UNIVERSITY OF TEXAS ARLINGTON

INSTITUTIONAL EFFECTIVENESS AND REPORTING

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

SPRING 2019 REPORT

Empirical & Quantitative Core Objective Assessment at UT Arlington

Empirical and quantitative skills allow an individual to understand information or raw data that is presented in tables, charts, graphs, or figures and evaluate it to draw accurate conclusions. Identifying applications of empirical and quantitative skills across academic disciplines is not hard to do. The ability to take information, analyze it, and predict outcomes is a common theme in the hard sciences such as engineering, physics, chemistry, and biology. In addition, quantitative literacy is utilized across disciplines, for instance, in nursing, business, and psychology.

An individual's comfort level and ability to evaluate data is a valuable skill, not only in academic pursuits; it is helpful in all areas of life. Data analysis without understanding the story that the data portrays is of minimal value and limits an individual, a business, or an organization from taking appropriate action. As such, educational objectives often emphasize elements of data analysis, as well as how to use the data to draw conclusions. In other words, individuals with empirical and quantitative skills see connections and systemic problems, but they don't stop there. They also use these skills to make data-driven decisions to find solutions. Action words typically connected with empirical and quantitative skills include identify, extract, validate and report.

Georgesen (2015) expanded the list of these verbs and ordered them as steps involved in empirical and quantitative processing (see Figure 1).

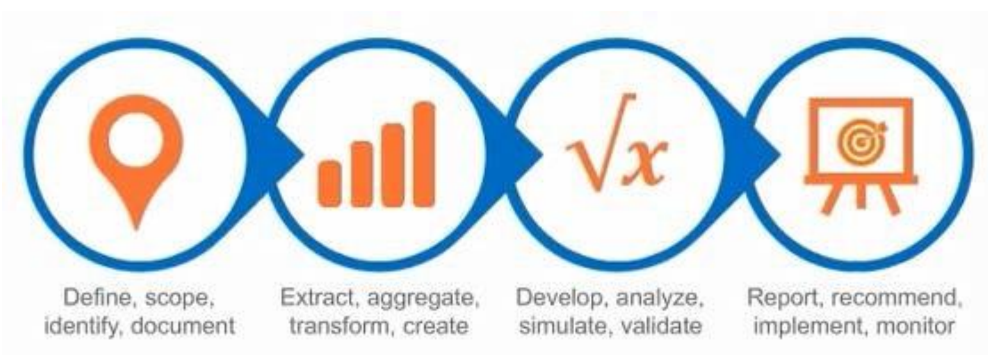

Figure 1. Action words that describe Empirical and Quantitative Skills.

Empirical and Quantitative Skill (EQS) is one of six core objectives selected by the Texas Higher Education Coordinating Board (THECB) when the current Core Curriculum was established in 2011 (THECB, 2019). The assessment of the EQS Objective is required in three of the eight Foundational Component Areas listed by THECB, thus, EQS is implemented within

Core Curriculum coursework at the undergraduate level in Life and Physical Sciences, Mathematics, and Social and Behavioral Sciences. The University of Texas at Arlington (UT Arlington) assesses the six THECB core objectives on a multi-year cycle to examine the extent of student achievement.

At UT Arlington, the EQS Objective was assessed using written samples of undergraduate student work from approved Signature Assignments embedded in the existing core courses. The quality of EQS in student work was rated by UT Arlington faculty and staff using a rubric developed by the Association of American Colleges and Universities (AAC&U; Rhodes, 2010). The purpose of this report is to present EQS ratings and information gleaned from student work samples collected during the 2019 spring semester among UT Arlington undergraduates.

Method

Participants

Written student work samples were obtained from undergraduates enrolled in Core Curriculum courses in either Life and Physical Sciences or Mathematics at UT Arlington. The demographic information that follows describes 335 students for which it was available (see Table 1). Over three-fourths of the participants were female (75.5%; *n* = 253); the remainder were male. In terms of race and ethnicity, the sample also reflected the rich diversity of students at UT Arlington. A little over one-fourth of the student participants identified as White (26.6%; *n* = 89), while slightly over one-third identified as Hispanic (35.2%; *n* = 118). Asians comprised 17.6% of this sample. Black/African American, American Indian/Alaskan Native, Foreign, Non Resident, and Mixed Race individuals made up the remainder. The majority of these students were either juniors or seniors (80.3%), and slightly under half (43.9% identified themselves as first-generation students (See Table 1).

Table 1
Student Demographics

Categorical Information	N	%
Gender		
Female	253	75.5%
Male	82	24.5%
Racial/Ethnic Description		
American Indian or Alaskan Native	1	0.3%
Asian	59	17.6%

Black, African American	33	9.9%
Foreign, Non-Resident Alien	15	4.5%
Hispanic, All Races	118	35.2%
Two or More Races/Ethnicities	17	5.1%
Unknown, Not Specified	3	0.9%
White, Caucasian	89	26.6%
Level		
Freshman	4	1.2%
Sophomore	56	16.7%
Junior	180	53.7%
Senior	89	26.6%
Fifth Year	5	1.5%
First Generation Student		
Yes	147	43.9%
No	188	56.1%
Pell Grant Eligible*		
Yes	170	50.8%
No	165	49.2%
Transfer Student		
Yes	98	29.3%
No	237	70.7%

*Eligibility as of Spring 2019

Procedure

Faculty currently teaching undergraduate courses in the Life and Physical Sciences, and Mathematics Foundational Component Areas agreed to submit the course Signature Assignment for this report. The syllabus for each Core Curriculum class at UT Arlington describes the Signature Assignment and the students enrolled in these courses complete it as they would other required course work. The samples submitted for this assessment process were ungraded, de- identified copies. Steps to redact personal and academic information are followed for two reasons:

1) to prevent any bias among rater scores in response to the grade the paper received from the professor and 2) 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, 2019) developed by the Association of American Colleges and Universities' (AAC&U). This rubric categorizes

EQS into six dimensions: *Interpretation, Representation, Calculation, Application/Analysis, Assumptions*, and *Communication*. The rubric describes each dimension and uses a four-point scale for determining scores (see Figure 2). Higher values indicate more evidence of EQS. Using the rubric, raters assigned a score to each of the six dimensions.

Typically, in student samples, the six dimensions are adequately represented in the narrative. It is important to note that in the EQS samples that visual communication in the form of charts, graphs, and figures enhanced the identification of the *Representation* and *Communication* dimensions. This not unexpected because communication (written and visual) is required for fleshing out and articulating ideas across all eight foundational component areas. However, it is the case that visual communication in the two Foundational Component Areas for this report, Life and Physical Sciences and Mathematics, is often essential for depicting information.

QUANTITATIVE LITERACY VALUE RUBRIC

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).

	4	3	2	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 orrepresent 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 & 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.)

Figure 2. Quantitative Literacy VALUE Rubric

Raters, Rater Calibration, and Scoring

Raters scored the student writing samples during a scheduled scoring day, so each paper was reviewed twice (two separate raters) in a group setting. The rater group included ten faculty members and professional staff with advanced degrees. Scoring day began with an orientation and description of the rating process. A qualified UTA staff facilitator led the raters in a review of the rubric and discussion of rating dimensions and scale. Then, the entire group read one anchor paper which was chosen beforehand by the facilitator. Following the sample paper review activity, the facilitator led a discussion using the anchor paper which was focused on reaching a common understanding of the EQS dimensions and finding exemplar indicators within the paper for the rubric levels of mastery.

When scoring began, raters read the papers and scored each dimension with the rubric on the four-point scale. Scores were gathered and analyzed to determine agreement. Each score was calculated as the average of the two rater scores as long as the values assigned by the raters differed by two points or less. In the case of differences that exceeded two points, a third rater read and scored the paper, then the average of the two most similar scores was used as the dimension score. In this report, a third rater was only needed once.

An estimate of inter-rater reliability was obtained to examine the agreement between raters, that is, to see how frequently the rater pairs agreed on the score when they were rating the same paper. This estimate is important because it allows the researcher to conclude that the dimension is measured consistently across multiple papers and ratings. The inter-rater agreement level was determined by calculating the intra-class correlation coefficient (ICC). High ICC values indicate more agreement between raters. Commonly accepted guidelines for the interpretation of ICC results suggest that values above 0.74 indicate excellent agreement, values below 0.40 indicate poor agreement, and values in-between are considered fair to good (Fleiss, 1986; Shrout & Fleiss, 1979). The reliability analyses were setup as a one-way random model that assessed consistency within the mean dimension values. Because the exact same pairs of raters did not rate each student sample, smaller ICCs were expected (Landers, 2015), however the ICC values for *Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication* indicated good inter-rater agreement. Table 2 contains the ICC values for each of the six dimensions.

Table 2
Intraclass Correlation Coefficient for Empirical and Quantitative Skills dimensions

Empirical and Quantitative Skills VALUE Rubric Dimension	
Interpretation	0.60
Representation	0.53
Calculation	0.67
Application/Analysis	0.60
Assumptions	0.56
Communication	0.55

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 one-way random effects model. Values in this type of model with random rater pairings are typically expected to be lower than models where rater pairings are fixed throughout rating day.*

Analysis and Results

The final data set contains rating scores on the six dimensions, and student samples for which demographic information was available (n = 335), however all the papers were rated on scoring day. Across the six dimensions, students scored highest in the *Calculation* (mean=3.02) and *Interpretation* (mean=3.01) categories; the *Assumptions* (mean=2.65) category had the lowest scores. The means for each dimension are presented in Table 3. As a rating above two indicates that dimension milestones were met, the fact that all averages were above 2.5 reflects well upon the students.

Table 3: Means for Empirical and Quantitative Skills Measure Scores

Measurement Dimensions	N	Mean	SD	Percent > μ-1σ
Interpretation	335	3.01	0.56	90.1%
Representation	335	2.94	0.64	85.3%
Calculation	335	3.02	0.65	89.3%
Application/Analysis	335	2.82	0.63	82.4%
Assumptions	335	2.65	0.63	74.3%
Communication	335	2.83	0.63	82.3%

Discussion

The report compiles information gleaned from student work to assess Empirical and Quantitative Skill mastery. Work was sampled from the Life and Physical Sciences and Mathematics Foundational Component Areas. Rubrics developed by the AAC&U to assess

Quantitative Literacy were used to rate the samples.

A pattern of strengths and weaknesses for this sample of undergraduates emerged from assessing the student work samples. According to the rating scores, student work exhibited strength in five areas: *Interpretation, Representation, Calculation, Application/Analysis, and Communication*. However, the student work was rated lower in the *Assumptions* dimension. Table 4 contains a complete breakdown of rubric values as assigned by the raters. This pattern may indicate an area in which the curriculum should directly address by adding activities to help students practice these skills. However, it may merely suggest an area in which Signature Assignments instructions from the faculty instructor for the course were not specific about their expectations for elements to include in the paper.

Table 4
Frequencies for Empirical and Quantitative Skills Dimension Rating Scores

		Rubric Values (Percent of Student Papers)							
	Total	1		2		3		4	
Measurement dimensions	N	N	%	N	%	N	%	N	%
Interpretation	670	11	1.6%	113	16.9%	405	60.5%	141	21.0%
Representation	670	26	3.9%	145	21.6%	341	50.9%	158	23.6%
Calculation	670	23	3.4%	112	16.7%	361	53.9%	174	26.0%
Application/Analysis	670	25	3.7%	179	26.7%	355	53.0%	111	16.6%
Assumptions	670	39	5.8%	230	34.3%	322	48.1%	79	11.8%
Communication	670	26	3.9%	176	26.3%	348	51.9%	120	17.9%

Note: Each paper was rated twice, therefore the number of ratings contained in this table is double the number of papers

Limitations

As the multi-year cycle unfolds, establishing a mastery threshold for each dimension will be important to guide understanding of whether to regard a dimension as a strength area. In addition, future samples should include representation from Social and Behavioral Sciences. The lack of information from that Foundational Component Area for EQS will enhance the scope of this study.

The leadership of a facilitator with quantitative experience and a background in calibrating raters was essential. Her expertise seemed to help raters during the calibration activities on rating day. Specifically, assistance was offered to identify discreet differences for the levels of mastery across dimensions. This aspect is particularly important for the rating of EQS Signature Assignments because the curriculum content was focused on topics that were typically outside the expertise of the general population. For example, the Biology lab reports contained equations for evaluating soil samples.

In addition, alignment between the Signature Assignments and the VALUE rubrics used for rating them is essential. Providing expert explanations of the assignment and identifying specific areas to look for the VALUE rubric dimensions improved the interrater reliability. However, in some cases, alignment between with the Signature Assignments was not straightforward. While the composition of the Signature Assignment is up to the faculty instructor, some tailoring suggestions may need to be considered. For example, suggestions could be offered to better align the Signature assignment with VALUE rubric.

Overall, this assessment of the EQS Core Objective built on previous studies that reported on the use of Signature Assignments as measures of student mastery at UT Arlington. The multi-year plan of assessing the six THECB Core Objectives continues through 2022. Evidence collected thus far suggests adequate mastery in five of six EQS dimensions at UT Arlington.

References

Association of American Colleges and Universities (2019). Quantitative Literacy VALUE rubric. Retrieved on 08/22/2019 from https://www.aacu.org/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/