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 2015 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).

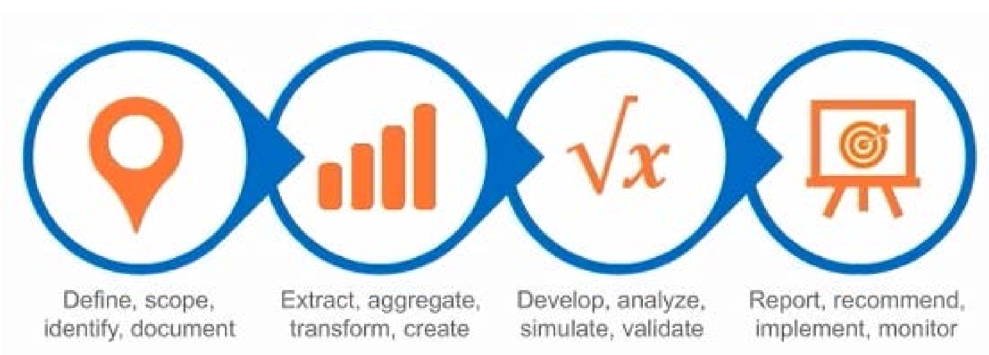

Figure 1. Action words that describe Empirical and Quantitative Skills.

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, 2015). 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 2015 spring semester among UT Arlington undergraduates.

Method

Participants

Written student work samples were obtained from 296 undergraduates enrolled in Core Curriculum courses in either Life and Physical Sciences or Mathematics at UT Arlington. The descriptive demographic data that follows represents 245 students for which information was available (see Table 1). Over half of the participants were female (61%; n = 149); the remainder were male. Thus, the sample closely matched the approximate demographic ratio found at the university. In terms of race and ethnicity, the sample also reflected the rich diversity of students at UT Arlington. About a third of the student participants identified as White (33%; n = 80), almost a third identified as Hispanic (27%; n = 67), and the balance were split between African American, Asian, Foreign, non-resident aliens, Multiple ethnicities, and Unknown, not specified. Students represented nine of ten colleges and schools at UT Arlington and a majority (73%) were not freshmen (See Table 1).

Table 1

Characteristics of students who submitted Signature Assignments

Categorical Variables	*N*	*%*
Gender		
Female	149	60.8
Male	96	39.2
Ethnicity		
African American	33	13.5
Asian	49	20.0
Caucasian	80	32.7
Foreign, non-resident alien	6	2.4
Hispanic	67	27.3
Multiple	5	2.0
Unknown, not specified	5	2.0
College/School		
College of Architecture	2	2
College of Business	24	9.8
College of Education	13	5.3
College of Engineering	15	6.1
College of Liberal Arts	26	10.6
College of Nursing & Health Innovation	61	24.9
College of Science	61	24.9
School of Social Work	15	6.1
University College	13	5.3
Undeclared	13	5.3
*Missing	2	0.8
Level		
Freshman	67	27.3
Sophomore	85	34.7
Junior	49	20.0
Senior and above	42	16.2
*Missing	2	0.8

*Note: N = 245 for each of the categorical variables, * information was missing. For a portion of the sample (n = 51) demographic information was not obtained.*

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, 2015) 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 Likert 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 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 of the eight foundational component areas. Visual communication is particularly important, and in many cases essential, for depicting information in the two Foundational Component Areas contained in this report, Life and Physical Sciences and Mathematics.

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. Then, the entire group read one anchor paper which was chosen beforehand by a facilitator. Next, 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 (four-point Likert scale). 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.

Analysis and Results

The final data set contains rating scores on the six dimensions, and student samples for which demographic information was available (n = 245), however all the papers (n = 296) were rated on scoring day. Across the six dimensions, students in this sample scored higher in *Interpretation, Representation, Calculation, Application/Analysis, and Communication* and lower in *Assumptions*. The means for each dimension are presented in Table 2. Separate analyses for the remaining papers (for which demographic information was not available; n = 51) revealed the same pattern of mean scores. UT Arlington students on average demonstrated some mastery in five dimensions as demonstrated by the average scores that exceed a value of two. A rating above two indicates that dimension milestones were met above the minimum benchmark level. A rating below a value of two reveals the opposite, that milestones were not met.

Table 2

Mean scores for Empirical & Quantitative Skill Dimensions

Empirical & Quantitative Skill Dimensions	Mean	SD
Interpretation	2.33	0.60
Representation	2.5	0.56
Calculation	2.59	0.49
Application / Analysis	2.22	0.63
Assumptions	1.8	0.63
Communication	2.37	0.59

Inter-rater Agreement

An estimate of inter-rater reliability was needed 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 3 contains the ICC values for each of the six dimensions.

Table 3

ICC Values by Dimension

Empirical and Quantitative Skill Dimension	**ICC Value**
Interpretation	.52
Representation	.51
Calculation	.47
Application/Analysis	.56
Assumptions	.51
Communication	.60

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. 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.

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. This sample is the first to examine EQS dimensions. Typically, the establishment of mastery thresholds should be examined within larger groups. In addition, future samples should include representation from Social and Behavioral Sciences. Information from that Foundational Component Area for EQS will enhance the scope of this study.

In terms of the distribution of the sample, the ratio of males to female students matched that of the overall undergraduate population at UT Arlington (40% male, 60% female). In addition, the ethnic representation was consistent with campus diversity. While the courses sampled were in the Life and Physical Science and Mathematics areas, a multi-disciplinary representation was achieved at the College/School level.

The leadership of a facilitator chosen from among the faculty within a quantitative field was essential. Their expertise seemed to help raters during the calibration activities on rating day. Specifically, they assisted by identifying discreet differences for the levels of mastery in each dimension. 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 determining concentrations of chemical elements and results from experimental manipulations involving light.

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 to better align the Signature assignment with VALUE rubric or visa-versa.

Overall, this assessment of the EQS Communication Core Objective built on the previous studies that reported on the use of Signature Assignments as measures of student mastery at UT

Arlington during 2014. This report expanded that work by including student work samples from across nine of the ten colleges and schools in the Life & Physical Sciences and Mathematics Foundational Component Areas. The multi-year plan of assessing the six THECB Core Objectives continues through 2017. Evidence collected thus far suggests adequate mastery in five of six EQS dimensions at UT Arlington.

References

Association of American Colleges and Universities (2015). Quantitative Literacy VALUE rubric. Retrieved on 02/08/2016 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 (2015). Texas Core Curriculum. Retrieved on 12/7/2015 from http://www.thecb.state.tx.us/index.cfm?objectid=417252EA-B240-62F7-9F6A1A125C83BE08.

Appendix A

Student EQS Dimension Scores by Gender

EQS Dimension	Total N = 245	Male n = 96 (39%)	Female n = 149 (61%)	Comparison
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*F*
Interpretation	2.33 (0.60)	2.36 (.055)	2.32 (.063)	0.27
Representation	2.50 (0.56)	2.45 (0.57)	2.53 (0.55)	1.21
Calculation	2.59 (0.49)	2.56 (0.48)	2.61 (0.50)	0.64
Application	2.22 (0.63)	2.25 (0.57)	2.20 (0.67)	0.39
Assumption	1.80 (0.63)	1.76 (0.61)	1.82 (0.65)	0.56
Communication	2.37 (0.59)	2.35 (0.57)	2.39 (0.61)	0.17

Note: *$p < .05$, **$p < .01$, ***$p < .001$.

Appendix B

Student EQS Scores by Classification Level

EQS Dimension	Freshmen n = 67 (27%)	Sophomores and above n = 178 (73%)	Comparison
	Mean (SD)	*Mean (SD)*	*F*
Interpretation	2.37 (0.65)	2.32 (.058)	0.25
Representation	2.42 (0.58)	2.53 (0.55)	2.10
Calculation	2.49 (0.49)	2.63 (0.49)	**3.94***
Application	2.20 (0.66)	2.23 (0.62)	0.07
Assumption	1.85 (0.66)	1.78 (0.62)	0.64
Communication	2.37 (0.58)	2.37 (0.60)	0.00

Note: *****$p < .05$, ******$p < .01$, *******$p < .001$.

Appendix C

Mean EQS Dimension scores by College

College	*n*	*M*	*SD*
College of Architecture, Planning, Public Affairs			
Interpretation	2	2.00	0.00
Representation	2	2.50	0.00
Calculation	2	2.25	0.35
Application	2	2.50	0.71
Assumption	2	1.50	0.71
Communication	2	2.00	0.00
College of Business			
Interpretation	24	2.38	0.49
Representation	24	2.56	0.66
Calculation	24	2.70	0.53
Application	24	2.31	0.59
Assumption	24	1.79	0.57
Communication	24	2.31	0.57
College of Education			
Interpretation	13	2.08	0.49
Representation	13	2.42	0.45
Calculation	13	2.46	0.32
Application	13	1.77	0.56
Assumption	13	1.31	0.25
Communication	13	1.77	0.33
College of Engineering			
Interpretation	15	2.17	0.49
Representation	15	2.30	0.53
Calculation	15	2.50	0.46
Application	15	2.10	0.47
Assumption	15	1.77	0.56
Communication	15	2.37	0.35
College of Liberal Arts			
Interpretation	26	2.29	0.64
Representation	26	2.69	0.55
Calculation	26	2.63	0.50
Application	26	2.25	0.51
Assumption	26	1.73	0.55
Communication	26	2.44	0.61
College of Nursing and Health Innovation			
Interpretation	61	2.31	0.57
Representation	61	2.43	0.51
Calculation	61	2.52	0.46
Application	61	2.21	0.64
Assumption	61	1.88	0.65
Communication	61	2.44	0.61

College of Science			
Interpretation	61	2.37	0.64
Representation	61	2.49	0.64
Calculation	61	2.62	0.54
Application	61	2.30	0.65
Assumption	61	1.81	0.73
Communication	16	2.41	0.67
School of Social Work			
Interpretation	15	2.30	0.70
Representation	15	2.50	0.33
Calculation	15	2.77	0.37
Application	15	2.23	0.75
Assumption	15	1.80	0.70
Communication	15	2.43	0.70
University College			
Interpretation	13	2.42	0.70
Representation	13	2.62	0.62
Calculation	13	2.69	0.63
Application	13	2.08	0.81
Assumption	13	1.69	0.43
Communication	13	2.31	0.66
Undeclared			
Interpretation	13	2.73	0.53
Representation	13	2.61	0.55
Calculation	13	2.58	0.53
Application	13	2.35	0.69
Assumption	13	2.12	0.62
Communication		2.54	0.56
Missing, unknown affiliation with School/College			
Interpretation	2	2.50	0.71
Representation	2	2.25	0.35
Calculation	2	2.25	0.35
Application	2	2.00	0.00
Assumption	2	2.00	0.00
Communication	2	2.25	0.35