UNIVERSITY OF TEXAS ARLINGTON

INSTITUTIONAL EFFECTIVENESS AND REPORTING

ASSESSMENT OF EMPIRICAL & QUANTITATIVE SKILLS USING AAC&U VALUE RUBRICS AT THE UNIVERSITY OF TEXAS AT ARLINGTON

Fall 2024 Report

Assessment of Empirical & Quantitative Skills, Fall 2024 Report

Empirical and Quantitative Skills (EQS) are essential competencies that enable students to analyze real-world problems through data interpretation, mathematical reasoning, and empirical inquiry. The Texas Higher Education Coordinating Board (THECB) recognizes EQS as one of six core objectives in the Texas Core Curriculum (TCC), emphasizing its integration within Life and Physical Sciences, Mathematics, and Social and Behavioral Sciences.

At The University of Texas at Arlington (UT Arlington), EQS is assessed as part of a structured, multi-year evaluation cycle for all core objectives. Student work is collected through embedded Signature Assignments in approved core courses and evaluated by trained faculty and staff using the AAC&U VALUE Rubric for Quantitative Literacy to ensure standardized, rigorous scoring.

This report presents the results of EQS assessment conducted in Fall 2024, offering insights from undergraduate work samples to inform faculty-led improvements in instruction and curriculum. As empirical and quantitative reasoning increasingly underpin academic and professional success, UT Arlington remains committed to equipping students with the analytical skills needed to address complex challenges across disciplines.

Method

Participants

Written student work samples were obtained from undergraduates enrolled in the fall 2022 semester. Data were collected from two high-enrollment core curriculum courses in mathematics that are offered every fall and spring semesters: MATH 1301 – Contemporary Math, MATH 1308 – Elementary Statistical Analysis, and MATH 1309 – Statistical Literacy. These were high-enrollment courses, with different majors enrolled in these courses. These courses use real-world examples and applications of mathematical concepts, making them suitable sample courses for evaluating EQS.

The EQS assessment sample for Fall 2024 included 234 undergraduate students, with a demographic profile reflecting the diversity of UT Arlington’s student body. Most participants were female (74.36%), while male students accounted for 25.64%. In terms of racial and ethnic representation, Hispanic/Latino students comprised the largest group (42.31%), followed by White (19.66%), Black/African American (14.53%), and Asian students (12.39%). Additional representation included students identifying as Multiple Ethnicities (5.98%), Foreign (4.27%), and Not Specified (0.85%).

Regarding first-generation college status, 51.28% of students identified as first-generation, 32.91% as non-first-generation, and 15.81% had unavailable data for this category. (see Table 1). This demographic distribution provides valuable context for interpreting the assessment results and identifying potential equity gaps in the development of empirical and quantitative reasoning skills.

Table 1
Student Demographics

Categorical Information	N	%
Gender		
Female	174	74.36%
Male	60	25.64%
Racial/Ethnic Description		
Hispanic/Latino	99	42.31%
White	46	19.66%
Black/African American	34	14.53%
Asian	29	12.39%
Multiple Ethnicities	14	5.98%
Foreign	10	4.27%
Not Specified	2	0.85%
First Generation Student		
First Generation	120	51.28%
Non-First Generation	77	32.91%
First Generation Determination Data Unavailable	37	15.81%

The majority of students assessed for Empirical and Quantitative Skills (EQS) in Fall 2024 were in the early stages of their academic careers. Sophomores represented the largest group (55.13%), followed by freshmen (20.51%), juniors (15.38%), and seniors (8.55%), with only one student (0.43%) classified as a fifth-year undergraduate. In terms of enrollment year, most students began their studies at UT Arlington in 2022–2023 (69.66%), and an additional 25.21% enrolled in 2021–2022. Earlier enrollment years accounted for a small portion of the sample, with only 5.13% of students having entered the university before 2021.

The academic load data shows that nearly all students were enrolled full-time (94.87%), with a small number attending part-time (5.13%). Additionally, 84.19% of students were continuing UT Arlington students who had not transferred from another institution, while 15.81% were transfer students (see Table 2). These enrollment characteristics suggest that the sample largely represents lower-division, full-time, non-transfer students who are relatively early in their academic journey—a factor that may influence their development of empirical and quantitative reasoning skills.

Table 2
Student Status at UT Arlington

Categorical Information	N	%
Academic Level		
Sophomore	129	55.13%
Freshman	48	20.51%
Junior	36	15.38%
Senior	20	8.55%
Fifth Year	1	0.43%
Enrollment Year		
2022 - 2023	163	69.66%
2021 - 2022	59	25.21%
2019 - 2020	5	2.14%
2020 - 2021	4	1.71%
2018 - 2019	2	0.85%
2014 - 2015	1	0.43%
Academic Load		
Enrolled Full-Time	222	94.87%
Enrolled Part-Time	12	5.13%
Transfer Student		
Non-Transferred	197	84.19%
Transferred	37	15.81%

*Eligibility as of Spring 2024

The EQS assessment sample in Fall 2024 included students from a range of academic colleges and schools at UT Arlington, with the largest representation coming from the College of Liberal Arts (36.32%) and the College of Nursing and Health Innovation (34.62%). Together, these two colleges accounted for over 70% of the total sample. The Division of Student Success also contributed a significant portion, representing 17.52% of assessed students, likely reflecting participation from first-year experience or core curriculum courses. Smaller proportions of students were drawn from the College of Science (3.85%), College of Architecture, Planning & Public Affairs (3.42%), School of Social Work (2.56%), College of Business (1.28%), and the College of Engineering (0.43%). This distribution illustrates broad interdisciplinary participation, while also reflecting a heavier concentration of student work from programs where empirical and quantitative reasoning is closely tied to general education or health-related coursework (see Table 3).

Table 3
Students by Colleges/Schools

College/School	No. of Students	Percentage
College of Liberal Arts	85	36.32%
Col Nurse & Health Innovation	81	34.62%
Division of Student Success	41	17.52%
College of Science	9	3.85%
Col Arch, Plan & Pub Affairs	8	3.42%
School of Social Work	6	2.56%
College of Business	3	1.28%
College of Engineering	1	0.43%

Procedure

Faculty currently teaching undergraduate courses in the Mathematics Foundational Component Area agreed to submit course Signature Assignment(s) for this report. The syllabus for each core curriculum class at UT Arlington describes the Signature Assignment(s). Students enrolled in core courses complete the Signature Assignment(s) as they would complete other required coursework and assignments. The samples submitted for this assessment process were ungraded, and the Office of Institutional Effectiveness and Reporting de-identified copies of student work before using them for the core curriculum assessment to mitigate rater bias and to protect the confidentiality of student information.

Assessment Instrument

The Signature Assignments were assessed using the Valid Assessment of Learning in Undergraduate Education (VALUE) Rubric for Quantitative Literacy (AAC&U 2009) developed by the Association of American Colleges and Universities (AAC&U). The rubric categorizes EQS into six dimensions: Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication. The rubric describes each dimension and uses a four-point scoring scale (see Appendix A). The rubric functions as a matrix that provides narrative descriptions of expected work quality and corresponding point values for scoring the six measures. The point values range from 1 to 4, with 1 indicating baseline performance (Benchmark-1), 2 indicating approaching milestone (Milestone-2), 3 indicating achieved milestone (Milestone-3), and 4 indicating the highest mastery (Capstone-4) of Empirical and Quantitative skills. AAC&U, the authors of the rubric, permit zero ratings if the paper does not meet the minimum content or quality standards defined in the rubric. The attainment target (numerical ratings) was set at a score of 2 (Milestone-2). The attainment target was set above the benchmark following recommendations from AAC&U research (Greenhoot & Bernstein, 2012) and standard acceptance criteria in the assessment community.

All raters assigned a score to each of the six dimensions in the rubric for each student work sample. Higher values indicate more evidence of EQS in student work and vice versa. Beginning this year, raters were advised to use zero per AAC&U recommendations if any dimension is absent in student work. Typically, in student samples, the six dimensions are adequately represented in the narrative. It is important to note that the EQS samples with visual components in the form of charts, graphs, and figures enhanced the identification of the Representation and Communication dimensions. Since written and visual communications are required across all eight FCAs, a visual component was expected in students' work in foundational mathematics courses.

Raters, Rater Calibration, and Scoring

Raters scored the student writing samples during a scheduled scoring day, and each paper was reviewed twice (two separate raters) in a group setting. A third "tiebreaker" rating was obtained when ratings diverged by more than one rating interval on a single dimension. In these cases, the two sets of most closely aligned ratings were used in calculating the mean scores. The rater group included thirteen faculty members and professional staff with advanced degrees.

The scoring day began with an orientation and description of the rating process. A qualified UTA facilitator led the raters through reviewing the rubric and discussing the rating dimensions and scale designed to calibrate the rater's understanding and use of the rubric in the rating process. Then, the entire group read and rated one practice anchor paper, which was chosen beforehand by the facilitator. Following the sample paper review, the facilitator led a discussion among all raters using the anchor paper to reach a common understanding of the EQS dimensions and to find exemplar indicators within the paper for the rubric levels of mastery. Following completion of the calibration activity, formal review and rating of the de-identified student papers began. During the formal review and rating of papers, raters read each paper and assigned scores for each dimension on the rubric using the four-point scale (plus the available "zero" rating). If the values of the skill measure scores for a paper from the two raters were identical or within one point difference, then the two scores were considered in agreement and averaged. For example, if Rater A scored the Calculation measure with a value of 2 and Rater B scored the same measure with a value of 3, then the rating was considered in agreement, and scores for that dimension were averaged, resulting in a score value of 2.5. If the scores from the two raters differed by over two points, a third rater was assigned the paper. In such cases, three scores were averaged together to determine the final score. For example, if Rater A scored the Calculation measure with a value of 1 and Rater B scored the same measure with a value of 3, the rating was not in agreement, and a third rater was asked to read and score the paper.

Analysis and Results

Inter-rater Agreement

To ensure consistency and fairness in scoring, inter-rater agreement was calculated across the six dimensions of the Empirical and Quantitative Skills (EQS) VALUE Rubric. An agreement was recorded when two raters scored within one point of each other, and the percentage was determined by dividing the number of agreements by the total number of ratings.

The results reflect strong agreement across most dimensions, with the highest levels observed in Communication (91%), Calculation (90%), and Application/Analysis (90%). Representation (89%) and Interpretation (88%) also demonstrated high levels of scoring consistency. However, the Assumptions dimension showed a notably lower agreement percentage (72%), suggesting that this area may involve more subjective interpretation or variability in how raters apply the rubric criteria. This pattern highlights the importance of further calibration or rubric clarification for dimensions where agreement is less consistent. Overall, the high agreement percentages across most categories indicate that raters were generally aligned in their evaluations, supporting the reliability of the assessment process (see table 4).

Table 4. Scoring Agreement Percentage Among Raters for Empirical and Quantitative Skills Dimensions

Dimension (EQS VALUE Rubric)	Percentages
Interpretation	88%
Representation	89%
Calculation	90%
Application/Analysis	90%
Assumptions	72%
Communication	91%

Note: If values assigned by the raters differed by the rating interval of one point or less, it was counted as agreement. The agreement percentage was computed by dividing the number of agreements by the total number of ratings.

Apart from the simple percentage agreements, researchers widely measure the reliability of rating agreements between different raters to eliminate chance agreements. All raters who participated in the scoring process had advanced degrees and work experience and attended the same training just before the scoring session. Hence, the probability of chance agreement was very low, but inter-rater agreement was computed to follow best research practices. Inter-rater reliability is the consistency among raters when scoring the same subjects independently. The extent to which different raters agree on their judgments establishes the validity and credibility of measurements or ratings.

The inter-rater agreement was determined to check the consistency level of the rating by calculating the Intraclass Correlation Coefficient (ICC). High ICC values indicate more reliability between rater scores. Commonly accepted guidelines were used to interpret the ICC results. These suggest that the range of 0.40 to 0.74 is considered fair to good inter-rater agreement, with results above 0.74 classified as excellent inter-rater agreement and results lower than 0.40 considered poor inter-rater agreement (Fleiss, 1986; Shrout & Fleiss, 1979). The ICC results indicated fair to good reliability across all dimensions. Communication (0.69) and

Calculation (0.67) exhibited the highest reliability, closely followed by Application/Analysis (0.65) and Representation (0.64). Interpretation also showed good consistency with an ICC of 0.62. The Assumptions dimension had the lowest ICC at 0.43, suggesting it was more difficult for raters to apply consistently (see table 5).

Table 5. Intraclass Correlation Coefficient for Empirical and Quantitative Skills Dimensions

Dimension (EQS VALUE Rubric)	Coefficient
Interpretation	0.62
Representation	0.64
Calculation	0.67
Application/Analysis	0.65
Assumptions	0.43
Communication	0.69

Note 1: *less than 0.40 = poor agreement; between .40 and .74 = fair to good agreement; greater than .74 = excellent agreement.*

Note 2: *The intra-class correlation coefficient (ICC) was calculated as a two-way random effects model. Values in this model type with random rater pairings are typically expected to be lower than those where rater pairings are fixed throughout the rating day.*

This lower reliability, along with the lower agreement percentage observed in Table 4 for the Assumption dimension, points to the need for further calibration or refinement of the rubric criteria for assessing assumptions. Overall, these results support the general reliability of the scoring process, while also identifying specific areas—particularly the Assumptions dimension—where greater rater alignment may be needed.

Students Performance

The analysis of student performance across the six dimensions of the EQS VALUE Rubric reveals varying levels of proficiency. Table 6 summarizes student performance across six dimensions of Empirical and Quantitative Skills (EQS), as measured by the VALUE rubric. The established attainment target was a mean score of 2.0 or higher, indicating that students have reached milestone-level proficiency.

Students met or exceeded the attainment target in five out of six dimensions. The highest performance was observed in Calculation (M = 2.68, SD = 0.53), Representation (M = 2.58, SD

= 0.51), and Interpretation (M = 2.55, SD = 0.53). These strong results indicate students were generally proficient in performing calculations accurately, representing data effectively, and interpreting quantitative information. The dimensions of Communication (M = 2.46, SD = 0.59) and Application/Analysis (M = 2.36, SD = 0.67) also exceeded the milestone attainment target, reflecting moderate to strong proficiency in effectively communicating quantitative information and applying analytical skills to real-world contexts.

Table 6. Means for Empirical and Quantitative Skills Measure Scores

Measurement Dimensions	N	Mean	SD
Interpretation	234	2.55	0.53
Representation	234	2.58	0.51
Calculation	234	2.68	0.53
Application/Analysis	234	2.36	0.67
Assumptions	234	1.56	0.64
Communication	234	2.46	0.59

Students notably scored below the attainment target in the Assumptions dimension (M = 1.56, SD = 0.64). This lower performance indicates significant challenges among students in identifying, articulating, and evaluating assumptions underpinning quantitative analyses. Overall, while students demonstrated strong proficiency across most EQS dimensions, targeted instructional support will be necessary to improve performance related to recognizing and evaluating assumptions.

Observations and Limitations

The Fall 2024 assessment of Empirical and Quantitative Skills (EQS) focused on student work from three high-enrollment core mathematics courses: MATH 1301 (Contemporary Math), MATH 1308 (Elementary Statistical Analysis), and MATH 1309 (Statistical Literacy). These courses provided real-world contexts and practical applications, making them ideal settings to assess EQS proficiency. The student sample (n = 234) reflected UT Arlington's diverse undergraduate population, primarily comprising first-generation students (51.28%) and a broad range of racial and ethnic backgrounds.

Several important observations emerged. Students successfully achieved the milestone attainment target (mean score ≥ 2.0) in five out of six dimensions—particularly excelling in

Calculation, Representation, and Interpretation. Strong scores in these areas indicate effective instructional practices and robust student comprehension of foundational quantitative concepts. Moderate proficiency in Communication and Application/Analysis suggests students effectively articulated quantitative findings and applied their skills practically, though additional refinement could enhance these dimensions further.

However, notably lower performance in the Assumptions dimension highlights a critical instructional gap. Students had difficulty identifying and evaluating underlying assumptions, suggesting the need for explicit instruction and practice opportunities in this challenging skill area. Several limitations should be considered. Student work samples were drawn exclusively from mathematics courses, limiting generalizability across disciplines. Future assessments should broaden the sample to include additional courses, particularly from Life and Physical Sciences and Social and Behavioral Sciences. Further, variation in assignment design and clarity of instructions may have influenced student performance in specific dimensions, particularly Assumptions. Ensuring that Signature Assignments explicitly address all rubric dimensions will be essential for improving alignment and scoring reliability.

Overall, this assessment indicates substantial strengths in foundational EQS areas and clearly identifies assumptions analysis as a key area for targeted improvement. Continued efforts to refine assignments and instructional methods are recommended to ensure comprehensive attainment of EQS objectives in future assessment cycles.

References

Association of American Colleges and Universities. (2009). Quantitative Literacy VALUE rubric. https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-quantitative-literacy

Fleiss J. L. (1986). The design and analysis of clinical experiments. New York: John Wiley & Sons.

Georgesen, J. (2015). Evolving from big data to smart data: New ways CX researchers predict customer behavior. Retrieved on 12/08/15 from http://mrweek.com/content

Landers, R.N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower, 2*:e143518.81744. DOI: 10.15200/winn.143518.81744.

Rhodes, T. (Ed.). (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Shrout, P., & Fleiss, J. L. (1979). Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

Texas Higher Education Coordinating Board (2019). Texas Core Curriculum. Retrieved on 08/22/2019 from http://www.thecb.state.tx.us/apps/tcc/

Appendix A: Quantitative Literacy VALUE Rubric

QUANTITATIVE LITERACY VALUE RUBRIC

for more information, please contact value@aacu.org

Association of American Colleges and Universities

Definition

Quantitative Literacy (QL) – also known as Numeracy or Quantitative Reasoning (QR) – is a "habit of mind," competency, and comfort in working with numerical data. Individuals with strong QL skills possess the ability to reason and solve quantitative problems from a wide array of authentic contexts and everyday life situations. They understand and can create sophisticated arguments supported by quantitative evidence and they can clearly communicate those arguments in a variety of formats (using words, tables, graphs, mathematical equations, etc., as appropriate).

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

	Capstone 4	**Milestones** 3	2	**Benchmark** 1
Interpretation *Ability to explain information presented in mathematical forms (e.g., equations, graphs, diagrams, tables, words)*	Provides accurate explanations of information presented in mathematical forms. Makes appropriate inferences based on that information. *For example, accurately explains the trend data shown in a graph and makes reasonable predictions regarding what the data suggest about future events.*	Provides accurate explanations of information presented in mathematical forms. *For instance, accurately explains the trend data shown in a graph.*	Provides somewhat accurate explanations of information presented in mathematical forms, but occasionally makes minor errors related to computations or units. *For instance, accurately explains trend data shown in a graph, but may miscalculate the slope of the trend line.*	Attempts to explain information presented in mathematical forms, but draws incorrect conclusions about what the information means. *For example, attempts to explain the trend data shown in a graph, but will frequently misinterpret the nature of that trend, perhaps by confusing positive and negative trends.*
Representation *Ability to convert relevant information into various mathematical forms (e.g., equations, graphs, diagrams, tables, words)*	Skillfully converts relevant information into an insightful mathematical portrayal in a way that contributes to a further or deeper understanding.	Competently converts relevant information into an appropriate and desired mathematical portrayal.	Completes conversion of information but resulting mathematical portrayal is only partially appropriate or accurate.	Completes conversion of information but resulting mathematical portrayal is inappropriate or inaccurate.
Calculation	Calculations attempted are essentially all successful and sufficiently comprehensive to solve the problem. Calculations are also presented elegantly (clearly, concisely, etc.)	Calculations attempted are essentially all successful and sufficiently comprehensive to solve the problem.	Calculations attempted are either unsuccessful or represent only a portion of the calculations required to comprehensively solve the problem.	Calculations are attempted but are both unsuccessful and are not comprehensive.
Application / Analysis *Ability to make judgments and draw appropriate conclusions based on the quantitative analysis of data, while recognizing the limits of this analysis*	Uses the quantitative analysis of data as the basis for deep and thoughtful judgments, drawing insightful, carefully qualified conclusions from this work.	Uses the quantitative analysis of data as the basis for competent judgments, drawing reasonable and appropriately qualified conclusions from this work.	Uses the quantitative analysis of data as the basis for workmanlike (without inspiration or nuance, ordinary) judgments, drawing plausible conclusions from this work.	Uses the quantitative analysis of data as the basis for tentative, basic judgments, although is hesitant or uncertain about drawing conclusions from this work.
Assumptions *Ability to make and evaluate important assumptions in estimation, modeling, and data analysis*	Explicitly describes assumptions and provides compelling rationale for why each assumption is appropriate. Shows awareness that confidence in final conclusions is limited by the accuracy of the assumptions.	Explicitly describes assumptions and provides compelling rationale for why assumptions are appropriate.	Explicitly describes assumptions.	Attempts to describe assumptions.
Communication *Expressing quantitative evidence in support of the argument or purpose of the work (in terms of what evidence is used and how it is formatted, presented, and contextualized)*	Uses quantitative information in connection with the argument or purpose of the work, presents it in an effective format, and explicates it with consistently high quality.	Uses quantitative information in connection with the argument or purpose of the work, though data may be presented in a less than completely effective format or some parts of the explication may be uneven.	Uses quantitative information, but does not effectively connect it to the argument or purpose of the work.	Presents an argument for which quantitative evidence is pertinent, but does not provide adequate explicit numerical support. (May use quasi-quantitative words such as "many," "few," "increasing," "small," and the like in place of actual quantities.)