UNIVERSITY OF TEXAS ARLINGTON

INSTITUTIONAL EFFECTIVENESS AND REPORTING

Measuring Written Communication

MAY 2018 REPORT

The University of Texas at Arlington

Measuring Written Communication, May 2018 Report

This report summarizes evidence of *Communication* in written student work. Assessment took place in Texas Core Curriculum (TCC) courses using embedded assignments. In order to reduce the burden of assessment on departments while maintaining consistent data collection, a multi-year schedule was used that covers the six TCC objectives within each of the eight Foundational Component Areas (FCA) specified by the Texas Higher Education Coordinating Board (THECB). As such, this report contains a summary of the findings from four FCAs: Communication; History; Language, Philosophy, and Culture; and Life and Behavioral Science. A copy of the schedule is available from the Office of Institutional Effectiveness and Reporting and displayed on their webpage.

Communication remains in the top five skills that hiring managers seek in potential employees (National Association of Colleges and Employers, 2016). The research described in this report assessed written *Communication* within embedded assignments using a rubric based on the Written Communication VALUE Rubric developed by the Association of American Colleges and Universities (AAC&U, 2015; Rhodes, 2010). UTA made minor changes to the rubric, consisting primarily of revising a title (the *Genre & Disciplinary Conventions* was changed to *Organization & Structure).* This change, suggested by UTA faculty, seemed to better operationalize the measure and increase assignment alignment. The purpose of this report is to present findings from the assessment of written *Communication* during the 2017-2018 academic year.

Method

Participants

The project gathered evidence of written *Communication* within a representative sample of undergraduates at UTA. Demographic data covering most of the sample (n = 250) was obtained and it indicated that female students represented more than half of the undergraduates (53%; n = 141) and the rest were male (42%, n = 109). While this collection of artifacts contained assignments from

students of several ethnicities, the top four ethnic groups represented were, Hispanic/Latino (34%, *n* = 85), White (30%, n = 72), Asian (17%, *n* = 43), and Black/African American (8%, *n* = 21). UTA admission materials indicated that many (36%) perceived themselves as first-generation college students and almost half (44%) were Pell Grant eligible (see Table 1). Students represented all nine UTA colleges and schools and the student artifacts were completed in a variety of approved TCC courses represented by four FCAs.

Procedure

Student essays were collected from TCC courses to measure evidence of written *Communication* attainment. Typically freshmen and sophomore-level students represent a majority, but TCC course rosters also contain upper division and transfer students who need to meet graduation criteria for the TCC. Courses that have larger proportions of sophomores or juniors are ideal. Typically, those students have completed more of their TCC courses. TCC course assignments varied by discipline. The student essays from the Language, Philosophy, and Culture FCA applied ethical frameworks to decision-making over a range of topics, whereas from the History FCA, students discussed government policies from an assigned reading. In the Life and Physical Science FCA, the students conducted an experiment, analyzed their data, and summarized the results in lab reports. From the Communication FCA, students wrote a research position paper that examined an issue from different viewpoints. After samples were collected, the essays were prepared for rating. Preparation consisted of assigning the papers a coded tracking number and then removing all personal identification information (e.g., the student's name, the faculty instructor's name) to prevent rater bias during the planned group "Scoring Day" activities.

Table 1
Student Demographics

Categorical Information	Number of Students	Percent
Gender		
Female	141	53
Male	109	41
Ethnic Description		
American Indian or Alaskan Native	2	<1
Asian	43	17
Black, African American	21	8
Foreign, Non-Resident Alien	16	6
Hispanic, All races	85	34
Multiple Ethnicities	7	3
Unknown or Not Specified	2	<1
White, Caucasian	74	30
Level		
Freshman	22	9
Sophomore	82	33
Junior	74	30
Senior	72	29
First generation college student (self-report)		
Yes	89	33
No	161	60
Pell Grant eligible upon admission (self-report)		
Yes	111	42
No	139	52
Transfer Student		
Yes	41	15
No	209	78

Note: Student demographics variables were available for 250 of the 267 essays. Six percent of the sample (n = 17 was not available.

Assessment Instrument

The assessment instrument used in this report was adapted from the AAC&U's Written Communication Rubric (AAC&U, 2015). A multi-disciplinary team of faculty experts developed the Valid Assessment of Learning in Undergraduate Education (VALUE) Rubrics under the direction of the AAC&U. Based on faculty feedback, UTA adapted the rubric in 2014 to operationalize and clarify one of the dimension titles and level descriptions (see Figure 1). The five rubric dimensions included: 1) *Context & Purpose*, 2) *Organization & Structure*, 3) *Content Development*, 4) *Sources & Evidence*, and 5) *Control of Syntax & Mechanics*. The rubric contained a narrative description of the expected

quality for each written communication paper and the corresponding point values for rating the five dimensions. Rating values ranged from 1 – 4, with 4 representing the highest observed levels of *Communication*. Raters read the student papers and rated each measure.

Communication Rubric

	Levels of Achievement			
Criteria	**Capstone 4**	**Milestone 3**	**Milestone 2**	**Benchmark 1**
Context & Purpose	Demonstrates a thorough understanding of context, audience, and purpose and a clear focus on the assigned task.	Demonstrates adequate consideration of context, audience, and purpose and is responsive to the assigned task	Demonstrates awareness of context, audience, and purpose and to the assigned task.	Demonstrates minimal attention to context, audience, purpose, and to assigned task.
Organization and Structure	Demonstrates detailed attention to successful organization, content presentation, formatting, and stylistic choices.	Demonstrates consistent use of organization, content presentation, formatting, and stylistic choices.	Follows expectations for basic organization, content presentation, formatting, and stylistic choices.	Attempts to use a consistent system for basic organization and presentation.
Content Development	Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's understanding, and shaping the whole work.	Uses appropriate, relevant, and compelling content to explore ideas within the context of the discipline and shape the whole work.	Uses appropriate and relevant content to develop and explore ideas through most of the work.	Uses appropriate and relevant content to develop simple ideas in some part of the work.
Sources and Evidence	Demonstrates skillful use of high quality, credible, relevant sources to develop ideas. Writer contextualizes sources and credits sources throughout the essay an in a works cited/bibliography page or other appropriate source documentation format.	Demonstrates consistent use of credible, relevant sources to support ideas. Writer clearly identifies sources in essay and in a works cited/bibliography page or other appropriate source documentation format.	Demonstrates an attempt to use credible and/or relevant sources to support ideas. Writer generally identifies sources in essay and in a works cited/bibliography page or other appropriate source documentation format.	Demonstrates a basic attempt to use sources to support ideas. Writer does not consistently credit borrowed material to its source in essay and/or in a works cited/bibliography page or other appropriate source documentation format.
Control of Syntax and Mechanics	Uses effective, virtually error-free, language that skillfully communicates meaning to readers with clarity and fluency.	Uses straightforward language with minimal errors to convey clear meaning to readers.	Uses language that conveys general meaning to readers although the language may contain some errors.	Uses language that sometimes impedes meaning because of errors in usage.

Adapted for The University of Texas at Arlington
from AAC&U's Written Communication VALUE Rubric
Last Revised 9/24/2014

UNIVERSITY OF TEXAS ARLINGTON

Association of American Colleges and Universities

Figure 1. UTA Communication rubric.

Raters, best practices for "Scoring Day" and inter-rater reliability goals

A multi-disciplinary group of raters was recruited from among UTA faculty. This "Scoring Day" provided training in the use of rubrics and was seen by most departments as a professional development opportunity. All raters had earned masters or doctoral degrees in their respective fields and six had professional certifications (see Table 2). On average, they had over nine years of teaching experience at the university level (M = 9.33, SD = 6.98, *Range* = 0 - 20). The multi-disciplinary group of raters represented the College of Liberal Arts (47%), College of Education, (20%), College of Nursing and Health Innovation (13%), College of Science, (7%), School of Social Work (7%), or other (7%).

Table 2
Rater Demographics

Categorical Information	Number of Raters	Percent
Gender		
Female	11	73
Male	4	27
Ethnic Description		
Hispanic, All races	0	0
White, Caucasian	15	100
Classification		
Faculty	9	60
Graduate Teaching Assistant	5	33
Staff	1	7
Highest Degree Received		
Masters	7	47
Doctoral	8	53
Additional Certifications		
IEEE	1	7
K-12 Teaching Certificate	1	7
Licensed Social Worker	1	7
Registered Nurse (RN) or RN, Clinical Nurse Specialist	2	14
Texas Principal Certificate	1	7

The raters gathered for scoring day in a group setting and began with a training/rater-calibration process led by a faculty expert from the Communications FCA. This facilitator guided a group discussion about the use of the rubric and the distinctions between rating and grading. For example, the facilitator described identifiable features for each level of the rubric and then all read a student work sample chosen by the facilitator for discussion. During this step in the calibration process, each rater read the essay and assigned ratings for each rubric dimension. After the facilitator tallied the dimension ratings using a show of hands, she led a discussion aimed at reaching a common understanding of each measure of *Communication* and the group discussed the elements that a paper must contain for awarding a score at each level. After sufficient consensus was reached, the scoring process began. A minimum of two raters individually read each paper and scored it independently using the rubric. To avoid biasing the second rater by letting them see the first rater scores, a “post-it”-type note was placed over the area of the rating sheet (Figure 3) containing the first rater scores

before passing the work to the second rater to read and score.

Besides facilitated calibration, other efforts were made to attain high inter-rater agreement. If the values awarded by the two raters were identical or within two points, then scoring was complete. However, if the scores from the two raters differed by more than two points, then a third rater would read the paper and an average of the three scores would be calculated. In this group of essays and raters, a third rater was not necessary.

	Rater 1				**Rater 2**				**Rater 3**			**(only if needed)**
Context & Purpose	4	3	2	1	4	3	2	1	4	3	2	1
Organization & Structure	4	3	2	1	4	3	2	1	4	3	2	1
Content Development	4	3	2	1	4	3	2	1	4	3	2	1
Sources & Evidence	4	3	2	1	4	3	2	1	4	3	2	1
Control of Syntax & Mechanics	4	3	2	1	4	3	2	1	4	3	2	1

Figure 3. Rater Score Sheet for UTA Communication Rubric

Analysis and Results

Inter-rater reliability

Inter-rater agreement analyses assessed whether the rater scores corresponded to each other for a particular student paper. Levels of agreement were determined by calculating the intraclass correlation coefficient (ICC). High ICC values (Fleiss Kappa) indicate more agreement between rater scores (Fleiss, 1986; Shrout & Fleiss, 1979). For this sample, ICC values indicated excellent agreement (see Table 3). These high values suggested that there is more than sufficient agreement to proceed in analyzing the data for student attainment trends that may emerge and using the data to guide university leadership with improvement decisions.

Table 3
Intraclass Correlation Coefficient (Fleiss' Kappa) for Communication dimensions

Communication VALUE Rubric Dimensions	*N* = 267
Context & Purpose	0.82
Organization & Structure	0.85
Content Development	0.77
Sources & Evidence	0.75
Control of Syntax & Mechanics	0.74

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.*

Scores from Signature Assignment ratings

The distributions of score frequencies for all but one of the dimensions closely followed standard normal curves with more student scores along the mean (rated values between 2 and 3) and fewer scores at the two tales of the curve (rated values between 1 and 4). *Sources and Evidence* was the one exception with a tri-modal distribution. Table 4 contains the score frequencies of all the ratings. Because each paper was rated twice there are twice as many ratings (*N* = 534) as papers (*N* = 267). The means for each dimension (see Table 5) show that three of the five dimensions had an average score that exceeded 2.5. The rest fell a little short, but importantly attained an average score of 2, which is the standard targeted threshold recommended by the AAC&U (Lederman, 2015). UTA follows the AAC&U recommendation and targets 2 as the target outcome. These results indicate that, on average, students exceeded the target in all five targeted dimensions.

Table 4
Frequencies for Communication Dimension Rating Scores

		Rubric Values (Percent of Student papers)							
Measurement dimensions	N	1		2		3		4	
Context & Purpose	534	38	(7%)	150	(28%)	252	(47%)	94	(18%)
Organization & Structure	534	84	(16%)	182	(34%)	203	(38%)	65	(12%)
Content Development	534	47	(9%)	224	(42%)	208	(39%)	55	(10%)
Sources & Evidence	534	164	(31%)	182	(34%)	155	(29%)	33	(6%)
Control of Syntax & Mechanics	534	44	(8%)	186	(35%)	238	(45%)	66	(12%)

Note: Each paper was rated twice, therefore the number of ratings contained in this table is double the number of papers (N=534).

Table 5
Means for Communication Measure Scores

Measurement Dimensions	N	Mean	SD	Percent> μ-1σ
Context & Purpose	267	2.75	0.76	92.9
Organization & Structure	267	2.47	0.84	84.3
Content Development	267	2.51	0.72	91.2
Sources & Evidence	267	2.11	0.82	69.3
Control of Syntax & Mechanics	267	2.61	0.72	91.8

Analyses probed the student scores further using standardized scores and the Empirical Rule (e.g., 68-95-99.7 Rule, first described by de Moivre in 1733) in order to answer the question "what percent of students score within one standard deviation of the mean or better?" The Empirical Rule drills deeper into the data to count the student scores that are above the mean or not statistically different from the mean. This step adds analytical value by distinguishing important differences that looking at a table of the means does not reveal about student attainment.

Thus the targeted threshold proposed from the Empirical Rule determined whether 84% of students would have a score that was greater than negative 1 standard deviation from the mean (84% > μ-1σ). For this sample, students met or exceeded that goal in four of the five dimensions, *Context & Purpose* (93%), *Organization & Structure* (84%), *Content Development* (91%), and *Control of Syntax and Mechanics* (92%). However, for *Sources and Evidence* (69%), almost a third scored less than negative 1 standard deviation from the mean (see Table 5), in other words, a large group scored statistically below the mean.

Summary

The current assessment of signature assignments used an adapted AAC&U Written Communication VALUE rubric. Results revealed some patterns in the evidence that indicated strengths and weaknesses in the written work samples collected from undergraduate students. In this sample of papers scored in May 2018, average student scores were strongest for the *Context &Purpose* dimension from the rubric, followed by *Control of Syntax and Mechanics*. The means

for the other dimensions exceeded the threshold value. Importantly, for all dimensions, the student's average scores met previous threshold criteria established by the university and standard use criteria set by the AAC&U (rubric values of two or better).

In addition, this written *Communication* report includes analyses that examine additional attainment criteria using standardized scores and the Empirical Rule. In doing so, this report continued the inquiry into a new target of having 84% of the students attain scores above or within one standard deviation of the mean for each dimension. Used in conjunction with the AAC&U threshold, which indicated attainment for all dimensions, this additional analysis drilled down a bit further to show that students did not meet the threshold of 84% for one dimension of the *Communication* Core Objective, *Sources and Evidence.* However, the other four: *Context & Purpose, Organization & Structure, Content Development,* and *Control of Syntax & Mechanics* met or exceeded the 84% target. While these analyses were exploratory in nature, they suggest that future studies continue this analytical approach to examine trends in student performance and improvement because they further differentiate strengths and weaknesses.

This report contains evidence from half of the THECB Foundational Component Areas (Communication; History; Language, Philosophy, and Culture; and Life and Physical Science). Measurement of authentic student work samples were completed as part of the multi-year plan to assess Communication. This report presents positive evidence of student attainment for Communication in the five AAC&U Communication VALUE Rubric dimensions using the student essays rated in May 2018. All of the reports developed by UTA to meet the THECB requirements are available from the Office of Institutional Effectiveness and Reporting.

References

Association of American Colleges and Universities (2015). *VALUE Rubrics*. Retrieved from https://www.aacu.org/value-rubrics/

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

Lederman, D. (2015). Are they learning? New effort aims to standardize faculty-driven review of student work. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2015/09/25/new-effort-aims-standardize-faculty-driven-review-student-work

National Association of Colleges and Employers. (2016). *Job Outlook 2016*. Bethlehem, PA.

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.