

# HPU Critical Thinking Assessment Report 2023-2024

## Overview of the Assessment Project

In Fall 2023, HPU launched its second, campus-wide initiative designed to assess critical thinking in general education and undergraduate capstone courses. Critical thinking is one of five institutional learning objectives; the others include written communication, oral communication, information literacy, and quantitative and symbolic reasoning. Critical thinking was last assessed in 2018-2019.

**HPU ILO: Students will identify and explain issues, analyze evidence, assess assumptions, define their own perspectives and positions, and present the implications and consequences of their conclusions at the associate's or bachelor's level.**

## Method

During the academic year 2023-2024, a total of twenty-eight General Education course sections (across 15 different courses and four general education curricular areas--Critical Thinking and Expression, Technology and Innovation, Writing and Information Literacy I and Quantitative Analysis and Symbolic Reasoning) participated in the project. Six other courses (1 from GECT, two from GETI and 3 from GEQA) were invited but no students submitted work. Nineteen undergraduate capstone course sections (across 16 different courses and 20 majors) participated in the project. Eight other UD courses were supposed to be included, but no students from those courses submitted work to the assessment.

The committee sent out instructions to participating instructors, asking them to identify an assignment within their course that would produce an artifact that fulfilled all requirements of the critical thinking common rubric. A total of 459 artifacts of authentic student work were collected, and these artifacts were archived in the Critical Thinking Assessment Project in OAP. After participating in a norming session, individual faculty from across the university scored the artifacts. Some artifacts were considered unscorable with the rubric and were excluded. Of the remaining 427 artifacts, 299 artifacts from GE courses and 85 UD artifacts were scored, including all UD artifacts submitted.

Faculty reviewers scored the artifacts on a scale of 0 to 4; 0= Evidence of this skill is **not present**, 1= Student demonstrates an **initial** ability, 2= Student's skills are **emerging** but still not fully developed, 3= Student's skills are **developed**, 4= Student's skills are **highly developed**.

The rubric describes levels of achievement over six criteria.

1. The student identifies and summarizes the problem or question at issue or the source's position if responding to a specific argument
2. The student identifies and presents the student's own perspectives and positions as important to the analysis of the issue
3. The student identifies and considers other salient perspectives and positions that are important to analysis of the issue
4. The student supports the student's opinion with evidence from sources
5. The student analyzes the issues in context
6. The student draws conclusions and identifies and assesses their implications and consequences.

## Assessment Findings

### General Education Critical Thinking Assessment Findings

The committee set the following target for the general education population participating in this project: 85% of students enrolled in general education courses with a critical thinking GELO will achieve an initial, emerging, developed, or highly developed score for each criterion and 60% will achieve an emerging, developed, or highly developed score. We also expected the mean score to be around 2.

The results are summarized in Table 1 below

**Table 1: Summary of General Education results for Critical Thinking ILO 2023-2024**

Critical Thinking Rubric Criteria	% Initial or higher (Target $\geq 85\%$ )	% Emerging or higher (Target $\geq 60\%$ )	Mean score (Target $\geq 2$ )
Problem/Question Summary	93.96	76.85	2.42
Student's Own Perspectives and Positions	97.64	82.89	2.52
Other Perspectives and Positions	88.93	67.45	2.05
Evidence from Sources	81.66	57.44	1.85
Contextual Analysis	93.27	73.06	2.21
Conclusions, Implications, Consequences	89.19	67.91	2.13
Overall	94	69.9	2.205

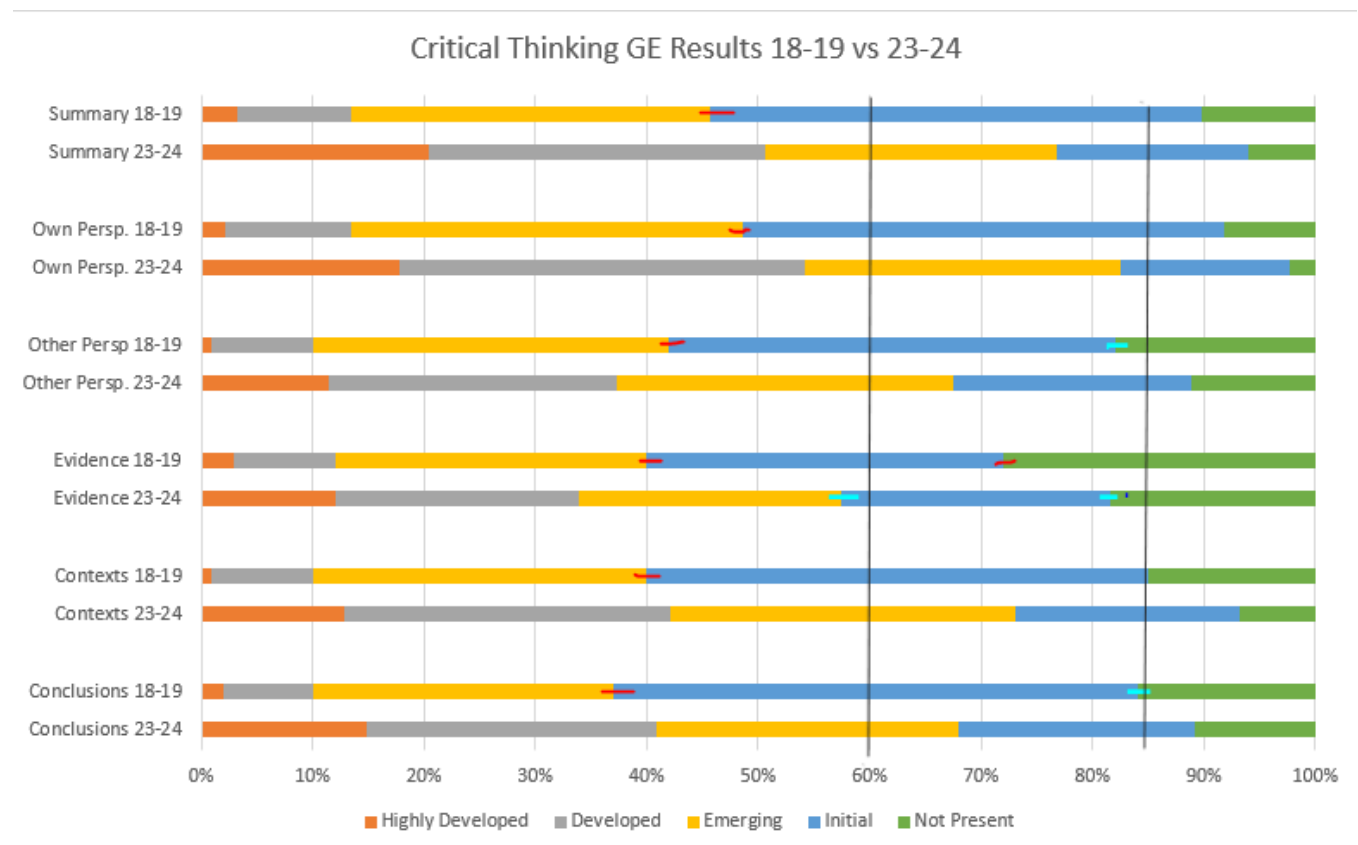
For the 85% target, these results exceeded expectations in three areas, met expectations in two areas, and approached in the area of Evidence from Sources. For the 60% target, results exceeded expectations in five areas but approached them in Evidence from Sources. The mean score rounded to 2 (Emerging) for all criteria except Student's Own Perspectives and Positions where it rounded to 3. The mean score was greater than or equal to 2 for all criteria except Evidence from Sources where it approached the target at 1.85.

A detailed breakdown of the results is provided in Table 2 below.

**Table 2: General Education Critical Thinking Assessment Results by Criterion 2023-2024**

Critical Thinking Rubric Criteria (N = 299)	Score				
	4	3	2	1	0
	Highly Developed	Developed	Emerging	Initial	Not Present
	%	%	%	%	%
Problem/Question Summary	20%	30%	26%	17%	6%
Student's Own Perspectives and Positions	18%	36%	28%	15%	2%
Other Perspectives and Positions	11%	26%	30%	21%	11%
Evidence from Sources	12%	22%	24%	24%	18%
Contextual Analysis	13%	30%	31%	20%	6%
Conclusions, Implications, Consequences	15%	26%	27%	21%	11%
Overall average across the rubric	9%	29%	31%	24%	6%

**Figure 1: Comparison between GE Critical Thinking Assessment Results in 18-19 vs. 23-24**



Red dashes show performance more than 5% below the target; aqua show performance approaching the target (less than 5% below). These results show noticeable improvement over the first assessment in 2018-2019, but in both cases Evidence from Sources was the weakest category and Presentation of the Student's Own Perspectives was the strongest.

In 2018-2019 we missed the goal of 85% scoring initial or higher goal for three criteria, but in 23-24 we only missed it in Evidence from Sources and by less than 4%. While we did not set a goal for the emerging or higher level in 18-19, we would have had 60% at or above that threshold for only one criterion, presentation of the Student's Own Perspectives. In 23-24 we met it for five out of six criteria, falling less than 3% short on Evidence from Sources. We also see many more students scoring at the developed and highly developed level in 24-25 across all criteria.

### Comparisons by course length and modality

In the report on the 2018-2019 assessment, the AAPRC recommended increasing the number of GE artifacts and also comparing results by modality and course length. These recommendations were followed.

To simplify comparisons, we use the mean scores for each criterion and overall across the rubric, followed by the percentages meeting the target levels across the rubric.

**Table 3 Comparison of 2023-2024 GE results by semester length and modality**

	ALL GE	16 week	8 week	OL	HYB	F2F
	N=299	N=259	N=40	N=84	N=20	N=195
Problem/Question Summary	2.42	2.37	2.725	2.14	1.7	2.6
Student's Own Perspectives and Positions	2.52	2.55	2.326	2.21	2.45	2.44
Other Perspectives and Positions	2.05	2.02	2.225	1.65	1.55	2.27
Evidence from Sources	1.85	1.81	2.125	1.33	1.47	2.10
Contextual Analysis	2.21	2.19	2.375	1.81	1.95	2.41
Conclusions, Implications, Consequences	2.13	2.09	2.35	1.79	1.89	2.30
<b>Overall average across the rubric</b>	<b>2.205</b>	<b>2.18</b>	<b>2.35</b>	<b>1.83</b>	<b>1.95</b>	<b>2.40</b>
Percent scoring initial or higher across rubric	94%	94.2%	92.5%	89.2%	90%	97.95%
Percent scoring emerging or higher across rubric	69.9%	68.7%	85%	56.4%	60%	77.4%

The 8-week students outperformed the 16-week students in all but one area in terms of mean scores. The 16-week students had a slightly higher percentage of students scoring initial or higher, but the 8-week students had a significantly higher percentage scoring at the emerging level or higher.

Results by modality show that face-to-face students had higher means than online or hybrid students and had nearly 100% scoring initial or higher across the rubric, while the other two modalities had about 90%. In terms of the percentage of students scoring at the emerging level or higher, online students ranked last and face-to-face students ranked highest. (In this assessment, hybrid means the course is conducted

partially online and partially in-person with at least 30% and no more than 70% in-person delivery. The 16-week hybrid courses typically had one 75-minute in-person meeting per week while the 8-week hybrid courses had one 3-hour in-person meeting per week).

In all modalities and lengths, the mean scores across the rubric showed students performing at the emerging level (1.6-2.5) in general education courses, which is the expected level. However, the online and hybrid means were below 2.0. Evidence from Sources was consistently the weakest area

### Undergraduate Upper-Division Critical Thinking Assessment Findings

The committee set the acceptable target for the undergraduate capstone or other upper-division class critical thinking assessment results as follows: 85% of students enrolled in undergraduate capstone or another selected upper-division course will achieve an emerging, developed, or highly developed score for each criterion. A secondary target is that 60% will achieve a developed or highly developed score. We expected the mean score to be around 3 (Developed).

**Table 4: Summary of UD Critical Thinking Results for 2023-2024**

Critical Thinking Rubric Criteria	% Emerging or higher (Target ≥ 85%)	% Developed or higher (Target ≥ 60%)	Mean score (Target ≥ 3)
Problem/Question Summary	94	83	3.1
Student's Own Perspectives and Positions	82.72	64.2	2.59
Other Perspectives and Positions	86.08	66.08	2.67
Evidence from Sources	85.71	63.12	2.78
Contextual Analysis	91.67	77.38	3.02
Conclusions, Implications, Consequences	85.37	69.51	2.81
Overall	90.59	69.41	2.84

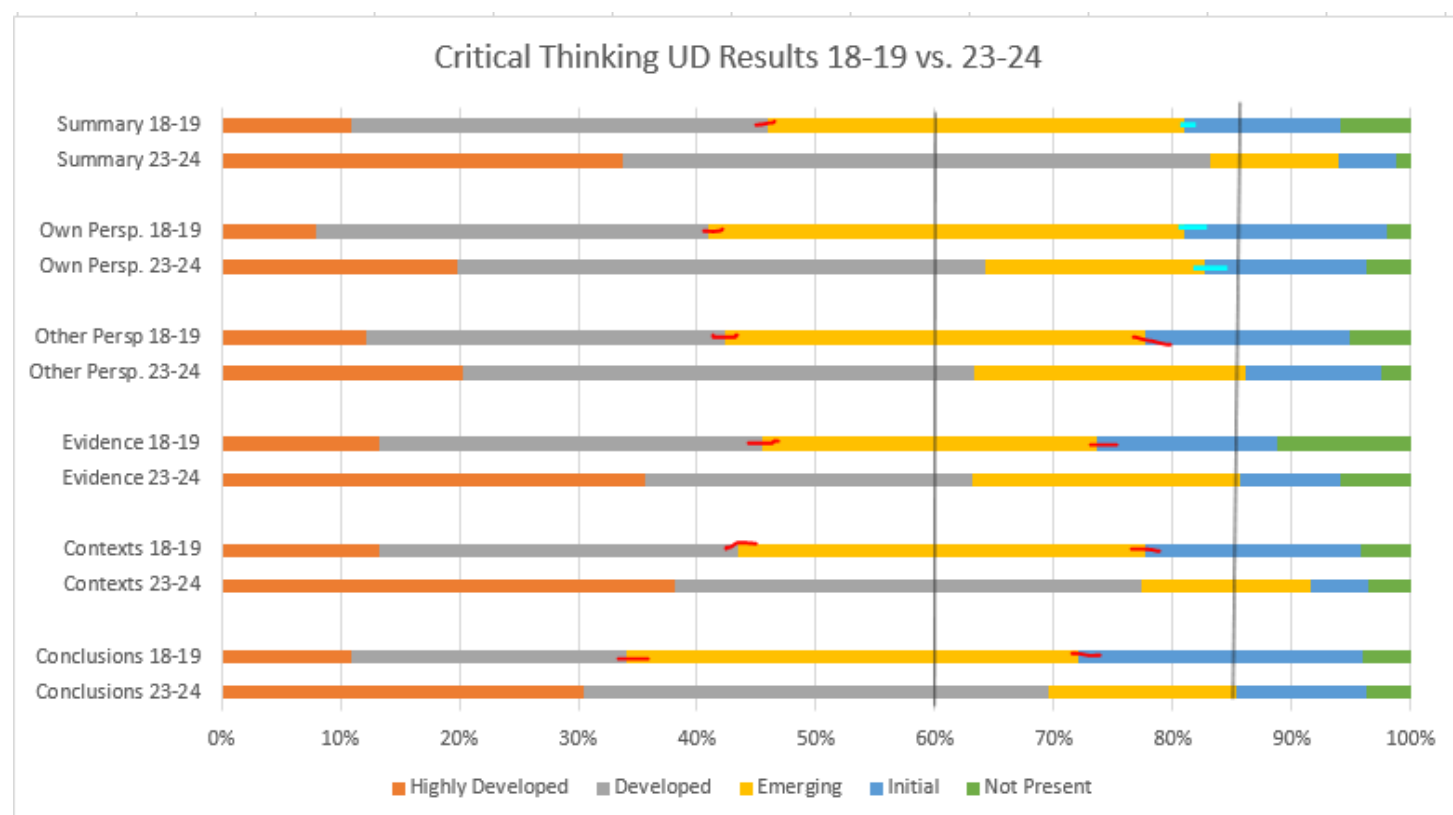
Students approached the emerging or higher target in the Student's Own Perspectives category, exceeded it by more than five percentage points in the Problem/ Question Summary and Contextual Analysis categories, and met it in the other three, including Evidence from Sources showing growth from the general education results where this was the weakest category. The 60% target for students scoring developed or higher was met in two categories and exceeded by more than 5% points in all the others. All of the mean scores rounded to 3, indicating students were in the developed range. However, the mean was greater than or equal to 3 in only the Problem/Question Summary and Contextual Analysis categories



**Table 5. Undergraduate UD Critical Thinking Results by Criterion 2023-2024**

Critical Thinking Rubric Criteria (N = 85)	Score				
	4	3	2	1	0
	Highly Developed	Developed	Emerging	Initial	Not Present
	%	%	%	%	%
Problem/Question Summary	33.7%	49.4%	10.8%	4.8%	1.2%
Student's Own Perspectives and Positions	19.75%	44.4%	18.5%	13.6%	3.7%
Other Perspectives and Positions	20.25%	43%	22.8%	11.4%	2.5%
Evidence from Sources	35.7%	27.4%	22.6%	8.3%	5.95%
Contextual Analysis	38.1%	39.3%	14.2%	4.8%	3.6%
Conclusions, Implications, Consequences	30.5%	39%	15.8%	11%	3.7%

**Figure 2. Comparison Upper Division Critical Thinking Results 2018- 2019 vs 2023-2024**



Red dashes show performance more than 5% below the target; aqua show performance approaching the target (less than 5% below).

With the exception of Student's Own Perspectives and Positions where the results were similar both times for the emerging or higher target, upper-division students improved in every area over 2018-2019. The 2023-2024 students met the 85% target in 5 of six categories. By contrast, in 2018-2019 students were ranked as either not meeting or approaching the target in every category. Looking at the target of 60% scoring developed or higher, in the 2023-2024 assessment it was met in three categories and exceeded in two. In 2018-2019 the percentages reaching the developed level or higher were only 35 to 46%.

While the performance was much stronger in 2023-2024, fewer UD samples were submitted (85 versus 246). The number of sections was about the same, but the level of participation by students in several courses was very weak, with many sections having three or fewer submissions. Participation was particularly low or non-existent for Nursing, Psychology, Marine Biology, Computer Science, and all CPS majors.

### **Comparisons by course length and modality**

In the review of the 2018-2019 assessment, the AAPRC recommended comparing results by modality and course length. These recommendations were followed.

To simplify comparisons, we use the mean scores for each criterion and overall across the rubric, followed by the percentages meeting the target levels across the rubric.

**Table 6 Comparison of 2023-2024 UD results by semester length and modality**

	ALL UD	16-week	8-week	OL	HYB	F2F
	N=85	N=68	N=17	N=9	N=9	N=67
Problem/Question Summary	3.1	3.08	3.18	3	3.44	3.06
Student's Own Perspectives and Positions	2.59	2.66	2.53	2.55	2.67	2.63
Other Perspectives and Positions	2.67	2.69	2.59	2.89	2.44	2.67
Evidence from Sources	2.78	2.88	2.41	2.67	2.33	2.87
Contextual Analysis	3.02	3.01	3.12	3.22	3.11	3
Conclusions, Implications, Consequences	2.81	2.86	2.65	3.11	2.33	2.84
Overall average across the rubric	2.84	2.88	2.75	2.91	2.73	2.87
Percent scoring emerging or higher across rubric	90.59%	89.71%	94.12%	88.89%	100%	90.9%

Percent scoring developed or higher across rubric	69.41%	70.59%	64.71%	66.67%	66.67%	69.41%
---	--------	--------	--------	--------	--------	--------

The 16-week classes had a higher percentage of students scoring developed or higher over the rubric, while the 8-week classes had a higher percentage at emerging or higher. The 16-week had a higher average score across the rubric and scored higher in 4 out of 6 categories. The online classes had the highest overall mean score, and the hybrid classes had the lowest, but no hybrid students scored below emerging overall. Overall means met both targets for all UD semester lengths and modalities and the overall means were all in developed range (2.6-3.5) but below 3.

**Table 7. Mean Critical Thinking scores (mean  $\pm$  stdev) as a function of student demographic data and course attribute**

		<i>GENERAL EDUCATION</i>		<i>UPPER DIVISION</i>	
		mean $\pm$ stdev	<i>n</i>	mean $\pm$ stdev	<i>n</i>
<i>Overall*</i>		13.1 $\pm$ 6.0	299	16.3 $\pm$ 5.5*	85
	<i>Gender</i>				
	Female	13.0 $\pm$ 6.3	198	16.9 $\pm$ 5.0	62
	Male	13.1 $\pm$ 5.5	101	14.8 $\pm$ 6.4	23
<i>Underrepresented minority</i>	YES	12.3 $\pm$ 6.4	99	17.8 $\pm$ 4.7*	29
	NO	13.4 $\pm$ 5.8	200	15.5 $\pm$ 5.7	56
<i>First generation college student</i>	YES	12.6 $\pm$ 6.1	71	16.4 $\pm$ 4.5	28
	NO	13.2 $\pm$ 6.0	228	16.3 $\pm$ 5.9	57
<i>Semester term</i>	8-week	14.1 $\pm$ 4.9	40	16.0 $\pm$ 4.0	17
	16-week	12.9 $\pm$ 6.2	259	16.4 $\pm$ 5.8	68
<i>Course modality</i>	Face-to-face	14.2 $\pm$ 5.8**	195	16.3 $\pm$ 5.8	67
	Hybrid	10.9 $\pm$ 5.9	20	15.6 $\pm$ 4.0	9
	Online	10.8 $\pm$ 5.9	84	17.4 $\pm$ 4.6	9
<i>Pell eligible</i>	YES	12.7 $\pm$ 6.2	95	16.8 $\pm$ 4.7	25
	NO	13.2 $\pm$ 5.9	204	16.1 $\pm$ 5.8	60

\* Significance determined by t-test, unequal variance,  $p < 0.05$

\*\* Significance determined by one-way ANOVA,  $p < 0.05$

## Discussion

- How accurately do we think these findings reflect the actual level of competence of our students?

We think this assessment gives us a reasonably accurate picture of student skills. The improvement probably resulted from efforts to familiarize faculty with the rubric and set expectations for a common assignment but with content determined by the instructor. The disseminated instructions and the workshop at Assessment Day encouraged instructors to think about how to make sure their assignments addressed elements such as reflections on context or implications of conclusions.

- Were there certain artifacts that were not appropriate for the kind of assessment conducted?

Yes, the artifacts submitted for some of the sections of MIS 2000 which posed a series of short answer questions about using Microsoft Excel did not match the rubric and were reported by most reviewers as unscorable or scored as N/A. These artifacts were dropped from the analysis.

Artifacts from GEQA courses such as MATH 1123 and PHIL 2090 may also have not matched the rubric well (high number of 0's given) and some courses had few or no artifacts submitted. We should consider whether a different rubric could be used for certain courses where the current rubric and suggested prompt do not match the content of the courses.

- Were there other problems with the process?

In general, the process went smoothly, but the number of artifacts submitted for the upper-division assessment was smaller than last time. We need to get better participation from certain majors if the results are to be useful for specific programs and to ensure that the sample is representative of our student body. On the other hand, scoring the assessment is very time-consuming and we were unable to score all of the artifacts or to use multiple readers, which implies that if we are able to get more UD artifacts from more majors next time, we will probably further reduce the number of gen ed artifacts scored. About 180 to 200 at each level would probably be adequate.

To make things easier for instructors, the AAPRC ensured a link was embedded in each section's Blackboard site through which students uploaded their assignments. A potential problem with this is that all work was then listed under one assignment and we could not identify which class or even which level an artifact came from in the scoring queue. We could not then assign scorers to specific courses, making it difficult to allow reviewers to score work in an area that matches their expertise, although there are mixed opinions about whether this is desirable at this level of assessment. We are also unable to see the instructions that go with the assignment while scoring and not being able to identify the course or level makes it difficult to ensure that when we randomly choose artifacts to assess, we don't end up accidentally picking too many or too few from a specific course. Perhaps it would be possible to insert the links for the faculty but create assignments based on courses or on grouping certain disciplines together rather than putting everything under one assignment.

Faculty participating in the norming exercise could not reach a consensus on how to score the sample artifacts and interpret the rubric and some argued that it needed revision, although this was the same rubric used in the previous assessment, and which was circulated in advance of the assessment. A particular issue was how to evaluate the use of evidence criterion. We may need to work on refining the rubric and spend additional time on norming next time.

- How shall we use these findings?
  - ☐ Results will be shared at Assessment Day and posted on our website.
  - ☐ Faculty who participated will be notified that the results are posted after Assessment Day.
  - ☐ Results will be shared with the General Education Learning and Assessment Committee and be broken out by curricular area. We will recommend that the GECLAC consider an assessment with a different rubric that would apply better to critical thinking as expressed in the GEQA courses
  - ☐ UD Results broken down by major (or by college when there was insufficient participation by major) have been shared with the deans for dissemination to program chairs for use in program assessment of aligned PLOs and for use in five-year self-studies

- Are we satisfied with the results?

Yes we are satisfied with the results and are quite pleased by the level of improvement over five years ago and between general education and upper-division work. We may consider pushing the targets up next time by 5-10%.