California State University MONTEREY BAY

ULO1 Written Communication Assessment Report Fall 2023

Purpose of Assessment Project:

The Written Communication Assessment Scholars have spent the last several cycles considering equity in writing classes--moving from A2 to GWAR to Capstone. For the 2022-2023 academic year, the scholars decided to return to lower-division writing to consider how students are doing now and how we're doing with regard to equity gaps in student performance. Instead of focusing on A2 courses, however, we moved to A3 to see how students are doing *after* A2 (for those who completed A2 before taking A3).

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):

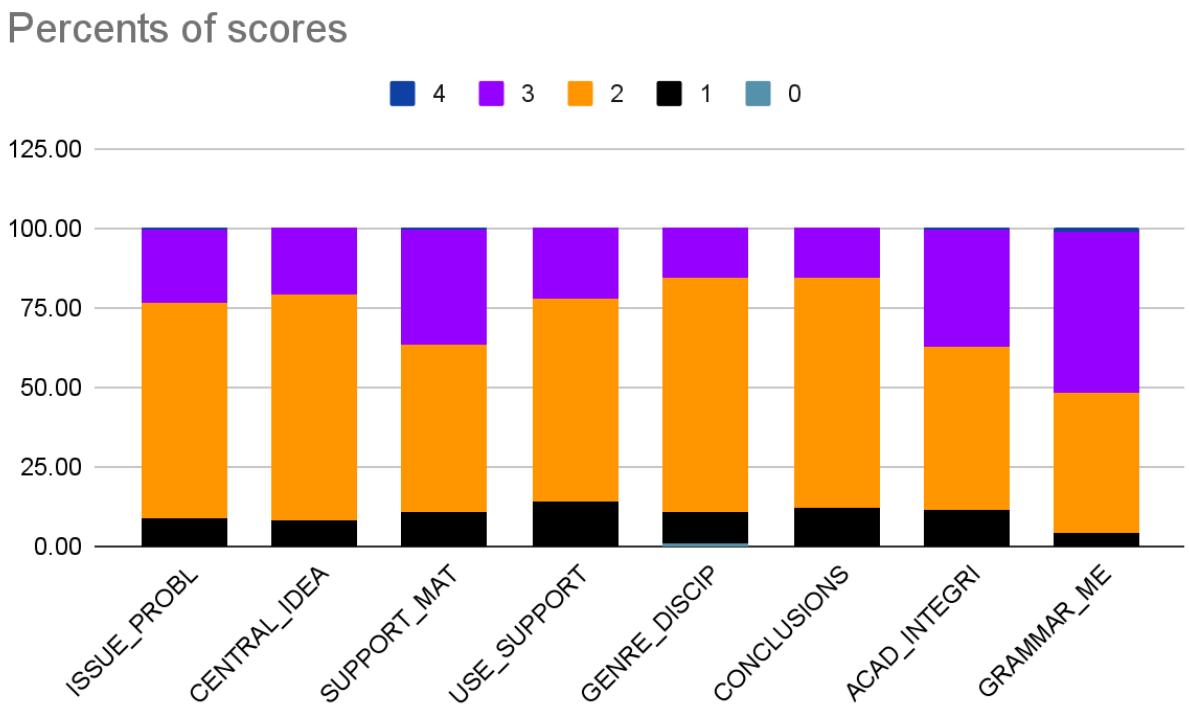
- How do students in A3 classes perform according to the CSUMB integrated WC rubric?
- To what extent do equity differences in performance line up with particular criteria on that rubric?

Procedure:

For summer assessment, TLA collected student work from A3 classes, from which two faculty scored a total of 134 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

Overall



	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
% Proficient	23.08	20.77	36.54	22.31	15.38	15.44	36.92	51.54
% benchmark or above	91.15	91.92	88.85	85.77	89.23	88.03	88.46	95.38

Interpretation

Because A3 is a lower-division class, we would not expect most samples to reflect achievement at the proficient level, though we would expect them to score at benchmark 2 or above. Fortunately, between 85 and 90% of the scores fall into that range. The lowest percent of scores at benchmark or above are in the area of use of support. Similarly, that criterion has a relatively low percentage of scores at the proficient

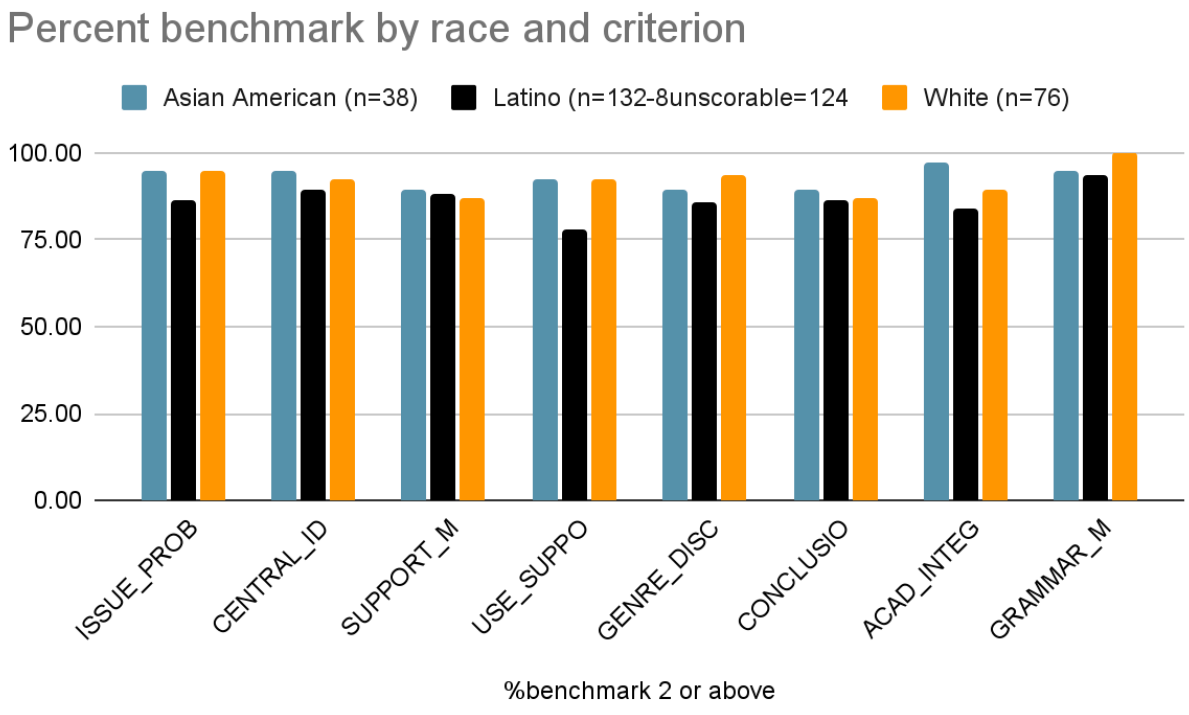
level. Synthesis is an advanced skill, and it is in synthesizing sources that students move their essays from benchmark 2 to proficient.

The lowest areas in terms of proficient scores are "genre and disciplinary conventions" and "conclusions." As we inferred from last year's results, students need more direct instruction in drawing conclusions from their analysis and composing those conclusions into effective essay endings. What distinguishes benchmark from proficient performance in most of the other criteria seems to relate to the lower score in "genre and disciplinary conventions." For instance, for "issue/problem," in order for an essay to score as proficient, the student must "explain[the central issue's] importance in a context appropriate to the discipline." A similar attention to context is required for proficient performance in terms of "central idea" and "supporting materials." This seems logical because A3 is a lower-division class, likely taken before students have engaged in much study in their majors, and the class itself is somewhat a-disciplinary, being less an introduction to a particular discipline than a general introduction to critical thinking.

One interesting observation is the distinction between the percentage of proficient scores at the A3 (51.5%) and capstone (57.5 %) in terms of grammar and mechanics. While it might seem problematic that the two numbers are so close to each other, the writing in capstone is much more complex than that in A3, so it is not surprising that students' grammar may not be as much improved as we might hope.

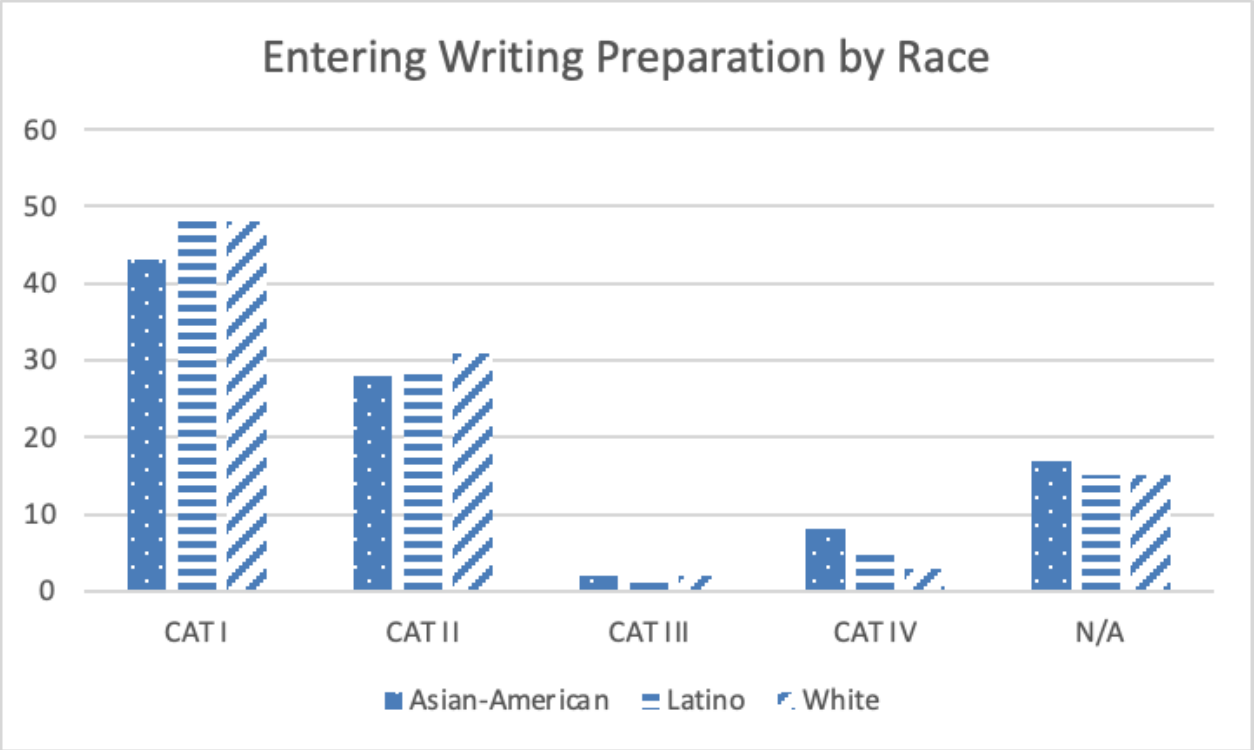
Student achievement disaggregated

Race



%benchmark 2 or above	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
Asian American (n=38)	94.74	94.74	89.47	92.11	89.47	89.47	97.37	94.74
Latino (n=132-8unscorable =124	86.29	89.52	87.90	78.23	85.48	86.29	83.87	93.55
White (n=76)	94.74	92.11	86.84	92.11	93.42	86.84	89.47	100.00

%proficient	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
Asian American (n=38)	36.84	31.58	63.16	31.58	21.05	26.32	52.63	63.16
Latino (n=132-8unscorable =124)	14.52	11.29	29.84	19.35	8.06	9.68	32.26	37.10
White (n=76)	23.68	18.42	31.58	21.05	23.68	13.16	34.21	60.53



	CAT I	CAT II	CAT III	CAT IV	N/A
Asian-American	**43**	**28**	**2**	**8**	**17**
Latino	**49**	**29**	**2**	**5**	**15**
White	**48**	**31**	**2**	**3**	**15**

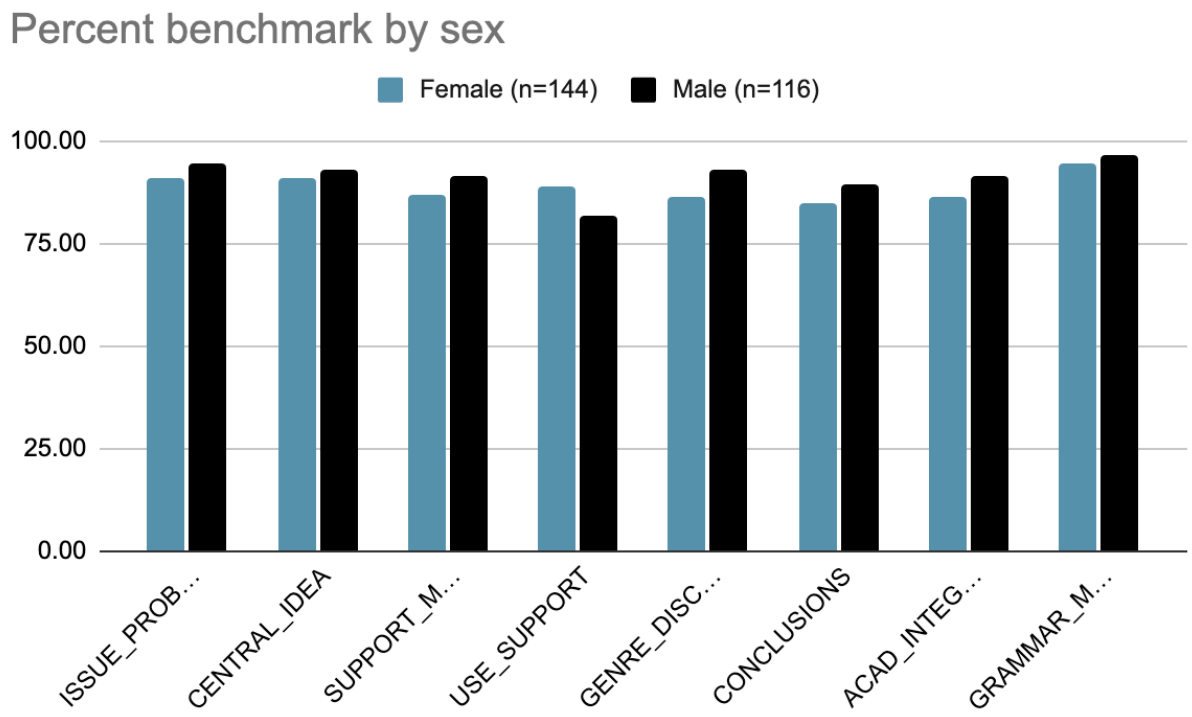
Interpretation

Only three "racial" groups had sufficient students for comparison, Asian American, Latino, and White. In itself, that communicates that CSUMB has been ineffective at attracting and retaining Black and Indigenous students.

On most criteria, Latino students' essays are scored substantially lower than Asian American or White students both at the benchmark 2 and proficient levels. At both levels, Asian American students' essays are scored the highest for most criteria except for grammar at the benchmark level. These findings raise questions about how well we are serving our Latinx students. Based on students' placement according to multiple measures, these differences do not reflect major differences in pre-college preparation. That suggests the issue is with our instruction, that we are somehow not reaching our Latinx students with our instruction. Of course, multiple measures is a crude classification and may not reflect differences in performance within the categories. The differences, however, do raise concerns.

The largest difference seems to be in terms of "use of support." Because the kind of synthesis expected in college writing may differ in meaningful ways from that expected in high-school writing, this is further evidence that the issue is with instruction in A2 and A3 classes. In our last assessment of A2 courses, we did not disaggregate the results by race, so we cannot determine in this analysis at what point in students' early career at CSUMB we are beginning to fail our Latinx students.

California Legal Gender



%benchmark 2 or above	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
Female (n=144)	90.97	90.97	86.81	88.89	86.11	84.72	86.11	94.44
Male (n=116)	94.69	93.10	91.38	81.90	93.10	89.57	91.38	96.55

%proficient	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS

Female (n=152-8 unscorable =144)	25.00	22.22	39.58	29.17	19.44	18.06	41.67	50.00
Male (n=116)	20.69	18.97	32.76	13.79	10.34	12.07	30.17	53.45

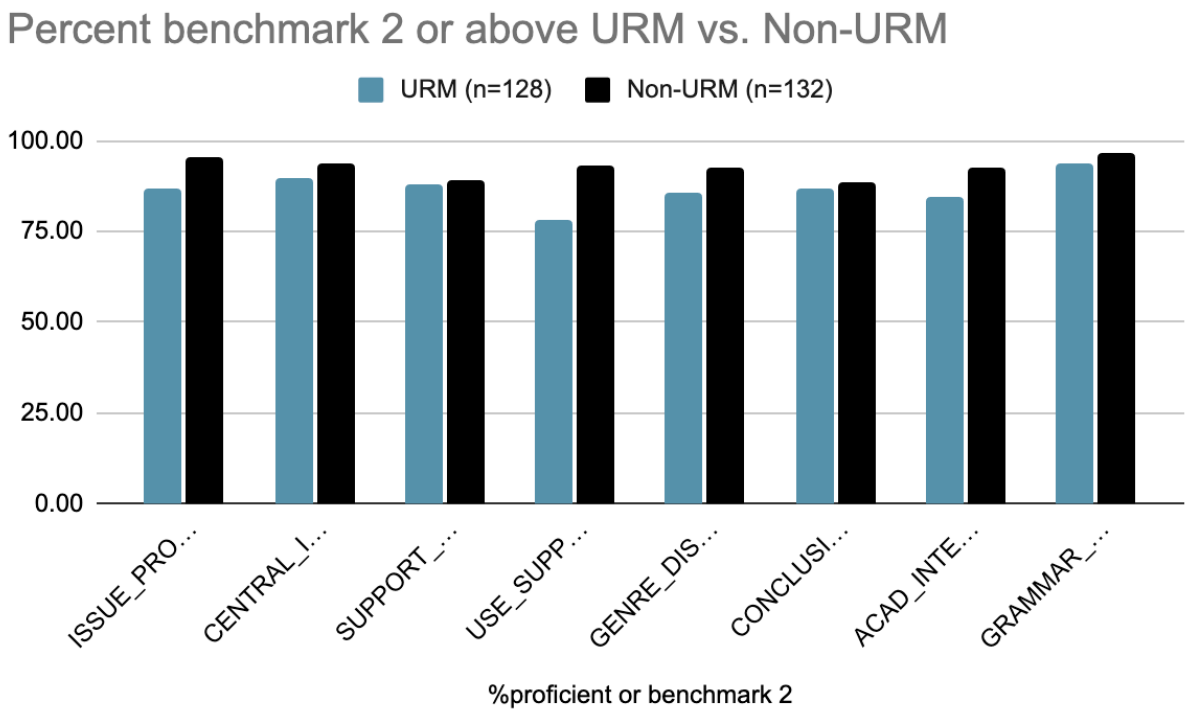
Equity Gaps

Academic Course Name	Students Graded	Female Course GPA	Female Students Graded	Male Course GPA	Male Students Graded	CA Legal Gender Course GPA Equity Gap
HCOM 200 - Critical Thinking & Ethics	560	2.95	328	2.81	232	**0.14**
HCOM 202 - Philosophy of Human Nature	168	2.75	102	2.81	66	**0.06**

Interpretation

Although female students are outperforming male students in terms of course GPA in one of the two A3 courses in HCOM and underperforming in the other, the differences are smaller in terms of the underperformance. This may reflect the higher percentage of female students than male students who are performing at the proficient level across criteria (which may result in higher grades on assignments). At the benchmark level, males are outperforming females in all criteria except use of support (in which females outperform males) and grammar (in which the two are fairly close). At the proficient level, the lowest percentage of scores for both groups is on the criteria of “use of support,” “genre and disciplinary conventions,” and “conclusions,” though the differences for male students is more dramatic in those criteria as well.

Under-represented Minority Status

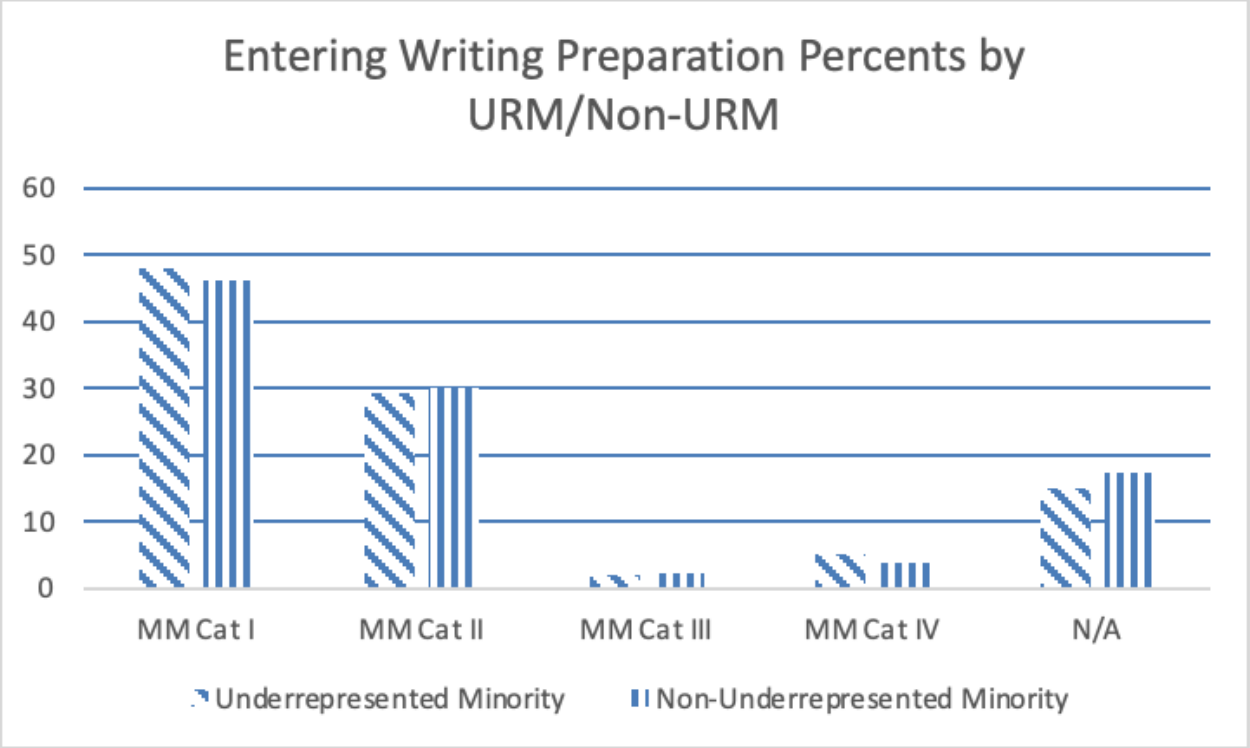


%proficient or benchmark 2	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
URM (n=128)	86.72	89.84	88.28	78.13	85.94	86.72	84.38	93.75
Non-URM (n=132)	95.45	93.94	89.39	93.18	92.42	88.64	92.42	96.97

%proficient	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
URM	13.24	10.29	30.15	17.65	7.35	8.82	30.88	36.76
Non-URM	31.82	30.30	40.91	25.76	22.73	21.37	40.91	63.64

Equity gaps

Academic Course Name	Students Graded	Non-URM Course GPA	Non-URM Students Graded	URM Course GPA	URM Students Graded	URM Status Course GPA Equity Gap
HCOM 200 - Critical Thinking & Ethics	561	3.04	299	2.72	262	**0.31**
HCOM 202 - Philosophy of Human Nature	169	2.89	89	2.66	80	**0.23**

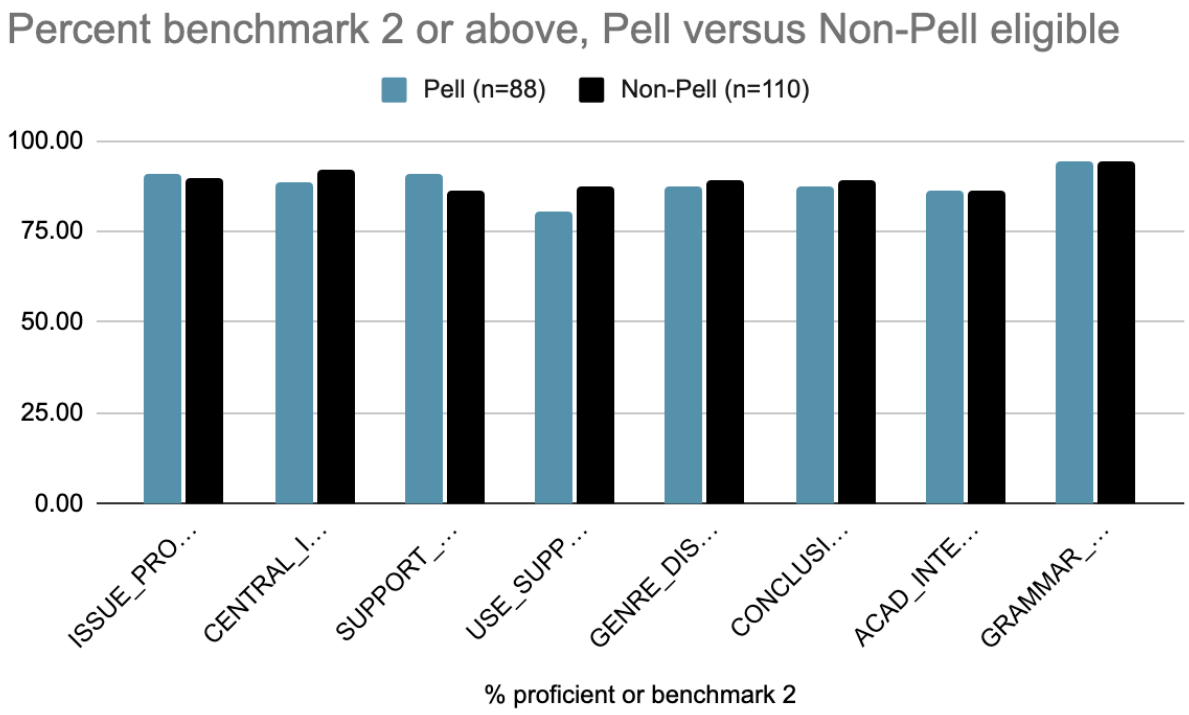


	MM Cat I	MM Cat II	MM Cat III	MM Cat IV	N/A
Underrepresented Minority	**48.06**	**29.41**	**2.08**	**5.27**	**15.06**
Non-Underrepresented Minority	**46.2**	**30.08**	**2.42**	**3.94**	**17.27**

Interpretation

Although large majorities of both groups (URM and non-URM students) are performing at the benchmark level or above, there are appreciable differences on most criteria at both the benchmark and proficient level that correspond with the equity gap by GPA. That suggests that the differences in course grades may relate closely to the difference in student performance in writing. Note that the differences in students' academic preparation according to the CSU's multiple measures classification is not as strong as the differences in students' performance according to the rubrics. Of course, multiple measures is a crude classification and will not reflect differences in student performance within the categories.

Pell Eligibility



% proficient or benchmark 2	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
Pell (n=88)	90.91	88.64	90.91	80.68	87.50	87.50	86.36	94.32
Non-Pell (n=110)	90.00	91.82	86.36	87.27	89.09	88.99	86.36	94.55

% proficient	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
Pell (n=88)	18.18	15.91	40.91	27.27	11.36	9.09	43.18	34.09
Non-Pell (n=110)	21.82	21.82	28.18	20.00	16.36	16.51	16.36	52.73

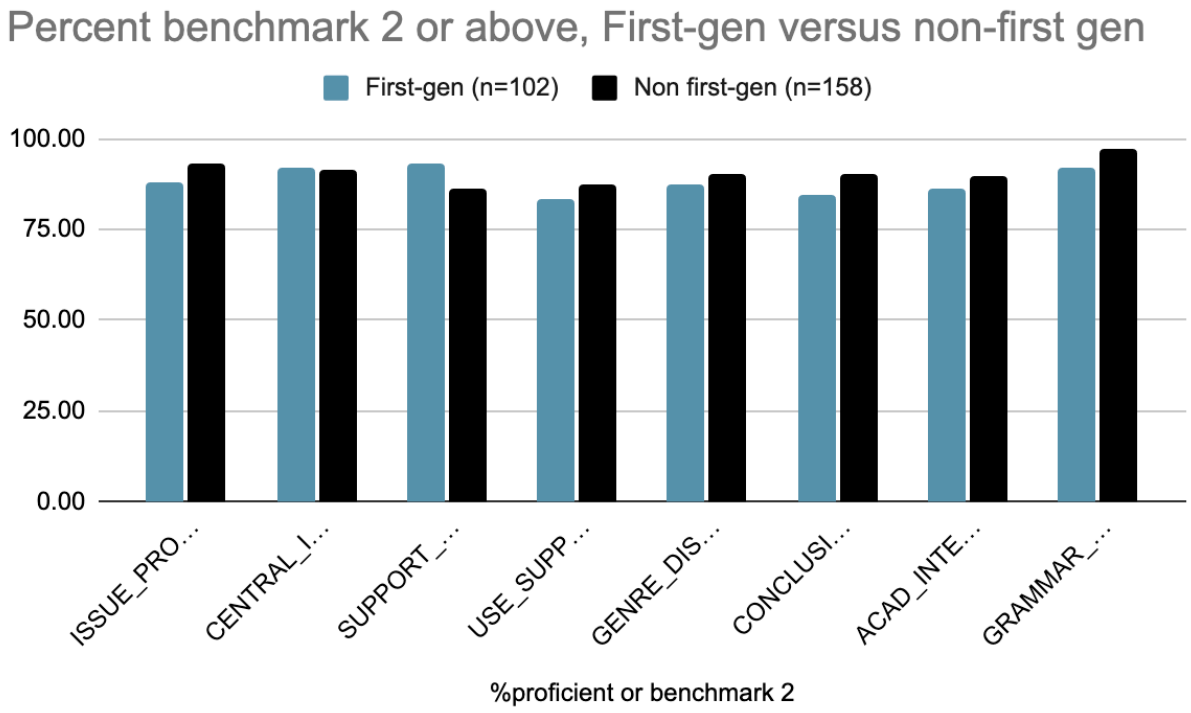
Equity Gaps

Academic Course Name	Students Graded	Did Not Receive Pell Aid Students Graded	Did Not Receive Pell Aid Course GPA	Received Pell Aid Students Graded	Received Pell Aid Course GPA	Received Pell Aid Course GPA Equity Gap
HCOM 200 - Critical Thinking & Ethics	561	336	3.04	225	2.67	**0.37**
HCOM 202 - Philosophy of Human Nature	169	102	2.98	67	2.48	**0.50**

Interpretation

With the exception of "central idea," "supporting materials," and "use of support," Pell- and non-Pell-eligible students score at benchmark two or above in close proportions. Interestingly, Pell-eligible students score in this range at higher percentages than non-Pell-eligible students on "supporting materials." Higher percentages of Pell-eligible students score proficient on "supporting materials," "use of support," and "academic integrity." These differences suggest that other factors than writing quality account for the equity gap in course grades for this group of students.

First-Generation



%proficient or benchmark 2	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
First-gen (n=102)	88.24	92.16	93.14	83.33	87.25	84.31	86.27	92.16
Non first-gen (n=158)	93.04	91.77	86.08	87.34	90.51	90.45	89.87	97.47

%proficient	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
First-gen	15.69	13.73	27.45	17.65	5.88	9.80	35.29	41.18
Non first-gen	26.51	24.10	40.36	24.10	20.48	18.07	36.14	55.42

%proficient	ISSUE_PROBLEM	CENTRAL_IDEA	SUPPORT_MATERIAL	USE_SUPPORT	GENRE_DISCIPLINARY	CONCLUSIONS	ACAD_INTEGRITY	GRAMMAR_MECHANICS
First-gen	15.69	13.73	27.45	17.65	5.88	9.80	35.29	41.18
Non first-gen (n=158)	27.85	25.32	42.41	25.32	21.52	18.99	37.97	58.23

Equity Gaps

Academic Course Name	Students Graded	Not First Generation Course GPA	Not First Generation Students Graded	First Generation Course GPA	First Generation Students Graded	First Generation Course GPA Equity Gap
HCOM 200 - Critical Thinking & Ethics	547	3.00	301	2.78	246	**0.22**
HCOM 202 - Philosophy of Human Nature	168	2.98	94	2.50	74	**0.48**

Interpretation

The percentage of first-generation students scoring either at the benchmark or proficient level across criteria are substantially lower than non-first-gen students. That may correlate with the lower course GPAs for those students. The greatest difference at both levels is in terms of "genre and disciplinary conventions," which may not be surprising for students who are the first in their family to attend college. But that suggests the need for more explicit attention to that criterion in instruction.

Overall interpretations and implications

A consistent message in all of our assessments is that the quality of student work closely tracks to the quality of the prompt evoking that work. The clearer and more detailed the prompt, the better the student writing. We encourage faculty to take advantage of resources to improve prompts such as the regular Assignment Design workshops, feedback from the CAD/CLC Assignment Review service, or informal sharing and norming with faculty.

The good news of this assessment is that a large majority of the essays scored were at or above the benchmark 2 level, which is the level we would anticipate for lower-division students. Among some groups, roughly a quarter of the scores were in the proficient range. At the lower-division level, that may reflect greater academic preparation before coming to college rather than reflecting our success in teaching students.

Disaggregating by various populations in which differences in performance might be expected, we did see differences in percentages of essays scoring both at the benchmark and proficient level across categories (with an arbitrary marker of 3 percentage points difference marked as notable). Sometimes, as with students identified as underrepresented minorities or first-generation versus those not so identified, the difference corresponds to a difference in course GPA. In other groups, such as Pell-eligible versus non-Pell-eligible students, the difference in course GPAs seems to be attributable to factors other than writing quality. Latino students also fare poorly in comparison to White and Asian American students in their scores on various criteria. We suspect that if we were looking at intersectional identities (e.g., Latino males versus White females), we might be able to more narrowly identify the populations we are least well serving and have better recommendations for faculty.

Across the board, students seem to struggle with information literacy and critical thinking criteria: “use of support,” “central idea,” and “conclusions”; and with “genre and disciplinary conventions.” Clearly, more instruction in synthesis and drawing legitimate inferences from evidence is warranted both at the lower and upper-division levels.

Instructors teaching GWAR and upper-division writing classes might explore the rubric guide to consider areas on which they can focus their instruction. While one of the criteria titles have changed (“position” is now “central idea”) the descriptions of the boundary between benchmark 2 and proficient continue to be useful.

We want to close by calling attention once again to a key finding in our disaggregation of student data. If we are truly a Hispanic-Serving Institution, not simply a Hispanic-Enrolling Institution, we should be very concerned by the lower scores of Latino students on all criteria. This reinforces the calls that others have made for culturally-sustaining pedagogy and assignment prompts.