



# ULO1 Written Communication Assessment Report

## Fall 2022

### **Purpose of Assessment Project:**

In the last assessment cycle, Written Communication assessment scholars examined equity in Capstone courses. For the academic year 2021-2022, the group of scholars decided to continue that work by examining courses meeting the Graduation Writing Assessment Requirement (GWAR) of the CSU. We wanted to explore equity differences in performance in GWAR classes during the fall 2020 semester and consider whether those equity differences mapped onto particular areas of the rubric that integrates written communication criteria with those associated with critical thinking and information literacy (referred to as the integrated written communication rubric).

### **Facilitator:**

Nelson Graff, ULO WC Assessment Coordinator

### **ULO Assessment Scholars:**

Aimee Escalante  
Arlene Haffa  
Corin Slown  
Patrick Belanger  
Ryan Eller  
Sunita Lanka  
Timothy Orme

### **Assessment question(s):**

- How do students in Capstone classes perform according to the CSUMB integrated WC rubric?
- To what extent do equity differences in performance line up with particular criteria on that rubric?



## **Procedure:**

We began our annual coop by considering the results from the GVAR assessment in summer of 2021. As one of the findings from that assessment was that grammar may have played too big a role in faculty evaluation of student work, we spent some time reading about and discussing issues of grammar instruction and assessment at the college level. Although these were fruitful discussions, they did not result in any changes to our assessment practices or the integrated written communication rubric. We did, however, revise the integrated written communication rubric to simplify it and make the language more student friendly. We piloted that revised rubric for assessment in summer 2022.

In spring, 2022, we began the assessment process by gathering data about course gpa equity gaps in capstone classes during the fall 2021 semester from the student data warehouse. Although there were many possible distinctions to explore, we chose as most convenient the distinction between students categorized as underrepresented minorities and those categorized as non-underrepresented minorities. We chose to focus on the three classes with the lowest (HDFS, CART, CST) and highest (CHHS, SBS, MSCI) equity gaps between URM and non-URM students.

For summer assessment, TLA collected student work from those classes, from which two scorers scored a total of 138 student samples. Scholars had access to the artifacts, assignment prompts, original rubrics if available, and model essays if available. Two scholars assessed student artifacts for each course independently using the revised [CSUMB Integrated WC rubric](#). In the case of splits larger than 2 levels or 2-3 level splits, the scholars met via Zoom to discuss and resolve the scores. Scholars met to reflect mid-day on the first day, at the beginning of the second day, and at the end of the day during both days of assessment.

## Key Findings

### Student achievement aggregate

#### Percents without N/A or Unscorable

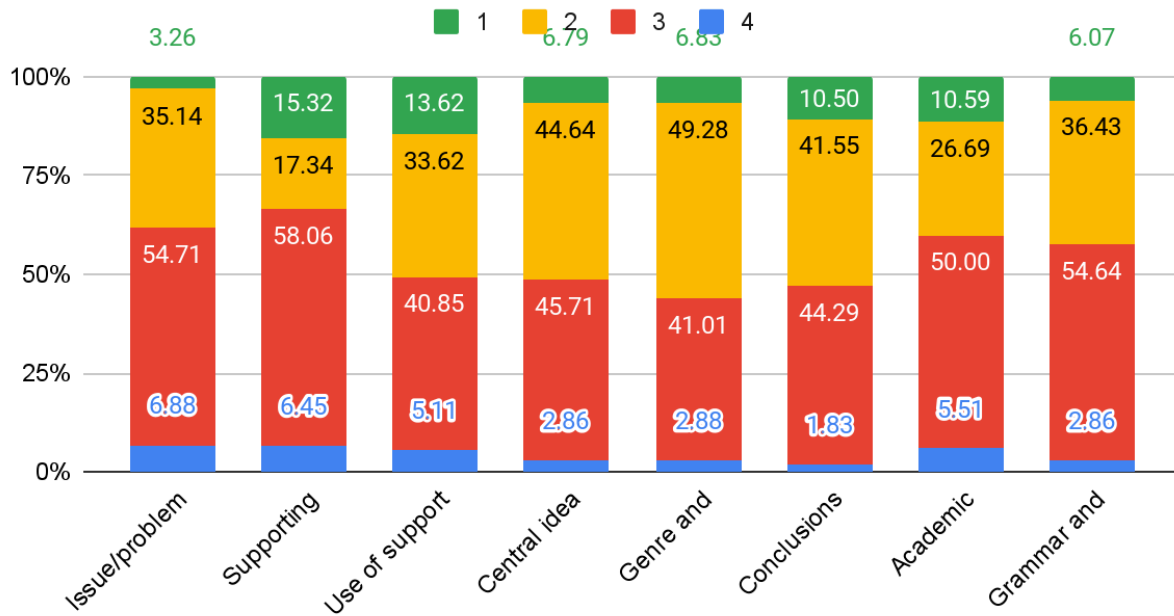


Figure 1. Percent of scores on capstone essays with N/A or unscorable scores removed.

	Issue/pr oblem (n=276)	Supporti ng material s (n=248)	Use of support (n=235)	Central idea (n=280)	Genre and disciplin ary conventi ons (n=278)	Conclusi ons and outcome s (n=219)	Academi c integrity (n=236)	Gramma r and mechani cs (n=280)
Percent of 3 & 4 scores	61.59	64.48	45.96	48.57	43.89	46.12	55.51	57.5

Table 1: Percent of scores in summer 2022 capstone that were 3 or 4 (proficient) after eliminating N/A and unscorable ratings. One class and a portion of another could not be scored on some criteria, which accounts for the different number of scores for different criteria.

	<b>Issue/problem (n=282)</b>	<b>Supporting materials (n=244)</b>	<b>Use of support (n=282)</b>	<b>Position (n=256)</b>	<b>Genre and disciplinary conventions (n=282)</b>	<b>Conclusions and outcomes (n=256)</b>	<b>Academic integrity (n=282)</b>	<b>Grammar and mechanics (n=282)</b>
Percent of 3 and 4 scores	61.7	76.23	60.64	51.57	58.51	46.1	75.89	57.09

Table 2: Percent of scores in summer 2021 GVAR that were 3 or 4 (proficient). One class and a portion of another could not be scored on some criteria, which accounts for the different number of scores for different criteria.

### Interpretation

Looking just at the percentage of scores without not-applicable or unscorable, a few observations stand out. First, on none of the criteria are 75% of students earning a 3 or 4, though nearly two-thirds of students earn in that range on supporting materials. The criteria on which fewer than 50% of students score as proficient or above are “use of support,” “central idea,” “genre and disciplinary conventions,” and “conclusions and outcomes.” We speculate that the reason for these low scores is that the types of writing that students do for their capstone are mainly only focused on in their GVAR classes, which means that they have few opportunities to practice and develop their skill over multiple iterations.

The differences among criteria are also suggestive. Judging from the difference in scores between “issue/problem” and “conclusions and outcomes,” students are better at writing essay introductions than they are at writing conclusions. The lower score on “use of support” suggests the need for more explicit instruction in information literacy, specifically synthesis, across the disciplines. The lower scores in “central idea” and “conclusions and outcomes” suggest the need for more critical thinking instruction. The lower scores in “genre and disciplinary conventions” suggests the need for greater explicit attention to genre in disciplinary instruction across the curriculum.

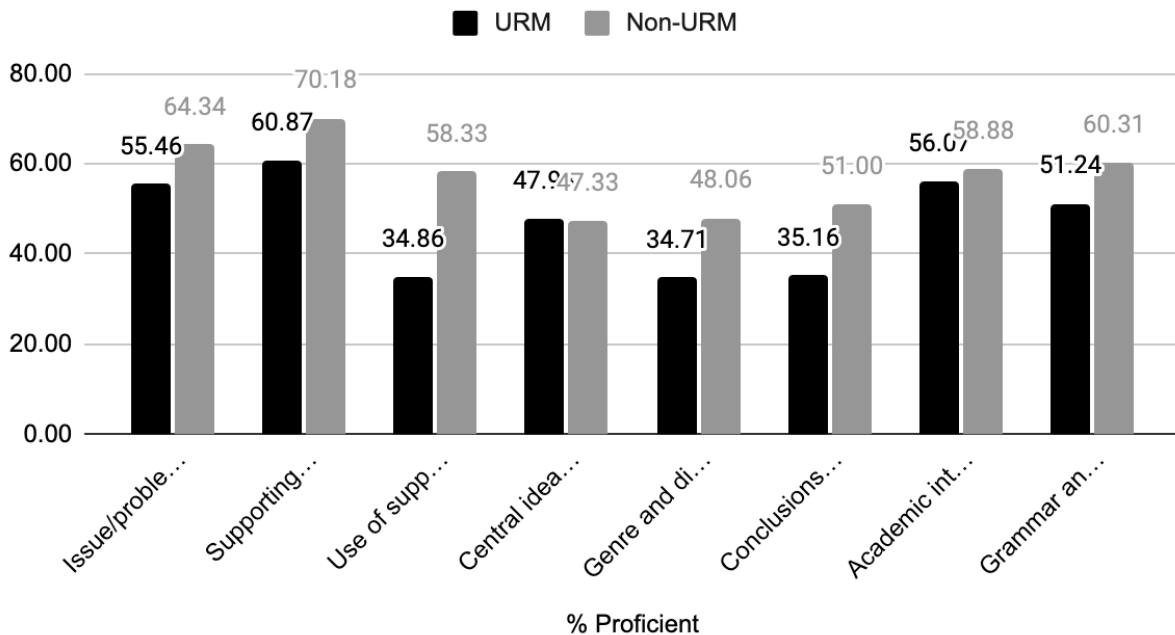
Contrasting tables 1 and 2 (scores for capstone this year versus GVAR last year) is striking. Some of the differences may relate to changes in the descriptors on the rubric, as we revised

the rubric between scoring sessions. However, it is notable that other than “issue/problem” and “grammar and mechanics,” all of the scores on capstone projects are markedly lower than those in GWAR. This is, of course, exactly the opposite direction we hope that scores will take between the two classes. That decrease in scores may support the inference that students are not sufficiently practicing key writing skills that they learn in GWAR between their GWAR classes and capstone. However, the variety of projects submitted for assessment may have also had an impact. While some courses sent forward traditional research papers, other artifacts were research or project posters, film descriptions, or paragraphs summarizing and justifying research methods and hypotheses.

### Student achievement disaggregated

#### Proficiency Comparisons

#### Percent proficient (without N/A or unscorable)



% Proficient	Issue/problem (CT)	Supporting materials (IL)	Use of support (IL)	Central idea (CT)	Genre and disciplinary (WC)	Conclusions and outcomes (CT)	Academic integrity (IL)	Grammar and mechanics (WC)
	55.46	60.87	34.86	47.94	34.71	35.16	56.07	51.24
	64.34	70.18	58.33	47.33	48.06	51.00	58.88	60.31

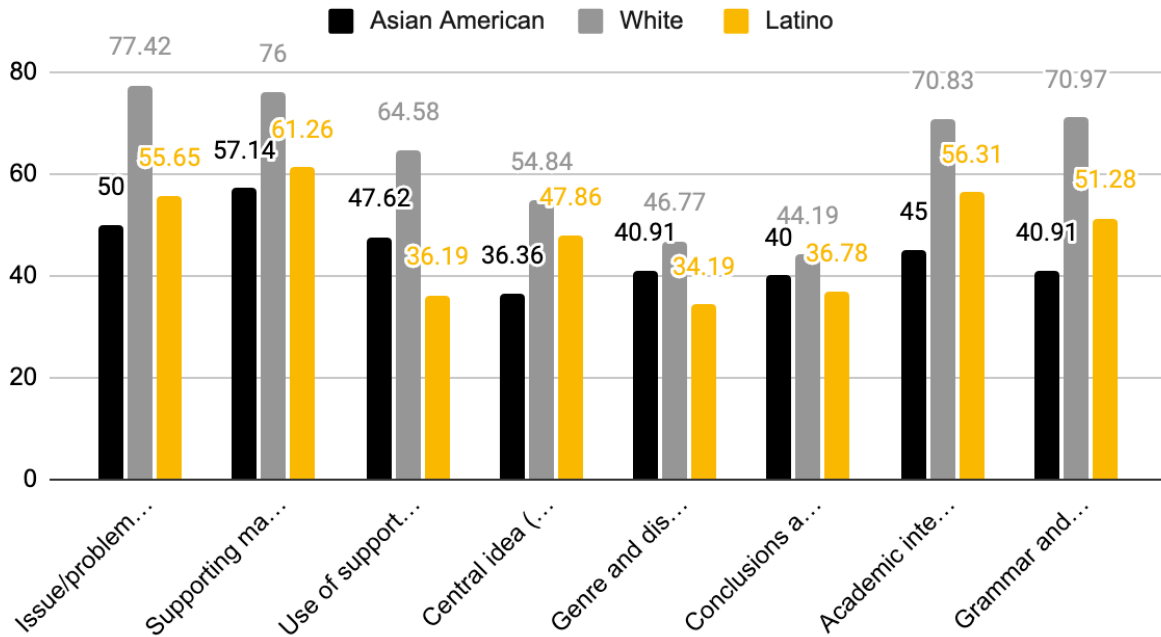
					conventions (WC)			
URM	55.46	60.87	34.86	47.93	34.71	35.16	56.07	51.24
Non-URM	64.34	70.18	58.33	47.33	48.06	51.00	58.88	60.31

Table 3: Percent of scores for students coded as underrepresented minorities and non-underrepresented minorities in summer 2022 capstone that were 3 or 4 (proficient) after eliminating N/A and unscorable ratings. One class and a portion of another could not be scored on some criteria, which accounts for the different number of scores for different criteria.

### Interpretation

Disaggregating between students labeled under-represented minorities and those not so labeled reveals substantial differences in the proportion of students who score proficient on all criteria except for Central Idea, for which the difference is both smaller and in the reverse direction (that is, more students labeled URM score proficient than non-URM). For “Use of Support” and “Conclusions and Outcomes,” the proportion of students labeled URM fall below the 50% threshold, while they are above that threshold for students labeled non-URM. These results suggest that we are not doing enough writing *instruction* across the curriculum, or such instruction has not been effective for students labeled as URM. Whether that is due to faculty instruction that does not meet the needs of such students (is not culturally sustaining) or differential preparation for which we are not accounting in our instruction is impossible to determine from these data.

## Percent 3 or 4 scores (without N/A or Unscorable) by Race



	Issue/pro blem (CT)	Supportin g materials (IL)	Use of support (IL)	Central idea (CT)	Genre and disciplina ry conventio ns (WC)	Conclusi ons and outcome s (CT)	Academi c integrity (IL)	Grammar and mechanic s (WC)
Asian American	50	57.14	47.62	36.36	40.91	40	45	40.91
White	77.42	76	64.58	54.84	46.77	44.19	70.83	70.97
Latino	55.65	61.26	36.19	47.86	34.19	36.78	56.31	51.28

### Interpretation

As with the difference between students labeled URM versus non-URM, it is clear that a higher proportion of White students is performing proficiently than either Asian American or Latino students at the capstone level. In some areas a higher proportion of Asian American students is performing proficiently than Latino students (Use of Support, Genre and Disciplinary Conventions, Conclusions and Outcomes); in other areas, a higher proportion of Latino students are performing proficiently (Central Idea, Academic Integrity, Grammar and Mechanics). It is



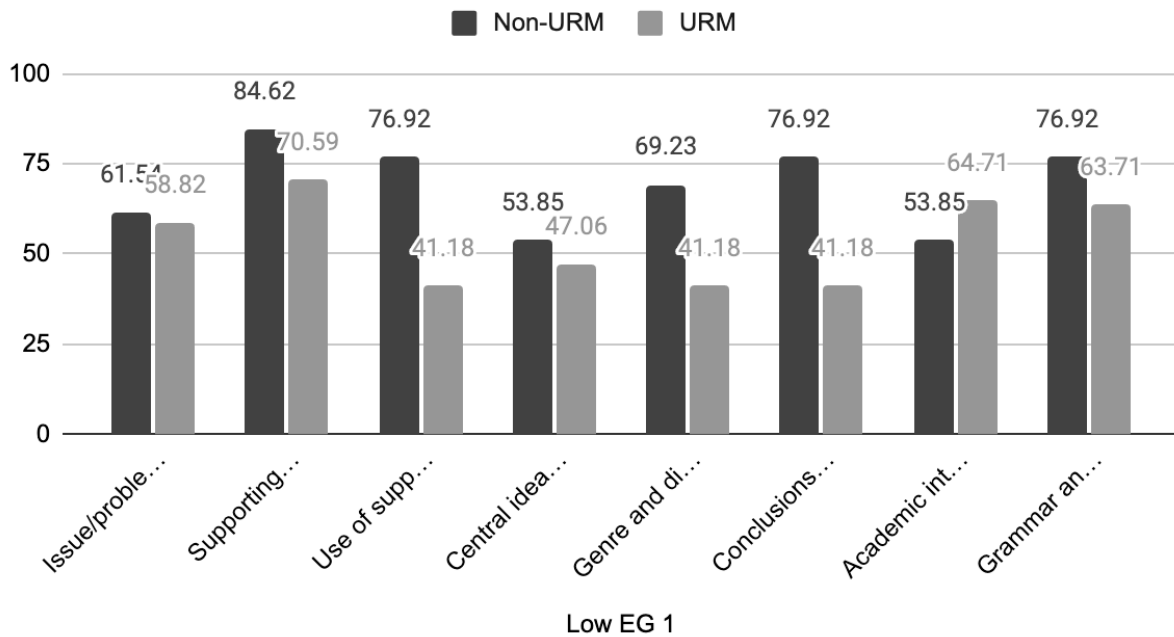
difficult with these data to distinguish whether these differences arise from disproportionate representation by one group or another in different majors (and the performances in those capstone classes), whether they represent a difference in preparation that has not been addressed at CSUMB, or whether they represent differential uptake of instruction because of lack of culturally responsive instruction or differential evaluative options/strategies.



### According to equity gaps in classes

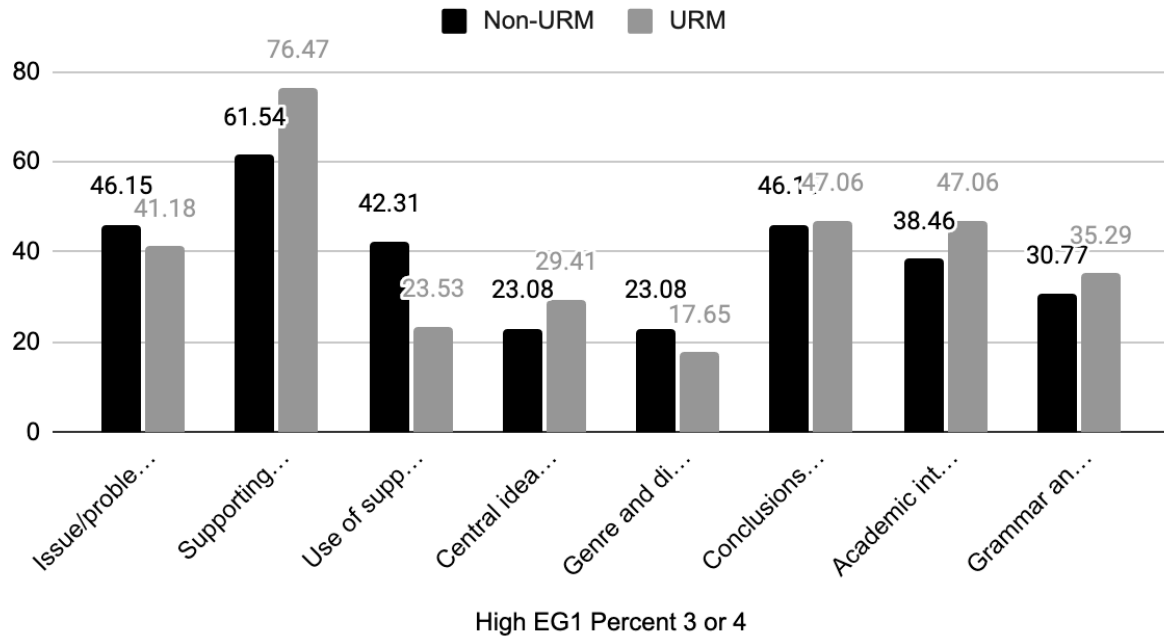
Below, we compare the percent of students scoring proficient (3 or 4) in classes ranked as having the lowest equity gap (low eg) between students labeled URM and non-URM and those in classes ranked as having the highest equity gap (high eg). Rather than identifying the specific classes, we identify them only as Low EG1, High EG1, and High EG2.

### Low EG 1 Percents Proficient or Above



Low EG 1	Issue/problem (CT) (n=26, 34)	Supporting materials (IL) (n=26, 34)	Use of support (IL) (n=26, 34)	Central idea (CT) (n=26, 34)	Genre and disciplinary conventions (WC) (n=26, 34)	Conclusions and outcomes (CT) (n=26, 34)	Academic integrity (IL) (n=26, 34)	Grammar and mechanics (WC) (n=26, 34)
3 or 4 (non-URM)	61.54	84.62	76.92	53.85	69.23	76.92	53.85	76.92
3 or 4 (URM)	58.82	70.59	41.18	47.06	41.18	41.18	64.71	63.71

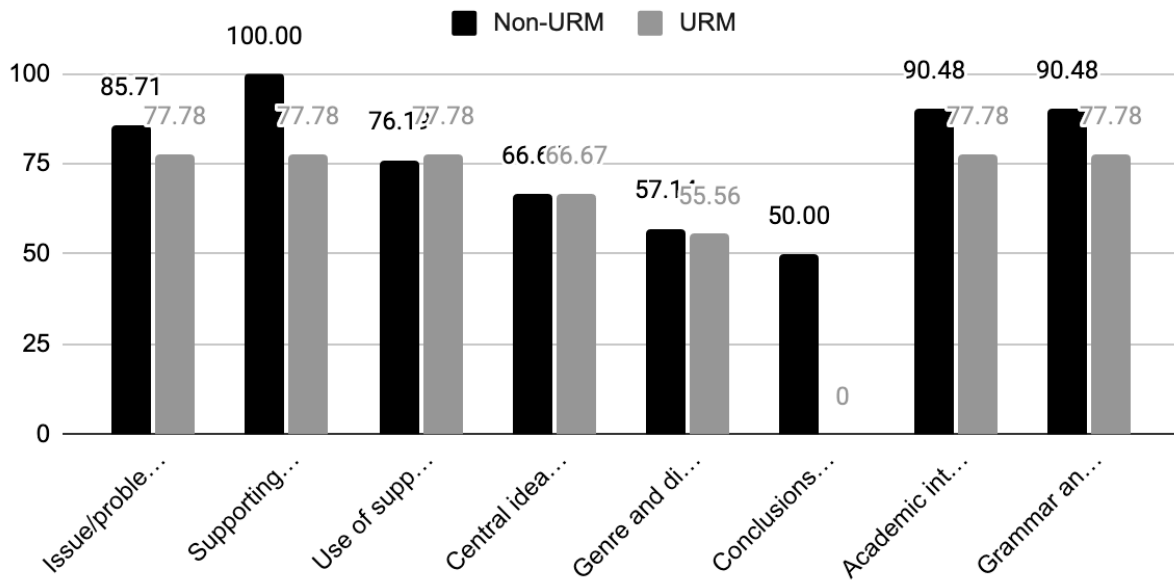
## High EG1 Percent 3 or 4



## High Equity Gap 1

High EG1 Percent 3 or 4	Issue/problem (CT) (n=26)	Supporting materials (IL) (n=26)	Use of support (IL) (n=26)	Central idea (CT) (n=26)	Genre and disciplinary conventions (WC) (n=26)	Conclusions and outcomes (CT) (n=26)	Academic integrity (IL) (n=26)	Grammar and mechanics (WC) (n=26)
Non-URM	46.15	61.54	42.31	23.08	23.08	46.15	38.46	30.77
URM	41.18	76.47	23.53	29.41	17.65	47.06	47.06	35.29

## High EG2 Percent 3 or 4



High EG2 Percent 3 or 4

High EG2 Percent 3 or 4	Issue/problem (CT) (n=42)	Supporting materials (IL) (n=42)	Use of support (IL) (n=42)	Central idea (CT) (n=42)	Genre and disciplinary conventions (WC) (n=42)	Conclusions and outcomes (CT) (n=16)	Academic integrity (IL) (n=42)	Grammar and mechanics (WC) (n=42)
Non-URM	85.71	100.00	76.19	66.67	57.14	50.00	90.48	90.48
URM	77.78	77.78	77.78	66.67	55.56	0	77.78	77.78

### Interpretation

The short and very self-contained writing submitted in high EG2 classes may account for the abundance of proficient scores in those classes, though it is impossible to rule out quality of writing instruction as a factor in those scores. (The absence of any proficient scores for students labeled URM on “Conclusions and related outcomes” stems from the fact that most of the scorers rated that criterion not applicable, though some scored it, apparently only for students who were labeled non-URM.) For this group, the only noticeable differences appeared in the criteria “Issue/problem” and “Supporting materials.”



In low-equity gap class 1, a higher proportion of students labeled URM scored proficient on academic integrity than those labeled non-URM. Otherwise, students labeled non-URM scored substantially better in all criteria, though the difference was least in “Issue/Problem.”

In high-equity gap class 1, the scores in general are lower than in low-equity gap class 1, and the pattern of differences in proportions of students scoring proficient reverses for supporting materials and academic integrity (and slightly for conclusions and related outcomes). That may suggest that writing instruction is less effective in this class for all students but that there is greater cultural responsiveness in that instruction. Or it may suggest that the task is significantly more difficult in this particular class.

### Limitations & Ongoing Work

The most obvious limitations of this work are two: lack of representativeness and variety of genres. While we analyzed artifacts from three colleges, we did not analyze artifacts from all of the colleges. As importantly, the low numbers of some groups of students--African Americans and Native Americans in particular--made it difficult to draw any inferences about the quality of instruction they are receiving and specifically whether that instruction is culturally sustaining or not. Those low numbers themselves, however, offer some indictment of CSUMB's current performance in the areas of recruitment and retention.

In addition to lack of representativeness, the samples we analyzed differed widely in genre. As noted above, while some courses sent forward traditional research papers, other artifacts were research or project posters, film descriptions, or paragraphs summarizing and justifying research methods and hypotheses. Because the tasks themselves called for applications of different skills, it is difficult to compare among courses and therefore to derive any solid inferences about student performance across the curriculum.

One important area for future work would be to try to understand *why* student performances differ. Some explanation may be open to statistical investigation--for instance, comparing student high school achievement and achievement at or near graduation. Such a comparison would help us to see whether we are perpetuating inequities in preparation with which students enter the university. Investigating the actual pedagogy in the classes would be more difficult, though faculty and students could respond to surveys to provide some information. We had hoped to correlate these assessment results with the Campus Climate Study, but those results are not yet available. In the future, it might be helpful to disaggregate using different criteria--such as first-generation/non-first-generation and Pell-eligible or not.



## Recommendations

The lower score on “use of support” suggests the need for more explicit instruction in information literacy, specifically synthesis, across the disciplines. The lower scores in “central idea” and “conclusions and outcomes” suggest the need for more critical thinking instruction. The lower scores in “genre and disciplinary conventions” suggests the need for greater explicit attention to genre in disciplinary instruction across the curriculum. That “issue/problem” typically earns higher scores than “conclusions and outcomes” suggests that faculty are doing a better job of explicitly teaching students strategies for introducing their essays than concluding their essays.

The differences in performance between those labeled under-represented minorities and those not so labeled raise important implications for us as faculty. More professional development in culturally responsive and sustaining instruction might help faculty better support minoritized students. Additionally, more explicit attention to how we compensate for differences in student preparation when they arrive in college is needed. In particular, studying these rubrics may help faculty consider specific interventions that can move students from scoring 2 to 3 on this rubric.

For populations of students whose numbers in graduating classes are low (Native American, African American), we may need to use additional measures to assess our instruction because we are limited in the conclusions we can draw from the data collected so far.

Finally, in spring 2023, the assessment scholars reviewed the assignments from the summer 2022 assessment and generated the following list of recommendations about assignment prompts:

- 1) We encourage faculty to use the [Purpose, Context, Task, Criteria for Success \(PCTC\) framework](#) adapted from the [Transparency in Learning and Teaching \(TiLT\)](#) project.
- 2) Composing assignment prompts with a supportive tone seems likely to improve performance particularly for students who doubt that they belong in college.
- 3) Clearly defining key terms on assignment prompts may improve performance.
- 4) Providing guidance for how sources should be used may improve students’ synthesis.
- 5) Assignments should be as brief as possible while still providing comprehensive instructions.
- 6) Examples are helpful.
- 7) It is worth considering the aesthetics of the assignment prompts and proofreading carefully.