Spring 2023 General Education Assessment Report

Introduction and Context:

The Spring 2023 report focuses on assessing Student Learning Outcome (SL2.O) 2a and 2b, which pertain to Scientific Inquiry and Quantitative Reasoning respectively, in courses falling under the C.1 and C.2 categories of the General Education curriculum. In Spring 2023, every course falling under SLO 2a and 2b categories was scheduled to be assessed.

In Fall 2022, the GEAC revised the rubrics from holistic to analytic rubrics for C.1 and C.2 courses and implemented them in Spring 2023. Faculty members who taught these courses in Fall 2022 provided feedback on the revised rubrics. The aim of using analytic rubrics was to help faculty understand more precisely where our students need to improve their performance to achieve the SLOs so that teaching interventions or curricular changes can be made. To facilitate the norming and rating of rubrics, a Spring General Education Retreat was scheduled on May 23, 2023, at the University's McFarland Student Union. Leading up to the retreat email reminders were sent out to highlight the changes from holistic to analytic rubrics, to announce the agenda for the retreat, which included norming and rating sessions, and to provide opportunities for Q&A.

The two SLOs are defined as follows:

SLO 2a

Student Learning Outcome 2a, Scientific Inquiry, involves students in the process of observing natural phenomena, organizing, and interpreting their observations, constructing explanatory models, and using those models to make predictions and evaluate their accuracy. Through this iterative process of refining and testing their models, students attempt to gain a deeper understanding of the physical and natural world around them. This SLO is met by courses falling under the C.1 category of Understanding Science and Technology.

SLO 2b

Student Learning Outcome 2b, Quantitative Reasoning, involves creating sophisticated arguments that are supported by quantitative evidence and communicated clearly using a variety of formats such as words, tables, graphs, mathematical equations, and other quantitative data representations. Mathematics plays an important role in decision-making and problem-solving that require quantitative reasoning, which can involve algebraic, numerical, symbolic, or graphical representations based on logical structures. The use of procedural skills is also an inherent aspect of quantitative reasoning, enabling the generalization and application of results to specific problems or decisions, including those related to the physical and natural world. This SLO is met by courses falling under the C.2 category of Understanding Science and Technology.

Methodology:

Faculty members teaching courses falling under the C.1 and C.2 categories of the General Education curriculum were contacted by email with instructions to submit student artifacts via D2L. After faculty submitted student artifacts for SLO 2a and 2b at the end of the Spring 2023 semester, the artifacts were prepared for independent rating at the Spring General Education Retreat. Each artifact was given an accession number to maintain student and instructor identity confidential. Faculty volunteers were to rate student artifacts using the revised analytic SLO 2a rubric, which focuses on scientific inquiry, to determine performance levels. Similarly, faculty volunteers were to rate student artifacts using the revised SLO 2b rubric, which focuses on quantitative reasoning, to determine performance levels. Both rubrics were developed so areas of strengths and needs for improvement, based on ratings, could be determined. Consequently, faculty teaching C.1 and C.2 general education classes will have a benchmark against which they can potentially adjust their teaching practice in a responsive way.

During the Spring General Education Retreat approximately 26 raters were in attendance consisting of faculty volunteers, GEAC and GEC members. On that day, only artifacts from courses in C.1 category were available for rating because of the limited number of volunteers, low compliance from C.2 courses, or lack of clarity about how C.2 artifacts related to the SLO. Raters were paired, with each pair consisting of one physical or life science professor and one non-science professor. A norming session was conducted in the morning with three anchors provided, *i.e.*, samples of student artifacts. The four individual skills or dimensions of the SLO 2a rubric were unpacked so that raters could have a common understanding of each skill and performance level measure. The skills include: *Dimension 1* – Demonstrate an understanding of how scientific principles can be used to answer a question, solve a problem, or test a hypothesis, *Dimension 2* – Properly organize and evaluate data from a scientific investigation, case study, observations, and/or experimentation, *Dimension 3* – Critically interpret results from a scientific investigation, case study, observations and/or experimentation to draw conclusions, and *Dimension 4* – Communicate findings.

Norming procedures included the following steps: *Step 1* – Read the entire artifact, *Step 2* – Re-read and highlight evidence, *Step 3* – Assign points using the rubric, *Step 4* – Discuss artifact and rubric scoring with partner, *Step 5* – Recalibrate and enter one score into online system as a pair. No half-points were permitted in this analytic rubric. Performance levels were either a 4 (excellent), 3 (good), 3 (fair), 1 (poor), or 0 (unable to rate). Rating pairs were expected to rate independently, then converse to come to a consensus. If a pairs' scoring differed by one point on more than two dimensions, a third rater was called in to participate. Rating of student artifacts occurred in the afternoon and all scores were uploaded to a Google Form using the artifact accession number.

Analysis of the scores assigned to each artifact are summarized as the overall frequency distribution of performance level scores for each of the four dimensions, the average performance level score for each dimension based on the courses' area of study, e.g., business, science or social science, and the courses' prefix, e.g., BIO, ANT, etc.

Results

The compliance rate for submitting student artifacts was 100% for C.1 courses (n= 23), and 70% for C.2 courses (n= 52). A total of 130 C.1 student artifacts were scored after the norming session during the Spring General Education Retreat. The number of items scored 0 (unable to rate) varied for each of the four dimensions, with *Dimension 1* and *Dimension 2* having highest number of items that were not rated (n= 41 and 57, respectively), and *Dimension 3* and *Dimension 4* having the lowest (n= 26 and 25, respectively; **Figure 1**). Of the artifacts that were rated, 88.8% of *Dimension 1* items (n= 89), 80.8% of *Dimension 2* items (n= 73), 87.5% of *Dimension 3* items (n= 104), and 80.9% of *Dimension 4* items (n= 105) were scored above 2 (fair) (**Figure 1**).

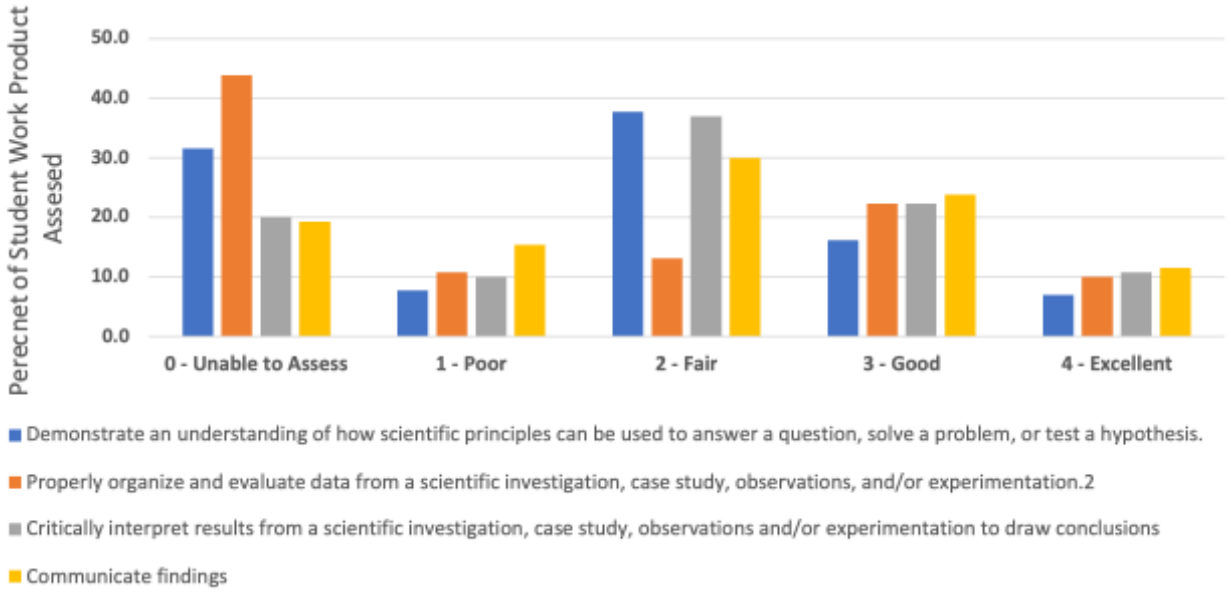

FIGURE 1. The percentage of student work (artifacts) submitted to KU GEAC for SLO 2a – C1 by courses taught in Spring 2023 that was scored as either 0 (unable to assess), 1 (poor), 2 (fair), 3 (good), or 4 (excellent) by pairs of volunteer raters. Bars are color-coded by the four dimensions of the analytic rubric.

Most of the student artifacts scored 0 (unable to rate) in all four *Dimensions* had a response assessed as not addressing the metric, while a smaller proportion had a response assessed as either addressing the wrong SLO or not addressing the *Dimension*. *Dimension 1* and *Dimension 2* had a higher proportion of artifacts scored 0 because the response was rated as not addressing the metric, than *Dimension 3* and *Dimension 4* (**Figure 2**). In contrast, *Dimension 3* and *Dimension 4* had a higher proportion of artifacts scored 0 because the response was rated as addressing the wrong SLO or not addressing the *Dimension*, than *Dimension 1* and *Dimension 2* (**Figure 2**).

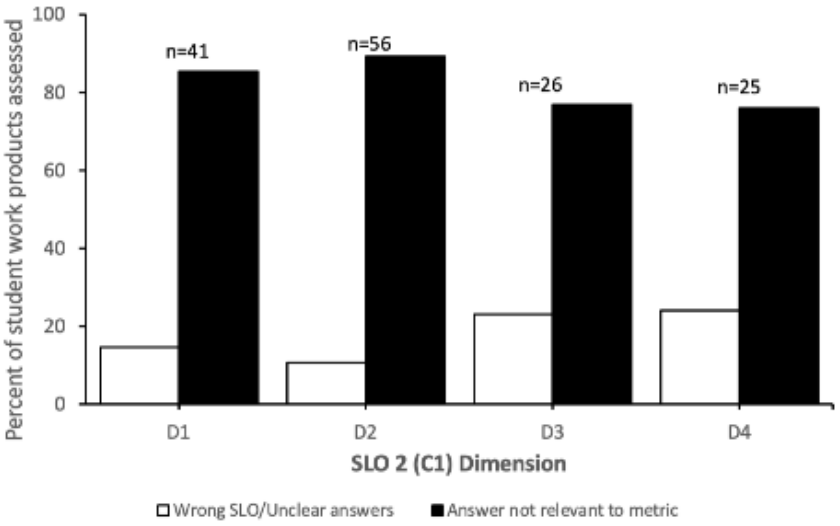

FIGURE 2. The percent of student work scored as 0 "unable to assess" for each of the four dimensions of the SLO 2 C1 rubric categorized as either addressing the wrong SLO/ equations and concepts leading to correct answers were unclear (white bars), or the response was thought to not address the metric (black bars). Data are from KU GE courses taught in Spring 2023.

The average score for student artifacts assessed was above 2 in all four *Dimensions* (**Figure 3**). The lowest average score (± Standard Deviation (SD)) was in *Dimension 1* (2.31 ± 0.78), followed by *Dimension 4* (2.37 ± 0.95), *Dimension 3* (2.41 ± 0.87), and *Dimension 2* (2.54 ± 0.99). *Dimensions 1* and *2* had fewer artifacts rated than the other two Dimensions because more of those artifacts were rated 0, i.e., assessed as addressing the wrong SLO, had unclear answers, or answers were not relevant to metric (**Figure 3**).

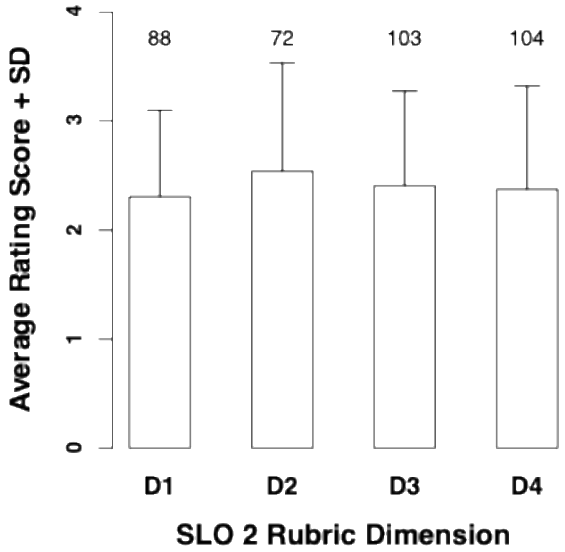

FIGURE 3. The average score (1-4) given to each of the four dimensions of the SLO 2 C1 rubric by pairs of raters to assessable artefacts submitted to KU GEAC by courses taught in Spring 2023. Scores given were either 1 (poor), 2 (fair), 3 (good), or 4 (excellent)

The average score for student artifacts according to the courses' area of study, i.e., science or social science, was above 2 in all four *Dimensions* (**Figure 4**). The average score ± SD was numerically higher in science courses for *Dimension 1* (2.49 ± 0.95 vs. 2.15 ± 0.59 for science and social science, respectively), *Dimension 2* (2.64 ± 1.2 vs. 2.44 ± 0.76), and *Dimension 3* (2.41 ± 0.99 vs. 2.39 ± 0.71), whereas the opposite pattern was measured for *Dimension 4* (2.32 ± 1.05 vs. 2.43 ± 0.82; **Figure 4**).

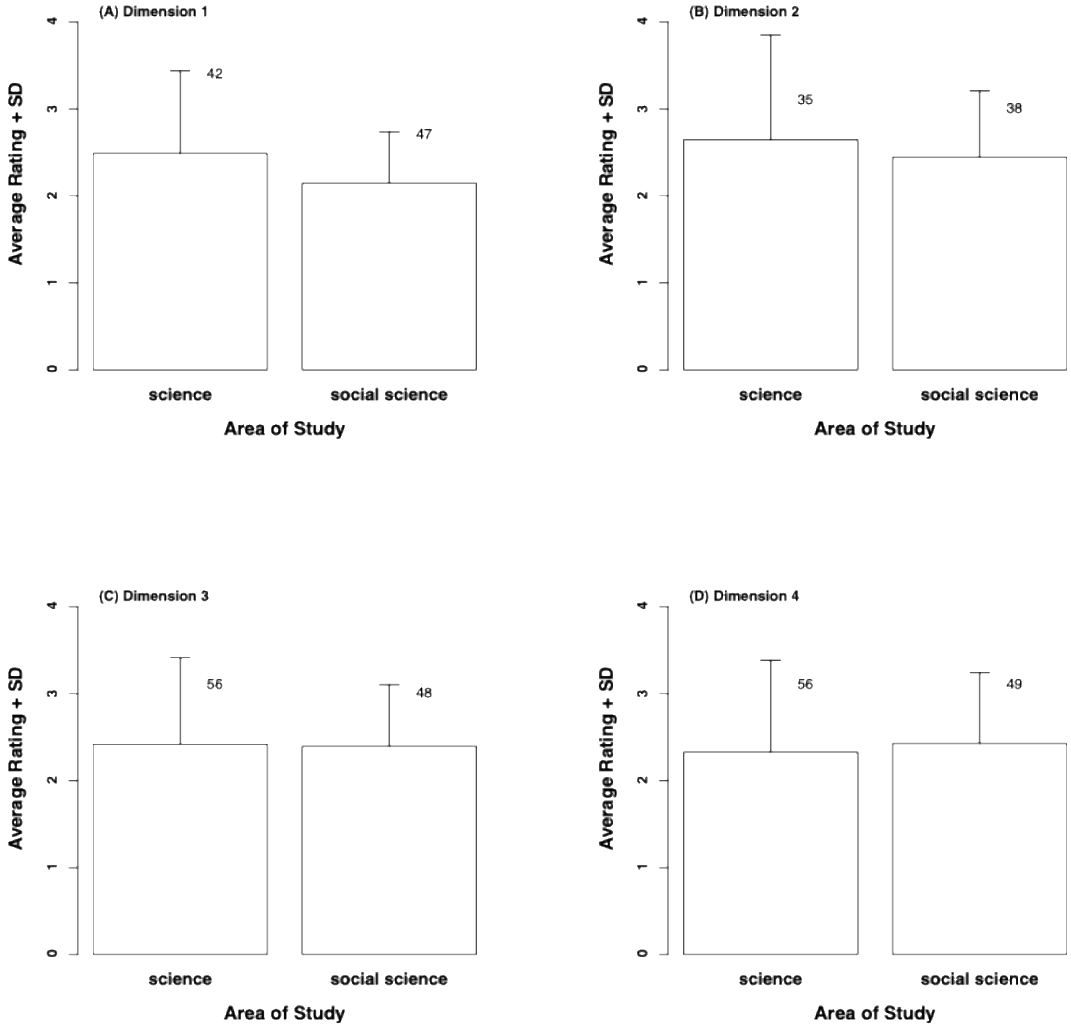

FIGURE 4. The average score in *Dimension 1* (A), *Dimension 2* (B), *Dimension 3* (C), and *Dimension 4* (D) of the SLO 2 C1 rubric for assessable artefacts submitted to KU GEAC that were rated by pairs of independent volunteers and arranged according to area of study (science or social science) in Spring 2023. Scores given were either 1 (poor), 2 (fair), 3 (good), or 4 (excellent)

The average score for student artifacts according to the courses' area of study, i.e., science or social science, and whether the courses were for majors or non-majors, was above 2 in all four *Dimensions* (**Figure 5**). The average score ± SD was numerically higher in major relative to non-major courses in both science and social science courses in *Dimension 1* (science major: 2.66 ± 0.77, science non-major: 2.41 ± 1.02, social science major: 2.16 ± 0.6, social science non-

major: 2.14 ± 0.59), and in *Dimension 4* (science major: 2.58 ± 1.08, science non-major: 2.26 ± 1.05, social science major: 2.45 ± 0.83, social science non-major: 2.41 ± 0.82; **Figure 5**). In *Dimension 2* and *Dimension 3* the average score ± SD was numerically higher in major relative to non-major science courses (*Dimension 2* - science major: 3.0 ± 0.95 vs. science non-major: 2.45 ± 1.3 and *Dimension 3* - science major: 2.92 ± 0.99 vs. science non-major: 2.28 ± 0.96), but higher in non-major relative to major social science courses (*Dimension 2* - social science major: 2.39 ± 0.85 vs. social science non-major: 2.5 ± 0.69 and *Dimension 3* - social science major: 2.37 ± 0.76 vs. social science non-major: 2.41 ± 0.68; **Figure 5**).

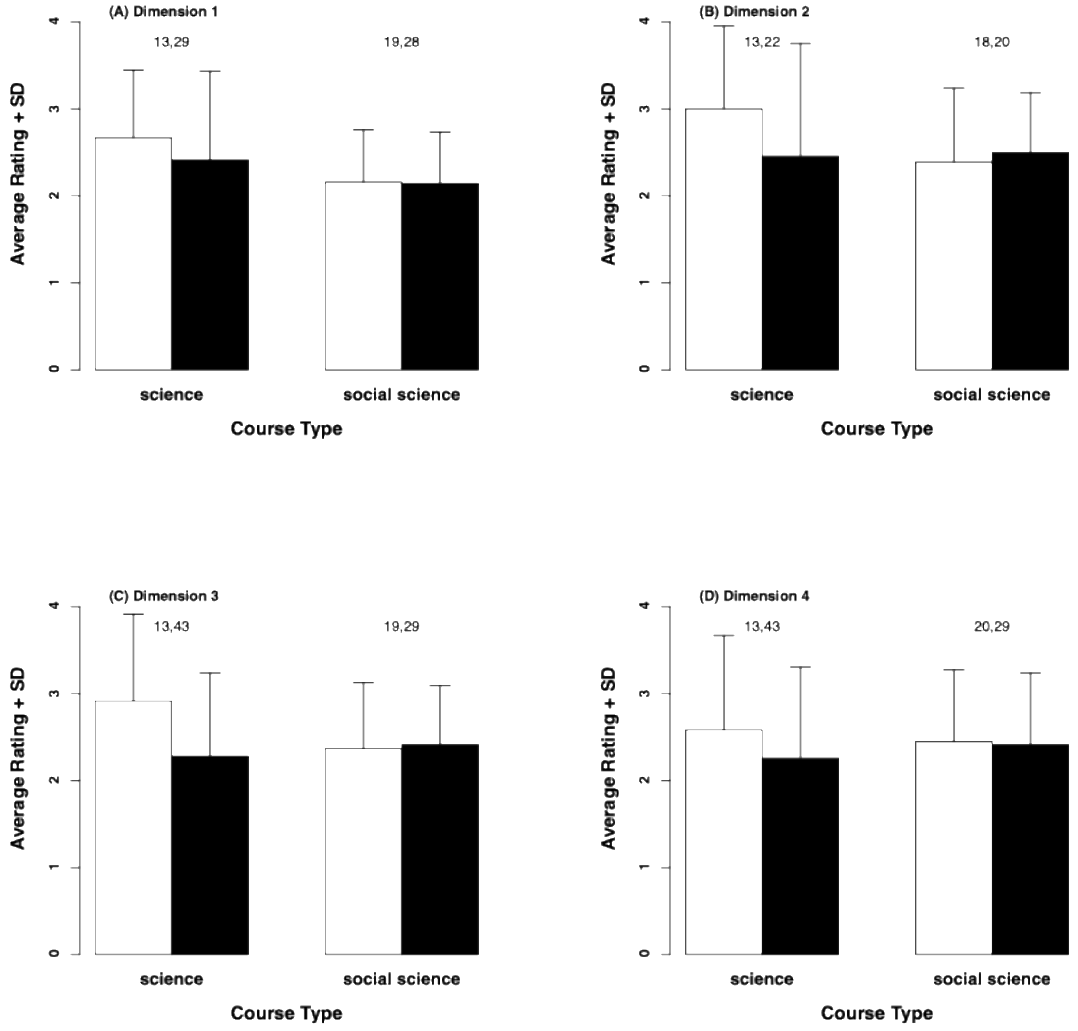

FIGURE 5. The average score in *Dimension 1* (A), *Dimension 2* (B), *Dimension 3* (C), and *Dimension 4* (D) of the SLO 2 C1 rubric for assessable artefacts submitted in Spring 2023 to KU GEAC and rated by pairs of independent volunteers arranged according to area of study (science or social science) and whether the course was for majors (white bars) or non-majors (black bars). Scores given were either 1 (poor), 2 (fair), 3 (good), or 4 (excellent)

The average score for student artifacts was above 2 for all course prefixes in *Dimension 3* and above 2 in all but one (*Dimension 2*) or two course prefixes (*Dimension 1* and *Dimension 4*; **Figure 6**). The average score ± SD was above 2 in *Dimension 1* for course prefix ANT (2.28 ±

0.68), BIO (2.25 ± 0.77), CHM (3.06 ± 0.93) and CRJ (2.2 ± 0.42), and below 2 for course prefix ASTWGS (1.89 ± 0.78) and GEG (1.83 ± 0.39; **Figure 6**). In *Dimension 2* the average score ± SD was above 2 for course prefix ANT (2.5 ± 0.73), CHM (3.56 ± 0.73), CRJ (2.5 ± 0.85) and GEG (2.33 ± 0.78), and below 2 for course prefix BIO (1.83 ± 0.92; **Figure 6**). In *Dimension 3* the average score ± SD was above 2 for course prefix ANT (2.5 ± 0.81), BIO (2.18 ± 0.83), CHM (3.0 ± 1.15), CRJ (2.2 ± 0.63) and GEG (2.33 ± 0.49; **Figure 6**). Finally, in *Dimension 4* the average score ± SD was above 2 for course prefix ANT (2.67 ± 0.78), CHM (3.19 ± 0.98) and CRJ (2.4 ± 0.84), and below 2 for course prefix BIO (1.97 ± 0.87) and GEG (1.92 ± 0.67; **Figure 6**).

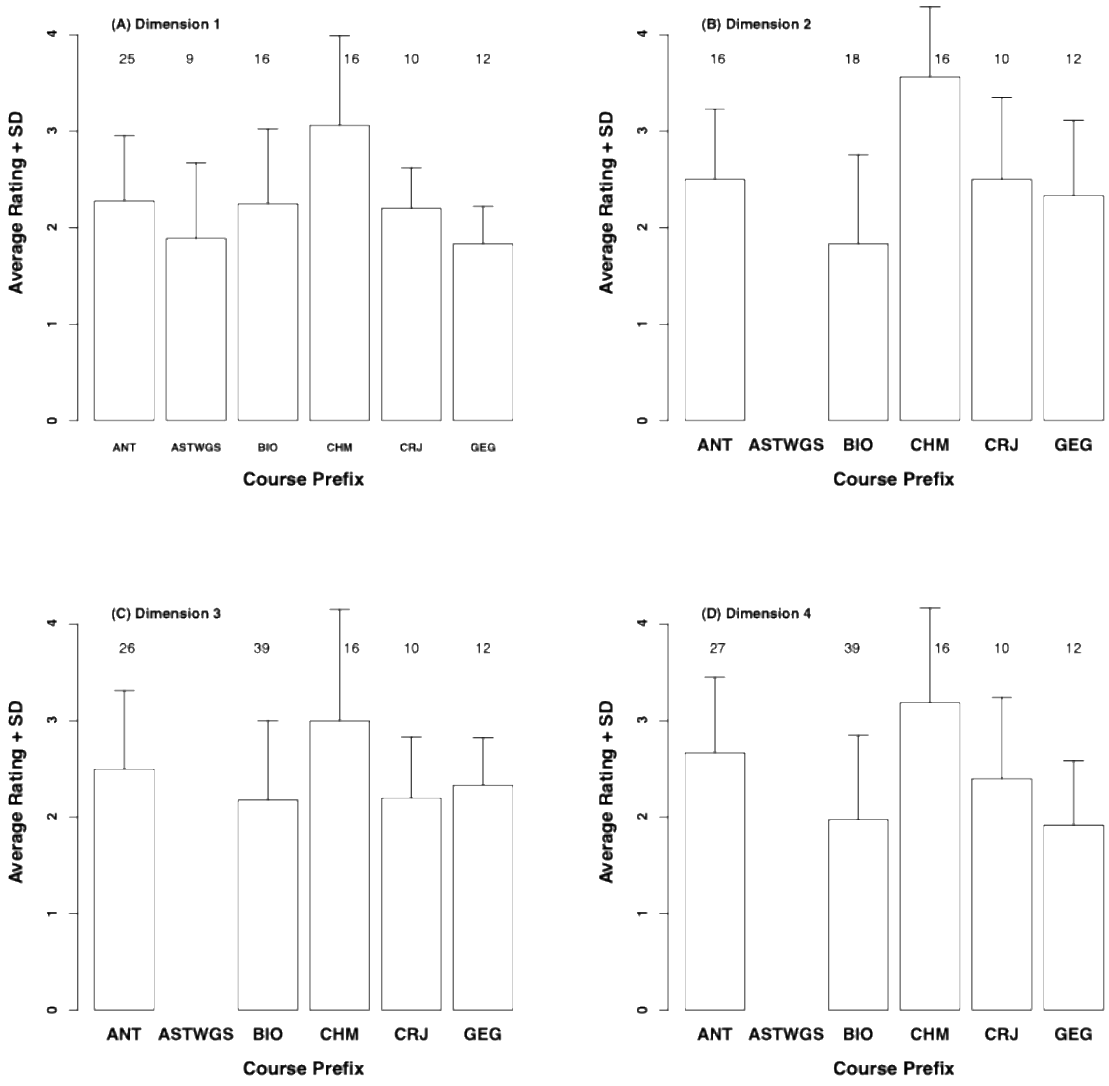

FIGURE 6. The average score in *Dimension 1* (A), *Dimension 2* (B), *Dimension 3* (C), and *Dimension 4* (D) of the SLO 2 C1 rubric for assessable artefacts submitted in Spring 2023 to KU GEAC and rated by pairs of independent volunteers arranged according to course prefix. Scores given were either 1 (poor), 2 (fair), 3 (good), or 4 (excellent)

Key Points and Recommendations

Based on this assessment of General Education Program Category C courses using the newly revised analytic rubric for SLO 2 in Spring 2023 semester, the GEAC makes the following key points and recommendations about the process:

- One key point is that modifying the rubric from holistic to analytic is a positive change in the sense that assessment of different skills through the multiple dimensions of the new rubric facilitates the identification of areas of strength, as well as areas in need of further development in student learning outcomes.
- A second key point is that not only is the revision of existing holistic rubrics positive, but also an evolving piece of work that should continuously be assessed and revised accordingly.
    - A case in point is that for Fall 2023, faculty instructing GE courses being assessed can now use multiple assignments to meet different dimensions of the rubric. Additionally, discussions during the Spring 2023 General Education Retreat, and since, are being had in relation to other things that faculty instructing GE courses can provide to improve independent assessment beyond the students' work. For example, faculty teaching all GE courses could be encouraged to provide context or background information for the assignment submitted.
- A final key point is that the Spring 2023 General Education Retreat was the first opportunity to provide independent rating of students' work using the new analytic rubrics, and the format followed that day serves as a good starting point for what these events will look like in the future, including developing recommendations for improving the process of independent rating, such as encouraging faculty to provide context or background information for the assignment submitted. Of course, the GEAC acknowledges that care should be taken to not have raters assess assignments but student responses, which is something achievable during norming sessions.
- Recommendations originating from this revised process of assessing GE courses to attempt improving student learning are summarized in **Appendix 2** and include:
    - For GEAC:
        - To collaborate with faculty and Departments on how multiple-choice assignments can be better aligned with the rubrics when this is the instrument of choice for assessment.
        - To facilitate the development of a program whereby faculty currently contributing to GE provide mentorship to individuals new to teaching a course in the university's core curriculum.
        - To develop a clear plan to communicate findings from assessment to faculty instructing GE courses, Department Chairs, Deans, and the general University community. For example, it's hoped that providing feedback on what assignments were not ratable, and why, will help faculty in those Departments to have their own discussions about the SLOs, rubrics and assignments used in GE assessment.

- For Departments and Programs:
  - To encourage faculty teaching GE courses to read SLO and rubric, and design assignments to align with both. This could be facilitated through an initiative to work with faculty to fully understand how they meet the specific dimensions of the SLO. This has also been achieved in the past for FYS courses, for example.
  - To encourage faculty teaching GE courses to share the rubric with students so that they understand the learning outcomes through the duration of their degree. For example, this could be done through the First Day Handout.

The GEAC notes that the results of the Spring 2023 assessment of Category C courses summarized above appear to be on par with results of the National Survey of Student Engagement (NSSE) 2019 report on Quantitative Reasoning (**Appendix 1**). The NSSE surveys undergraduates in the first and final years of their college experience to assess, among other things, participation in activities that matter to student learning. Students gave KU a mean score of 28.4 in the Quantitative Reasoning engagement indicator, which statistically, was not significantly different when compared to mean score across all PASSHE schools ($p=0.08$), and significantly higher than the average for Carnegie Class institutions ($p=0.05$), as well as the mean in the 2018 and 2019 NSSE ($p=0.04$). The NSSE data support the results from the KU Spring 2023 assessment summarized above in that KU students' performance in General Education courses that emphasize quantitative reasoning skills is at a level expected for students completing general education courses at comparable universities, and student perception of this aspect of their university education is equivalent to how students at other institutions perceive their learning quantitative reasoning.

Trend Analysis:

The GEAC would like to highlight the changes in student performance to achieve SLO 2a relative to its last assessment in 2019. In 2019, 65% of students sampled performed at or above the GEAC's benchmark level of 2, and this is lower than the performance level of students assessed in this report (2023), which was at least 80% of students in all four dimensions of the newly implemented analytic rubric performing at a level 2 or above. The comparisons do not completely correspond because assessment shifted in two ways: a) from holistic to analytic rubrics and b) from the use of means to the use of both means and frequencies.

Appendix 1

NSSE Surveys – Quantitative Reasoning

On the NSSE Surveys for 2019 and 2022, Quantitative Reasoning is the metric where KU has consistently had the lowest scores. The scores are achieved by assigning numerical values to Likert-scale responses as follows: 0 = Never, 20 = Sometimes, 40 = Often, and 60 = Very Often. These values are applied to all responses, and then a mean is generated and used for comparison purposes.

The 2019 NSSE showed KU having the following scores for QR:

First year students scores gave KU a mean score of 28.4. Senior students scores gave KU a mean score of 29.4.

First Year Students Mean Comparisons table:

Mean Comparisons	Kutztown University	*Your first-year students compared with*					
		PASSHE		Carnegie Class		NSSE 2018 & 2019	
Engagement Indicator	*Mean*	*Mean*	*Effect size*	*Mean*	*Effect size*	*Mean*	*Effect size*
Higher-Order Learning	37.8	37.9	-.01	37.8	.00	38.0	-.01
Reflective & Integrative Learning	36.5	35.0 *	.14	35.0 *	.13	35.2 *	.11
Learning Strategies	39.4	39.0	.03	38.5	.07	38.1	.09
Quantitative Reasoning	28.4	27.2	.08	27.6	.05	27.8	.04

Notes: Results weighted by institution-reported sex and enrollment status (and institution. size for comparison groups); Effect size: Mean difference divided by pooled standard deviation; Symbols on the Overview page are based on effect size and p before rounding; *$p < .05$, **$p < .01$, ***$p < .001$ (2-tailed).

Senior Students Mean Comparisons table:

Mean Comparisons	Kutztown University	*Your seniors compared with*					
		PASSHE		Carnegie Class		NSSE 2018 & 2019	
Engagement Indicator	*Mean*	*Mean*	*Effect size*	*Mean*	*Effect size*	*Mean*	*Effect size*
Higher-Order Learning	40.3	40.3	.00	40.5	-.02	40.0	.02
Reflective & Integrative Learning	38.8	38.6	.02	38.1	.06	38.0	.07
Learning Strategies	37.8	38.2	-.03	39.4 *	-.11	38.5	-.05
Quantitative Reasoning	29.4	28.4	.07	29.6	-.01	29.8	-.03

Notes: Results weighted by institution-reported sex and enrollment status (and institution. size for comparison groups); Effect size: Mean difference divided by pooled standard deviation; Symbols on the Overview page are based on effect size and p before rounding; *$p < .05$, **$p < .01$, ***$p < .001$ (2-tailed).

The 2022 NSSE showed KU having the following scores for QR:

First year students scores gave KU a mean score of 26.8. Senior students scores gave KU a mean score of 29.7. Below is a table showing the means of responses from Senior students compared to comparison groups:

Mean Comparisons	Kutztown University	Your seniors compared with					
		PASSHE		Carnegie Class		NSSE 2021 & 2022	
Engagement Indicator	*Mean*	*Mean*	*Effect size*	*Mean*	*Effect size*	*Mean*	*Effect size*
Higher-Order Learning	40.0	40.7	-.05	40.5	-.04	40.2	-.02
Reflective & Integrative Learning	39.2	39.1	.01	38.1	.09	38.1	.09
Learning Strategies	38.8	38.9	-.01	39.7	-.07	39.0	-.02
Quantitative Reasoning	29.7	30.5	-.05	30.7	-.06	30.9	-.07

Notes: Results weighted by institution-reported sex and enrollment status (and institution size for comparison groups); Effect size: Mean difference divided by pooled standard deviation; Symbols on the Overview page are based on effect size and p before rounding; *p < .05, **p < .01, ***p < .001 (2-tailed).

Comparing the two surveys, it is apparent that despite some small changes there has not been significant change in student responses to move answers from “Sometimes” to “Often.” When compared with PASSHE students overall, KU scores remain within the averages, but at the low end of the average. For example, in the 2022 NSSE survey, PASSHE’s mean score for first year students was 27.2 while KU’s was 26.8.

There is also a noticeable difference in the percentage of students responding “Very Often” or “Often” to questions about Quantitative Reasoning compared to the categories of Higher-Order Learning, Reflective & Integrative Learning, and Learning Strategies. While those last three categories consistently show that the percentage of students responding “Often” or “Very Often” to their questions is above 60%, the Quantitative Reasoning questions experienced rates between 35% and 48% for the 2022 NSSE Survey. As a result, the mean scores for Quantitative Reasoning are 10 points lower than the other three categories.

Appendix 2 – Summary of GEAC's Spring 2023 Assessment Recommended Tasks

Task	University Group(s) Leading Task	Target Due Date	Date Completed	Notes
Collaborate with faculty and Departments on how multiple-choice assignments can be better aligned with the rubrics when this is the instrument of choice for assessment.	GEAC	Jan 30		
Facilitate the development of a program whereby faculty currently contributing to GE provide mentorship to individuals new to teaching a course in the university's core curriculum.	GEAC	Fall 2024		
Develop a clear plan to communicate findings from assessment to faculty instructing GE courses, Department Chairs, Deans, and the general University community.	GEAC	Spring 2024		
Encourage faculty teaching GE courses to read SLO and rubric, and design assignments to align with both. This could be facilitated through an initiative to work with faculty to fully understand how they meet the specific dimensions of the SLO.	Departments and Programs	Fall 2024		
Encourage faculty teaching GE courses to share the rubric with students so that they understand the learning outcomes through the duration of their degree.	Departments and Programs	Spring 2024		