General Education Assessment Report (2019-2020): Critical Inquiry

Prepared by: Jacob M. Held, Ph.D.

Assistant Provost for Academic Assessment

and General Education

September 2020

Table of Contents

I. Prefatory

The following report covers assessment efforts for the Critical Inquiry Core competency over the academic year of 2019-2020. This report includes a narrative account of all components of the assessment cycle as well as selected data informing those narratives. The intention of this report and the recommendations herein is to be advisory to the UCA Core Council and all relevant stakeholders as stewards of the general education program at UCA.

II. Summary

The UCA Core is assessed on a four-year cycle. Each year one competency area is addressed. For AY 2019-2020, Critical Inquiry was the area scheduled to be assessed.

Assessment in higher education ought to be driven by the idea that reliable data can be used to inform curricular changes to improve student learning. The focus is always on student performance. If you want to improve learning you must know where your students are and whether or not your curricula and teaching is impactful. Thus, there must be moments of assessment where student performance is measured consistently, according to an objective standard, and across time.

During the AY 19-20 assessment cycle observations were made regarding student performance as well as the process of assessment itself as regards the Critical Inquiry competency of the UCA Core. Below are several key takeaways.

- Faculty participation continues to be an issue. **AY 19-20 survey response rate = 46.45%**
- Poorly chosen or designed assignments was a problem frequently noted by the score teams (See Appendix B). Pre-cycle training needs to focus on assignment design and needs to be readily accessible and more widely used by faculty.
- With respect to student learning: **significant growth** was noted in some areas.
- Only **50-60%** of students at the upper division scored **"accomplished" or higher**, with markedly **less than 20%** of students scoring **"exemplary"**.
- The **rubrics ought to be revisited** in order to clarify language and allow scorers to develop and impose a standard set of expectations.

The following report provides a detailed presentation and analysis of the assessment process and results for the Critical Inquiry competency of the UCA Core during AY 19-20. This report provides an initial interpretation of selected data of the assessment of the Critical Inquiry competency.

III. Critical Inquiry

The UCA Core is assessed on a four-year cycle. Each year one competency area is addressed. For AY 2019-2020, Critical Inquiry was the area scheduled to be assessed. The semester prior to the academic year scheduled for assessment training sessions were offered for all faculty scheduled to be teaching a course in the Critical Inquiry area during AY 19-20. Multiple sessions were scheduled for each rubric area, with times being scattered throughout the week to offer several opportunities for faculty to attend. Dr. Held facilitated all sessions. Topics included rubric interpretation, assignment design and selection, as well as a briefing on the process of artifact collection and scoring.

Rubric A (Inquiry and Analysis)

Sessions: March 6th and 7th 2019

Total attendance: 16 faculty

Rubric B (Scientific)

Sessions: March 25th and 26th 2019

Total attendance: 6 faculty

Rubric C (Quantitative)

Sessions: April 17th and 20th 2019

Total attendance: 3 faculty

Overall, attendance at these sessions was sparse. Far too few faculty who teach in these areas attended these sessions. Considering that Scientific and Quantitative encompass all hard sciences at UCA, including the most populous gen ed courses such as Math 1390 and BIOL 1400, that fact that between both goals we only had 9 faculty participants is worrisome. Poor attendance at these sessions may translate into poor participation later on. In addition, poor attendance at these information sessions may foment confusion during the academic year during which assessment takes place. Faculty may be unprepared to participate leading to greater work on behalf of assessment staff, and poorer results. Since these sessions cover the assessment process as well as assignment design, poor attendance may mean that the artifacts that are received may be poorly designed to register student performance across the learning outcomes the rubrics were designed to measure. In fact, poorly aligned artifacts is a common issue reported by the scoring team when assessing student work. Poor attendance at these pre-cycle trainings translates into flawed data. During spring 2020, the pre-cycle training was conducted on-line in webinar format, recorded, and posted online, as a result of Covid-19. Attendance at these webinars was significantly higher than previous pre-cycle trainings, and having the materials posted and accessible is also an added benefit to faculty who may wish to review them at their own pace or a more convenient time. Whether on-line webinar format will replace in person or supplement is yet to be determined, but on-line resources will continue to be used.

After pre-assessment training, the office of assessment prepared to collect student artifacts during AY 19-20.

Artifact Collection:

During AY 19-20, the Office of Assessment attempted to collect artifacts from all courses under the Critical Inquiry Core competency including all Lower and Upper division courses so designated as well as capstone courses. Faculty teaching these courses were identified through ARGOS. All identified faculty were contacted multiple times by means of email. Faculty were provided with a link to a google form. The form asked for information regarding what artifact would be chosen, when it would be administered to students, and when and how it would be delivered to the Office of Assessment.

Survey Response Rate:

	# of unique instructors teaching courses	# of unique instructors responding to survey	% response rate[1]
Fall 2019	258	96	**37.21**
Spring 2020	254	102	**40.16**
Total AY 19-20	310	144	**46.45**

Response rate indicates percentage of respondents in relation to total number of faculty identified as teaching a course identified as assessing under the Critical Inquiry competency.

Survey Yield rate = **80.74%**

Survey yield rate was calculated by comparing the total number of surveys received against the number of "assignments" created in AQUA, where an "assignment" is a unique course indicated by CRN. Presuming one survey per course per instructor a 100% yield rate would indicate that we received at least one usable artifact for every course a respondent instructor taught. Less than 100% indicates that there were instructors who responded who did not subsequently turn in student artifacts for their course, or the artifacts returned were not usable.

The total response rate was an improvement over the AY 18-19 response rate but still sub-optimal. A low response rate raises several issues for assessment. Since we had a low rate of response the resultant assessment data is problematic. In addition, responses were not random, but exhibited patterns. Thus, the resultant data is not reasonably generalizable. However, we did review all the collected material and provide an analysis while remaining aware of the problematic nature of the data.

Review of Artifacts:

Evaluation of the artifacts took place between August 10-19th, 2020. The evaluation team was recruited from faculty who had participated in the assessment process. The evaluation team consisted of:

- Rubric A (Inquiry and Analysis)

[1] Response rate for the survey. This would not reflect faculty that participated by submitting artifacts but did not complete the survey.

    - Michael Rosenow (History)
    - Jacob Bundrick (EFIRM)
    - Ramón Escamilla (LLLC)
- Rubric B (Scientific)
    - Kari Naylor (Biology)
    - Debra Burris (Physics and Astronomy)
- Rubric C (Quantitative)
    - Jeffrey Beyerl (Mathematics)
    - Monica Lieblong (FACS)
    - Ahmad Patooghy (Computer Science)

Evaluators were remunerated $250 per day. During the three day sessions evaluators participated in calibration exercises as well as artifact scoring. Days consisted of routine evaluation work from 8:00 am until 4:30 pm with intermittent breaks as evaluators deemed appropriate.

	# of artifacts available	# of artifacts processed[2]	% of artifacts processed
Goal A	1590	624	39.25
Goal B	620	620	100.00
Goal C	945	945	100.00
Total	3155	2189	69.38

The teams for Goals B and C were able to score the entire population of artifacts. This is due to two factors. First, we had a smaller pool of artifacts due to poor survey response. There should have been significantly more artifacts to score under Goals B and C. For example, we received only one course worth of artifacts for all Upper Division Core Critical Inquiry (Goal B: Scientific) courses. Secondly, given the nature of the rubrics and the artifacts received scoring was often “formulaic” allowing scorers to process artifacts significantly quicker than Goal A, where lengthy papers are the norm. In the future, it may be prudent to enlist more scorers for Goal A in order to process a greater number of artifacts while holding Goals B and C to two scorers each.

Reliability:

The score teams spent the first half of their first day together engaged in norming exercises. The teams reviewed the rubric and proceeded to evaluate anchor assignments. After each assignment is evaluated the team discussed the results and then proceeded to the next assignment. By the close of the

[2] Disregards the number of artifacts receiving a second score. At least 20% of artifacts received a second score in order to calculate inter-rater reliability.

calibration exercise, the teams expressed a shared understanding of the rubric and shared expectations. Teams also conferred regularly during scoring to continuously “re-calibrate.”

These norming exercises are intended to insure that regardless of team member the score an artifact receives is consistent. If scorer expectations are consistent, then the data will be consistent and generalizable. Calibration is crucial to reliable data, that is, data that reflects the nature of the artifact, in this case student performance, and not the idiosyncrasies of the scorer. Below various measures of reliability are provided.

Percent Agreement and Interrater Reliability

	% agreement	% disagree at 1 pt.[3]	Weighted Kappa[4]	Reliability[5]
Goal A	53.52	86.87	.253	Fair
Goal B	52.93	89.80	.262	Fair
Goal C	71.50	82.04	.623	Substantial

Although percent agreement is not often accepted as a reliable statistic when judging inter-rater reliability, of note in this case is the fact that when scorers did disagree well over 80% of the time that disagreement was only one point in variance. That indicates that even when scorers disagreed it was minor, indicating a slight disagreement in student performance, not a major incongruity between scorer expectations. This would suggest that the data reliably reflect trends. Using a more standard measure of inter-rater reliability for ordinal values, Weighted Kappa, we find “fair” reliability in the teams scoring goals A and B, and “substantial” reliability in the team for goal C. The high measure of reliability in goal C may be attributable to the nature of the rubric and artifacts, that is, mathematical/quantitative artifacts that are more standardized across courses and disciplines than the artifacts provided for goals A and B. Regardless, reliability among the teams was good; the data should therefore reflect accurately the relative student performance on the rubric.

In addition, data is provided in terms of where disagreements occurred most frequently. This data may help evaluate the score teams performance and better understand where, and perhaps why, members disagreed when they did so.

[3] When scorers did disagree, this is the percent of disagreements between a single level, for example, scorer A = 1, scorer B = 2, or scorer A = 3 and scorer B = 4.

[4] Weighted Kappa used given the nature of the data as ordinal variables rated by two different evaluators. Calculated using SPSS.

[5] Based on Landis JR, Koch GG. The measurement of observer agreement for categorical data, *Biometrics* (1977); 33:159-74.

(Dis)Agreement Goal A

GOAL A	Total # of disagreements	% of total disagreements	Most common disagreement
Analysis	37	37.37	Between 2 and 3
Info	26	26.26	Between 1 and 2
Knowledge	36	36.36	Between 2 and 3

(Dis)Agreement Goal B

GOAL B	Total # of disagreements	% of total disagreements	Most common disagreement
Define	35	17.86	Between 1 and 2
Evaluate	13	6.63	N/A
Propose	82	41.84	Between 2 and 3
Method	66	33.67	Between 1 and 2

(Dis)Agreement Goal C

GOAL C	Total # of disagreements	% of total disagreements	Most common disagreement
Communication	40	23.95	Between 1 and 2
Info	58	34.73	Between 1 and 2
Method	68	40.72	Between 1 and 2

The data above allows us to understand which student learning outcomes are more problematic in terms of consistent scoring. This may be due to rubric language or it might be due to scorer expectations in terms of interpreting rubric language. Of note is the fact that overall the most common disagreement across all teams was between levels 1 and 2. This indicates that although team members may have disagreed about a student's performance in terms of where in the developmental/emerging stage the work was, they did not disagree that it was in fact emerging and not a proficient performance. Given that the teams were reliably normed, disagreements varied by one point over 80% of the time, and that when disagreements were present they were between the two developmental levels of the rubric, the data presents a reliable picture of student performance and development in the aggregate, disregarding issues with sampling and poor response rate from faculty.

IV. Results

Assessment in higher education ought to be driven by the idea that reliable data can be used to inform curricular changes to improve student learning and assist faculty in developing pedagogies that are more effective. The focus is always on student performance, the goal is learning. If you want to improve learning you must know where your students are and whether or not your curricula and teaching is impactful. Thus, there must be moments of assessment where student performance is measured consistently, according to an objective standard, and across time.

When interpreting assessment data in higher education it is important to note several points. Firstly, the methodology used is often derived from the behavioral and social sciences. However, the higher education environment makes it difficult, if not impossible, to maintain the conditions necessary for reliable statistical analysis using these methods. Samples are small, or in isolated communities, there are myriad factors influencing any variable, most of which cannot be controlled for, nor is it possible to offer control groups as withholding educational opportunities from students for experimental purposes is unethical. The data collected, therefore, must be interpreted in light of these structural barriers, which are endemic to the nature of the study. But while these barriers cannot be removed, they can be ameliorated.

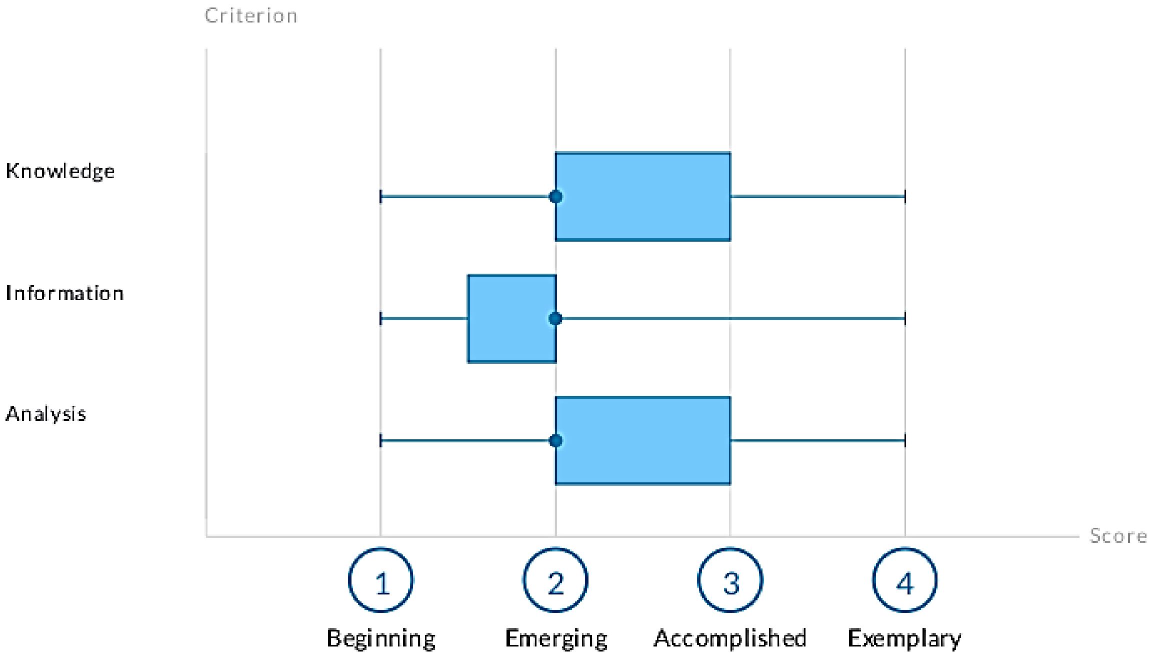
We can get reliable data in terms of identifying trends so long as we know wherein the problems lie and work intentionally to mitigate them. With Core assessment, we have striven to lessen these barriers where possible. We collect student work from the entire population in order to derive a representative sample. Artifacts are all scored on the same rubric, by a single team of calibrated, trained, faculty scorers, thus increasing interrater reliability. We offer training to faculty on assignment design prior to artifact collection, thus allowing faculty to use individual assignments, not standardized ones, while maintaining a consistency of expectation.

If a general education program is to be assessed for common student learning outcomes at a university the size of UCA, the means by which we are doing so addresses, as well as can be addressed, the limitations inherent in assessment in higher education.

Goal A (Inquiry and Analysis)

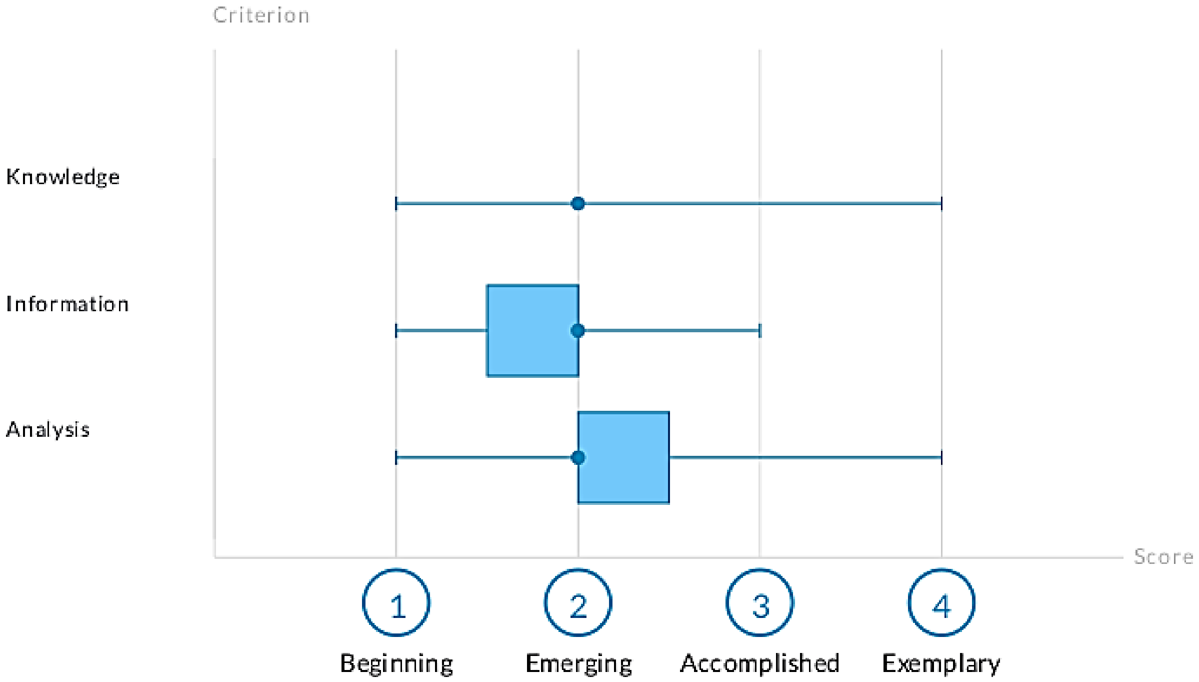
Goal A presented us with the largest pool of artifacts, as well as the most representative, providing a significant selection from both the lower and upper division Core, as well artifacts from a variety of disciplines. Given the disparate disciplines presented in this area, as well as faculty scorers originating from different disciplines it is not surprising that reliability was lower in this area than the others. Regardless, the data does demonstrate some notable trends.

Overall Score Distribution by Outcome



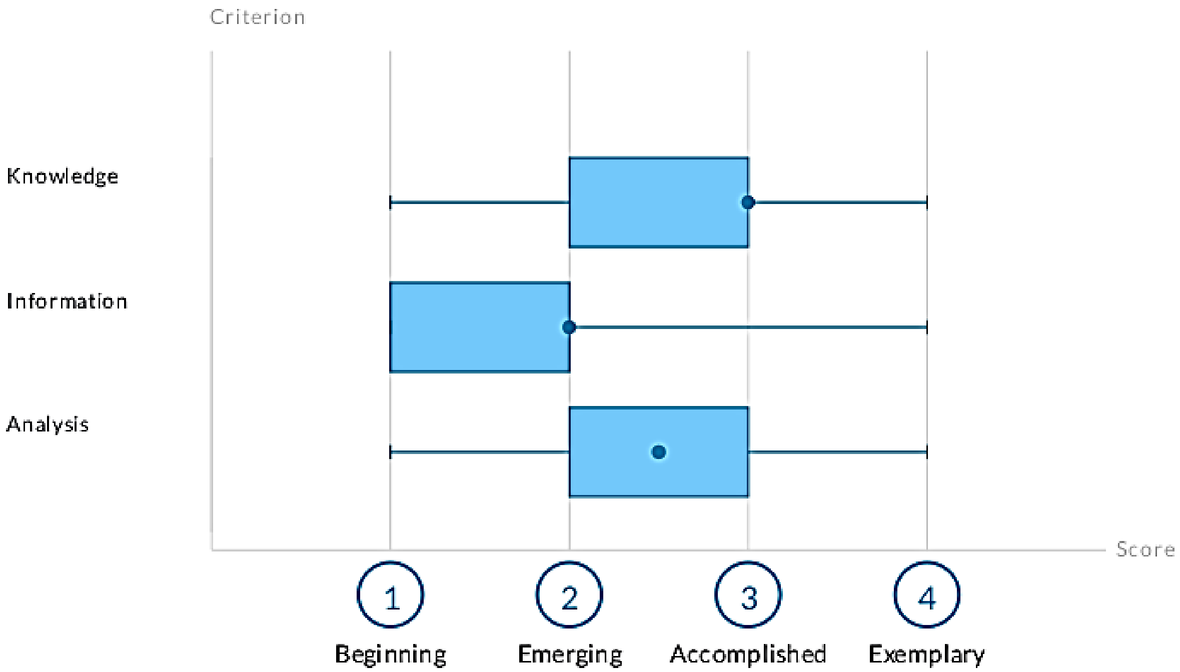
	Max Score	Min Score	Median Score	# of submissions
Knowledge	4	1	2	574
Information	4	1	2	555
Analysis	4	1	2	573

Lower Division by Outcome



	Max Score	Min Score	Median Score	# of submissions
Knowledge	4	1	2	378
Information	3	1	2	372
Analysis	4	1	2	378

Upper Division by Outcome



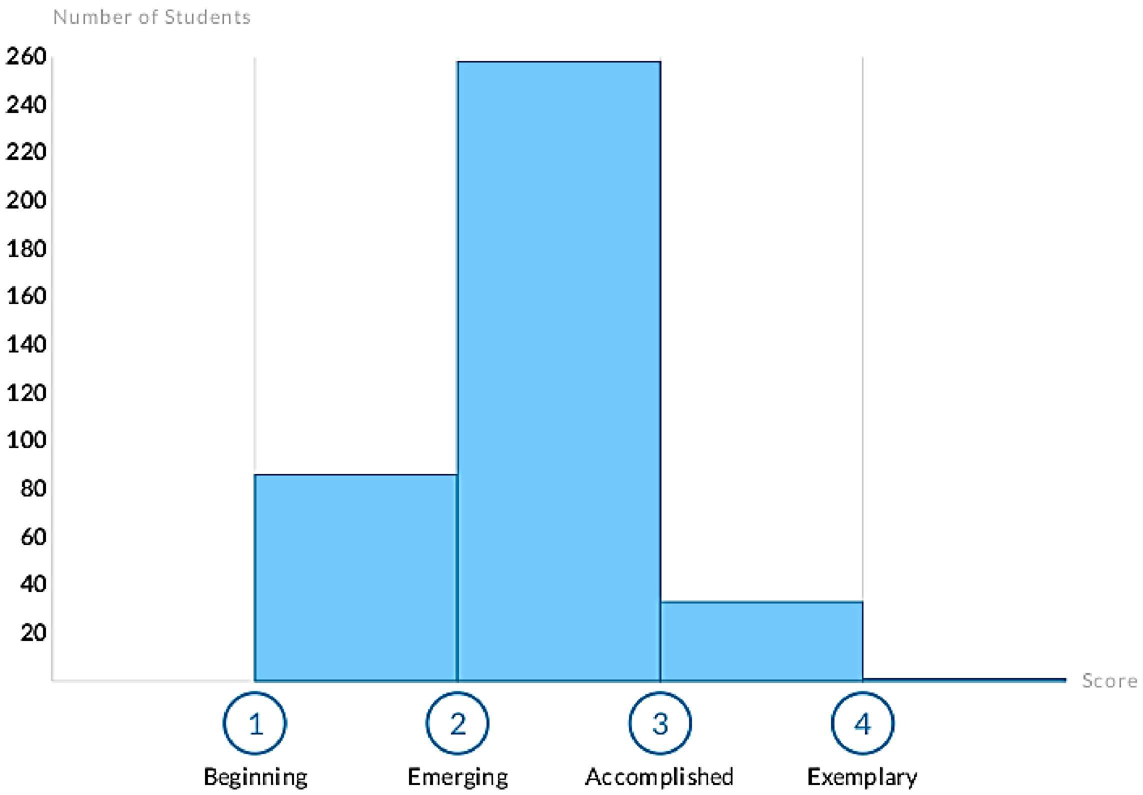
	Max Score	Min Score	Median Score	# of submissions
Knowledge	4	1	3	196
Information	3	1	2	183
Analysis	4	1	2.5	195

In general, the data reflects trends to be expected. The median score across all outcomes overall was 2, and the median score increased across all outcomes, except for "Information," from the lower division to the upper division indicating an increase in frequency of proficient scores as students progressed through the curriculum. The only anomaly is the conspicuous lack of any scores of 4 (exemplary) in the Information outcome. However, considering the outcome itself and discussions of the score team, the lack of scores of 4 is not surprising. The Goal A rubric defines a score of 4 under "Information" as: "Selects information from the most relevant and credible sources, without critical omissions of key sources." As the score team observed, unless a faculty scorer were well acquainted with the discipline from which the artifact was generated, and the specific area within that discipline that the artifact addressed, adjudging “most relevant,” “credible,” and whether “critical omissions” occurred is impossible to discern. Thus, a lack of scores of 4, is indicative of limitations inherent in the scoring process, not necessarily indicative of student competence.

SLO 1: Knowledge

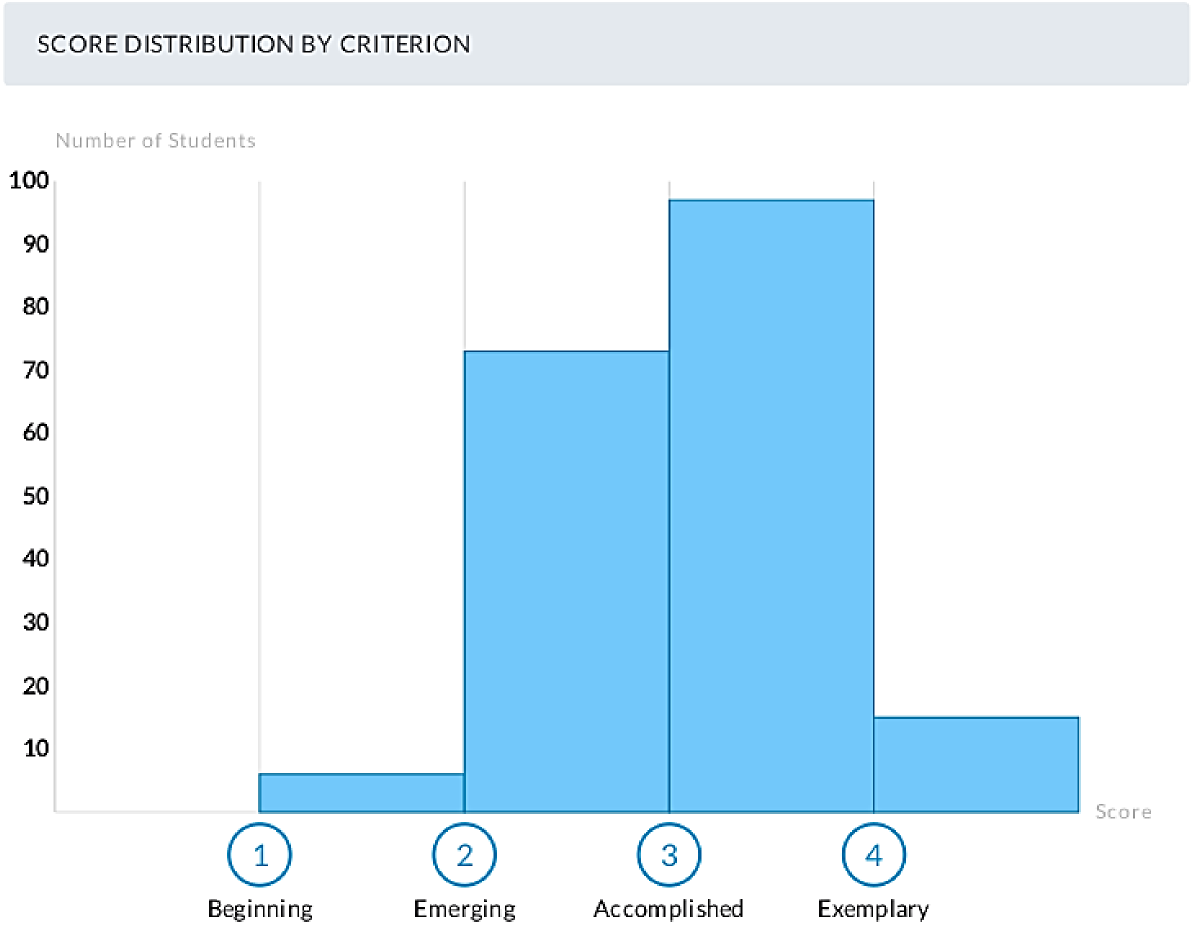
Knowledge: An understanding of the concepts and/or principles in the discipline and how they relate to important questions.

Lower Division Frequency



	1- Beginning	2- Emerging	3- Accomplished	4- Exemplary
# of Scores	86	258	33	1
% of Scores	22.75	68.25	8.73	0.26

Upper Division Frequency



	1- Beginning	2- Emerging	3- Accomplished	4- Exemplary
# of Scores	6	73	97	15
% of Scores	3.14	38.22	50.79	7.85

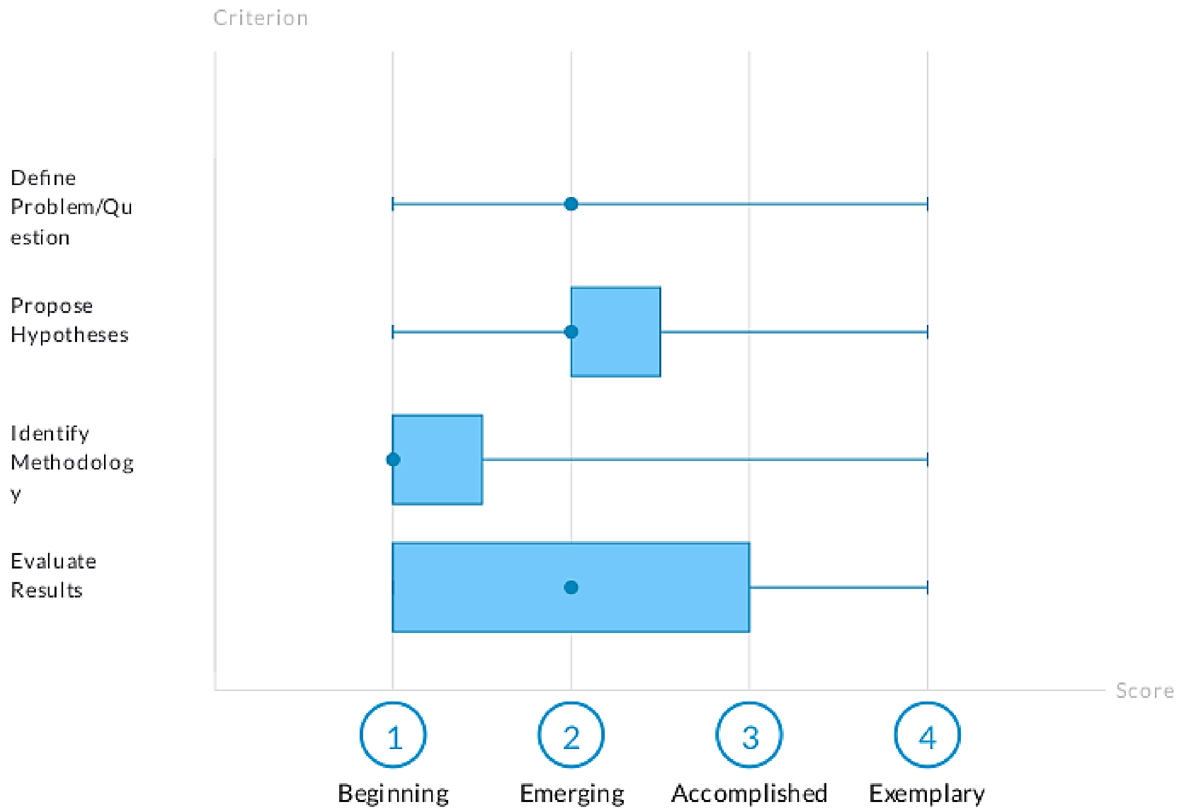
The knowledge outcome showed growth from the lower to upper division Core. We see a shift towards 3 and 4 scores. At the lower division, less than 10% of student artifacts scored at the accomplished and exemplary levels combined. However, at the upper division almost 60% of student artifacts so scored. This is a significant increase. This trend suggests that near the end of our curriculum a healthy majority of students are accomplished in this outcome, when they are not demonstrating significant competency early in the curriculum. Various explanations are possible for this. If students were not prompted or otherwise provided the opportunity to demonstrate mastery early on, then that would explain why the artifacts failed to demonstrate mastery. Scorers did note that some student artifacts were designed in a way that prohibited the student from scoring higher on the rubric due to assignment design and expectations. Thus, the lower scores at the lower division may be reflective of assignment design and not student skill level. Regardless, at the upper level we still see a healthy majority scoring at accomplished or above. This shows that students do possess this competency at a proficient level at the end of their programming. That is reassuring. Although we may not be able to discern the whys and wherefores of student growth, we are able to see where students are at the end of their engagement with the Core curriculum and set a benchmark against which we can measure future student assessment.

Assessment data for the “Information” and “Analysis” outcomes of Goal A can be located in Appendix C. Given that the data, as well as scorer commentary, suggests that the scores for the “Information” outcome at the LD level were a result of narrowly defined course assignments, that data did not appear to accurately reflect student acumen and so was not included in the full report. The UD level data shows a marked lack of accomplished or exemplary scores, but given the nature of the rubric, as noted in scorer comments, these scores likely are more reflective of limitations on the scorers’ ability to evaluate this area than student skill. In addition, for brevity’s sake, the body of this report focuses on marked or notable trends when present. The data for “Analysis” represented trends to be expected. However, it is located in Appendix C for reference.

Goal B (Scientific)

Goal B presents a disappointing low in UCA Core assessment. The response rate was significantly poor and lopsided insofar as we had some sizable areas not participate. In addition, we received usable artifacts for only one upper division course, thus making assessing growth impossible. Unfortunately, in this type of scenario the resources used to conduct assessment, such as faculty time, are squandered. It is an inefficient use of faculty time and university resources to engage in assessment efforts when the resultant data is unusable. Below, only the overall scores are presented since a lower/upper division Core comparison is impossible, and even lower division scores would not be generalizable due to the lopsided nature of artifacts collected.

Overall Score Distribution by Outcome



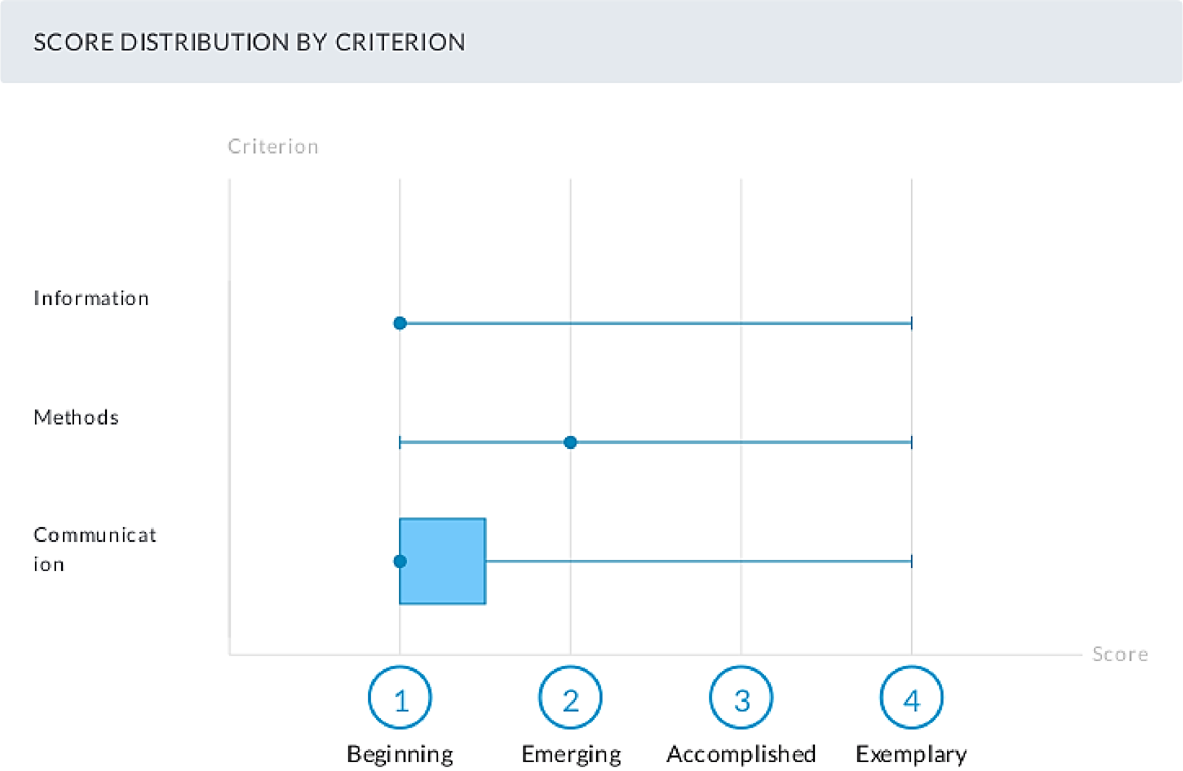
	Max Score	Min Score	Median Score	# of submissions
Define Problem	4	1	2	574
Propose Hypotheses	4	1	2	566
Identify Methodology	4	1	1	574
Evaluate Results	4	1	2	99

If we interpret these scores presuming they reflect only lower division courses, and if we assume they reflect lower division students in our Core science courses as a whole, we can see that the only notable point might be that the median score for methodology (defined as: Selecting the appropriate set of procedures to test the hypotheses.) was a one, with the upper quartile not even reaching a 2. This would suggest that either students aren't being asked to select methods for scientific inquiry, or can't do so. Given that the majority of lower level science courses are about introducing students to scientific method, it would not be surprising if they were being asked to perform best laboratory practices without having to discern between options. That would explain the median score of 1, and would also be reasonable pedagogy given students' need to be able to perform best practices before being asked to discern between options.

Goal C (Quantitative)

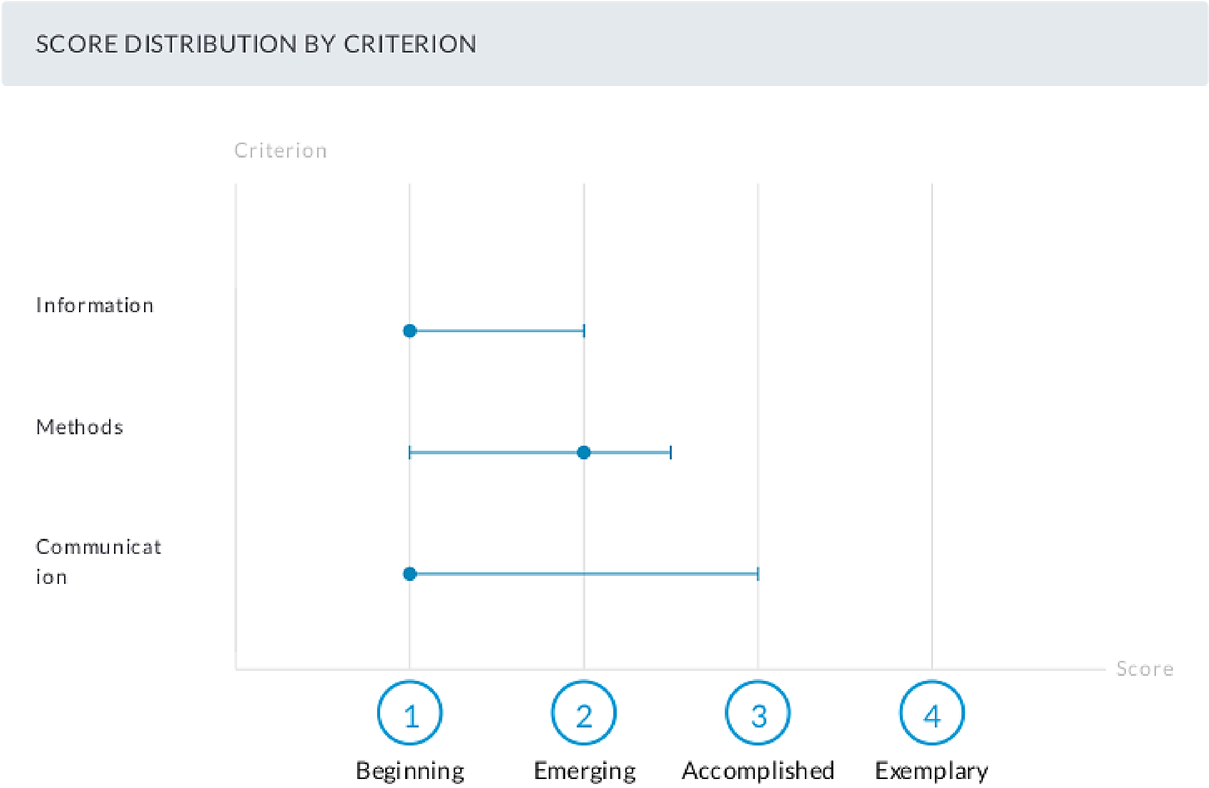
Goal C was unique insofar as a great deal of artifacts from the lower division were standard style assignments, or similar insofar as they were from college algebra or intro level math courses. Thus, the score team was able to process them quickly, and agree at a significant rate.

Overall Score Distribution by Outcome



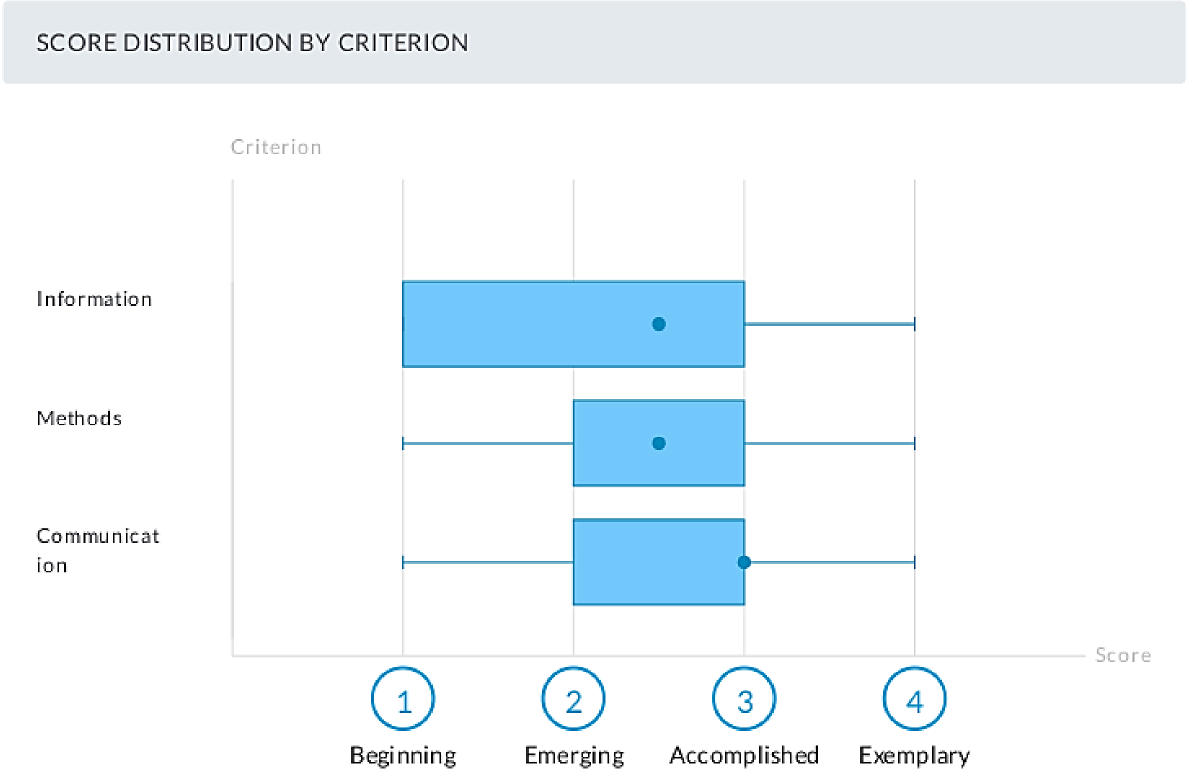
	Max Score	Min Score	Median Score	# of submissions
Information	4	1	1	941
Methods	4	1	2	941
Communication	4	1	1	936

Lower Division by Outcome



	Max Score	Min Score	Median Score	# of submissions
Information	2	1	1	669
Methods	2.5	1	2	669
Communication	3	1	1	665

Upper Division by Outcome



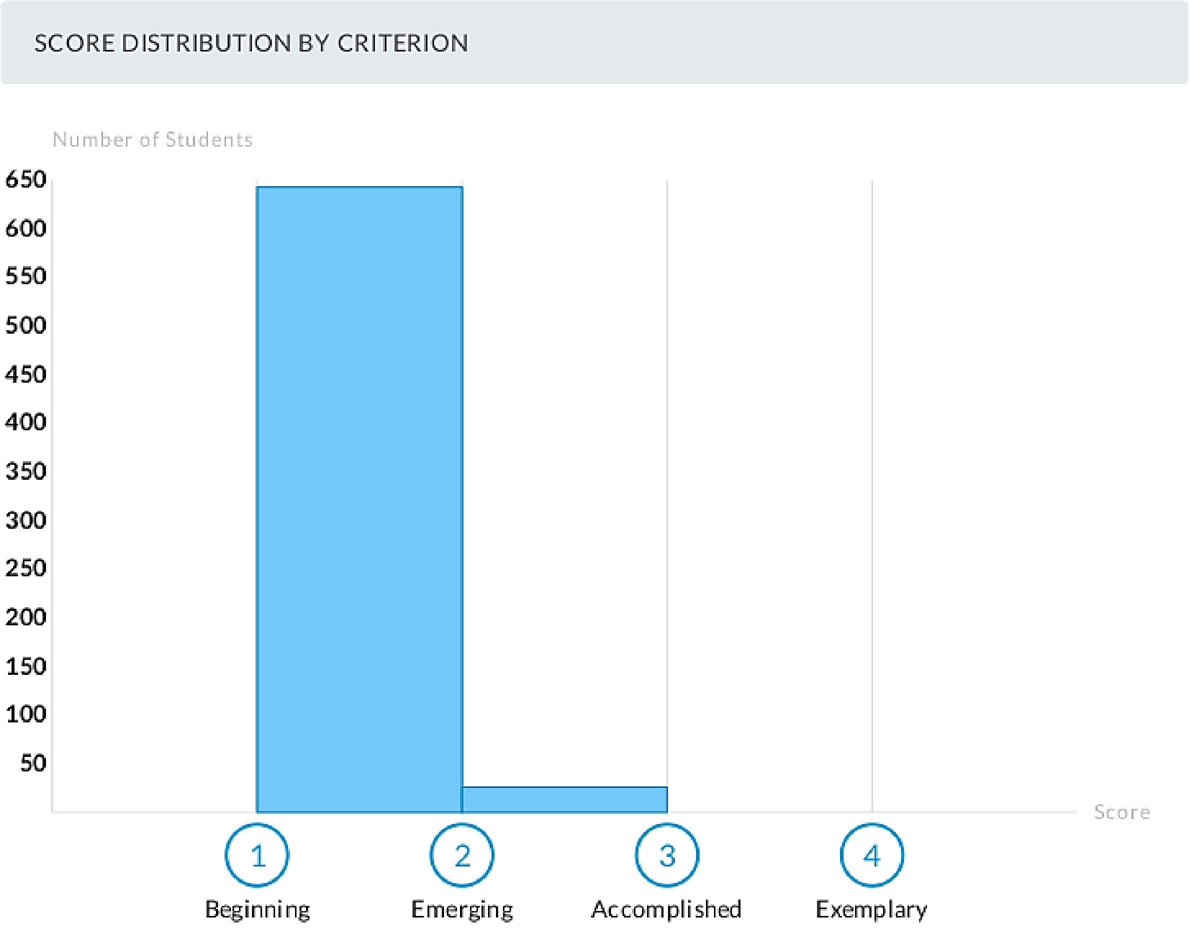
	Max Score	Min Score	Median Score	# of submissions
Information	4	1	2.5	272
Methods	4	1	2.5	272
Communication	4	1	3	271

Most notable is the growth across the information and communication outcomes. The lower division median scores were 1, which is to be expected given the nature of artifacts received. But the subsequent median scores at the upper division of 2.5 and 3, respectively, suggests growth.

SLO 1: Information

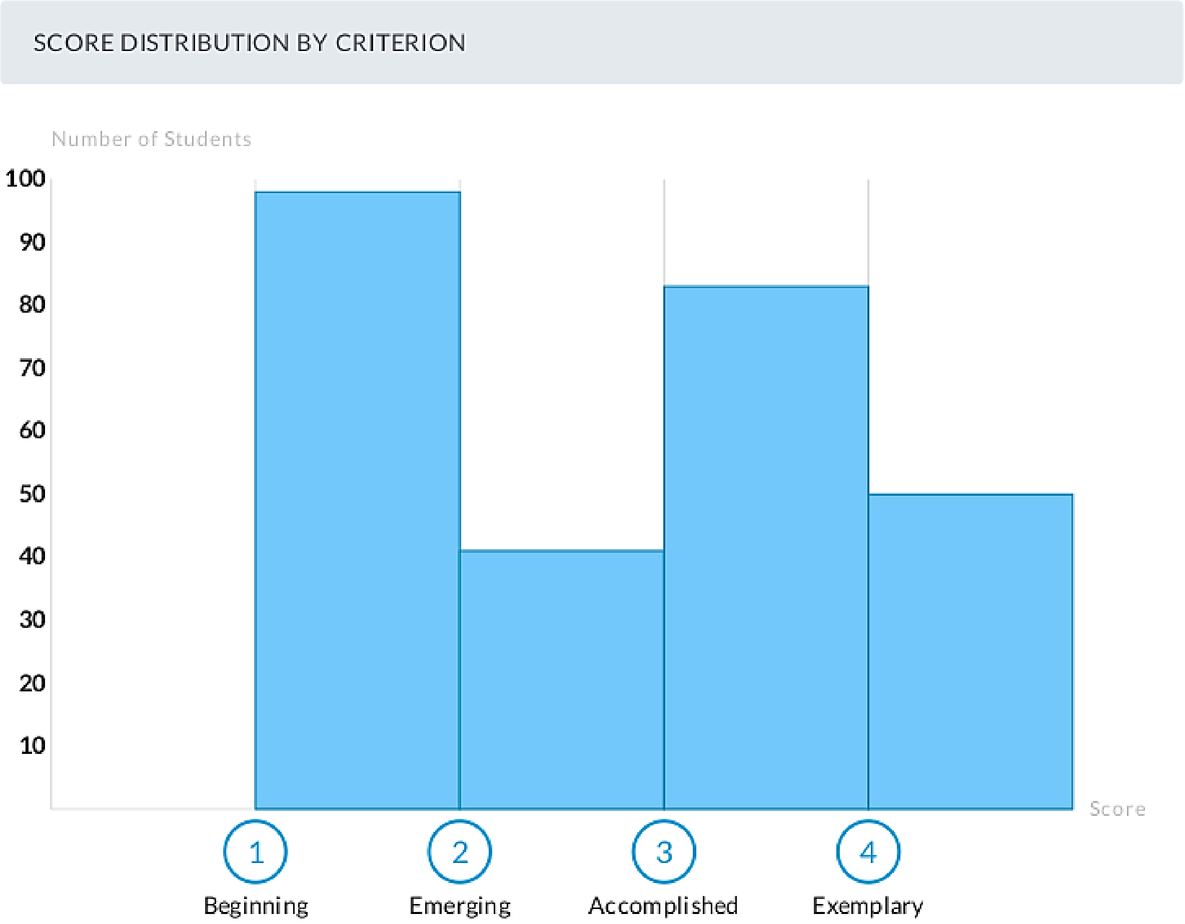
Information: Identifying and extracting relevant information needed to solve the problem.

Lower Division Frequency



	1- Beginning	2- Emerging	3- Accomplished	4- Exemplary
# of Scores	643	26	0	0
% of Scores	96.11	3.89	0	0

Upper Division Frequency


SCORE DISTRIBUTION BY CRITERION

	1- Beginning	2- Emerging	3- Accomplished	4- Exemplary
# of Scores	98	41	83	50
% of Scores	36.03	15.07	30.51	18.38

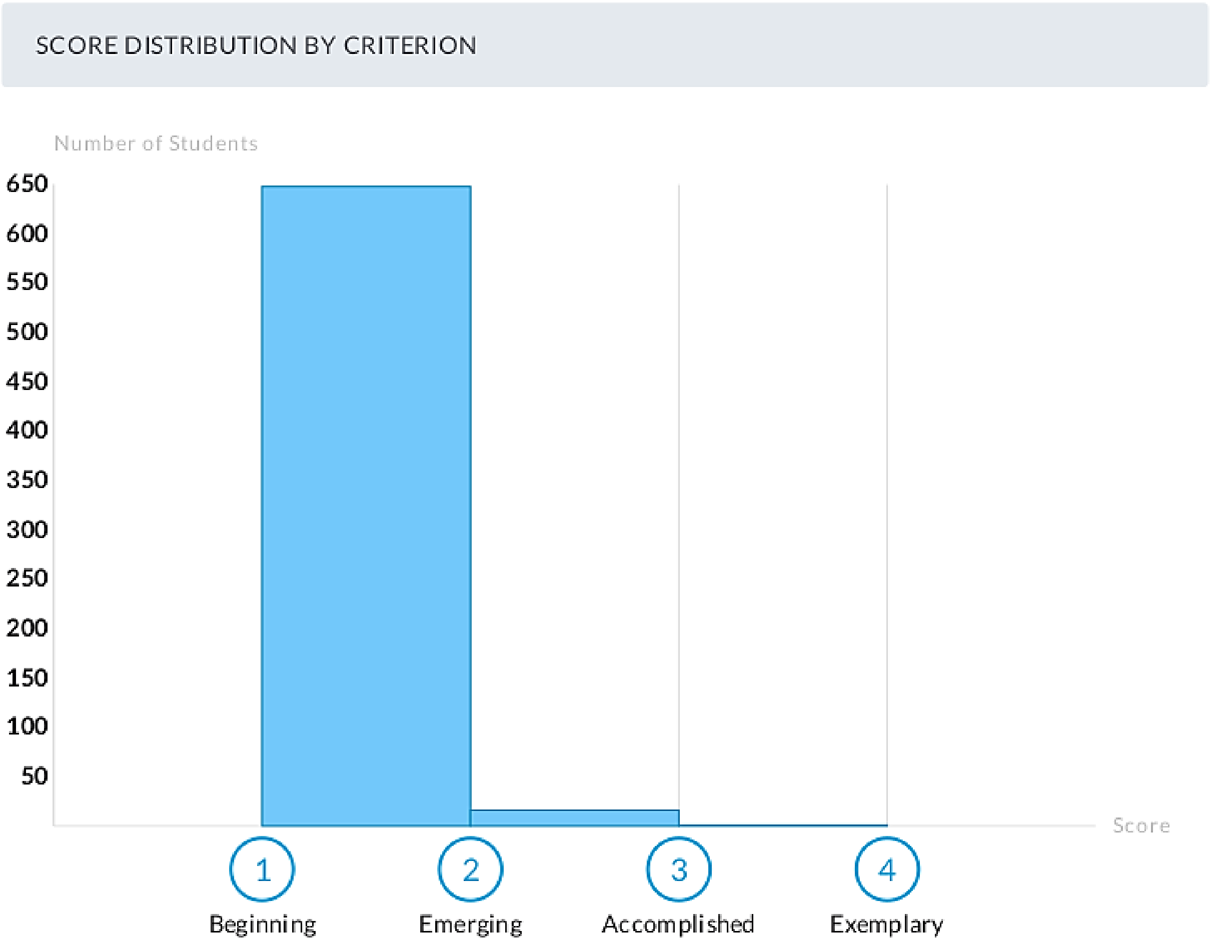
At the lower division, we note 96.11% of all artifacts receiving a score of 1. Clearly, the assignments, as designed, offered students a limited opportunity to demonstrate competence at this level. This interpretation is borne out by scorer comments presented in Appendix B.

At the upper division, we do see growth, which is to be expected when 96% of students' artifacts score a 1 at the lower division. Although growth is promising, it should be noted that less than 50% of our students score "accomplished" or above at the end of their curriculum in this outcome. The 50% mark presents us a benchmark for future iterations of Critical Inquiry assessment.

SLO 3: Communication

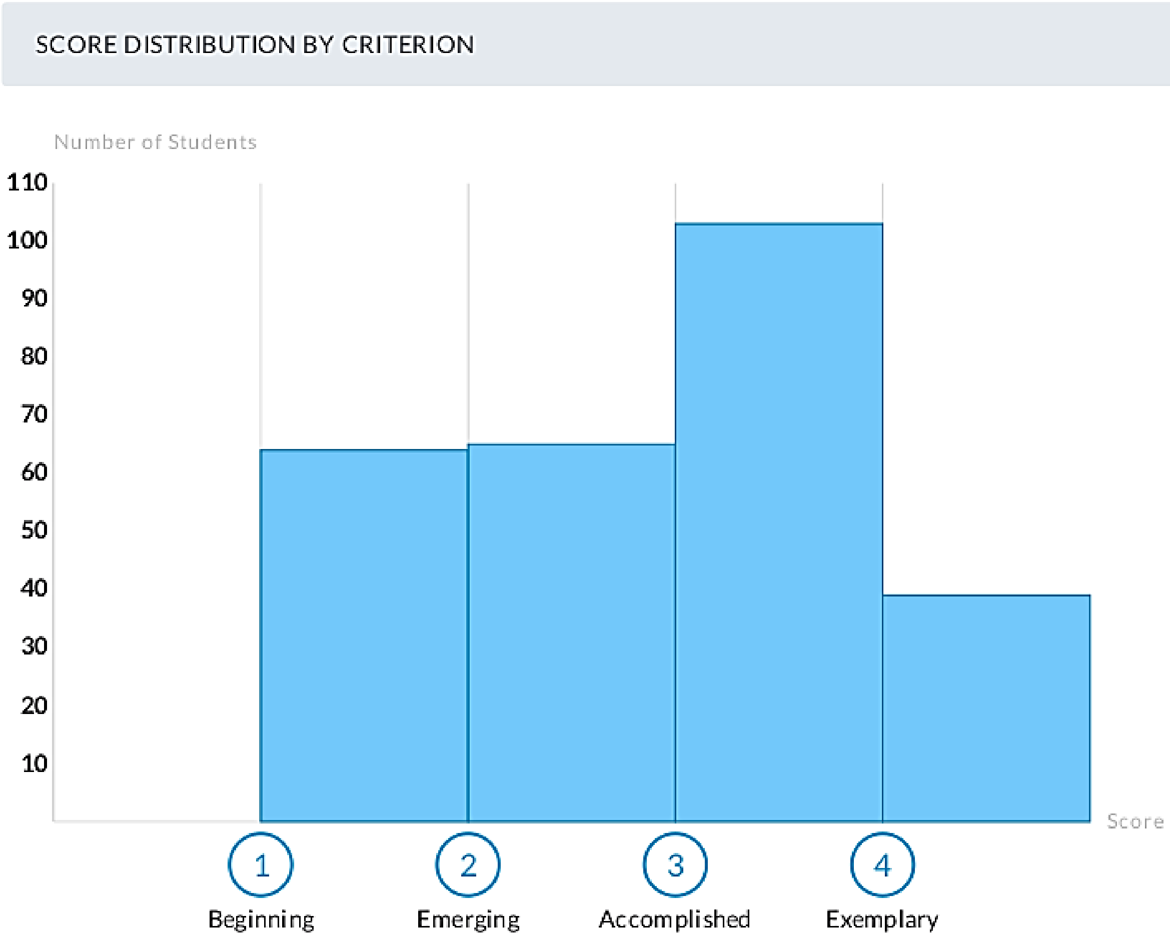
Communication: Effectively communicating quantitative concepts or evidence consistent with the purpose of the assignment.

Lower Division Frequency



	1- Beginning	2- Emerging	3- Accomplished	4- Exemplary
# of Scores	648	16	1	0
% of Scores	97.44	2.41	0.15	0

Upper Division Frequency



	1- Beginning	2- Emerging	3- Accomplished	4- Exemplary
# of Scores	64	65	103	39
% of Scores	23.62	23.99	38.01	14.39

At the lower division, we note 97.44% of all artifacts receiving a score of 1. Clearly, the assignments, as designed, offered students only an opportunity to demonstrate competence at this level. This interpretation is borne out by scorer comments presented in Appendix B.

At the upper division, we do see growth, which is to be expected when 97% of students' artifacts score a 1 at the lower division. Although growth is promising, it should be noted that only 52% of our students score "accomplished" or above at the end of their curriculum in this outcome. There is clear room for improvement here, and the 52% mark presents us a benchmark.

Assessment data for the “Methods” outcome is included in Appendix C. Given that the data, as well as scorer commentary, suggests that the scores at the LD level were a result of narrowly defined course assignments, that data did not appear to accurately reflect student acumen and so was not included in the full report. The UD level data shows similar results to the other UD data in this Goal, but trends are not discernible given the interpretation of the LD Data.

V. Conclusions and Recommendations

The data is suggestive of several things and supports the following observations and recommendations:

1) Faculty participation continues to be an issue. Without participation from faculty, assessment of the UCA Core cannot be done. The office of assessment has attempted to engage faculty in myriad ways. Perhaps chairs and deans can assist in supporting faculty as they participate in the assessment process as the owners and producers of the general education curriculum. The office of assessment has taken the following measures to address these issues:
    a. The survey instrument has been revised to be of easier use.
    b. Faculty are identified through ARGOS and contacted multiple times via UCA Inform and targeted emails.
    c. Chairs are contacted through "Academic Council" to inform them about general education assessment efforts and to ask their assistance.
2) Given that poorly chosen or designed assignments pose a problem in terms of generating representative data, a problem frequently noted by the score teams (See Appendix B), pre-cycle training needs to focus on assignment design and needs to be readily accessible and more widely used by faculty. In an attempt to address this issue, as well as a response to Covid-19, in spring 2020, pre-cycle trainings were offered on-line and posted on the UCA Core website for ease of access along with various educational materials. (see "Assessment" at https://uca.edu/core/for-faculty/). This practice will be continued in the future. Attendance was promising, and having resources readily available is prudent.
3) With respect to student learning: significant growth was noted in some areas. These numbers afford us the opportunity to set benchmarks against which to evaluate the data from the second cycle of assessment of the critical inquiry competency. However, the fact that only 50-60% of students at the upper division scored "accomplished" or higher, with markedly less than 20% of students scoring "exemplary" is worrisome. Students at the upper division should demonstrate a greater amount of mastery at higher rates if our curriculum is to be adjudged impactful. As one scholar notes, "Without improvement in how students are taught and how much they learn, many young people may graduate without the knowledge and skills to obtain the jobs they covet, become engaged and enlightened citizens, or gain the insights and interests that will help them live fuller and more rewarding lives."[6] If our programming is intended to assist students in developing various intellectual competencies, we must do better than graduating students with less than 1:5 achieving a level of mastery. The office of assessment will communicate these findings with relevant stakeholders and work to promote and develop improvement measures in the curriculum to increase student performance and teacher effectiveness in this regard.
4) Given scorer feedback, the rubrics need to be revisited. Issues to be addressed include the ability of non-experts to apply the rubric in the assessment exercises as carried out at UCA and more precise and consistent wording in the rubrics to better standardize expectations among scorers.

[6] Derek Bok, *Higher Education in America* (Princeton, Princeton University Press, 2013), pp. 408-9. In this work, research is also cited that indicates moderate, at best, increases in student learning during their time in the higher education system.

Appendix A: Scorer Comments in AQUA

Goal A

Assignment may have been structured to not require or depend on external sources, but to center human subjects data only.

Better fit for Rubric B?

Good fit for Rubric B

I marked the information category NA (Not Applicable) because the artifact was a self-critique of a piece of art and did not use any sources.

Rubric B?

Student's assignment is blank.

Suitable for Rubric B?

The assignment reads as though it was intended to analyze only one work (the student's original work) using only Aristotle's principles, so earning higher than a score of 2 on Information may have been impossible.

The author did not acknowledge any sources in the artifact.

The author did not acknowledge any sources in this artifact, which made it impossible to assess the information goal.

The author did not acknowledge sources in the artifact.

The author did not acknowledge sources in this artifact, which made effective assessment of the information goal impossible.

The author did not acknowledge sources in this artifact.

The author did not cite sources. It was impossible to score the information section of the rubric.

The author did not provide information in the text or did they provide a works cited page. It was impossible to evaluate the information portion of the rubric.

The nature of this assignment all but precludes earning a score of higher than 2 on Information.

The nature of this assignment might have precluded earning a score of higher than 2 on Information.

The nature of this assignment precludes earning a score of higher than 2 on Information.

The nature of this assignment probably precluded earning a score of higher than 1 or 2 on Information.

The nature of this assignment probably precluded earning a score of higher than 2 on Information.

The nature of this assignment seems to have precluded earning a score of higher than 2 on Information.

The nature of this assignment seems to preclude earning a score of higher than 2 on Information.

The nature of this assignment seems to preclude earning higher than a 1 or a 2 for Information.

This appears to be a summary/analysis of a single source.

This appears to be an analysis of one text (the student's own, original piece). It appears to apply Aristotle, though this student did not mention Aristotle.

This assignment appears to require the analysis of a single source, perhaps assigned by instructor.

This assignment does not seem to align with the rubric for Critical Inquiry Goal A (Inquiry and Analysis) and seems to be more appropriate for Goal B (Scientific).

This assignment does not seem to engage external information/sources.

This assignment may have precluded earning higher than 1 for Information.

This assignment may have precluded earning higher than 1 or 2 for Information.

This assignment may have precluded earning higher than a 1 or 2 for Information.

This assignment might preclude earning higher than 1 or 2 for Information.

This assignment requires analysis of one source only, and looks to be assigned by the professor.

This assignment requires discussion of only one source.

This assignment seems designed to preclude earning higher than 1 or 2 for Information.

This assignment seems designed to preclude earning higher than a 1-2 for Information.

This assignment seems designed to preclude earning higher than a 2 for Information.

This assignment seems to be designed around a single source provided by the professor.

This assignment seems to be designed to preclude earning higher than 1 in Information.

This assignment seems to be designed to preclude earning higher than 1 or 2 in Information.

This assignment seems to preclude a student's earning higher than a score of 2 for Information.

This assignment seems to require analysis of only one source.

This assignment seems to require the analysis of a single source.

This assignment seems to require the analysis of just one piece/passage.

This assignment seems to require the application of Aristotle's principles to just one work: the student's original piece.

This assignment was an analysis of one source, and precludes earning higher than 1 or 2 for Information.

This should be Critical Inquiry Rubric B rather than Critical Inquiry Rubric A

This should be Critical Inquiry Rubric B.

This should be scored on Critical Inquiry Rubric B rather than Critical Inquiry Rubric A.

This should be scored under Critical Inquiry Rubric B rather than Critical Inquiry Rubric A.

This assignment does not seem to align with the rubric for Critical Inquiry Goal A (Inquiry and Analysis) and seems to be more appropriate for Goal B (Scientific).

More appropriate for Rubric B?

Goal B

essentially same experiment (student provided no new info)

I believe this should go under quantitative rubric

In this assignment students are given a method, thus cannot access student ability to identify a method to test question/hypothesis.

lab protocol given to the students

method is determined by instructor

method was given to the students

methodology simply referenced a paper so unable to assess

since it referenced a paper there is no way to know what the methods actually were to evaluate them based on rubric

this should go with the quantitative rubric I believe

this should go with the quantitative rubric I believe

We followed the procedure from Lab 6: Enzymes Part II in the lab manual

lab protocol given to the students

Goal C

None

Appendix B: Sundry Observations from Relevant Stakeholders

Re: Goal B

Re: Define Problem - Without writing a test it's very hard to assess if a student can COMPARE problem statements, most assignments request that they create one. How this rubric is currently written it is hard to use ONE assignment that will work for all statements, especially in large classes. In terms of lower division courses, we have found this rubric would be easier to use if all statements were expressed similarly (Communicates.... understanding of the problem) with differentiation between scores being how thoroughly they communicate. An example is Communicates a rudimentary understanding....

Re: Propose Hypotheses - In practice we have found it's easier to say develops a hypothesis that relates to the problem statement

Re: Identify Methodology - Again it's hard to create an assignment that distinguishes between options

Re: Evaluate Results - This one is again hard to assess. Most assignments for assessing critical inquiry require the students to actually interpret the results, thus they may interpret results incorrectly but we aren't truly assessing if they can recognize an accurate interpretation. In the classroom with ALL of the discussions and assignments we can make a solid judgment, but it won't be possible from a single assignment scored by an outside observer.

Re: Goal C

Post-assessment scorer observations

1) Need to know the assignment directions, and for numerical problems a key or solutions guide would be most helpful since they are assessing also for accuracy. (Even the grades were helpful)
2) For Information, extract (as opposed to recognize) is different and a lot of the lower level math assignments didn't provide any opportunity to extract. They were formulaic plug and chug style assignments.
3) The assignment design is extremely important in this rubric since the difference between the levels is about discrete skills and jumps.
4) Standard assignments in the courses would be helpful.
5) Because of the difference between the goals, B and C go more quickly and don't need as many resources as Goal A. So devote more to A (which is paper grading) than B and C which are formulaic, standard assignments.
6) Group projects are really hard and inaccurate to assess in this way. So we need to focus on assignment design ahead of time to get artifacts that accurately present students' abilities.
7) Need to see where students are and where they can get so the assignments have to be well designed and really prompt students to give us their best. Maybe lower levels should provide chances for scores of 1-3, and upper levels have to really show they can achieve a 4. (Assignment design can prejudice the results.)
8) Reinforce in training that grading and assessment are different as our expectations are different in courses depending on level.

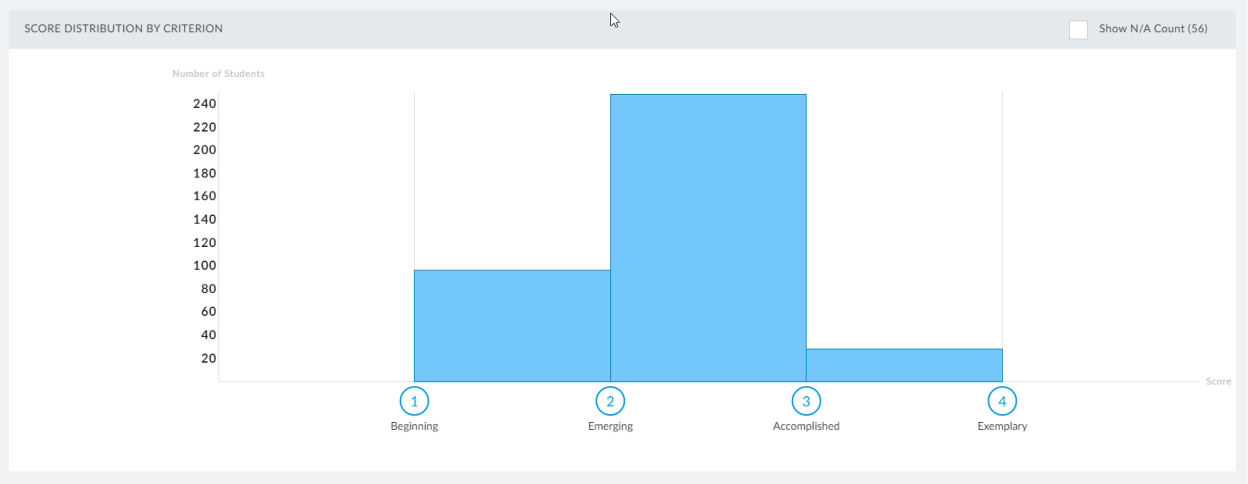
9) In training perhaps give homework or exercise where they look at their assignment, align it to the rubric and see or anticipate where their students would or could get and then see if the assignment needs to be revised in some manner.
10) Get feedback and results to departments and faculty from the actual results. Faculty need to see this as informative.

Appendix C: Additional Assessment Data

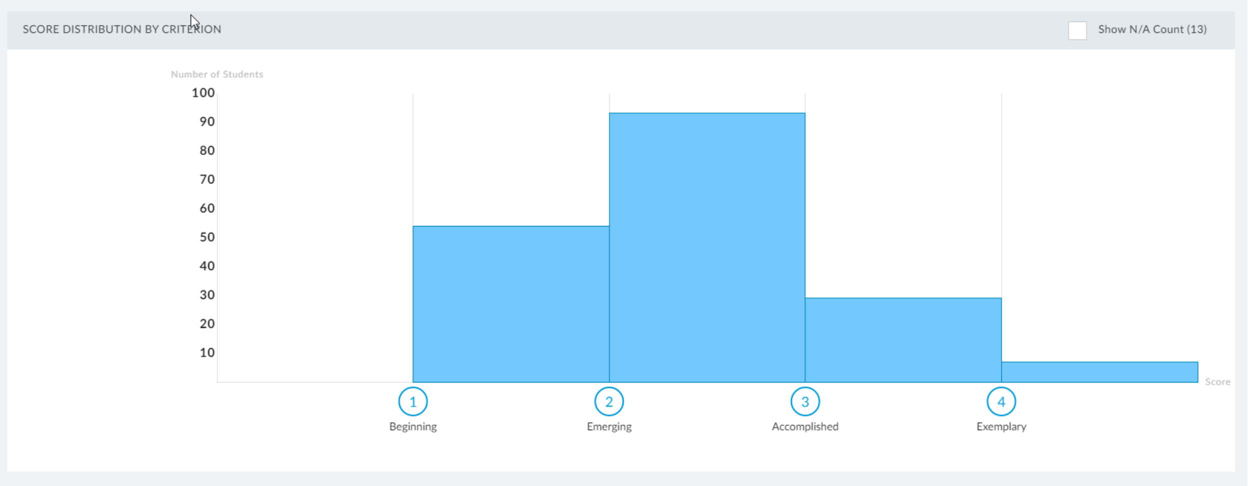
Goal A:

Information

LD Frequency



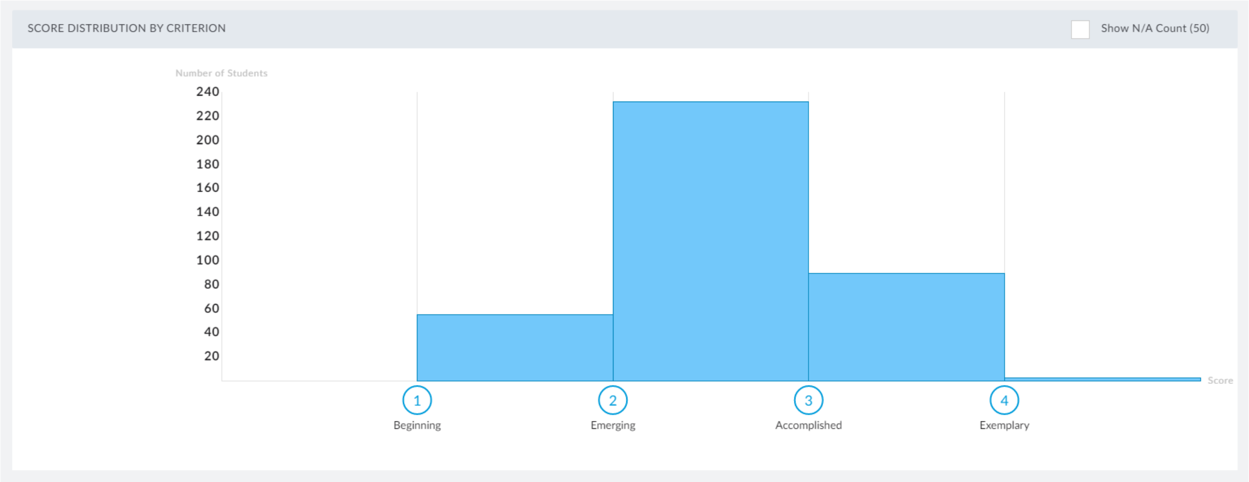
UD Frequency



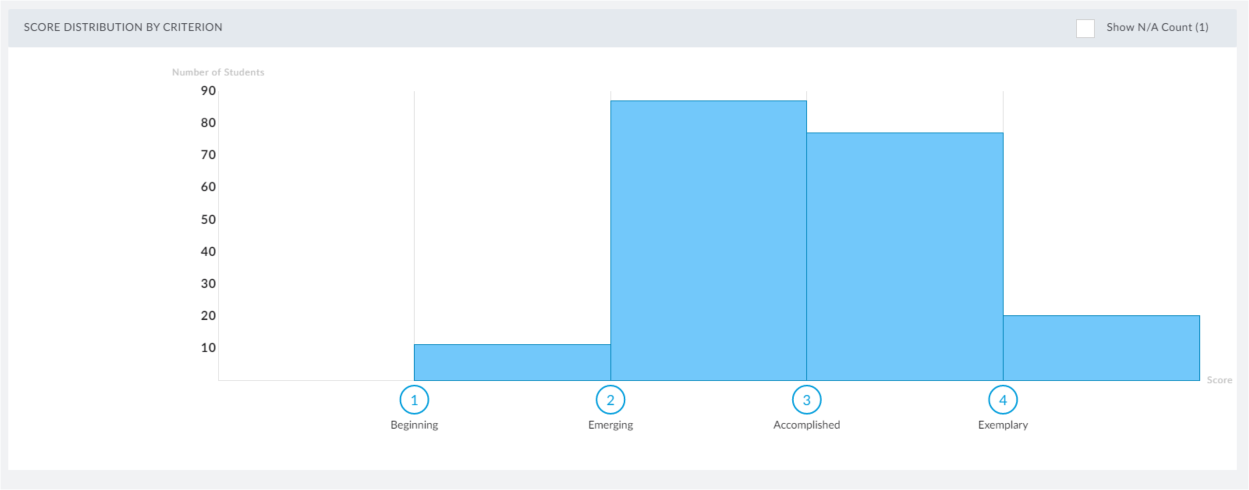
Goal A:

Analysis

LD Frequency



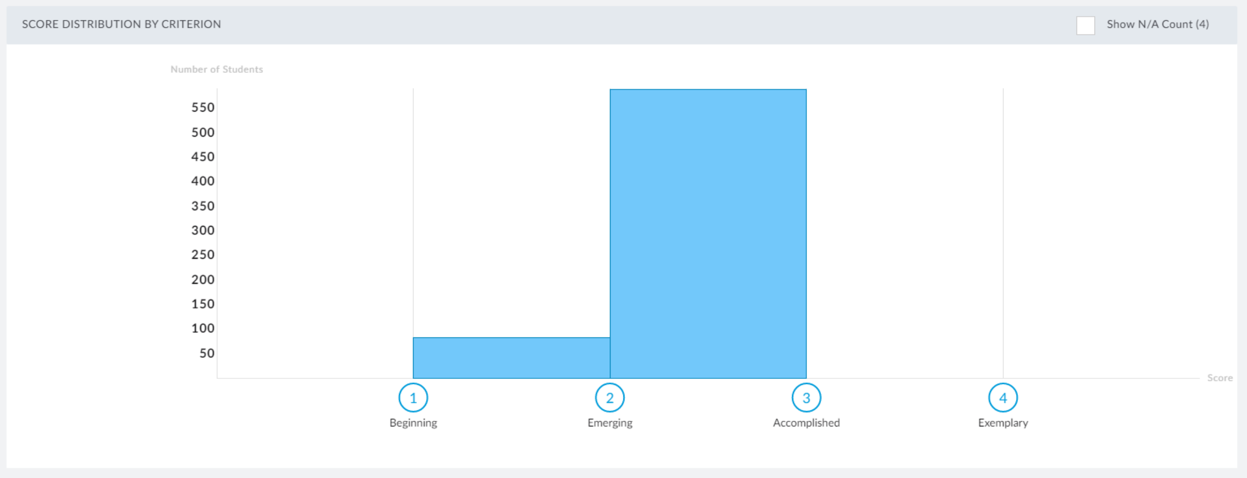
UD Frequency



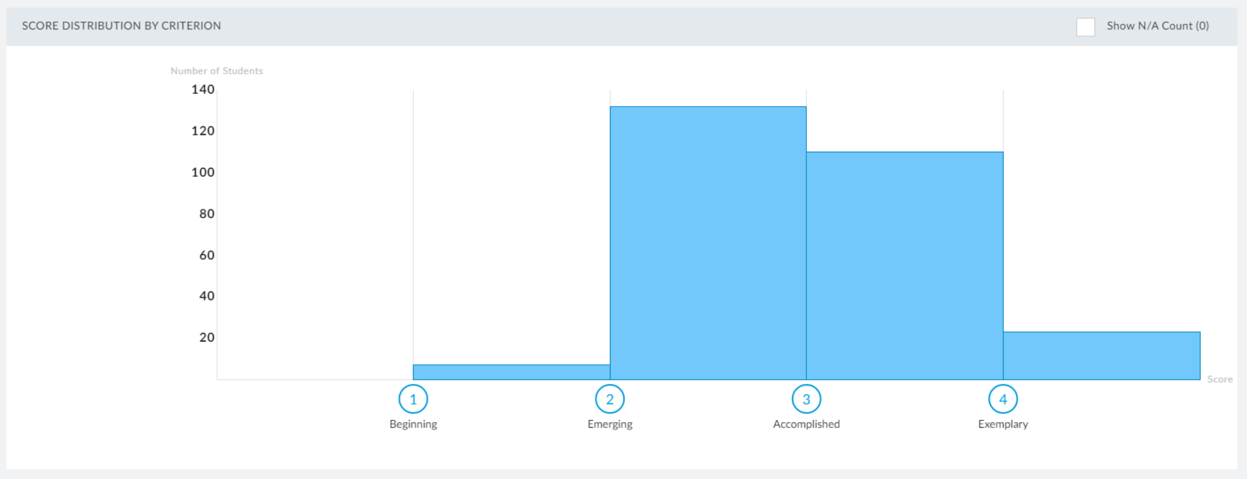
Goal C:

Methods

LD Frequency



UD Frequency


SCORE DISTRIBUTION BY CRITERION
□ Show N/A Count (0)
Number of Students
Score
Beginning
Emerging
Accomplished
Exemplary