



UNIVERSITY OF
TEXAS
ARLINGTON

INSTITUTIONAL EFFECTIVENESS AND REPORTING

Measuring Written Communication

SPRING 2018 REPORT

The University of Texas at Arlington

Measuring Written Communication, Spring 2018 Report

Communication remains in the top five skills that hiring managers seek in potential employees (National Association of Colleges and Employers, 2016). Not by coincidence, it is among the six Texas Core Curriculum (TCC) objectives defined by the Texas Higher Education Coordinating Board (THECB, 2015). This report assesses student attainment of written *Communication*. Evidence was obtained through examining a set of work samples that undergraduate students completed in TCC courses at The University of Texas at Arlington (UTA).

The research described in this report assessed written *Communication* within work samples using a rubric based on the Written Communication VALUE Rubric developed by the Association of American Colleges and Universities (AAC&U, 2015; Rhodes, 2010). The changes that UTA made to the rubric were minor and consisted primarily of revising the *Genre & Disciplinary Conventions* dimension title to *Organization & Structure* 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 2018 spring semester at UTA.

The university sustains assessment using a multi-year schedule that covers the six TCC objectives within each of the eight Foundational Component Areas (FCA). It represents an effort to reduce the burden of assessment on departments while maintaining consistent data collection. This report contains a summary of the findings from three FCAs: Communication, Creative Arts, and Social and Behavioral Science.

Method

Participants

The project gathered evidence of written *Communication* within a representative sample and recruited qualified and engaged raters to read/score each written student artifact. More than half of the students were female (59%; $n = 118$), the rest were male (41%, $n = 83$). While this collection of

artifacts contained assignments from students of several ethnicities, the top four ethnic groups represented were, White (46%, $n = 92$), Hispanic (22%, $n = 44$), Black/African American (15%, $n = 31$) and Asian (7%, $n = 14$). UTA admission materials indicated that many (42%) perceived themselves as first-generation college students and almost half (49%) were Pell Grant eligible (see Table 1). Students represented nine UTA colleges and schools. A majority of the students represented three colleges, the College of Nursing and Health Innovation (30%), the College of Liberal Arts (29%) or the College of Science (15%), and the student artifacts were completed in a variety of course types.

Most students completed the work in a traditional on-campus setting in which they met in a classroom face-to-face with their instructor several times a week (63%). Others only interacted with the course instructor and course materials in an on-line setting (37%) via a curriculum-management system over the internet. Some of the on-line students ($n = 38$) were enrolled in accelerated, dynamically dated course sections, typically eight weeks in duration. The remainder of the on-line students ($n = 36$) followed the traditional schedule in which they were engaged in course curriculum activities over a 16-week semester schedule.

Procedure

Student essays were collected from TCC courses to measure evidence of written *Communication* attainment. Students enrolled in TCC courses are typically freshmen and sophomore-level, but course rosters also contain upper division and transfer students who need to meet graduation criteria for the TCC. An ideal sample would contain larger proportions of sophomores or juniors than the other levels as they presumably would have completed more TCC courses at UTA.

Table 1
Student Demographics

Categorical Information	Number of Students	Percent
Gender		
Female	118	59
Male	83	41
Ethnic Description		
American Indian or Alaskan Native	1	<1
Asian	14	7
Black, African American	31	15
Foreign, Non-Resident Alien	10	5
Hispanic, All races	44	22
Multiple Ethnicities	8	4
Unknown or Not Specified	1	<1
White, Caucasian	92	46
Level		
Freshman	43	21
Sophomore	82	41
Junior	56	28
Senior	20	10
Type of course		
Traditional 16-week on-campus	127	63
Traditional 16-week on-line	36	18
Accelerated, dynamic-dated on-line	38	19
First generation college student (self-report)		
Yes	84	42
No	117	58
Pell Grant eligible upon admission (self-report)		
Yes	98	49
No	103	51
Transfer Student		
Yes	103	51
No	98	49
College or School		
College of Nursing and Health Innovation	61	30
College of Liberal Arts	58	29
College of Science	30	15
College of Business	20	10
University College	15	8
College of Engineering	8	4
College of Architecture, Planning, & Public Affairs	5	3
School of Social Work	3	2
College of Education	1	<1

Some TCC courses assigned students a research position paper that examined an issue from different viewpoints. Another assignment from the Creative Arts courses directed the students to attend and analyze performance art. Work samples from the Social and Behavioral Sciences consisted of the student's reflections after interviewing a person who was born outside our country about the similarities and differences between their homeland and their new home. 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.

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 with guidance from the AAC&U. UTA adapted the rubric in 2014 to operationalize and clarify one of the dimension titles and level descriptions. 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

Criteria	Levels of Achievement			
	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 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



Figure 1. Adapted Communication rubric.

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

Providing training in the use of rubrics as a professional development opportunity, among other things, seemed to enhance recruitment efforts to gather a multi-disciplinary group of raters from within the UTA faculty. All had earned masters or doctoral degrees in their respective fields and three had professional certifications (see Table 2). On average, they had eight years of teaching experience at the university level ($M = 8.06$, $SD = 7.27$, $Range = 0 - 22$). The multi-disciplinary group of raters represented the College of Liberal Arts (56%), University College, (13%), College of Business (6%), or other (25%, e.g., Center for Distance Education, Office of Institutional Effectiveness and Reporting).

Table 2
Rater Demographics

Categorical Information	Number of Raters	Percent
Gender		
Female	10	62.5
Male	6	37.5
Ethnic Description		
Asian	0	0
Black, African American	1	6.3
Hispanic, All races	3	18.8
White, Caucasian	13	81.3
Classification		
Faculty	5	31.3
Graduate Teaching Assistant	6	37.4
Staff	5	31.3
Highest Degree Received		
Masters	10	37.5
Doctoral	6	62.5
Additional Certifications		
IEEE	1	6.3
Licensed Mediator	1	6.3
Texas Teaching Certificate	1	6.3

The raters gathered for scoring day in a group setting and began with a training/rater-calibration process led by a faculty expert. This facilitator guided a group discussion about the distinctions between rating and grading, and the use of the rubric. For example, the facilitator described identifiable features for each level of the rubric and then all the raters 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 simple 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. After rating, Rater A placed an adhesive “post-it”-type note as a covering over their ratings on the score sheet to avoid biasing Rater B with their scores. Then Rater A passed the paper to Rater B to read and score.

Achievement of inter-rater agreement was a high priority. If the values awarded by the two raters were identical or within two points, then scoring was completed and during analyses the scores were averaged. For example, if Rater 1 scored the *Context & Purpose* measure with a value of 2 and Rater B scored *Context & Purpose* with a value of 4, then the average of the two scores for *Context & Purpose* was 3. 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. For this group of essays and raters, a third rater was unnecessary. Figure 3 displays an image of the rater score sheet.

	Grader 1				Grader 2				Grader 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 a trend of good to excellent agreement (see Table 3). High values suggest that there is 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	<i>N</i> = 201
Context & Purpose	0.71
Organization & Structure	0.66
Content Development	0.69
Sources & Evidence	0.68
Control of Syntax & Mechanics	0.70

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 each 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 tails of the curve (rated values between 1 and 4). Table 4 contains the score frequencies of all the ratings. Because each paper was rated twice there are twice as many ratings (*N* = 402) as papers (*N* = 201). The means for each dimension (see Table 5) show that one of the five dimensions, *Context & Purpose*, 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 their 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

Measurement dimensions	N	Rubric Values (Percent of Student papers)			
		1	2	3	4
Context & Purpose	402	40 (10%)	133 (33%)	196 (49%)	33 (8%)
Organization & Structure	402	68 (17%)	151 (38%)	171 (43%)	12 (3%)
Content Development	402	60 (15%)	173 (43%)	149 (37%)	20 (5%)
Sources & Evidence	402	66 (16%)	138 (34%)	186 (46%)	12 (3%)
Control of Syntax & Mechanics	402	66 (16%)	134 (33%)	191 (48%)	11 (3%)

Note: Each paper was rated twice, therefore the number of ratings contained in this table is double the number of papers (N=201).

Table 5
Means for Communication Measure Scores

Measurement Dimensions	N	Mean	SD	Percent $> \mu - 1\sigma$
Context & Purpose	201	2.55	0.69	87.1
Organization & Structure	201	2.32	0.68	79.1
Content Development	201	2.32	0.69	80.1
Sources & Evidence	201	2.36	0.69	80.1
Control of Syntax & Mechanics	201	2.37	0.69	80.6

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?” These analyses assume a standard normal curve (e.g., bell-shaped) and analyses found that these data were skewed negatively with more rating values of 1 than rating values of 4. That said, 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 to the evidence by examining meaningful target thresholds for student attainment. The targeted threshold proposed from the Empirical Rule determines whether 84% of students would have a score that was greater than negative 1 standard deviation from the mean ($84\% > \mu - 1\sigma$). For this sample, students exceeded that goal in one of the five dimensions, *Context & Purpose* (87%). The goal was unmet for *Organization & Structure* (79%), *Content Development* (80%), *Sources and Evidence* (80%) and *Control of Syntax and Mechanics* (81%). That said, more than eighty percent of the students scored greater than negative 1 standard deviation of the mean (see Table 5) in the majority of the measures, in other words, above the mean or statistically no different than the mean.

Further examination of the relationships between the student characteristics and the five written *Communication* dimensions used analysis of variance (ANOVA). Among course types (e.g. on campus vs. online), a significant effect was not found for any of the five rubric dimensions. That said, mean scores were higher for *Control of Syntax & Mechanics* and *Sources & Evidence* dimensions in

dynamically dated on-line sections ($M = 2.41$, $M = 2.36$, respectively) and in on-campus traditional 16-week sections ($M = 2.41$, $M = 2.45$, respectively) compared to on-line traditional 16-week sections ($M = 2.18$, $M = 2.25$, respectively). Independent sample t tests were used to look for significant differences in any of the five rubric dimensions: by gender, by Pell eligibility, by transfer, or by residing on-campus or off campus and none were found. However, analysis of variance suggested a significant effect for first generation students in *Organization & Structure*, $F(1) = 5.67$, $p = .02$, in which those who do consider themselves the first from their family to attend college scored lower on average than those who do not. As expected, significant differences were found between scores for upper level (juniors and seniors) vs lower level (freshmen and sophomores) on four of five dimensions, with upper level scoring higher than lower level, with the exception of *Organization & Structure*, $p = .24$, where underclassmen scored higher. Linear regression revealed that higher numbers of semester hours completed did not significantly predict higher dimension scores.

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 *Communication* in a sample of undergraduate students. In addition, analyses included an examination of student characteristics in order to identify trends and comparisons by groups.

In this sample of papers scored in the spring of 2018, average student scores were strongest for the *Context & Purpose* dimension from the rubric. The means for the other four 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 met the threshold of 84% for one dimension of the THECB *Communication Core Objective, Context & Purpose*. However, *Organization & Structure, Content Development, Sources and Evidence, and Control of Syntax & Mechanics* did not. 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.

An examination of student characteristics indicated that the sample was generally descriptive of the university. Continued evidence of quality instruction in the dynamically dated on-line courses is extremely encouraging, as students in these courses interact an accelerated schedule outside of the traditional brick-and-mortar institution. That said, this evidence is limited by the size of the sample, and plans to continue this line of inquiry should span all six TCC objectives.

This report contains evidence from three of the eight Foundational Component Areas (Communication, Creative Arts, and Social and Behavioral Sciences). Authentic student work samples were collected for this measurement 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 the spring 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.