August 2024

**NM** OFFICE OF  
ACADEMIC AFFAIRS



# Overview

## Introduction

This report provides 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 2023-24 academic year, the OAAPR assessed the Critical Thinking (CT) skill. This essential skill is assessed alone because CT spans all seven GE areas (I: Communication, II: etc.).

## Key Takeaways

- This cycle, CT had decreased ratings in every rubric dimension, compared to CT ratings in 2021. Evidence evaluation is the most challenging dimension for alignment of course assignments such as exams, papers, or homework, and for measurement of student performance overall. Not only are ratings highly dependent on assignments being aligned to the CT essential skill rubrics, but [critical thinking appears to be decreasing in students nationwide following on the 2020 pandemic \(Insight Assessment\)](#).
- Artifact submissions have increased this cycle, compared to the previous CT cycle. In 2021, 841 student artifacts, representing the work of 781 students, were submitted. In 2024, 894 student artifacts (up 6.3%), representing the work of 871 students (up 11.5%), were submitted.
- There was an emergence of student artifacts that raters suspected were generated using AI while analyzing the student submitted work (4.8% of coded responses). The use of AI in assignments is something to keep in mind as student artifacts continue to be rated in the upcoming years.

## Assessment Implications

It is important to look for ways to sustain assessment participation in the GE program overall, and among individual colleges, schools, and branches. The representation of student artifacts is crucial to evidence of the development of essential skills across diverse courses and content areas.

Units should explore ways to raise their overall scores and ratings within specific skill dimensions (e.g. Utilize their CARCS to review rubrics and GE customized reports). Instructors should ensure alignment between the essential skill rubric and the submitted artifacts. The difference in scores between aligned and unaligned artifacts, along with qualitative data from graduate student raters, show the importance of overall alignment and selecting the correct dimensions for review. A higher alignment means a more authentic student performance rating.

## Tips for Improvement

- Strengthen the connections between essential skills dimensions and assignment prompts. Consult the UNM GE assignment collection to explore robust examples of assignments with this dimension embedded into them. Attend an

annual GE workshop facilitated by the OAAPR and CT. Revisit the CT rubric to see which dimensions align with individual assignments.

- Use UNM's [GE assignment collection](#) of exemplary aligned assignments that embed essential skills and promote student skill attainment.
- Submit samples from a diverse population of students to offer a more representative picture of skill development.

## Methodology

### *Sampling Methods*

Each year, the OAAPR solicits student artifacts from instructors' assignments relating to the essential skill(s) designated for that year. Each college, school, and branch with GE certified courses is asked to sample artifacts randomly from multiple sections of GE courses in the content area(s) (depending on the size of the college, school, or branch) to which the essential skills being assessed are aligned [add link to NMHED alignment or to your own]. Instructors provide the two dimensions of the essential skill rubric for rating that the artifact was meant to demonstrate.

The OAAPR assesses evidence of GE skills in these artifacts using the relevant essential skills rubric. OAAPR hires graduate student raters to review the artifacts. These raters are trained in FERPA, GE assessment, and interrater reliability. To ensure reliability, raters are paired into groups of two and are assigned artifacts that were relevant to the rater's academic and content background.

Raters used the Critical Thinking essential skills rubric to rate submitted artifacts on a scale where 0 = No Evidence, 1 = Emerging, 2 = Developing, and 3 = Proficient. Each group discusses their ratings with their partners, and then meets with OAAPR staff to further discuss their ratings and rating process. Raters also annotate artifacts with qualitative comments to provide additional insights into their rating process. The OAAPR then conducts a quantitative analysis of these ratings and a qualitative analysis of the graduate students' process comments.

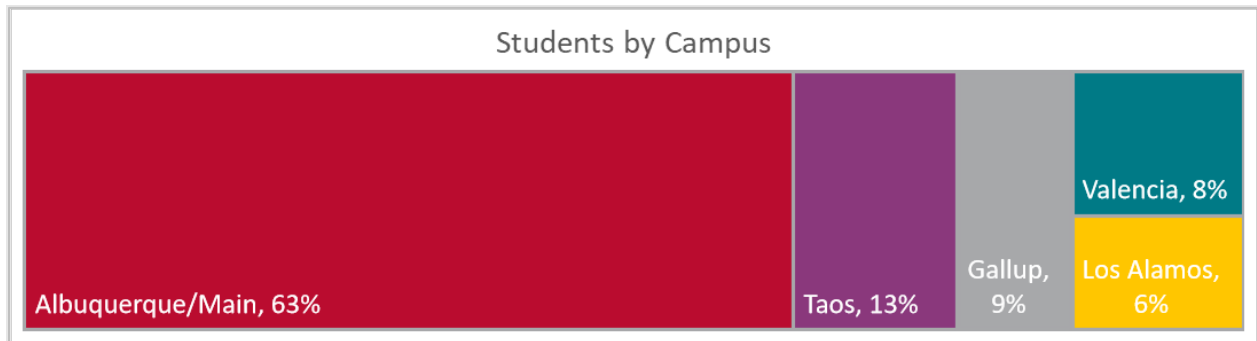
Ratings and comments reflect several variables: student performance, assignment alignment with the essential skills rubric, rubric alignment with student work, and the rubric dimension selection by each instructor.

### *Student Demographics*

When submitting student artifacts for the GE assessment process, the OAAPR requests that instructors provide the students' Banner ID. The following student demographic information is collected from Banner IDs.

This year, there were 871 students whose assignments were rated. Different instructors sometimes submit artifacts from the same student. Some artifacts may have come from groups of students. Some IDs could not be linked to a student due to IDs being entered incorrectly. The following data is for all unduplicated students who could be linked to a submitted artifact and Banner ID.

Graph 1. Percent students sampled by registered campus



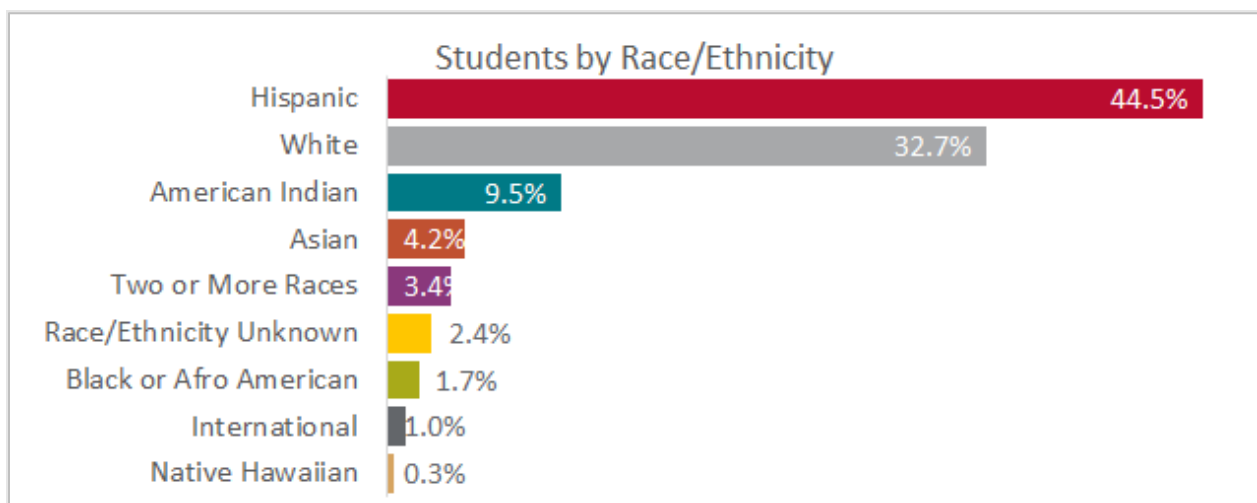
Student campus is a result of the submitting instructor. Frequently, students registered at branch campuses are enrolled in Albuquerque courses and registered Albuquerque campus students are enrolled in branch campus classes.

Based on the data collected from student Banner IDs:

- **Gender:** 64% female, 36% male
- **Student level:** 13% freshmen, 38% sophomores, 22% juniors, 14% seniors
- **Student years enrolled:** 54% first year, 25% second year, 8% third year, 3% fourth year, 3% over four years, 4% no data

While 37% of students are sophomores, 54% of students are in their first year at UNM. Many students come into UNM with college credits. This may be why more sophomores are represented in GE submissions, even though half of the sampled students are in their first year.

Graph 2. Artifact submissions by race and ethnicity of students

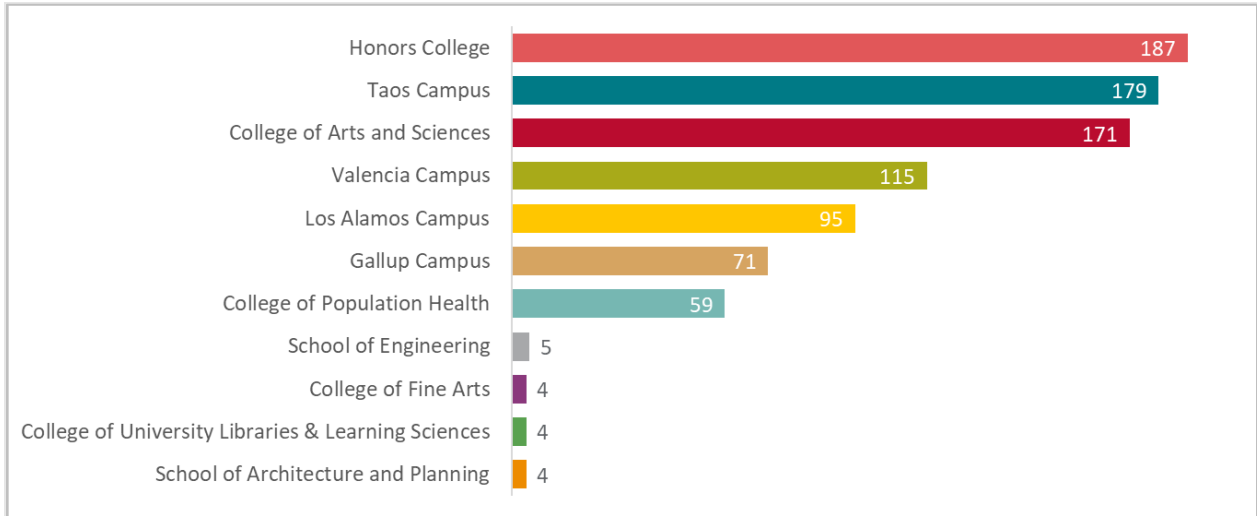


The sampled demographics are representational of UNM's AY23-24 overall demographic with only some minor differences.

# Results

## Artifact Submissions

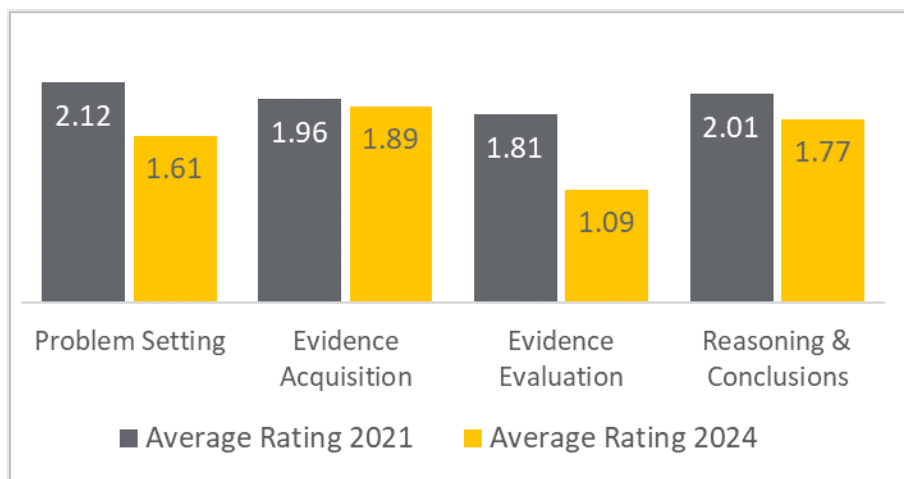
Graph 3. Artifacts submitted by college/branch



The OAAPR received 894 artifacts from eleven colleges, schools, and branches. Of the 894 artifacts, 885 were rated. Reasons for not rating submissions include duplicate submissions and inaccessible or unratable files.

This year saw a high number of submissions, with a large increase of artifacts from Honors College, Taos Campus, and College of Arts & Sciences.

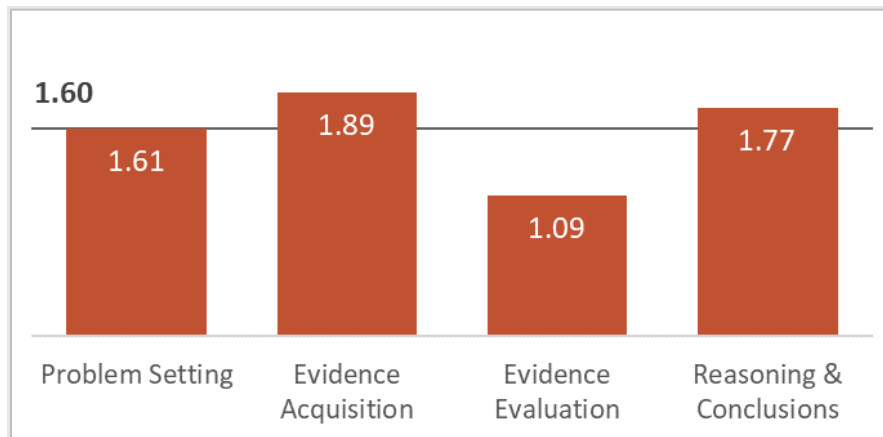
Graph 4. Comparison of average artifact ratings for Critical Thinking dimensions by years



The last Critical Thinking cycle was in 2021. Comparing this year's average ratings to the previous cycle, CT has decreased ratings in all dimensions.

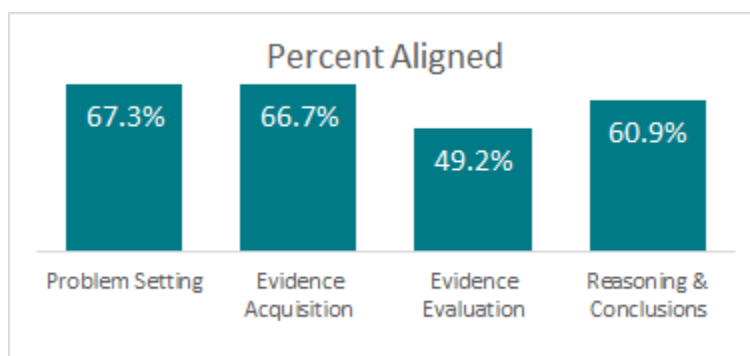
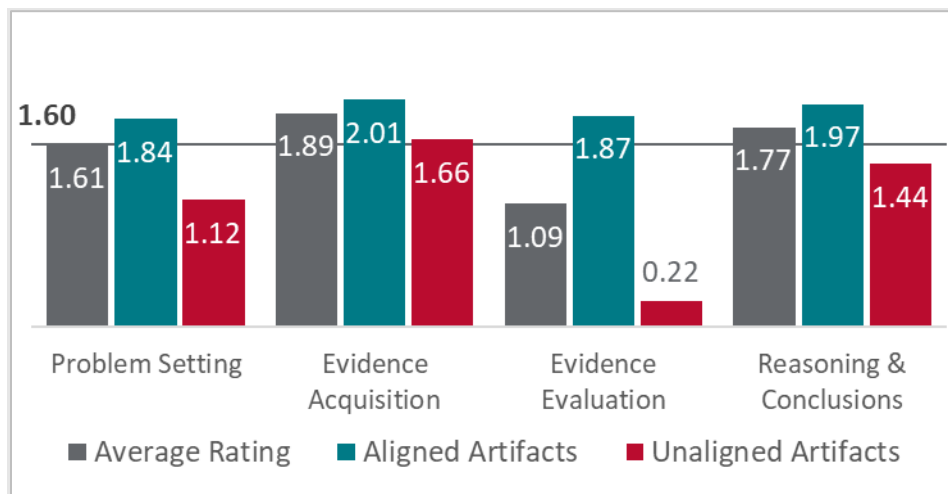
## Artifact Ratings

Graph 5. Average CT artifact ratings by rubric dimension



Evidence Acquisition (1.89) and Reasoning & Conclusions (1.77) had average ratings above the overall CT average of 1.6. Problem Setting was right above the average overall rating, and Evidence Evaluation had the lowest rating.

Graphs 6 and 7. Average artifact ratings by CT dimension for all artifacts



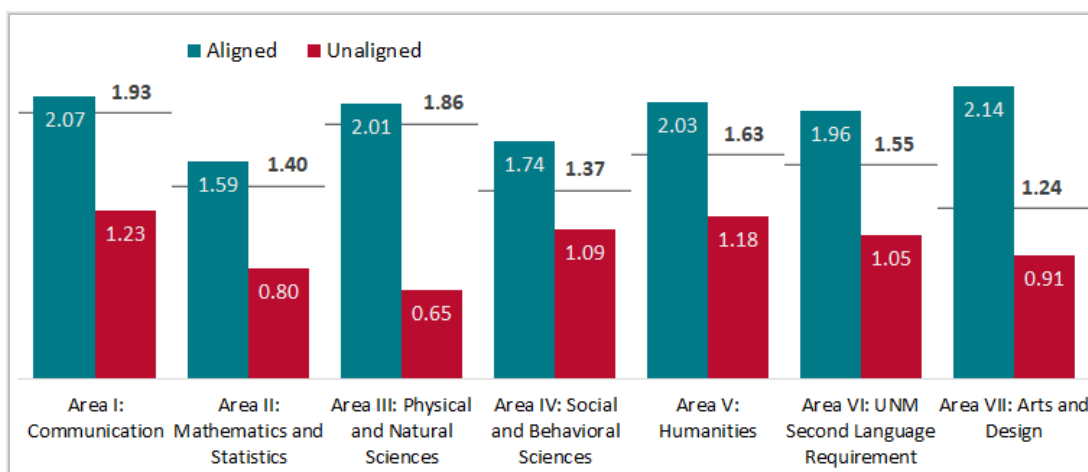
Problem Setting, Evidence Acquisition, and Reasoning & Conclusions had 60% or higher alignment, while Evidence Acquisition had the lowest alignment percentage. Though three of the four dimensions have >60% alignment percentages, there is still work that needs to be done to ensure even better alignment percentages. Those that were not aligned were rated significantly lower and lowered overall ratings. Differences among the specific skill dimensions are shown above. This cycle, CT had decreased ratings in every rubric dimension, compared to CT ratings in 2021. Evidence evaluation is the most challenging dimension for alignment and for student performance overall.

The overall average CT rating is 1.6, which means students are “developing” in this skillset.

*Relevant rater quotes regarding alignment:*

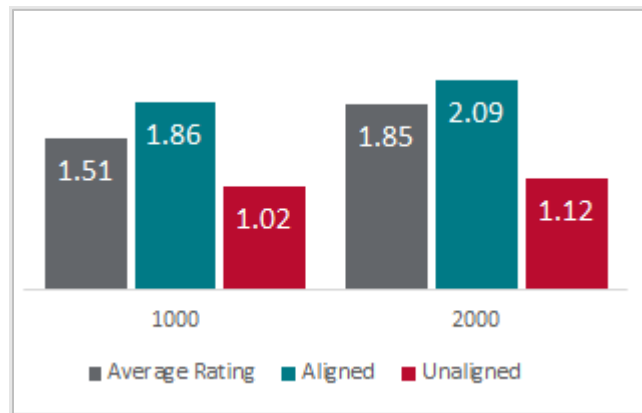
- "Assignment does ask for the domains to be shown, but not in a way that asks the students to provide a substantial amount of information that can reflect their critical thinking."
- "This assignment does not sound well aligned for the dimensions measured... It does not seem to me that this project would be effective for EE dimension at all, and problem setting seems mostly built into the mechanics of the assignment rather than requiring original student thinking. Perhaps reasoning/conclusion is the most aligned, but in general, this assignment seems unaligned with the dimensions."
- "These were exemplary artifacts for Critical Thinking in my opinion. Almost all the essays focused on emergent media and AI and posed nuanced arguments about the supposed or potential benefits, while also articulating the potential negative outcomes and impacts of these media forms."

Graph 8. Average artifact ratings by Critical Thinking and content area



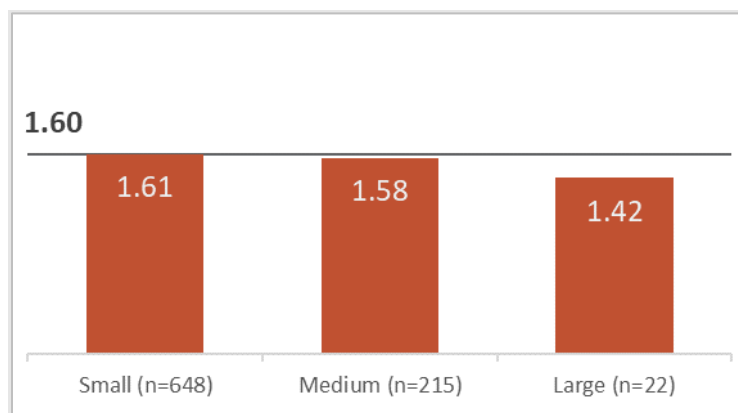
Arts and Design, Communication, Humanities, and Physical and Natural Sciences had the highest CT aligned ratings.

Graph 9. Average artifact ratings by Critical Thinking and course level



2000-level courses had higher CT ratings in both aligned and unaligned, and overall average rating, compared to 1000-level courses. An increase like this is expected as students move to higher level courses.

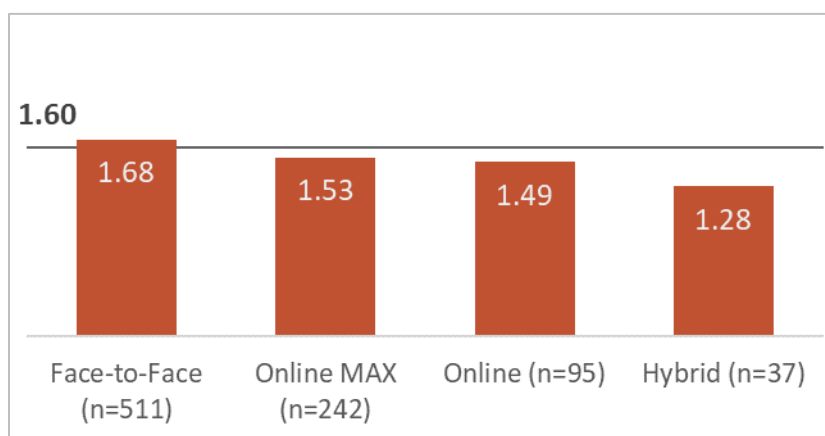
Graph 10. Average artifact ratings by Critical Thinking and course size



Courses were grouped into small (<25), medium (25-100), or large (>100) sizes. Most submissions (73.2%) came from small courses, and very few (2.5%) came from large courses. Small courses had an average rating of 1.61, just above the overall average rating of 1.6.



Graph 11. Average artifact ratings by Critical Thinking and course modality



57.7% of artifacts came from traditional face-to-face courses. Online MAX course had the next highest amount (27.3%). Face-to-face courses rated at 1.68, above the overall average rating of 1.6. The OAAPR recognizes that some fully online courses are offered as part of accelerated online programs (AOPs). For the purpose of this report, which seeks to capture skills development in the General Education Program, fully online courses are grouped in Online MAX with no separate category for AOPs.

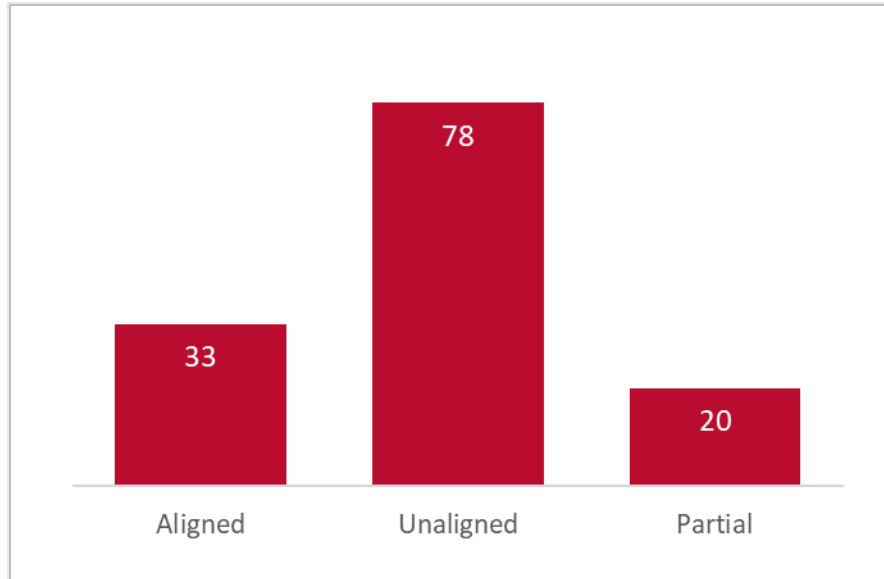
## Qualitative Themes

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 patterns earlier in this report and are further expanded on below. There were 230 coded responses.

### Alignment

#### *Code breakdown:*

- **Aligned** - comment mentions that the assignment and/or artifact aligns with Critical Thinking rubric criterion and selection.
- **Unaligned** - comment mentions that the assignment and/or artifact does not align with Critical Thinking rubric criterion and selection.
- **Partial** - comment mentions that the assignment and/or artifact is partially aligned with Critical Thinking rubric criterion.



Of 131 responses coded on alignment, 78 were unaligned, 33 were aligned, and 20 were partially aligned.

### **Relevant Quotes**

- Aligned
  - "Each of the artifacts from Honors 2112 are exemplary. They thoroughly meet the proficiency rating of the dimensions evaluated. Further, the dimensions of Evidence Evaluation and Reasoning/Conclusion seem applicable and met with proficiency; however, the latter 2 dimensions were not selected for consideration by the instructor."
  - "The prompt for these artifacts really set students up for success in measuring the listed content areas. It provided clear guidelines and structures which I think helped students complete the project thoroughly with regards to meeting the proficient dimensions of the content areas."
- Did not align
  - "This assignment does not sound well aligned for the dimensions measured. The assignment is to create a job resume and a one-page cover letter (the latter is the artifact we are given). It does not seem to me that this project would be effective for EE dimension at all, and problem setting seems mostly built into the mechanics of the assignment rather than requiring original student thinking. Perhaps reasoning/conclusion is the most aligned, but in general, this assignment seems unaligned with the dimensions"
  - "Beginning of a lot of Inherit the Wind assignments; instructor should have selected problem setting or evidence acquisition instead of evidence evaluation, especially with the 1st essay prompt not requiring any evaluation or comparison/contrast."
- Partial alignment
  - "The assignment is to suggest potential Wikipedia edits. This could in theory be aligned with the dimensions, but the brevity of the assignment

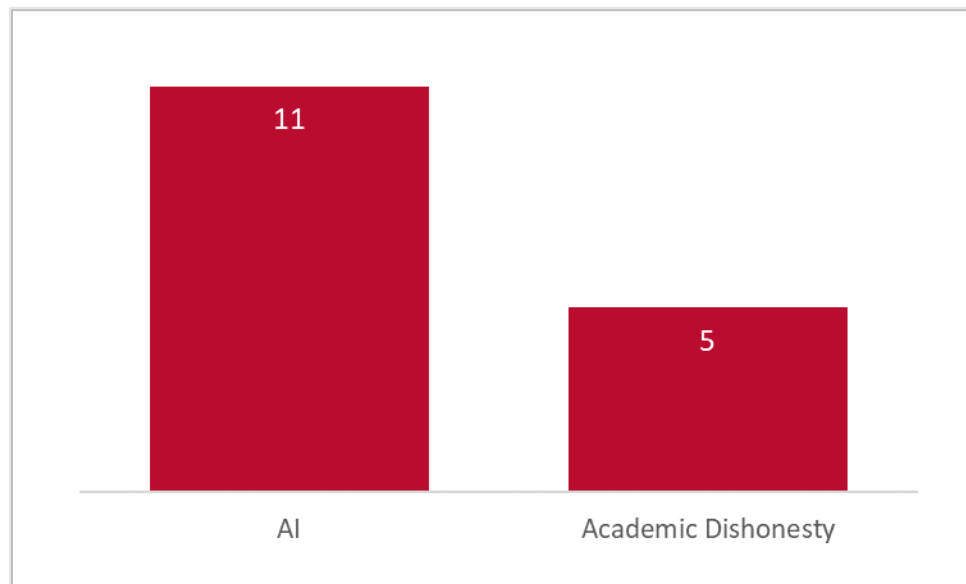
structure makes it feel less aligned and difficult to rate because of the sparing content in the artifacts that students actually created themselves. Not much evidence of critical thinking in general because of the assignment."

- "Can this be interpreted as aligned with evidence evaluation? I think they all necessarily meet the requirement for a 1 but the instructor should have chosen evidence acquisition. Students evaluate the credibility of claims by agreeing or disagreeing with statements, however, those statements were provided by the instructor and are not actually evidence."

## Academic Dishonesty

### *Code breakdown:*

- AI - rater believes that the student used AI to complete their assignment when assignment did not instruct them to use AI.
- Academic integrity- rater questions the academic integrity of the student's work for an assignment. (The artifact may not reflect original work produced by the student).



Of the 16 comments coded to academic dishonesty, 11 were on the use of AI and 5 were on academic dishonesty.

### *Relevant Quotes*

- "This one feels like it could likely be an AI response! It is an outlier from the rest of the group in terms of length and quality of response. The way information is listed, and bullet pointed resembles the way an AI would present information. Can't be positive of course but have a very strong feeling based on comparison to the others from this group."
- "These two assignments are identical. Plagiarism?"

## Assignment (Role of instructor)

### *Code breakdown:*

- **Conflicting assignment information** – parts of the assignment had conflicting information or instructions (i.e., not specifying needing a conclusion, but including a rubric that grades for a conclusion).
- **Missing assignment information** – parts of the assignment provided were missing, based on what reviewers saw from the artifact.



Of the 23 responses coded on assignment (instructor), 20 were on conflicting information given and 3 were about missing assignment information.

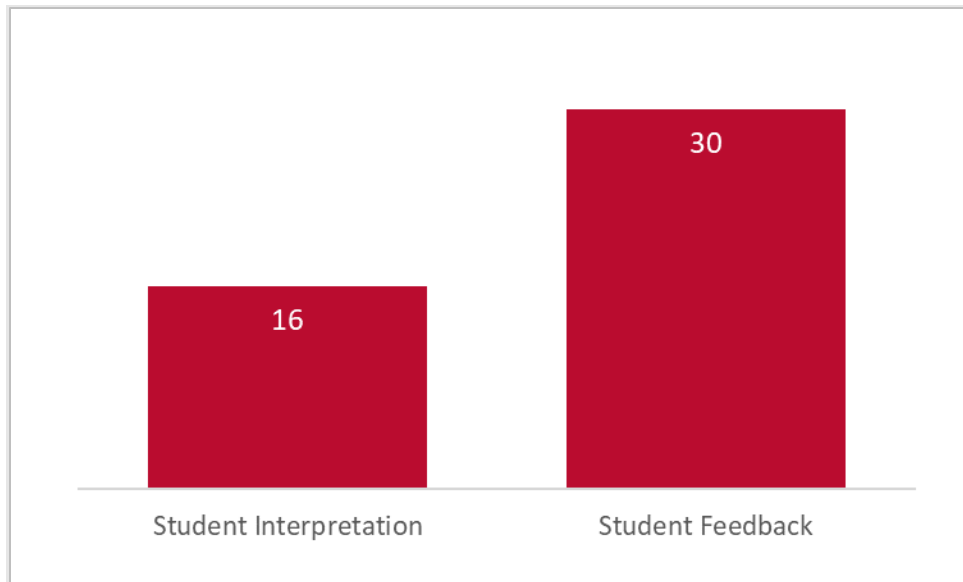
### *Relevant Quote(s)*

- "Prompt does not require that they have a conclusion, but it is evident that the outline they were given for the assignment has conclusion section."
- "Instructors need to provide more than just a simple question"

## Assignment (Role of student)

### *Code breakdown:*

- **Student interpretation** – Rater had a difficult time interpreting what a student wrote or submitted, whether it was because the student had poor handwriting or because the student themselves misinterpreted the assignment.
- **Student feedback** – Rater gave feedback on the student's work.



Of the 46 comments coded to assignment (student), 16 comments were on student interpretation and 30 were on student feedback.

#### ***Relevant Quote(s)***

- "Had to consider the drawings provided as the evidence [acquisition], but it was very difficult to read because it was journal drawings as well as the students putting notes in (reading students handwriting was difficult)."
- "The artifacts are photos from handwritten journals. It is very hard to read and discern what students are submitting... So these are in fact aligned at this point in my opinion, but the nature of the assignment and artifact made it very difficult to rate due to legibility."

## **Contact Us**

For more information on UNM's general education assessment, please go to our website: [assessment.unm.edu](https://assessment.unm.edu). For specific requests and support to help guide your assessment, please reach out to [assess@unm.edu](mailto:assess@unm.edu).