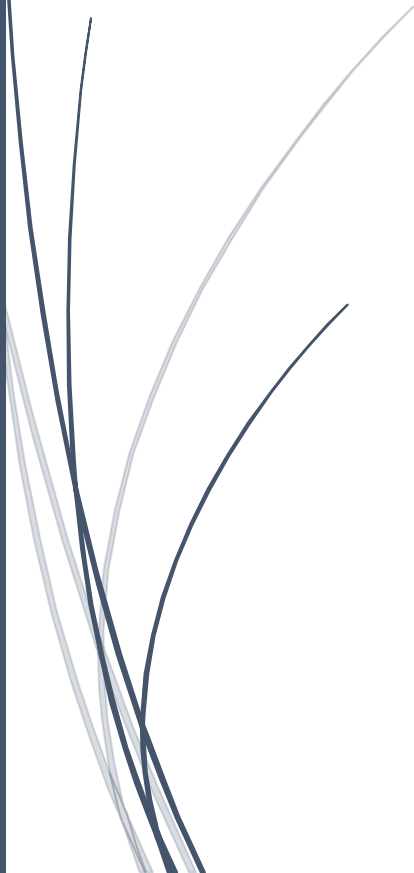




Report for the Texas Assessment Collaborative

Larry J. King & Chris Duke



Report for the Texas Assessment Collaborative

Introduction

The Texas Assessment Collaborative (TAC) is a project sponsored by LEAP Texas as part of the Faculty Collaboratives project of the Association of American Colleges and Universities (AAC&U). The project was funded through a grant from the Lumina Foundation. LEAP Texas selected two Assessment Fellows as part of the project, Chris Duke and Larry J. King. One Assessment Fellow was funded through the Faculty Collaboratives grant and the other was funded through funds provided by LEAP Texas. The Assessment Fellows outlined a project aimed at leveraging the newly redesigned Texas Core for the overall improvement in undergraduate education, advancing the authentic assessment of student work in the Texas Core, and promoting the development and refinement of the capacity of authentic assessment of student work in the Texas Core. Texas two-year and four-year colleges and universities were invited to participate in the TAC by providing samples of student work from their institutions which were then assessed by teams of assessors representing the participating institutions. The TAC was designed to be a state level complement to the Multi-State Collaborative (SHEEO, n.d.), and was intended to provide a foundation for understanding student performance on the Texas Core Objectives (Texas Higher Education Coordinating Board, n.d.). The Assessment Fellows determined that written communication would be the focus for this initial round of assessment.

Participants

Three 2-year institutions participated in the TAC (El Centro College, San Jacinto College, and Weatherford College) and four 4-year institutions participated (Stephen F. Austin State University, the University of Houston—Victoria, the University of North Texas, and The University of Texas at Arlington). The seven participating institutions collected and submitted 514 pieces of student work to LiveText, the ePortfolio system used for the TAC. Institutions removed all identifiable student and institutional information from the work before the institutional representative uploaded the work to LiveText. Institutions drew student work from the spring, summer, or fall 2016 semesters. Work was collected from students who were at the sophomore or junior level (or 42-72 completed semester credit hours). Institutions were allowed to select up to 25 documents from a single course section, but institutions were asked to diversify the sections from which student work was drawn. No more than one piece of student work was collected from an individual student. Institutions submitted a total of 514 pieces of student work. A random sample of 400 pieces of work was selected from the using the build in sampling tool in LiveText from the 514 pieces of work submitted by the seven participating institutions. Following the scoring it was determined that one piece of student work had not been assessed by one of the two original scorers. This piece of student work was removed from the sample leaving a total sample of 399 artifacts. It is important to note, because the student work selected for submission by the participating institutions was not selected through a random sample, the results of the TAC should not be generalized beyond the sample of work selected for this project.

Sample Demographics

The majority of the 399 students in the sample were 22-24 years old (N=194, 49%), sophomores (N=253, 63.4%), females (N=229, 57.4%), White (N=179, 44.9%), Not Hispanic or Latino (N=252, 63.2%), and from one of the 4-year institutions (N=276, 69%). Table 1 provides a summary of the demographic data for the sample.

Table 1

Category		N	Percent
Age			
	18 & Under	5	1.2
	19-21	86	21.5
	22-24	194	49
	25-30	53	13
	31 & Older	61	15.3
Classification			
	Sophomore	253	63.4
	Junior	141	35.3
	Unclassified	5	1.3
Gender			
	Female	229	57.4
	Male	170	42.6
Race			
	American Indian or Alaskan Native	3	.8
	Asian	14	3.5
	Black or African-American	71	17.8
	Hispanic or Latino Origin	62	15.5
	International	5	1.3
	Native Hawaiian or Other Pacific Islander	1	.3
	Two or More	12	3
	Unknown or Not Reported	52	13
	White	179	44.9
Ethnicity			
	Hispanic or Latino Origin	123	30.8
	Not Hispanic or Latino	252	63.2
	Not Answered	24	6
Institution			
	2-Year Institution	123	31
	4-Year Institution	276	69

Scorers

A volunteer team of faculty and experienced assessment staff from the seven participating institutions scored the student work. The team included 26 faculty and 4 staff members, and 12 of the team members were from 2-year institutions and 18 were from 4-year institutions. Members of the scoring team met for a four and a half hours rubric norming session and training session, lead by Chris Duke and Larry King, on the use of LiveText on February 19, 2017, in Dallas, Texas, before the 4th Annual LEAP Texas Forum. During this session, team members reviewed and comprehensively discussed the rubric they would be using to assess student work. Team members also scored samples of student work and discussed their questions and concerns. From this discussion, Duke and King led the group in the development of a formalized set of scoring rules and shared understanding of how scorers would

interpret and apply the rubric to scoring student work. Team members were also trained on the use of the LiveText ePortfolio System for assessing the sample of student work.

Scoring

Each piece of student work in the sample was sent to two scorers through the LiveText system. Scorers assessed the student work in LiveText using a copy of the rubric built into LiveText and following the rules developed in the norming session. If the two scorers had a disagreement on a rubric criterion, the artifact was sent to a third scorer to score only the criterion for which there was disagreement. For this project, disagreement was defined as a difference of greater than one integer between the first two scorers. A complete list of the rules for agreement/disagreement can be found in Appendix A. This procedure was drawn from the procedures used by Texas A&M University and Stephen F. Austin State University. Members of the scoring teams were given roughly a month to complete the assessment of the student work assigned to them by LiveText.

Rubric

The rubric used to assess student work in this round of the TAC was the Written Communication VALUES Rubric (AAC&U, 2009). Teams of faculty from over 100 colleges and universities from across the U.S. developed the VALUE rubrics. The faculty teams examined numerous rubrics used on campuses around the nation, and developed the 16 VALUE rubrics from the most commonly identified characteristics or criteria of learning for each learning outcome. Following the development process, faculty at over 100 college and universities around the country tested the rubrics. The Multi-State Collaborative also uses the VALUE rubrics to “produce valid data summarizing faculty judgments of students’ own work, and also seeks to aggregate results in a way that allows for benchmarking across institutions and states” (SHEEO, n.d.). The Written Communication VALUE Rubric has five criteria that are measured using a 4-point Likert scale, and raters are allowed to give 0 points if they feel the student work does not meet the lowest level of scale (see Appendix B).

Inter-Rater Agreement

Inter-rater reliability (IRR) analysis was conducted to validate the results of the scoring. Intra-class correlations coefficient (ICC) analysis was conducted on the scores given by the team members following the rules for disagreement (Appendix A). Guidelines for ICC scores are as follows: scores below .40 are considered “poor”, scores between .40 and .59 are considered “fair”, below .60 and .74 are considered “good”, and scores between .75 and 1.00 are considered “excellent” (Cicchetti, 1994, p. 286). ICC scores were calculated for each rubric criteria and an overall ICC was calculated for the scores for all criteria (see Table 2).

Table 2

Rubric Criteria	ICC Scores
Audience, Context, and Purpose	.79
Content Development	.78
Genre and Disciplinary Conventions	.75
Sources and Evidence	.85
Control of Syntax and Mechanics	.71
All Criteria	.80

Results

The mean and mode were calculated for each rubric criteria and for all criteria (see Table 3). The mean and mode indicate that students in this sample generally scored in the “2” or “Milestones” range on the Written Communication VALUE Rubric. One criteria, “Sources of Evidence,” deviated from this trend with a mode of 1 and a mean on 1.57.

Table 3

Rubric Criteria	Mode	Mean
Context and Purpose for Writing	2	2.13
Content Development	2	1.92
Genre and Disciplinary Conventions	2	1.91
Sources and Evidence	1	1.57
Control of Syntax and Mechanics	2	2.01
All Criteria	2	1.91

Frequencies were also calculated for the average scores by rubric criteria and for all criteria (see Table 4 and Figure 1). Average scores are based on the rules found in Appendix A. Figure 1 illustrates the trend of lower performance on the “Sources of Evidence” criteria. Assuming that students at the sophomore or junior level should perform at the 2 or “Milestones” level or higher on the rubric, we can see that the students in this sample performed highest on “Context of and Purpose for Writing,” second on “Control of Syntax and Mechanics,” third highest on the “Genre and Discipline Conventions,” fourth on “Content Development,” and fifth on “Sources of Evidence” (see Figure 2).

Table 4

Rubric Criteria	Frequency of Average Scores by Rubric Criteria									
	0	.5	1	1.5	2	2.5	3	3.5	4	
Audience, Context, and Purpose	3	5	46	80	98	84	43	27	13	
Content Development	3	17	59	93	106	66	29	20	6	
Genre and Disciplinary Conventions	4	11	58	94	115	73	24	13	7	
Sources and Evidence	32	50	82	78	54	51	34	13	5	
Control of Syntax and Mechanics	0	14	37	95	111	83	43	14	2	
All Criteria	42	97	282	440	484	357	173	87	33	

Figure 1

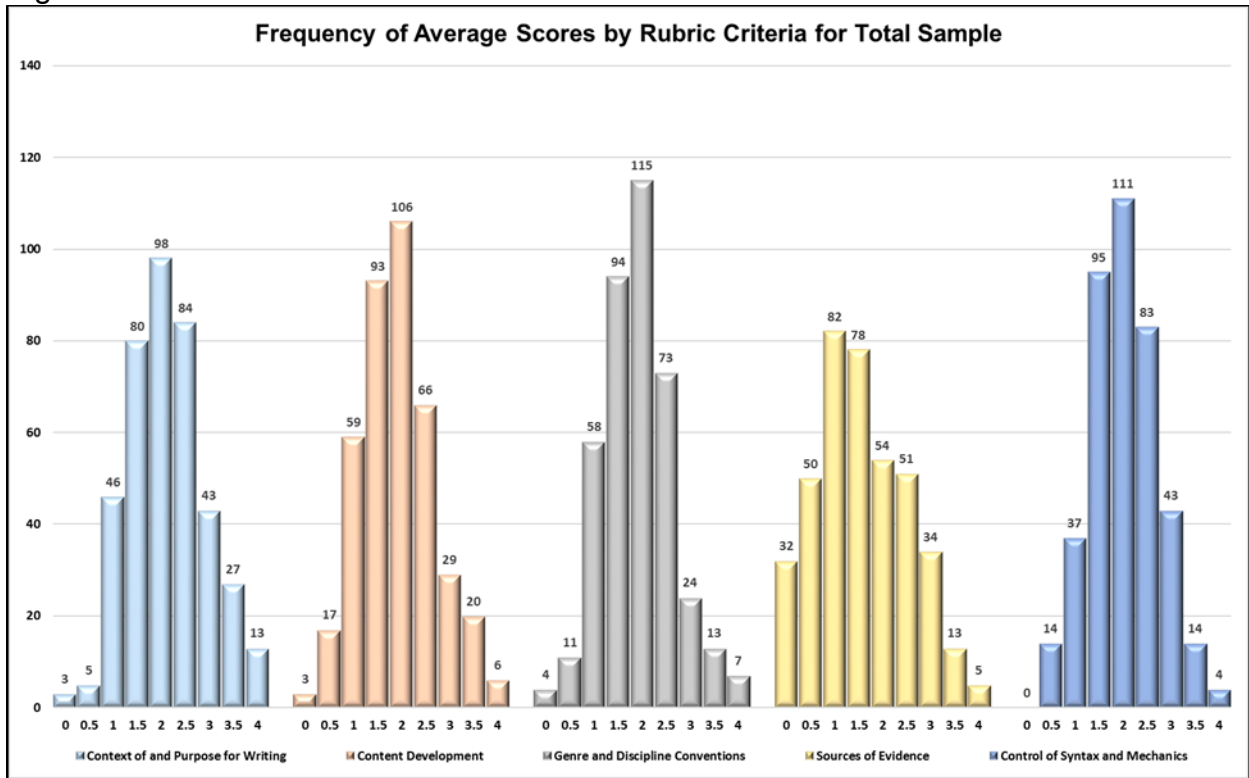
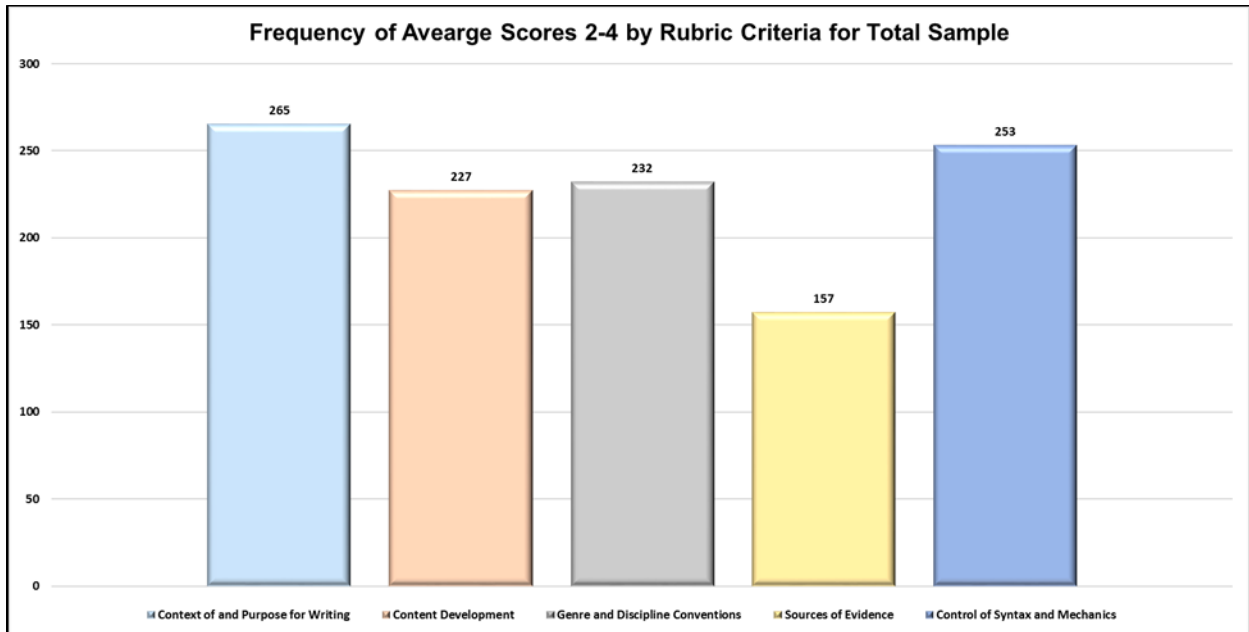


Figure 2



Conclusion

This initial venture of the TAC demonstrates the viability of the collection of student work from a number of institutions and the assessment of that student work at the state level by faculty and staff representatives of those institutions. While we have cautioned that generalizing from the results of this preliminary study is inappropriate, we can say that for the sample of student work assessed here, students scored lower on the criteria of “Sources of Evidence” than the other criteria include in the Written Communication VALUE Rubric. We can also say that the Written Communication VALUE Rubric appeared to be appropriate for the assessment of the student work collected from the participating institutions. We also believe that participation in this project allowed for training and experience in this type of authentic assessment by representatives of institutions where this type of assessment is not currently used for assessment of the core objectives.

Future work of the TAC should focus on the assessment of other Texas Core Objectives. If possible, future TAC assessment projects should attempt to collect a sample of student work that will allow for generalizing to students throughout the state. The TAC should also tackle the difficult task of developing rubrics for the assessment of the Texas Core Objectives that do not have a logically associated rubric among the AAC&U VALUE Rubrics (i.e., Social Responsibility and Personal Responsibility). We firmly believe that the work of the TAC should continue.

Reference List

- AAC&U. (n.d.). *Essential Learning Outcomes*. Retrieved from <http://www.aacu.org/leap/essential-learning-outcomes>
- AAC&U. (2009). *Written Communication VALUE Rubric*. Retrieved from <https://www.aacu.org/value/rubrics/written-communication>
- Cicchetti, D. V. (1994). Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284-290.
- SHEEO. (n.d.). *MSC: A Multi-State Collaborative to Advance Learning Outcomes Assessment*. Retrieved from <http://www.sheeo.org/projects/msc-multi-state-collaborative-advance-learning-outcomes-assessment>.
- Texas Higher Education Coordinating Board. (n.d.). *Elements of the Texas Core Curriculum*. Retrieved from <http://www.thecb.state.tx.us/index.cfm?objectid=427FDE26-AF5D-F1A1-E6FDB62091E2A507>

Appendix A

Rules for Scoring Student Work

Procedures for assessment of student work:

1. Each piece of student work will be initially assessed by two raters.
2. If the two raters agree on their rating on any element/criterion of a rubric then there is no need for a third rater on that element/criterion.
3. If the first two raters are no more than one integer apart on their ratings on an element/criterion of a rubric, then the two ratings are averaged together and there is no need for a third rater on that element/criterion.

For example, if Rater A gives a piece of student work a 2 on element/criterion of Audience, Context, and Purpose, and Rater B gives the piece of student work a 3 on Audience, Context, and Purpose, then the two ratings are averaged together to give a 2.5 on the Audience, Context, and Purpose element/criterion.

If the two raters are more than one integer apart on their ratings on any element/criterion of a rubric, a third rater is asked to rate only the element(s)/criteria where there was disagreement.

For example, if Rater A gives a piece of student work a 1 on the element/criterion Audience, Context, and Purpose, and Rater B gives the piece of student work a 3 on Audience, Context, and Purpose. Also, rater A also gives the same piece of student work a 4 on Sources and Evidence, and Rater B gives that same piece of student work a 2. Then a third rater (Rater C) is asked to rate the student work only on the elements/criteria of Audience, Context, and Purpose and Sources and Evidence.

4. If Rater C's rating agrees with one of the other two ratings, then that rating is used and the rating that is not in agreement is discarded.

For example, if Rater C and Rater A each rate a piece of student work a 2 on Content Development, but Rater B rates the work a 4, then Rater B's rating is discarded and the student work received a rating of 2 on Content Development.

5. If Rater C's rating does not agree with one of the other two ratings, and is no more than one integer from only one of the other ratings, then the rating that is more than one integer from the other ratings is discarded, and the two ratings that are no more than one integer apart are averaged.

For example, if Rater C rates a piece of student work 2, Rater A rated the work a 1, and Rater B rated the work 4 on Content Development. Rater B's rating of 4 is discarded and the ratings of Rater C and Rater A are averaged to get a rating of 1.5.

6. If Rater C's rating is no more than one integer from the other two ratings, then all of the ratings are averaged.

For example, if Rater C rates a piece of student work 3, Rater A rated the work a 2, and Rater B rated the work 4 on Content Development. All of the ratings are averaged for a rating of 3.

7. If Rater C's rating does not agree with one of the other two ratings and is more than one integer apart from the other two ratings, then Rater C's rating is discarded, and the other two ratings are averaged.

For example, if Rater C rates a piece of student work 4, Rater A rated the work a 0, and Rater B rated the work a 2 on Content Development. Rater C's rating of 4 is discarded, and the other two ratings are averaged to get a rating of 1.

Appendix B

WRITTEN COMMUNICATION VALUE RUBRIC
for more information, please contact value@accu.org



Definition

Written communication is the development and expression of ideas in writing. Written communication involves learning to work in many genres and styles. It can involve working with many different writing technologies, and mixing text, data, and images. Written communication abilities develop through iterative experiences across the curriculum.

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
Context of and Purpose for Writing <i>Includes considerations of audience, purpose, and the circumstances surrounding the writing task(s).</i>	Demonstrates a thorough understanding of context, audience, and purpose that is responsive to the assigned task(s) and focuses all elements of the work.	Demonstrates adequate consideration of context, audience, and purpose and a clear focus on the assigned task(s) (e.g., the task aligns with audience, purpose, and context).	Demonstrates awareness of context, audience, purpose, and to the assigned task(s) (e.g., begins to show awareness of audience's perceptions and assumptions).	Demonstrates minimal attention to context, audience, purpose, and to the assigned task(s) (e.g., expectation of instructor or self as audience).
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 parts of the work.
Genre and Disciplinary Conventions <i>Formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields (please see glossary).</i>	Demonstrates detailed attention to and successful execution of a wide range of conventions particular to a specific discipline and/or writing task (s) including organization, content, presentation, formatting, and stylistic choices	Demonstrates consistent use of important conventions particular to a specific discipline and/or writing task(s), including organization, content, presentation, and stylistic choices	Follows expectations appropriate to a specific discipline and/or writing task(s) for basic organization, content, and presentation	Attempts to use a consistent system for basic organization and presentation.
Sources and Evidence	Demonstrates skillful use of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline and genre of the writing	Demonstrates consistent use of credible, relevant sources to support ideas that are situated within the discipline and genre of the writing.	Demonstrates an attempt to use credible and/or relevant sources to support ideas that are appropriate for the discipline and genre of the writing.	Demonstrates an attempt to use sources to support ideas in the writing.
Control of Syntax and Mechanics	Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.	Uses straightforward language that generally conveys meaning to readers. The language in the portfolio has few errors.	Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.	Uses language that sometimes impedes meaning because of errors in usage.