

General Education Assessment Report – AY 2016-17

Author: Joseph F. van Gaalen, Ph.D., Director, Assessment & Effectiveness

1 INTRODUCTION

This report details the results of Florida SouthWestern State College's (FSW) General Education assessment for AY 2016-2017. The intent of FSW's General Education Program is to foster lifelong learning and establish academic excellence, interdisciplinary dialog, and a social responsibility among students. Each of the five competencies (Communication, Critical Thinking, Technology/Information Management, Global Socio-cultural Responsibility, and Scientific and Quantitative Reasoning) was assessed through assignments identified by faculty as fitting the criteria of the competency (Braselton, 2011; Rhodes and Finley, 2013). General Education assessment will continue in AY 2017-2018, however, the newly adopted competencies of Communicate, Research, Evaluate, Analyze, Think, Investigate, Visualize, and Engage, will replace the previous five competency model.

The General Education Assessment Subcommittee of the Learning Assessment Committee adopted (see [June 9, 2014 GEAS Subcommittee Meeting Minutes](#)) the Association of American Colleges & Universities (AAC&U) Value Rubric Model (Rhodes and Finley, 2013) after an extensive review of General Education assessment models employed throughout higher education. The study employed the use of the AAC&U rubrics for a comprehensive review of the Communication (CT) and the Scientific and Quantitative Reasoning (QR) competency. Note that the AAC&U Value Rubric was used for the CT competency, but an FSW developed rubric was used for QR.

As aligned with the AAC&U Value Rubric Model and Value Rubric Case Studies, FSW faculty from across disciplines voluntarily submitted assignments aligned with the competencies. Assignments do not have to be uniform if outcomes, rating, and the rationale for rating (rubric interpretation) are uniform (Rhodes & Finley, 2013). Outcomes are identified by the competency definition at FSW. Calibration sessions were conducted before scoring in each competency. Inter-rater reliability studies were performed on the results.

The purpose of the program is to: 1) measure against baseline data for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric, 2) measure against baseline data for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric across sites (Online, Dual Enrollment, and Traditional), 3) establish a baseline for the number of student artifacts receiving a score of 3 or higher on relevant aspects of the rubric across credit achievement level (e.g. achievement with respect to number of credits earned), 4) establish a baseline for the number of student artifacts receiving a score of 3 or higher on relevant aspects of the rubric across pre-requisite definition (e.g. achievement with respect to pre-requisite courses), and 5) determine the suitability of the AAC&U Value rubrics for assignments/assessments administered at FSW.

For additional detail on further analysis not provided in this report, please contact Dr. Joseph F. van Gaalen, Director of Assessment & Effectiveness, Academic Affairs (jfvanqaalen@fsw.edu; x16965).

2 CRITICAL THINKING (CT)

2.1 AY 2016-2017 STUDY

The CT competency goal is by completion of the general education requirements, students will be able to demonstrate skills necessary for analysis, synthesis, and evaluation. The FSW Learning Assessment Committee will measure the number of artifacts scored a 3 or higher on relevant dimensions of the rubric against the pilot results (AY 2014-2015). Figures 1 through 9 below depict achievement and inter-rater reliability for the CT competency in college-wide, Associate of Arts (AA) cohorts, as well as value-added studies as they relate to outcome goals and objectives.

Feedback from scorers regarding the AAC&U Value rubric for the CT competency included two main trends regarding rubric suitability. First, multiple scorers noted that not all CT assignments require sources. The varying interpretation of this dimension based on the incorporation of it in the assignment caused concerns over consistent scoring habits. This comment was noted in the pilot study (AY 2014-2015) as well. In response, scorers coordinated to leave the scoring for any rubric dimension blank if that area was not explicitly described in the assignment guidelines provided by the instructor. Second, scorers noted that the rubric was better suited for written assignments and that mathematical assignments was difficult to apply. For this study, no modifications to the AAC&U Value rubrics has been made to tailor achievement levels.

One third area noted by feedback unrelated to rubric suitability is that of plagiarism. This point was also noted in the AY 2014-2015 study. In response to this, all artifacts were reviewed using TurnItIn to confirm originality before being sent out to the scoring teams. Documented cases of plagiarism were removed from scoring.

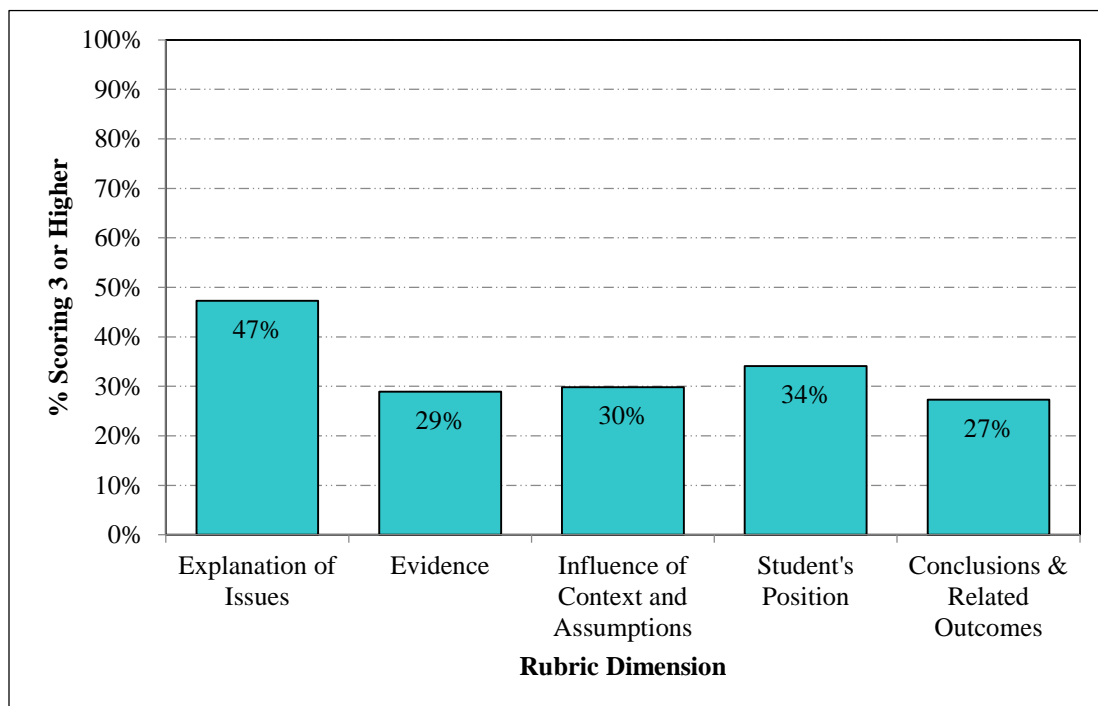


Figure 1. CT Achievement at 3 or higher across all rubric dimensions for 153 artifacts from 30 volunteered assignments.

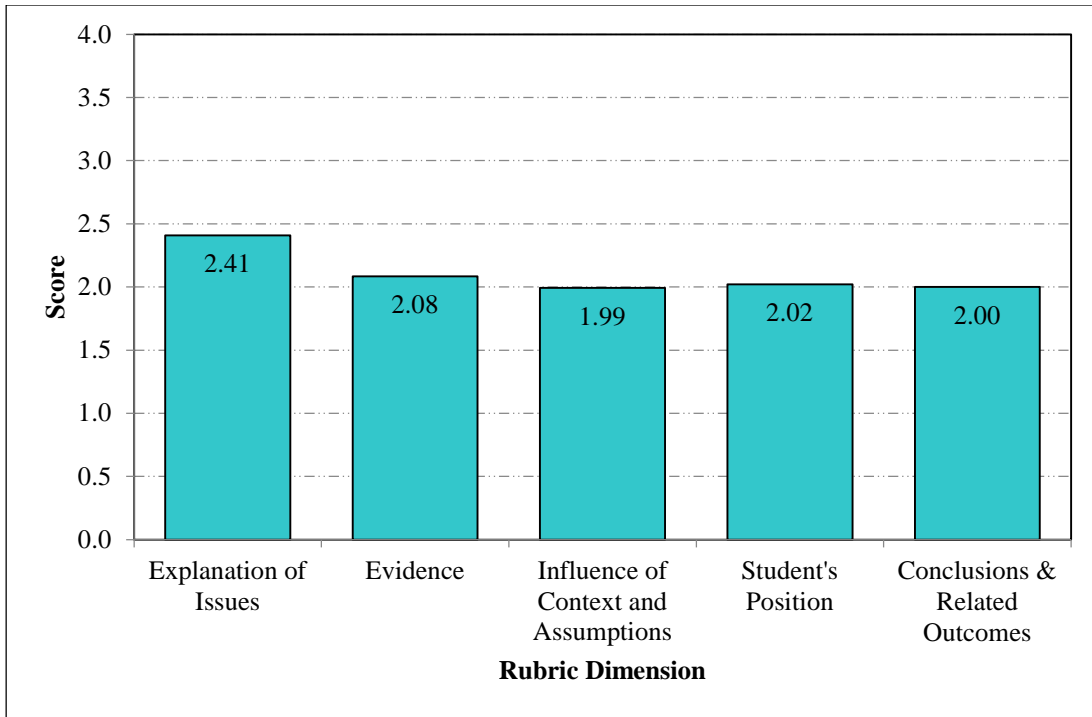


Figure 2. Mean score by rubric dimension for CT for 153 artifacts from 30 volunteered assignments.

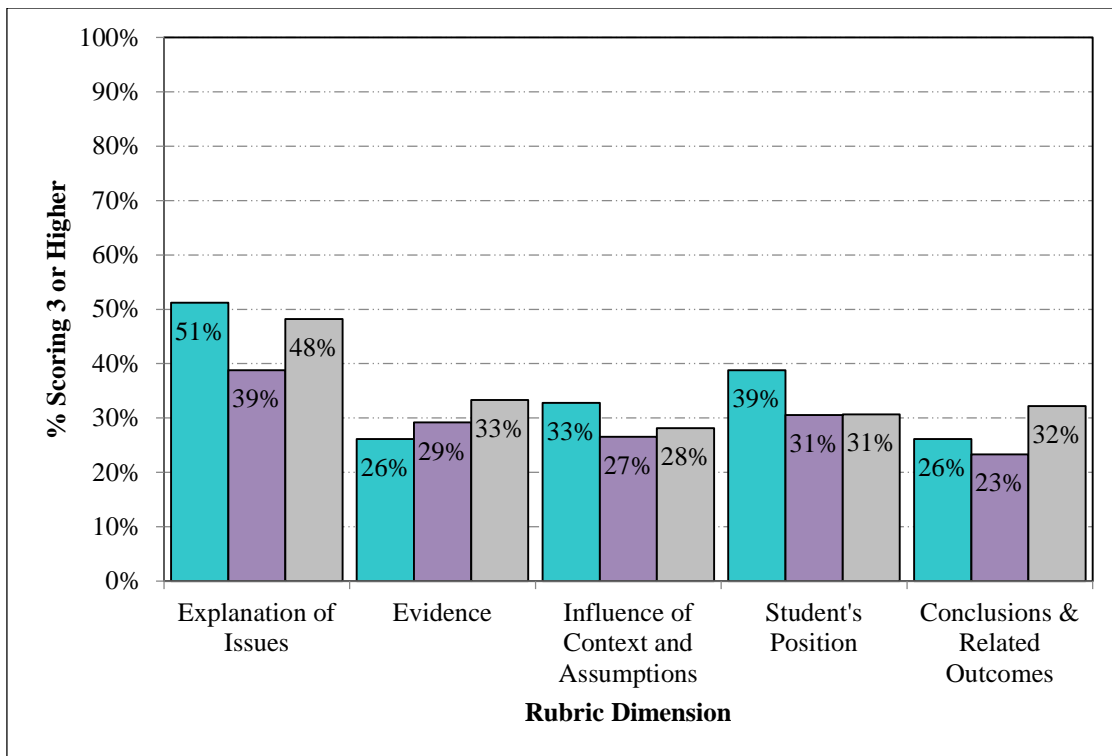


Figure 3. Comparison of CT achievement at 3 or higher across all rubric dimensions for 153 artifacts from 30 volunteered assignments. Traditional (aqua), n=81, Online (purple), n=38, Dual Enrollment (gray), n=51.

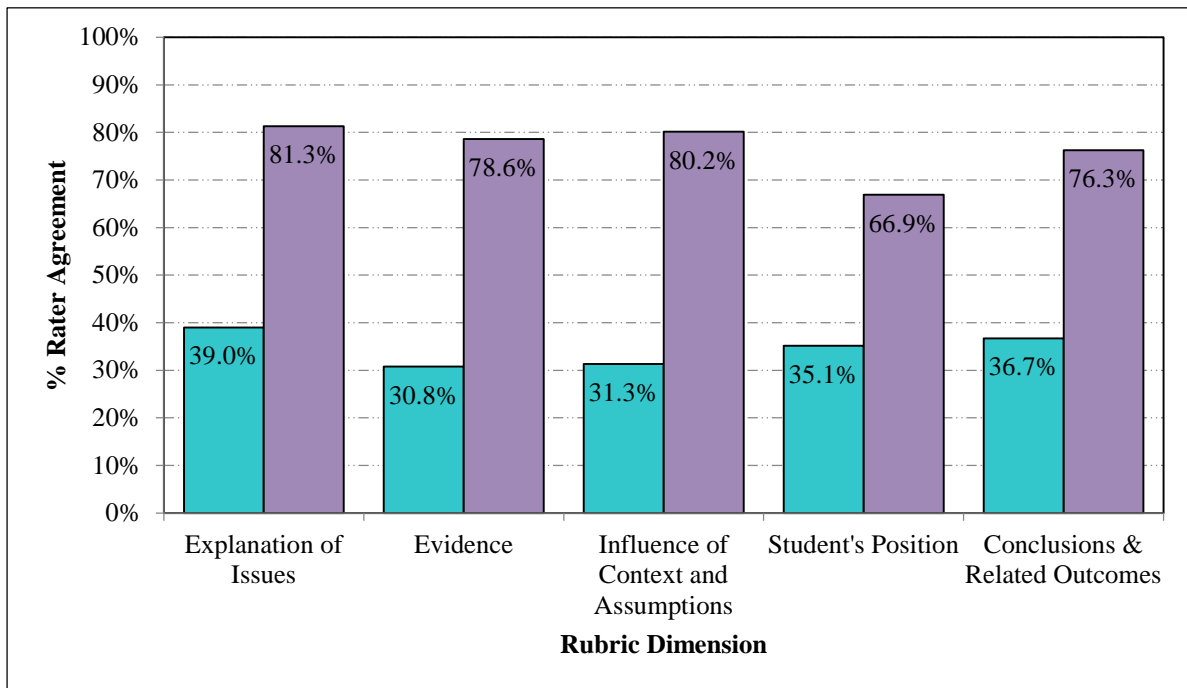


Figure 4. Inter-rater reliability (as %) for CT competency. Each artifact was scored by two scorers. Percentage (%) of agreement (aqua) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.

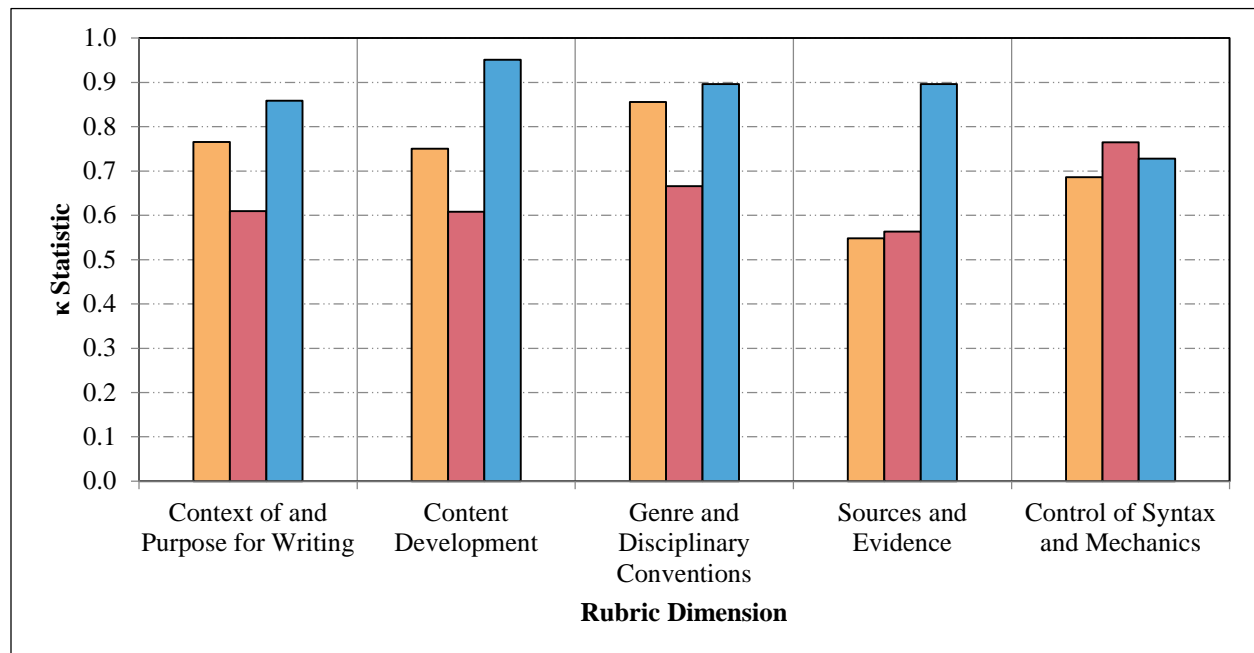


Figure 5. κ -statistic for scorer pairs of the CT competency artifacts (Scorers 1A & 1B – orange, Scorers 2A & 2B – red, Scorers 3A and 3B – blue). The κ -statistic is evaluated for the percentage (%) +/- 1 agreement shown in Figure 4 above and takes into account the random chance that scorers would agree (Cohen, 1960; Gwet, 2002), thus the κ -statistic results will be slightly lower than a straightforward percentage of agreement. The κ -statistic was calculated for the CT study to be use as a guide to interpret percentage agreement in an academic sense.

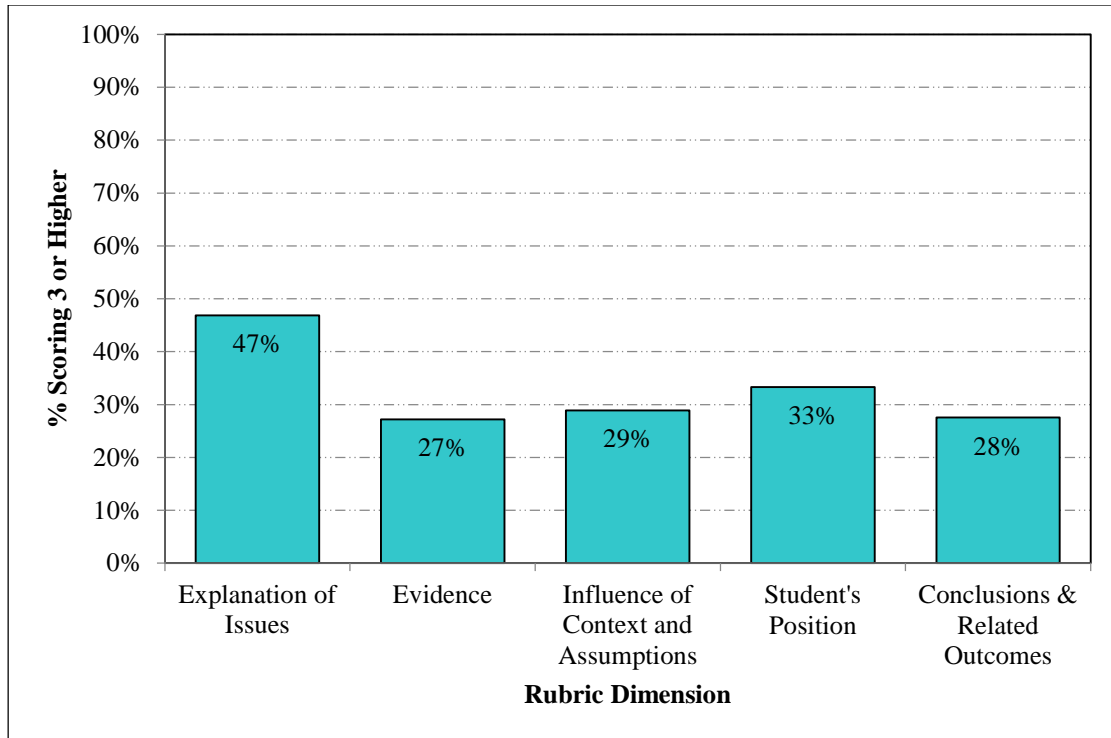


Figure 6. CT Achievement at 3 or higher for AA courses only across all rubric dimensions for 128 artifacts from 19 volunteered assignments.

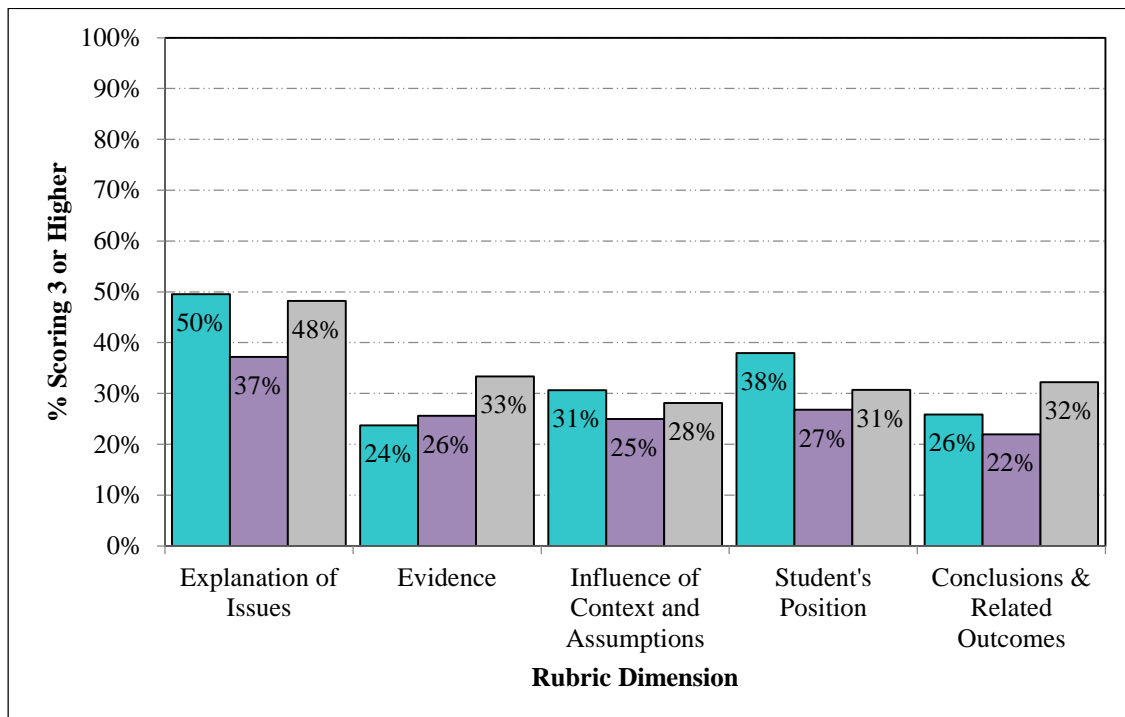


Figure 7. CT achievement at 3 or higher for AA courses only across all rubric dimensions for 128 artifacts from 19 volunteered assignments. Traditional (aqua), n=65, Online (purple), n=22, Dual Enrollment (gray), n=51.

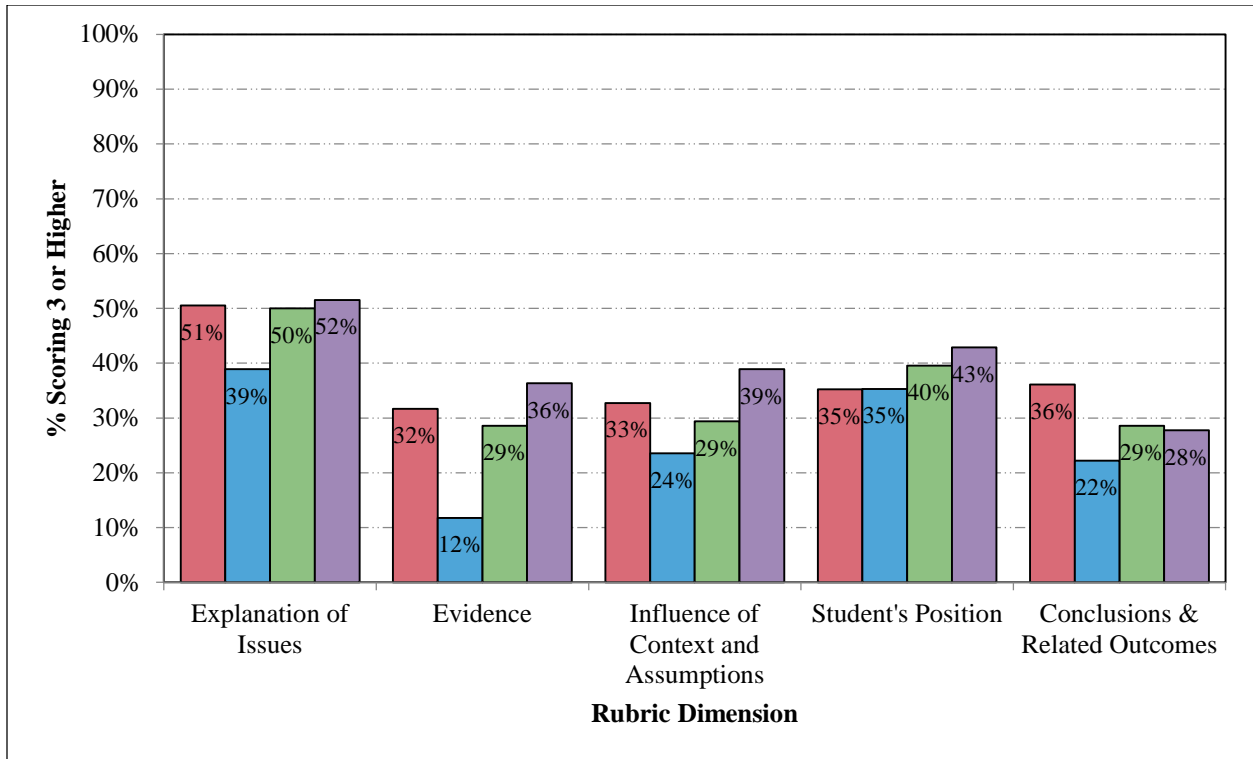


Figure 8. Comparison of CT achievement at 3 or higher across all rubric dimensions for 153 artifacts in which credit information could be matched to artifact score. 0-15 credits earned (red) n=74, 16-30 credits earned (blue) n=9, 31-60 credits earned (green) n=22, > 60 credits earned (purple) n=18. *Credits earned based on number of credits earned entering fall 2016 term.

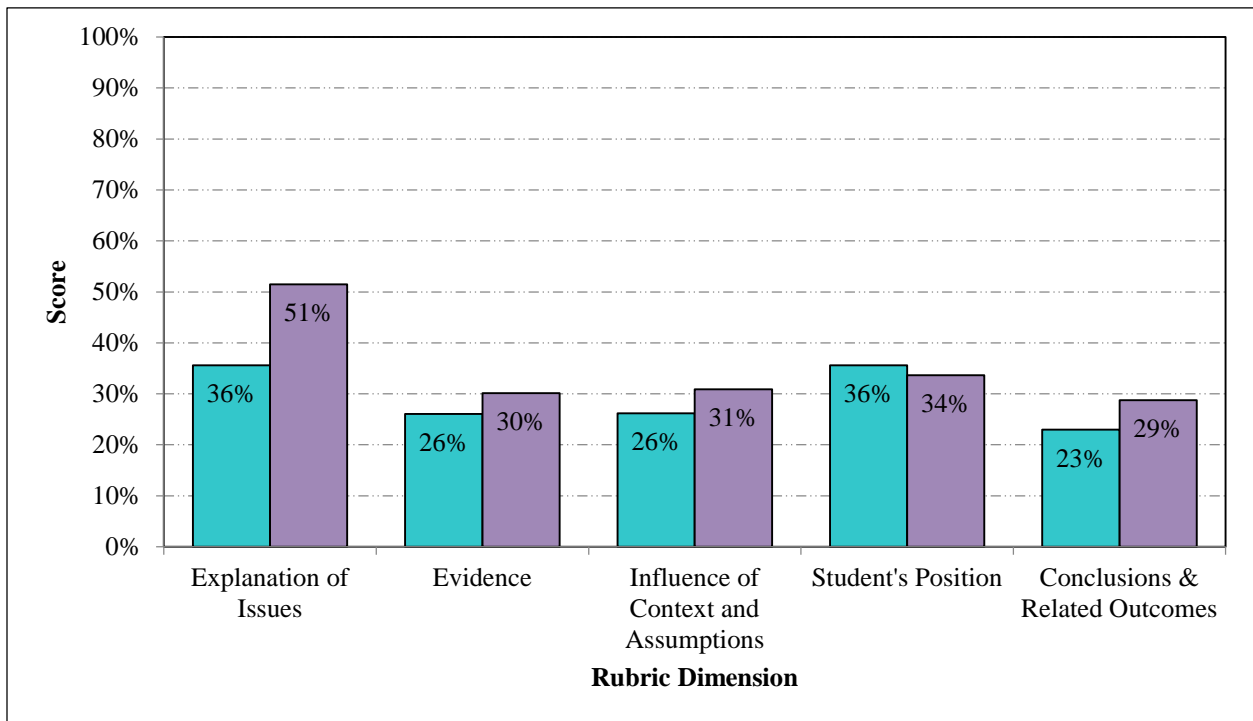


Figure 9. Comparison of CT achievement at 3 or higher across all rubric dimensions based on courses with pre-requisites. Courses with pre-requisites (aqua, n=38), courses without pre-requisites (purple, n=132).

2.2 LONGITUDINAL STUDY

The mean scores by dimension show increases from AY 2014-2015 to AY 2016-2017 in four of five rubric dimensions. These increases are likely not due to any increase in achievement but rather a solidification of the scoring procedure. For example, in AY 2014-2015, cases where assignment guidelines did not require one or more of the rubric dimensions the scorer may have scored a 0. Under new scorer protocol these assignments would simply be removed from analysis for those particular dimensions. The “Explanation of Issues” and “Evidence” dimensions, ones which were reported that some assignments did not require material, exhibit the largest increase (Figure 10). This is the clearest indicator that a change in scorer protocol is the most likely candidate for the cause of the increases. These protocol changes are also evident in scorer agreement which exhibits improvements in all but one dimension (Figure 11).

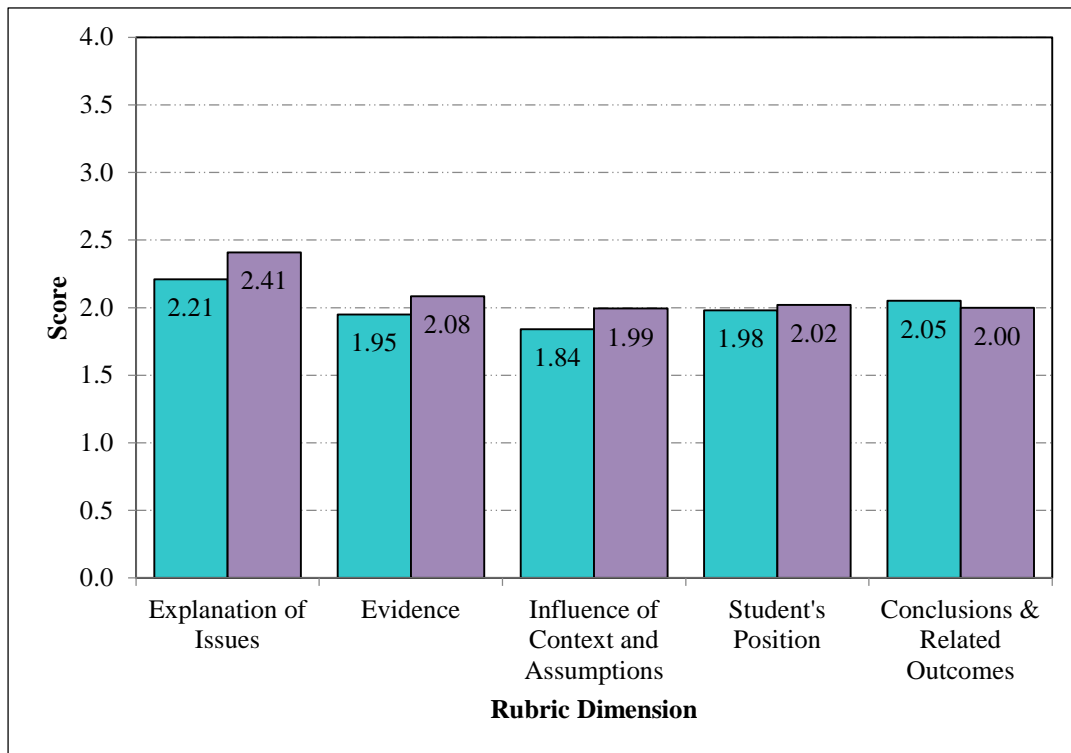


Figure 10. CT achievement over time based on mean rubric score by dimension. AY 2014-2015 (aqua), AY 2016-2017 (purple).

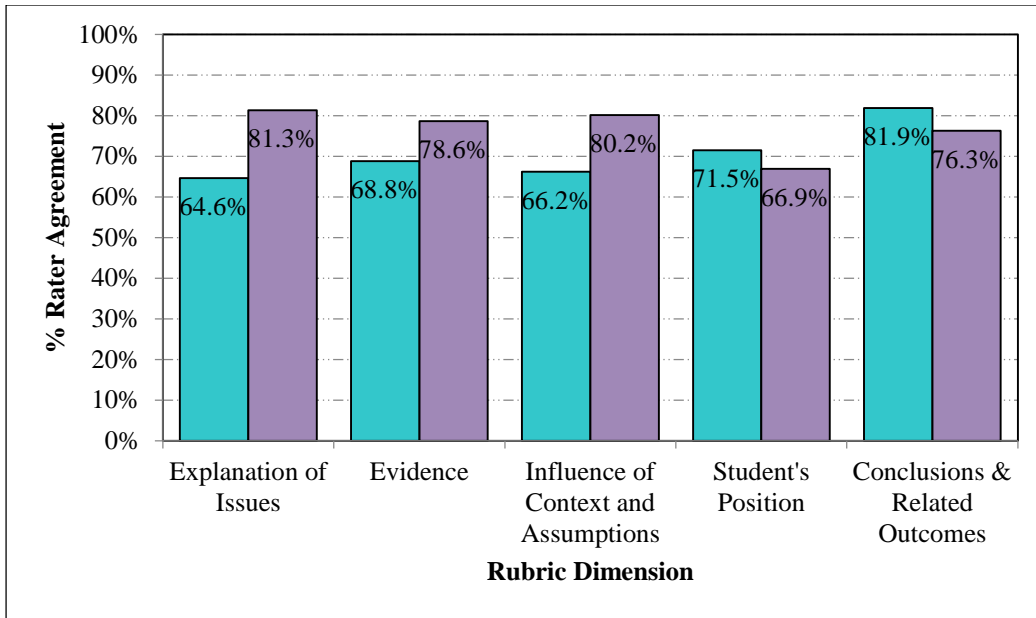


Figure 11. Inter-rater reliability (as %) for CT competency for AY 2014-2015 assessment (aqua) and AY 2016-2017 assessment (purple). Inter-rater reliability here is reported as percentage (%) +/- 1 agreement.

An additional longitudinal study comparing achievement by credits earned is shown in Figure 12. Note that in this study, the commonality is not the competency, but rather how the attribute of credits earned influences achievement in varying competencies. Here, results exhibit a consistent rise in achievement with credits earned for the COM (written). However, for CT, results initially drop in the 16-30 credits earned range before rising again. This is not uncommon as a case study by Braselton (2011) has pointed out that scores have often been initially higher for freshman-level courses than sophomore.

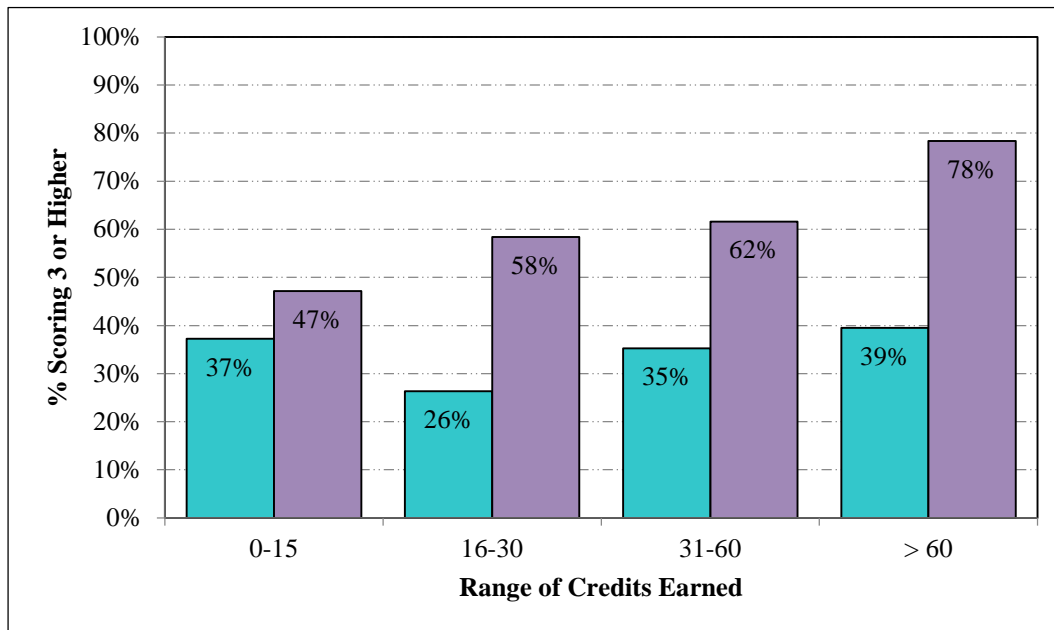


Figure 12. Comparison of CT achievement (AY 2016-2017) (aqua) and COM achievement (ANY 2015-2016) (purple) based on credits earned (average of all dimensions combined for each competency).

3 SCIENTIFIC AND QUANTITATIVE REASONING (QR)

3.1 AY 2016-2017 STUDY

The Scientific and Quantitative Reasoning (QR) competency goal is that by completion of the general education requirements, students will be able to identify and apply mathematical and scientific principles and methods. The FSW Learning Assessment Committee will measure the number of artifacts scored a 3 or higher on relevant dimensions of the rubric against the pilot results (AY 2014-2015). Figures 12 through 19 below depict achievement and inter-rater reliability for the QR competency in college-wide, Associate of Arts (AA) cohorts, as well as value-added studies.

Feedback from scorers regarding the FSW rubric for the QR competency included two main concerns. First, depending on the assignment, scorers reported difficulty in distinguishing between the “Interpretation” and “Evaluation” dimensions. And second, scorers also reported confusion between “Analysis” and “Evaluation” at times, again depending on the assignment. Scorers reported it is sometimes difficult to distinguish between these dimensions as well as that at times, assignments do not require much “Analysis” or “Evaluation” as described in the dimensions. For this study, no modifications to the FSW rubric has been made to tailor achievement levels.

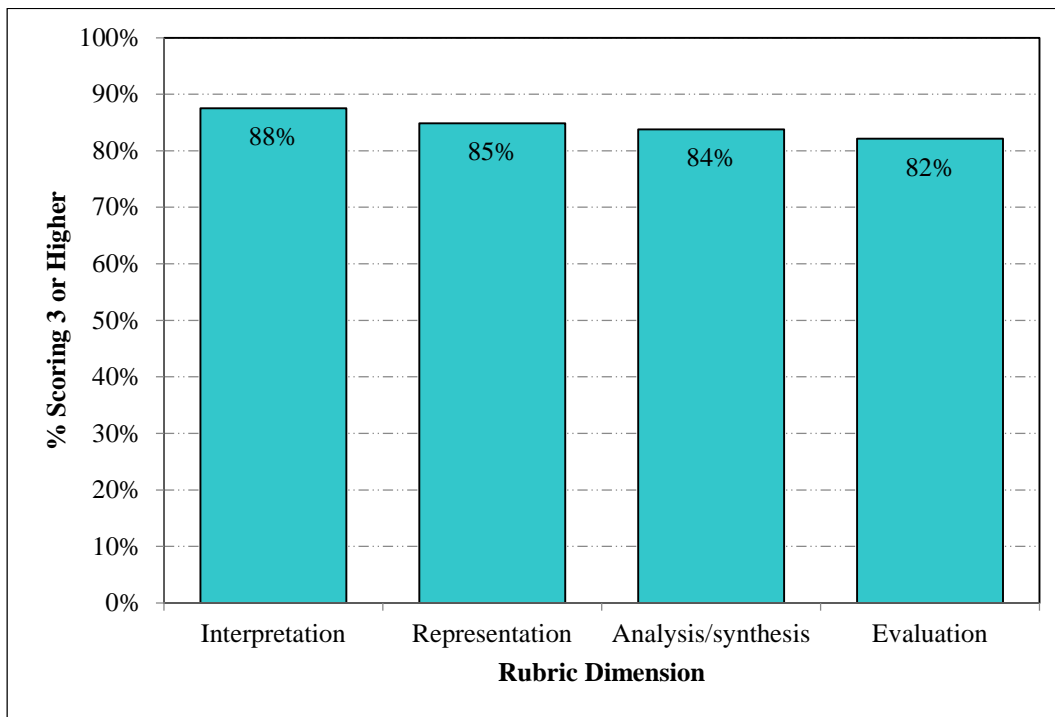


Figure 13. QR Achievement at 3 or higher across all rubric dimensions for 206 artifacts from 17 volunteered assignments.

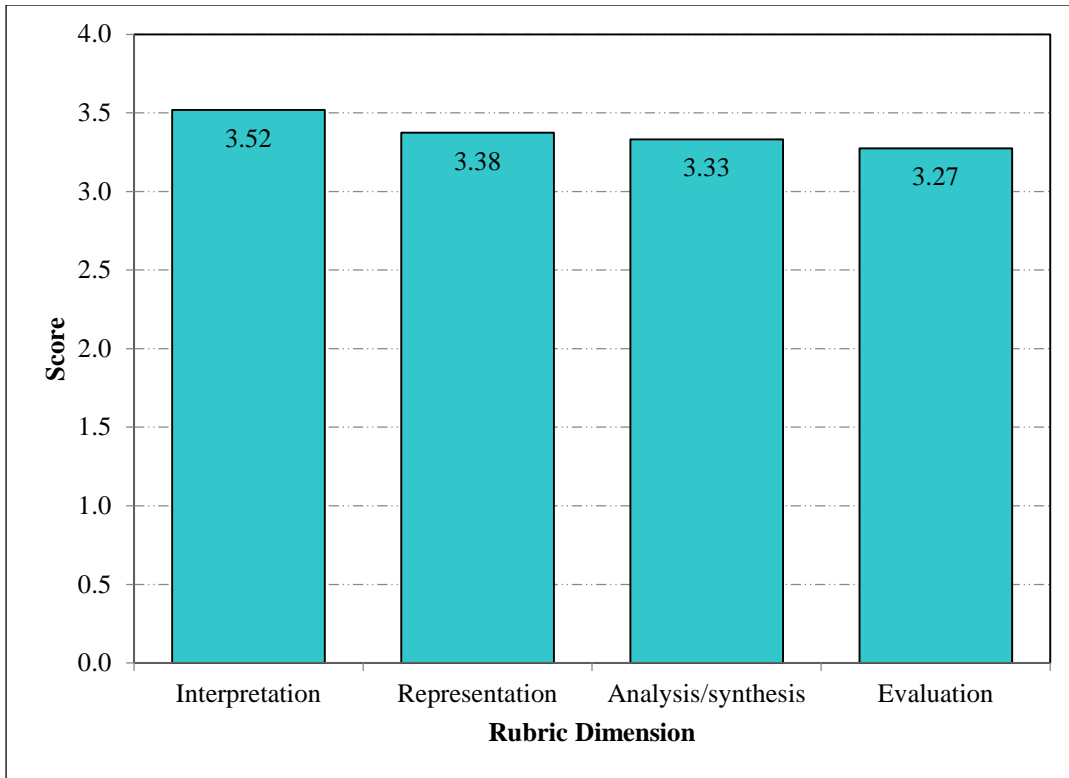


Figure 14. Mean score by rubric dimension for QR for 206 artifacts from 17 volunteered assignments.

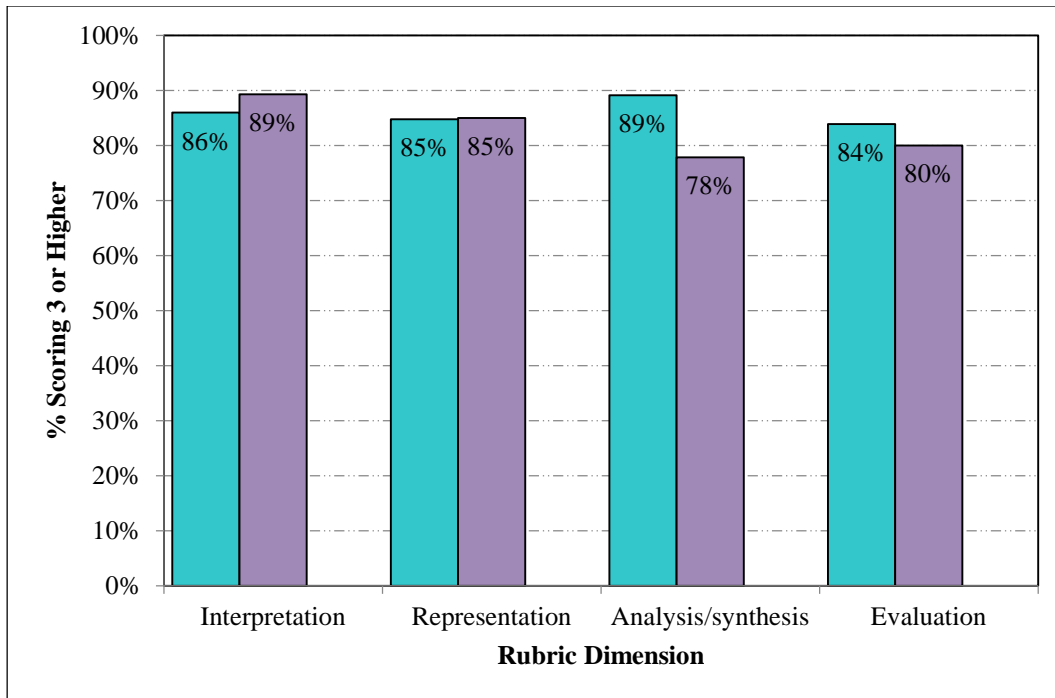


Figure 15. Comparison of QR achievement at 3 or higher across all rubric dimensions for artifacts by assignment. Traditional (aqua), n=108, Online (purple), n=98. Note that no assignments from dual enrollment course sections were volunteered for the study.

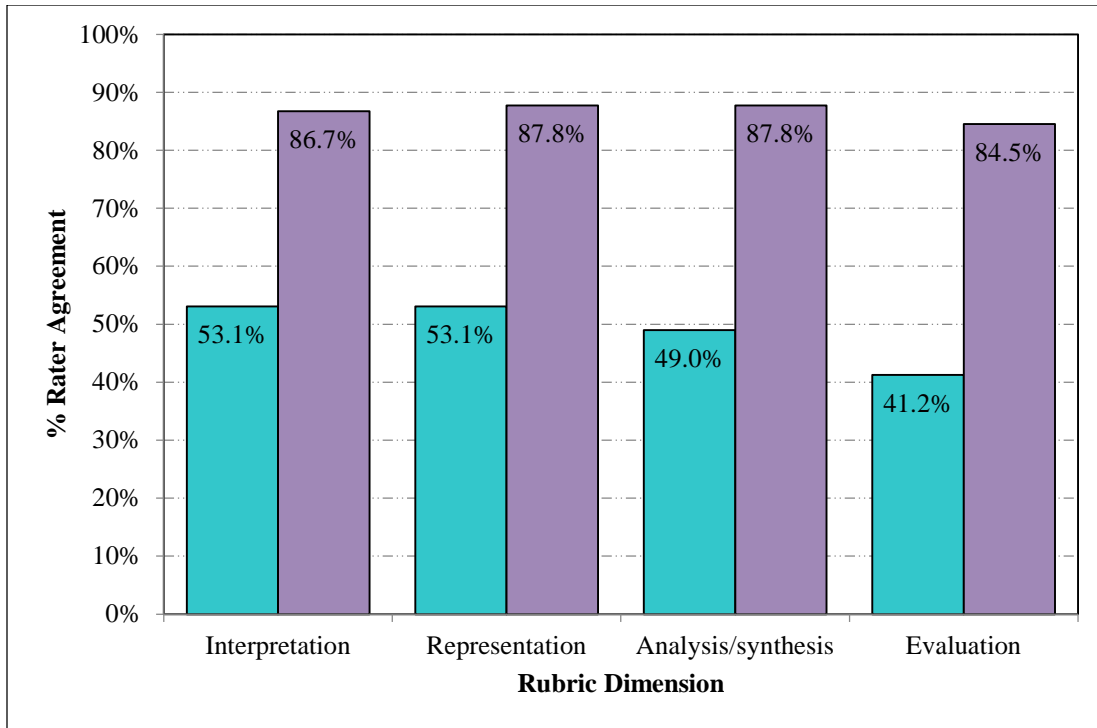


Figure 16. Inter-rater reliability (as %) for QR competency. Each artifact was scored by two scorers. Percentage (%) of agreement (aqua) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.

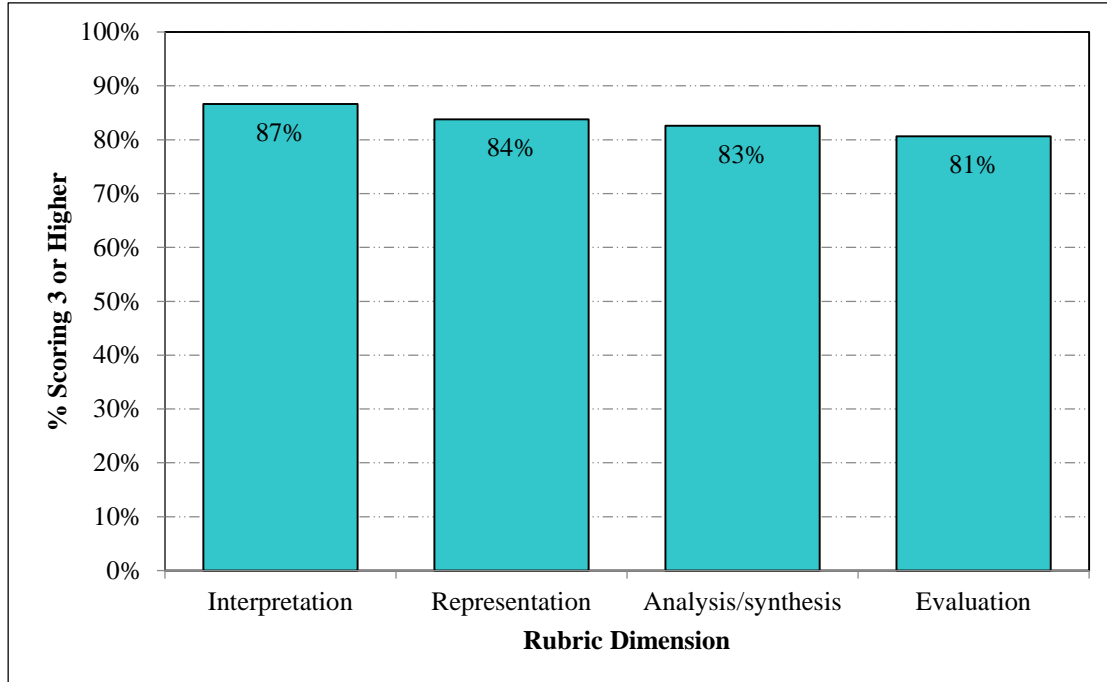


Figure 17. QR Achievement at 3 or higher for AA courses only across all rubric dimensions for 196 artifacts from 16 volunteered assignments.

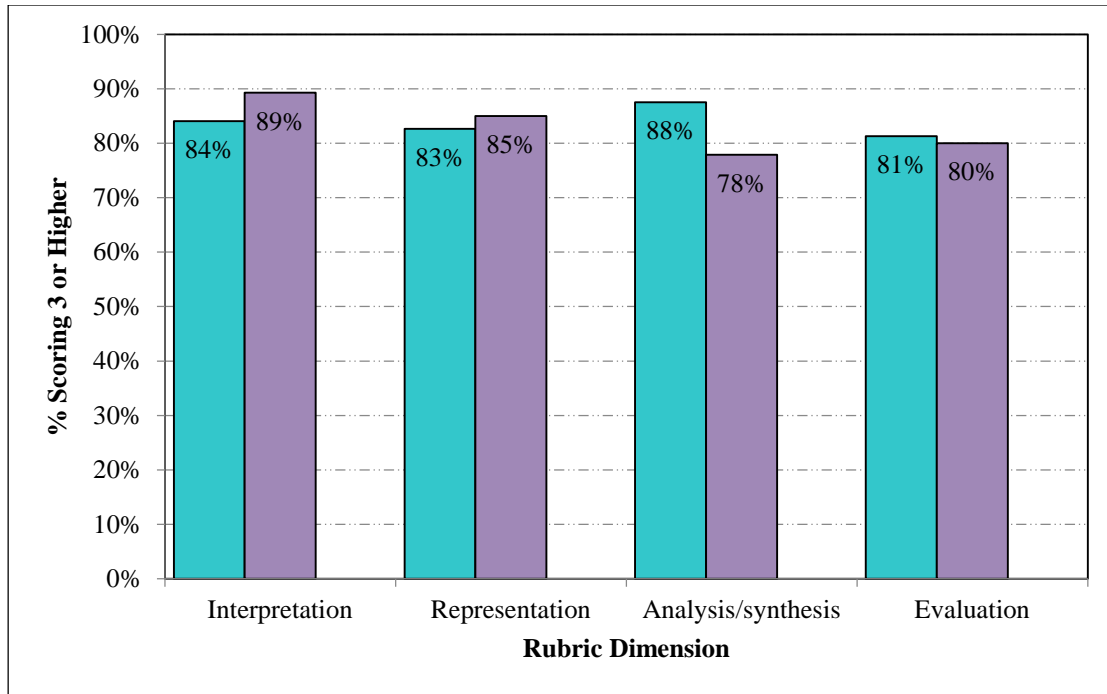


Figure 18. QR achievement at 3 or higher for AA courses only across all rubric dimensions for 196 artifacts from 16 volunteered assignments. Traditional (aqua), n=98, Online (purple), n=98. Note that no assignments from dual enrollment course sections were volunteered for the study.

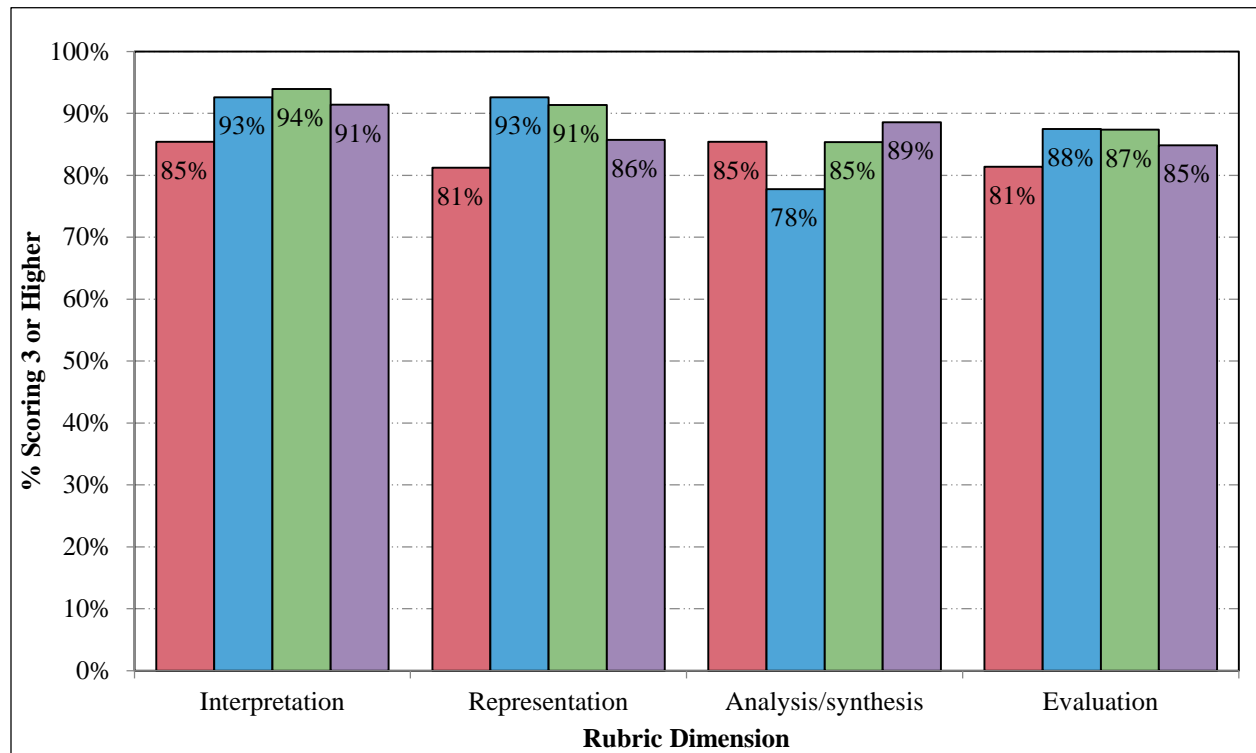


Figure 19. Comparison of QR achievement at 3 or higher across all rubric dimensions for 154 artifacts in which credit information could be matched to artifact score. 0-15 credits earned (red) n=34, 16-30 credits earned (blue) n=23, 31-60 credits earned (green) n=76, > 60 credits earned (purple) n=21. *Credits earned based on number of credits earned entering fall 2016 term.

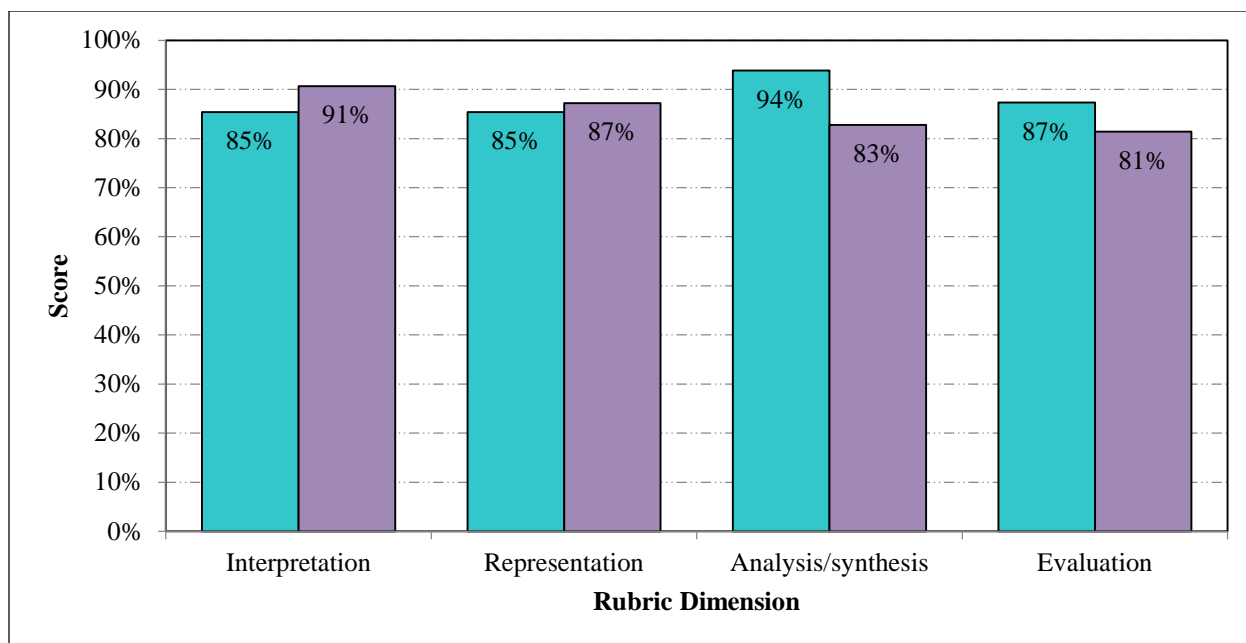


Figure 20. Comparison of QR achievement at 3 or higher across all rubric dimensions based on courses with pre-requisites. Courses with a pre-requisite (aqua, n=59), courses without pre-requisites (purple, n=147).

3.2 LONGITUDINAL STUDY

The mean scores by dimension show increases from AY 2014-2015 to AY 2016-2017 in all four rubric dimensions (Figure 21). In cases where improved scoring procedure may have helped to reduce ambiguities among scoring (“Analysis/Synthesis” and “Evaluation”) increases were present, but not substantial. To be clear, under new scorer protocol these assignments would simply be removed from analysis for those particular dimensions rather than potentially scored ‘0’. In point of fact, inter-rater reliability actually decreased, though not substantially, in three of four dimensions, including those with potential ambiguities based on scorer feedback (Figure 22).

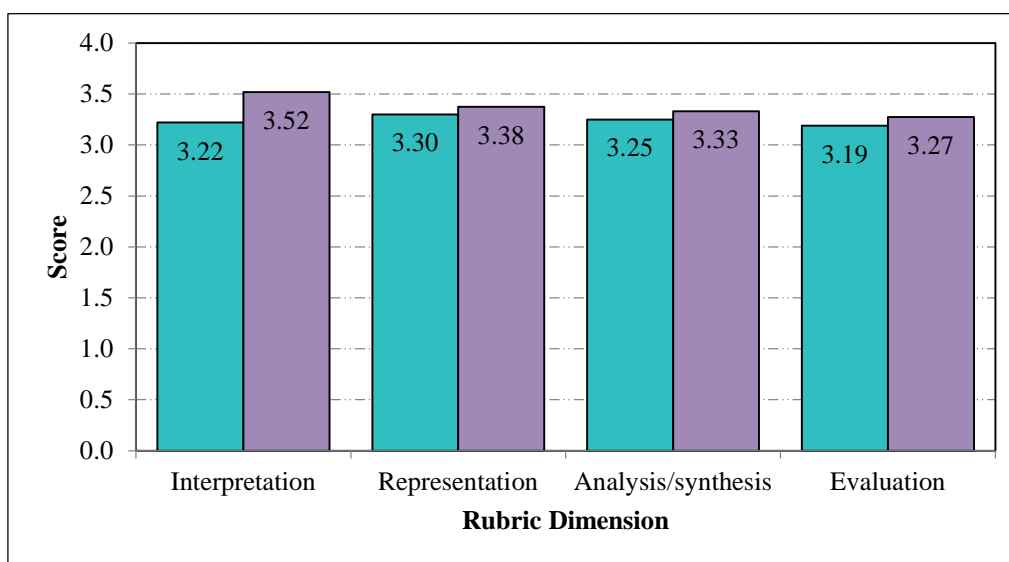


Figure 21. QR achievement over time based on mean rubric score by dimension. AY 2014-2015 (aqua), AY 2016-2017 (purple).

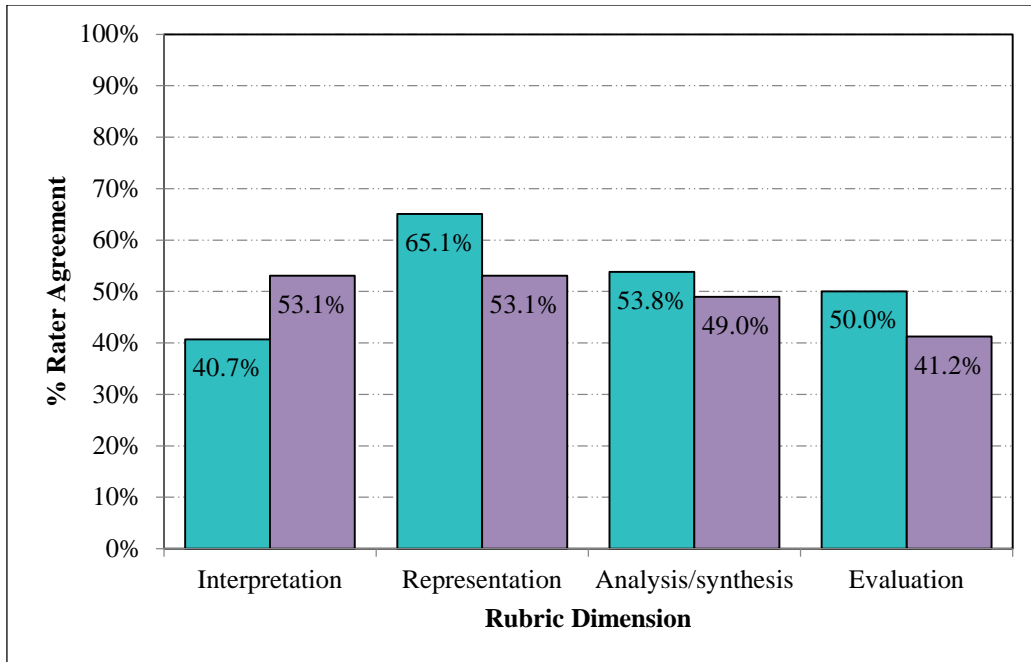


Figure 22. Inter-rater reliability (as %) for QR competency for AY 2014-2015 assessment (aqua) and AY 2016-2017 assessment (purple). Inter-rater reliability here is reported as percentage (%) +/- 1 agreement.

An additional longitudinal study comparing achievement by credits earned is shown in Figure 23. Note that in this study, the commonality is not the competency, but rather how the attribute of credits earned influences achievement in varying competencies. Here, results exhibit a consistent rise in achievement with credits earned for the COM (written). However, for QR, results exhibit a peak at 31-60 credits before dropping slightly for those artifacts originating from students with above 60 credits. This is again, not uncommon as a case study by Braselton (2011) has pointed out that scores have often been initially higher for freshman-level courses than sophomore.

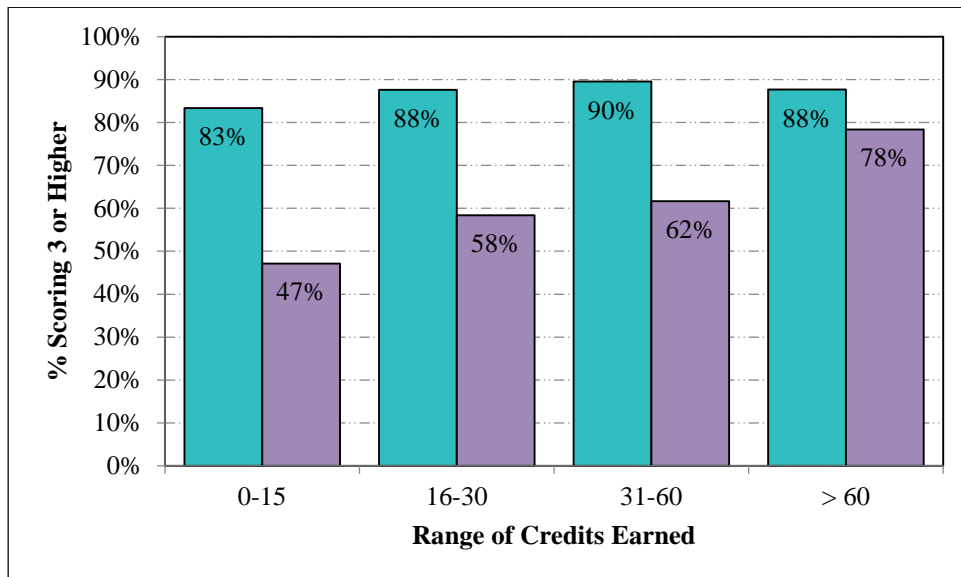


Figure 23. Comparison of QR achievement (AY 2016-2017) (aqua) and COM achievement (ANY 2015-2016) (purple) based on credits earned (average of all dimensions combined for each competency).

4 CONCLUSIONS

FSW's General Education Program was assessed through voluntary submission of assignments identified by faculty as fitting the criteria of both the Written and Oral Communication competency. The study employed the use of the Association of American Colleges & Universities (AAC&U) rubric for Critical Thinking along with a QR rubric developed by FSW. The study continued the study for the CT and QR from the pilot study in AY 2014-2015 for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric. Results also included these same outcomes with respect to courses included in the AA program and value-added studies based on credits earned and pre-requisites.

A drilldown of Critical Thinking (CT) results are as follows:

1. Zero of five rubric dimensions exhibit greater than 60% achievement at level '3'. The highest scored dimension is "Explanation of Issues" at 47% scoring '3' or higher.
2. Mean achievement levels for each of the five rubric dimensions of CT range from 1.99 to 2.41 on a 4-point scale.
3. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 23% to 51% (n = 81/38/51, traditional/online/dual enrollment). The traditional modality exhibits the highest in three of five dimensions. The dual enrollment modality exhibits the highest in the remaining two. No differences across modalities were statistically significantly different.
4. An inter-rater reliability study exhibits rubric scoring agreement ranging from 31% to 39% with a +/- 1 agreement ranging from 67% to 81%.
5. With respect to AA courses, zero of five rubric dimensions exhibit greater than 60% achievement at level '3'. The highest scored dimension is "Explanation of Issues" at 47% scoring '3' or higher.
6. In a study comparing AA courses with online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 22% to 50% (n = 65/22/51, traditional/online/dual enrollment). The traditional modality exhibits the highest in three of five dimensions. The dual enrollment modality exhibits the highest in the remaining two. No differences across modalities were statistically significantly different.
7. In a study comparing achievement at 3 or higher across rubric dimensions based on credits earned, achievement increases with increasing credits earned consistently for "Student's Position". The other dimensions exhibit some variation which is consistent with other case studies.
8. In a study comparing achievement at 3 or higher across rubric dimensions based on pre-requisites, in one of five dimensions artifacts stemming from courses which had pre-requisites exhibit a higher achievement. This pattern is consistent with AY 2015-2016's COM (written) study.
9. In a review of scorer feedback, multiple scorers noted that not all CT assignments require sources. The varying interpretation of this dimension based on the incorporation of it in the assignment caused concerns over consistent scoring habits. Second, scorers noted that the rubric was better suited for written assignments and that mathematical assignments was difficult to apply.
10. In a longitudinal study comparing mean rubric scores over time, mean scores by dimension show increases from AY 2014-2015 to AY 2016-2017 in four of five rubric dimensions. These

increases are likely not due to any increase in achievement but rather a solidification of the scoring procedure.

11. In a longitudinal study comparing achievement by credits earned, results exhibit a consistent rise in achievement with credits earned for the COM (written). However, for CT, results initially drop in the 16-30 credits earned range before rising again.

A drilldown of Scientific and Quantitative Reasoning (QR) results are as follows:

1. Four of four rubric dimensions exhibit greater than 60% achievement at level '3' with percentages ranging from 82% to 88%.
2. Mean achievement levels for each of the four rubric dimensions of QR range from 3.27 to 3.52 on a 4-point scale.
3. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 78% to 89% (n = 108/98/0, traditional/online/dual enrollment). The traditional modality exhibits the highest in two of four dimensions. The online modality exhibits the highest in the remaining two. Note that no assignments from dual enrollment course sections were volunteered for the study. No differences across modalities were statistically significantly different.
4. An inter-rater reliability study exhibits rubric scoring agreement ranging from 41% to 53% with a +/- 1 agreement ranging from 85% to 88%.
5. With respect to AA courses, four of four rubric dimensions exhibit greater than 60% achievement at level '3' with percentages ranging from 81% to 87%.
6. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 78% to 89% (n = 98/98/0, traditional/online/dual enrollment). The traditional modality exhibits the highest in two of four dimensions. The online modality exhibits the highest in the remaining two. Note that no assignments from dual enrollment course sections were volunteered for the study. No differences across modalities were statistically significantly different.
7. In a study comparing achievement at 3 or higher across rubric dimensions based on credits earned, achievement increases and peaks before declining again at higher credit counts. This is the case for all but "Analysis/Synthesis", which exhibit some variation which is consistent with other case studies.
8. In a study comparing achievement at 3 or higher across rubric dimensions based on pre-requisites, in two of four dimensions artifacts stemming from courses which had pre-requisites exhibit a higher achievement.
9. In a review of scorer feedback, scorers noted that depending on the assignment, scorers reported difficulty in distinguishing between the "Interpretation" and "Evaluation" dimensions. And second, scorers also reported confusion between "Analysis" and "Evaluation" at times, again depending on the assignment.
10. In a longitudinal study comparing mean rubric scores over time, mean scores by dimension show increases from AY 2014-2015 to AY 2016-2017 in all four rubric dimensions. In cases where improved scoring procedure may have helped to reduce ambiguities among scoring ("Analysis/Synthesis" and "Evaluation") increases were present, but not substantial. To be clear, under new scorer protocol these assignments would simply be removed from analysis for those particular dimensions rather than potentially scored '0'. In point of fact, inter-rater reliability

actually decreased, though not substantially, in three of four dimensions, including those with potential ambiguities based on scorer feedback.

11. In a longitudinal study comparing achievement by credits earned, results exhibit a consistent rise in achievement with credits earned for the COM (written). However, for QR, results exhibit a peak at 31-60 credits before dropping slightly for those artifacts originating from students with above 60 credits.

5 REFERENCES

- Braselton, M. "Using AAC&U Value Rubrics. Case Study: Midland College." <https://www.aacu.org/VALUE/casestudies/midland.pdf>
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. XX(1), 37-46.
- Florida SouthWestern State College, 2014, General Education Assessment Subcommittee Meeting Minutes, <https://docs.fsw.edu/docnew/view.php?fDocumentId=298910>.
- Gwet, K. 2002. Inter-rater reliability: Dependency on trait prevalence and marginal homogeneity. *Series: Statistical methods for Inter-Rater Reliability Assessment*, No. 2, 1-9.
- Rhodes, T. & Finley, A. (2013) Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment. AAC & U.