General Education Assessment Report – AY 2014-15

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

1 INTRODUCTION

This report details the results of Florida SouthWestern State College's General Education pilot assessment for AY 2014-2015. 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).

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 all but one competency (Scientific and Quantitative Reasoning, QR) in the assessment of artifacts. 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 pilot program is to: 1) establish a baseline for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric, 2) establish a baseline for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric across sites (Online, Dual Enrollment, and Traditional), and 3) 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 Academic Assessment, Academic Affairs (jfvangaalen@fsw.edu; x16965).

2 COMMUNICATION (COM)

The COM competency goal is by completion of the general education requirements, students will be able to communicate effectively using standard English (written or oral). The FSW Learning Assessment Committee will use the pilot results to establish a baseline for the number of artifacts scored a 3 or higher on relevant dimensions of the rubric. Figures 1 through 9 below depict achievement and inter-rater reliability for the COM competency in college-wide and Associate of Arts (AA) cohorts.

Feedback from scorers regarding the AAC&U Value rubric for the COM competency included two main trends regarding rubric suitability. First, multiple scorers noted that "Sources and Evidence" is not

relevant in all COM assignments. The varying interpretation of this dimension based on the incorporation of it in the assignment caused concerns over consistent scoring habits. For example, scorers noted in some cases removal of the rubric dimension might be warranted for the assignment while in others a best approximation of the dimension based on the assignment may be most useful. Second, scorers noted the use of "some" and "few" at differing achievement levels when interpreting "Control of Syntax and Mechanics" is at times very difficult to distinguish.
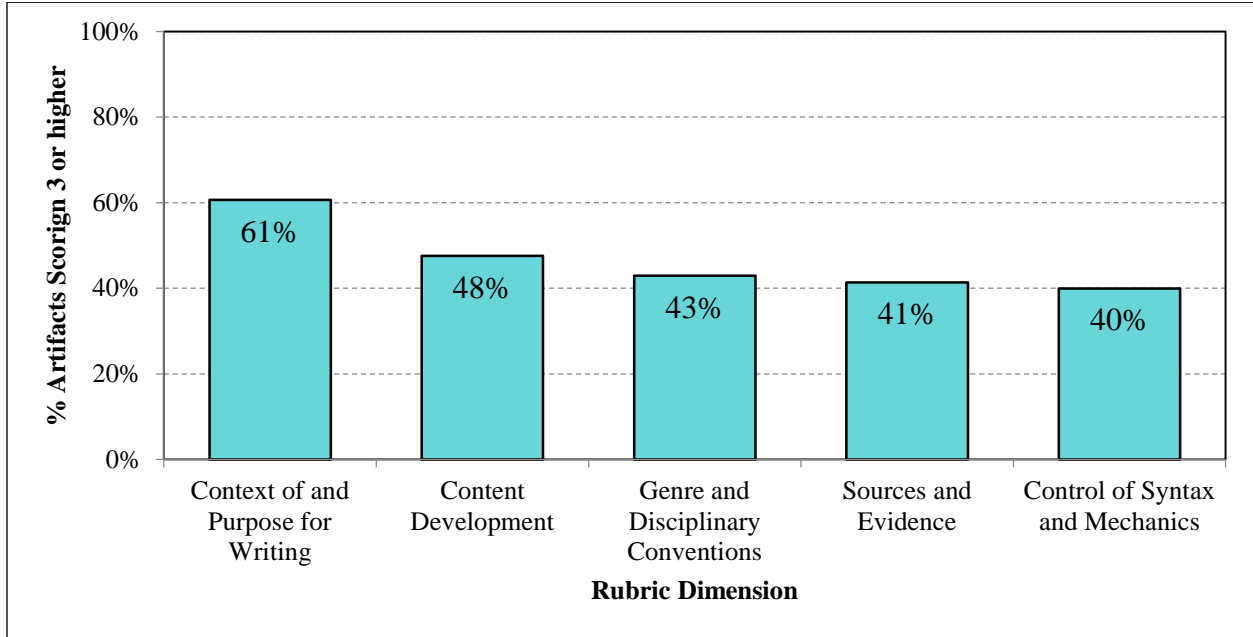
Figure 1. COM Achievement at 3 or higher across all rubric dimensions for 220 artifacts from 25 volunteered assignments.
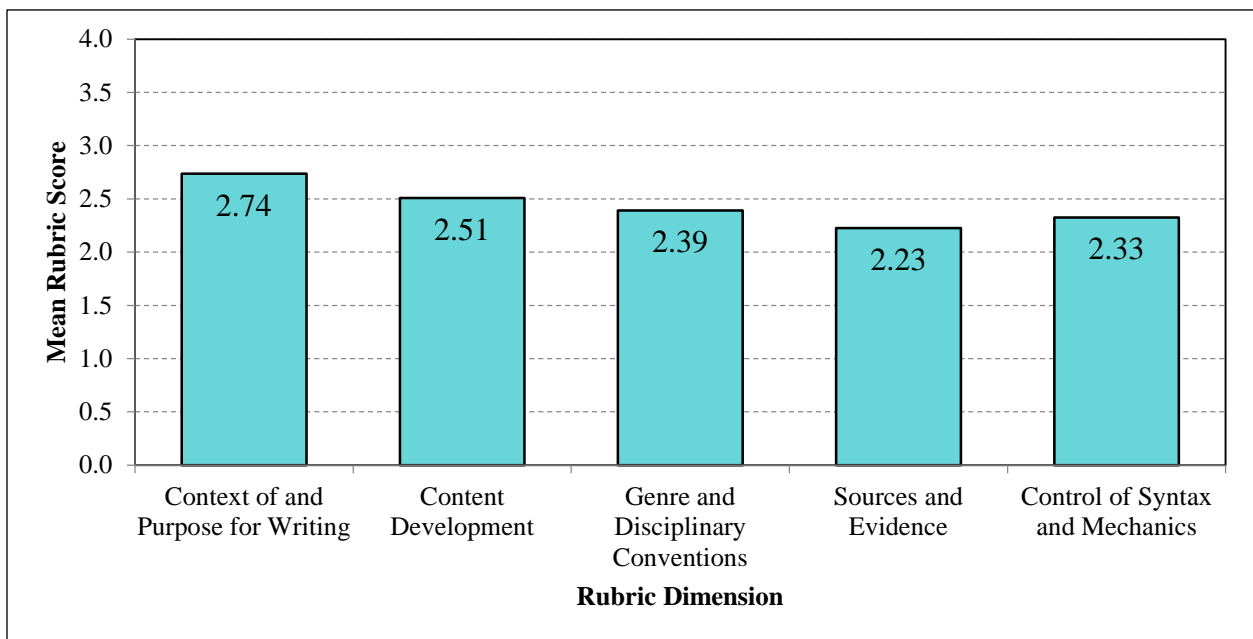

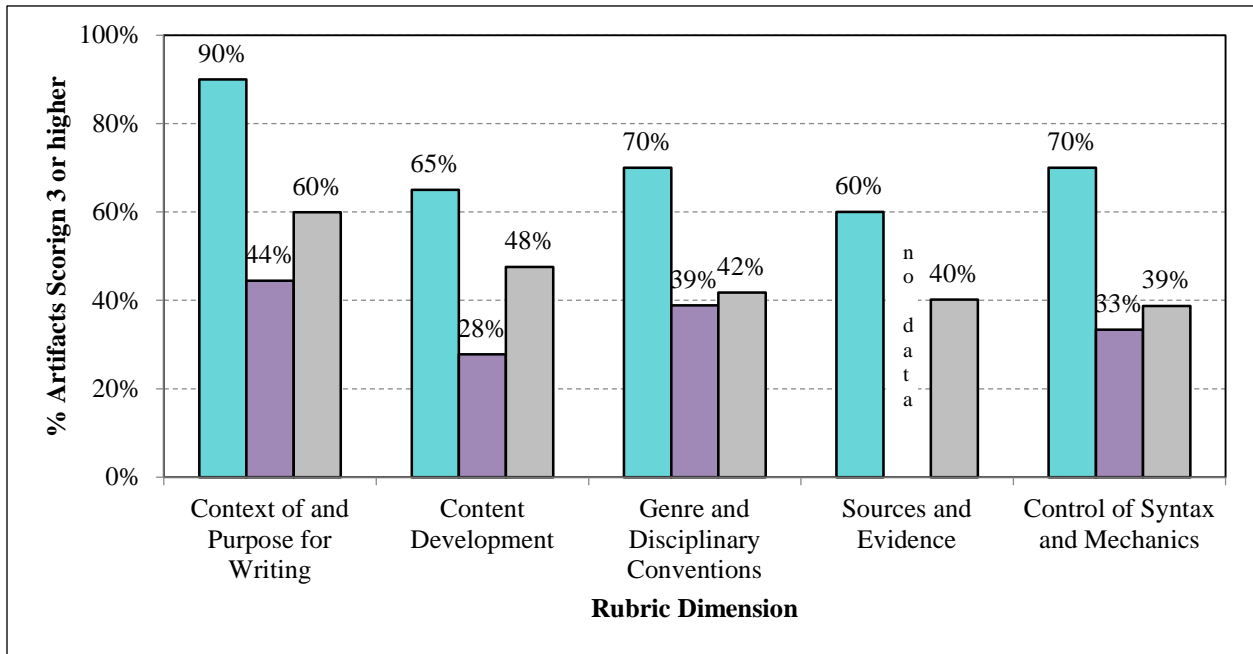*Figure 2. Mean score by rubric dimension for COM for 220 artifacts from 25 volunteered assignments.*

Figure 3. Comparison of COM achievement at 3 or higher across all rubric dimensions for 220 artifacts from 25 volunteered assignments. FSW Online (teal), n=10, Dual Enrollment (purple), n=9, Traditional (gray), n=201.
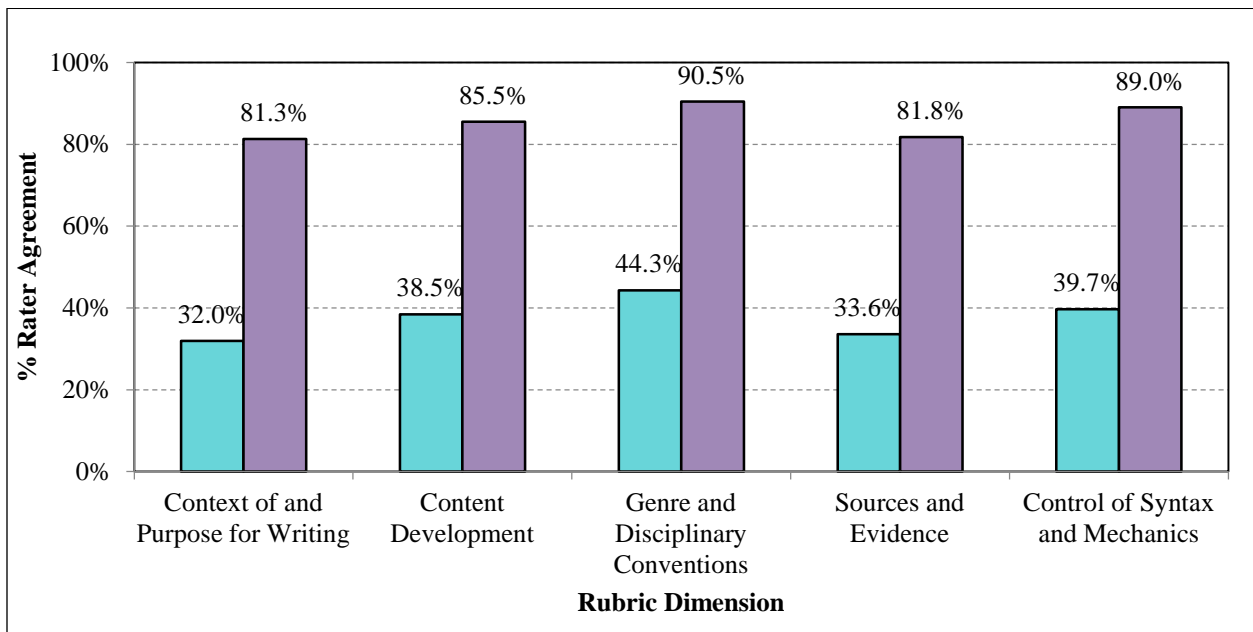

Figure 4. Inter-rater reliability (as %) for COM competency. Each artifact was scored by two scorers. Percentage (%) of agreement (teal) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.
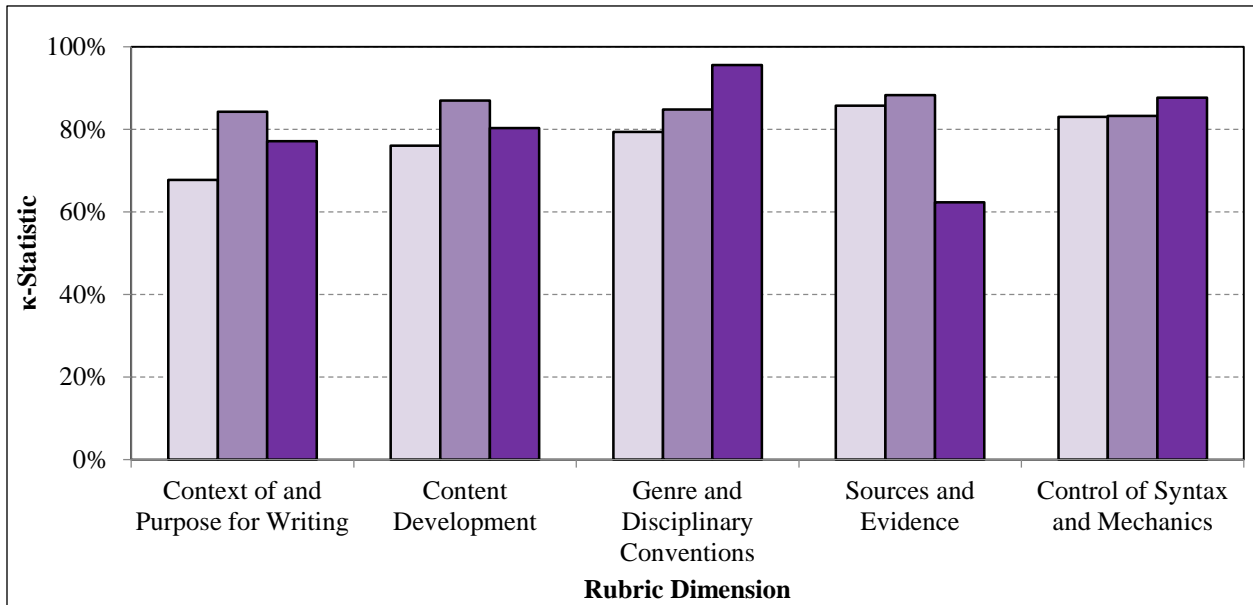
Figure 5. κ-statistic for scorer pairs of the COM competency artifacts (Scorers 1A & 1B – light purple, Scorers 2A & 2B – moderate purple, Scorers 3A and 3B – dark purple). The κ-statistic is evaluated for the percentage (%) +/- 1 agreement shown in Figure 4 above and takes into account the random chance 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 only calculated for the COM study and used as a guide to interpret percentage agreement in an academic sense.
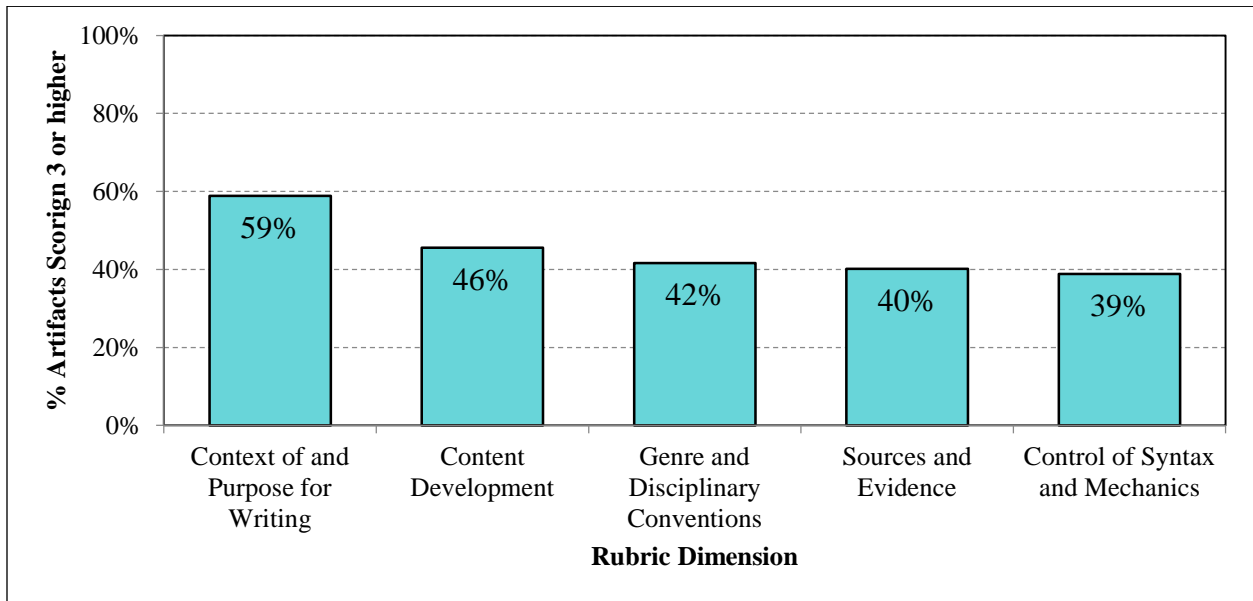

Figure 6. COM Achievement at 3 or higher for AA courses only across all rubric dimensions for 214 artifacts from 23 volunteered assignments.
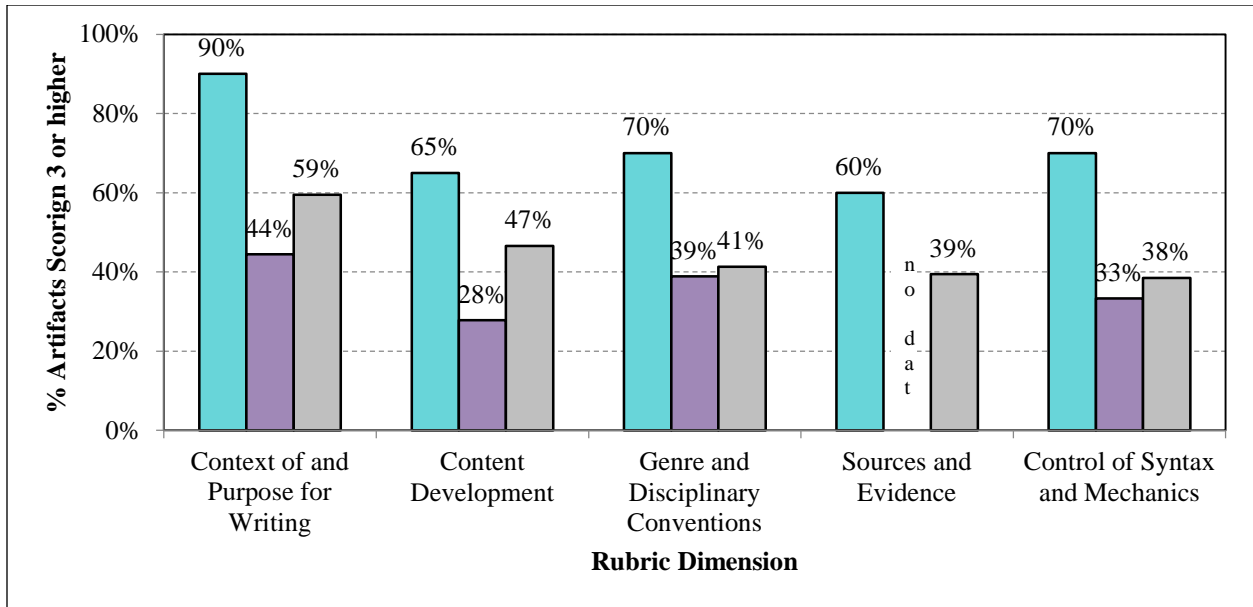
Figure 7. Comparison of COM achievement at 3 or higher for AA courses only across all rubric dimensions for 214 artifacts from 23 volunteered assignments. FSW Online (teal), n=10, Dual Enrollment (purple), n=9, Traditional (gray), n=195.
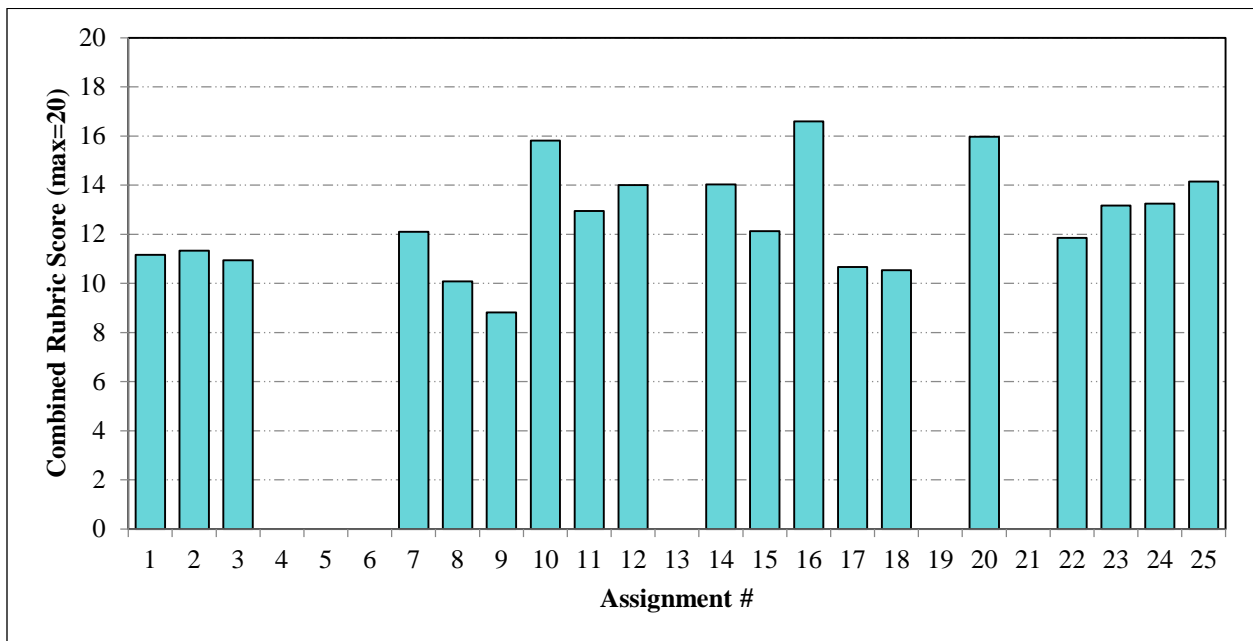

Figure 8. Overall rubric score (cumulative score of all five COM dimensions) by assignment. In cases where one or more dimensions was left unscored an overall score would not be a comparable statistic and is therefore left blank in this representation. Assignments: 1) AML 2010, 2) POS 2041, 3) AMH 2020, 4) DEP 2004, 5) PLA 1003, 6) MUL 1110, 7) WHO 1030, 8) BSC 1010, 9) DEP 2004, 10) SPC 1017, 11) PHI 2010, 12) DEP 2004, 13) REA 0019, 14) AML 2020, 15) PHI 2010, 16) ENC 1102, 17) AMH 2010, 18) SPC 1017, 19) HUM 2235, 20) RED 4519, 21) SLS 1515, 22) PHI 2600, 23) ENC 1102, 24) MCB 2010, 25) WHO 1030.
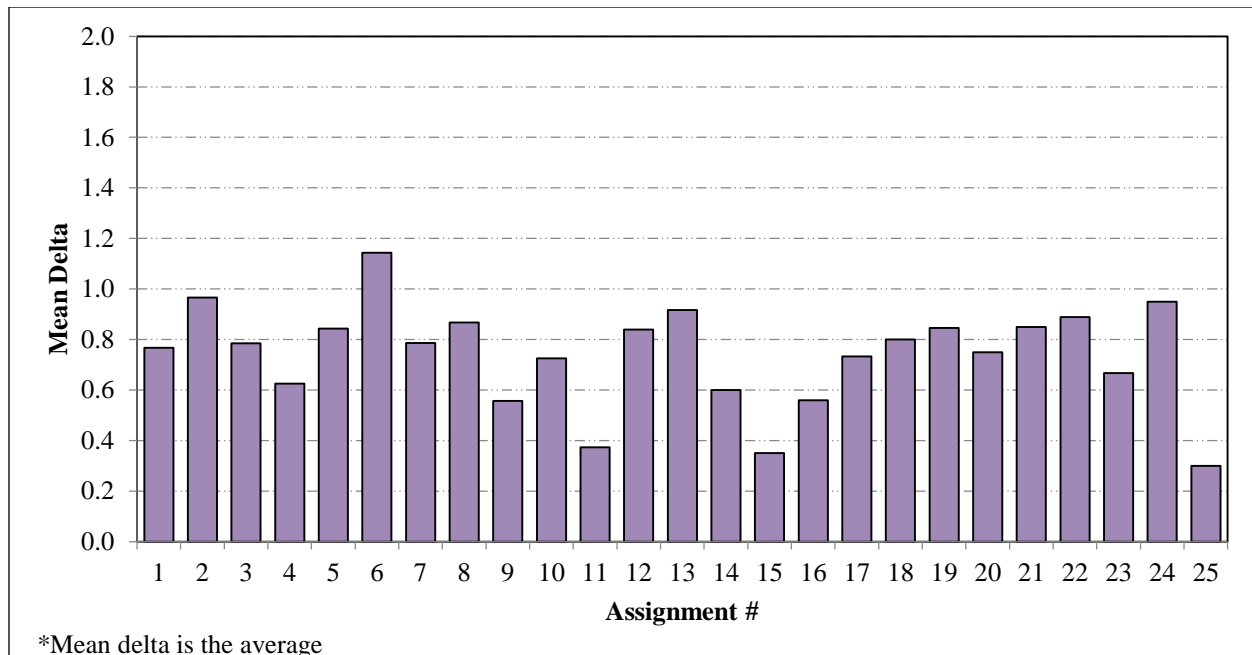
Figure 9. Inter-rater reliability of COM scoring by assignment using mean difference in scoring per dimension by assignment. Example: A mean delta of 1.0 means that, on average, the scoring pair disagrees by 1 achievement level for each rubric dimension scored. Assignments: 1) AML 2010, 2) POS 2041, 3) AMH 2020, 4) DEP 2004, 5) PLA 1003, 6) MUL 1110, 7) WHO 1030, 8) BSC 1010, 9) DEP 2004, 10) SPC 1017, 11) PHI 2010, 12) DEP 2004, 13) REA 0019, 14) AML 2020, 15) PHI 2010, 16) ENC 1102, 17) AMH 2010, 18) SPC 1017, 19) HUM 2235, 20) RED 4519, 21) SLS 1515, 22) PHI 2600, 23) ENC 1102, 24) MCB 2010, 25) WHO 1030.

3   CRITICAL THINKING (CT)

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 use the pilot results to establish a baseline for the number of artifacts scored a 3 or higher on relevant dimensions of the rubric. Figures 10 through 17 below depict achievement and inter-rater reliability for the CT competency in college-wide and Associate of Arts (AA) cohorts.

Feedback from scorers regarding the AAC&U Value rubric for the CT competency included several trends concerning rubric suitability. With regard to clarity in achievement, scorers noted the rubric seemed to require greater distinction between '2' and '3' and less distinction between '3' and '4'. Scorers described level '4' as being far more detailed than '3' and '3' being not all that different than '2'. Other concerns were with regard to determining achievement across varied assignments types. Some scorers felt the rubric was not fluid enough to accommodate the varied types of CT assignments produced in the voluntary submission, a result common to other studies using the AAC & U Value rubrics (Braselton, 2011). These types of concerns may or may not have gone unnoticed for some assignments meaning results may be affected.
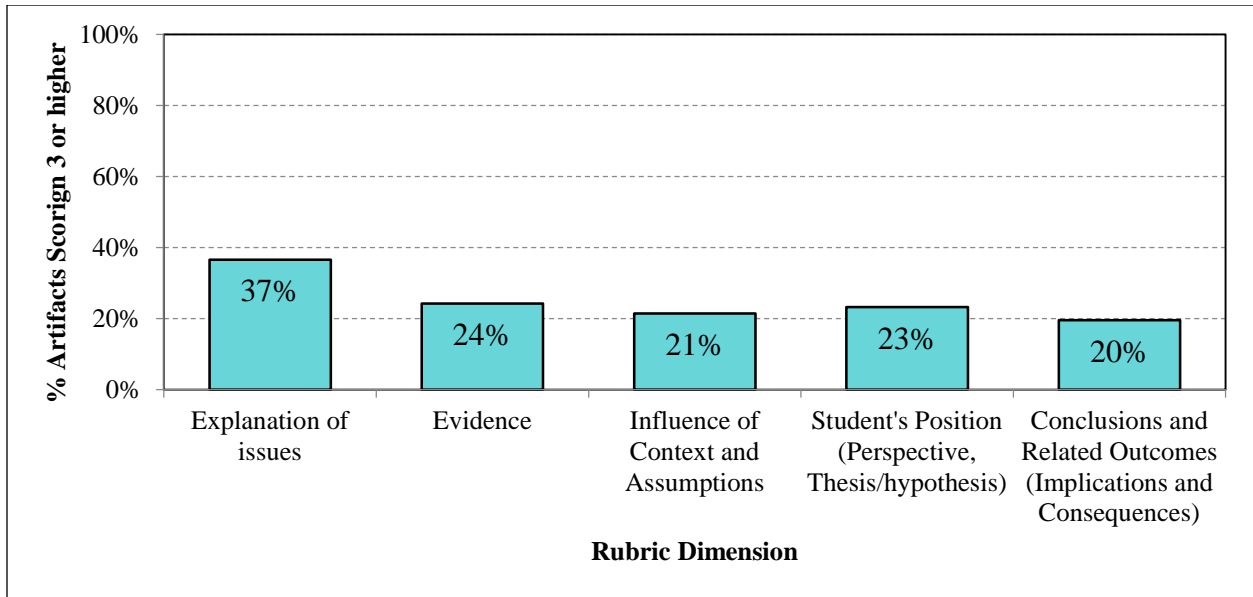
Figure 10. CT Achievement at 3 or higher across all rubric dimensions for 272 artifacts from 26 volunteered assignments.
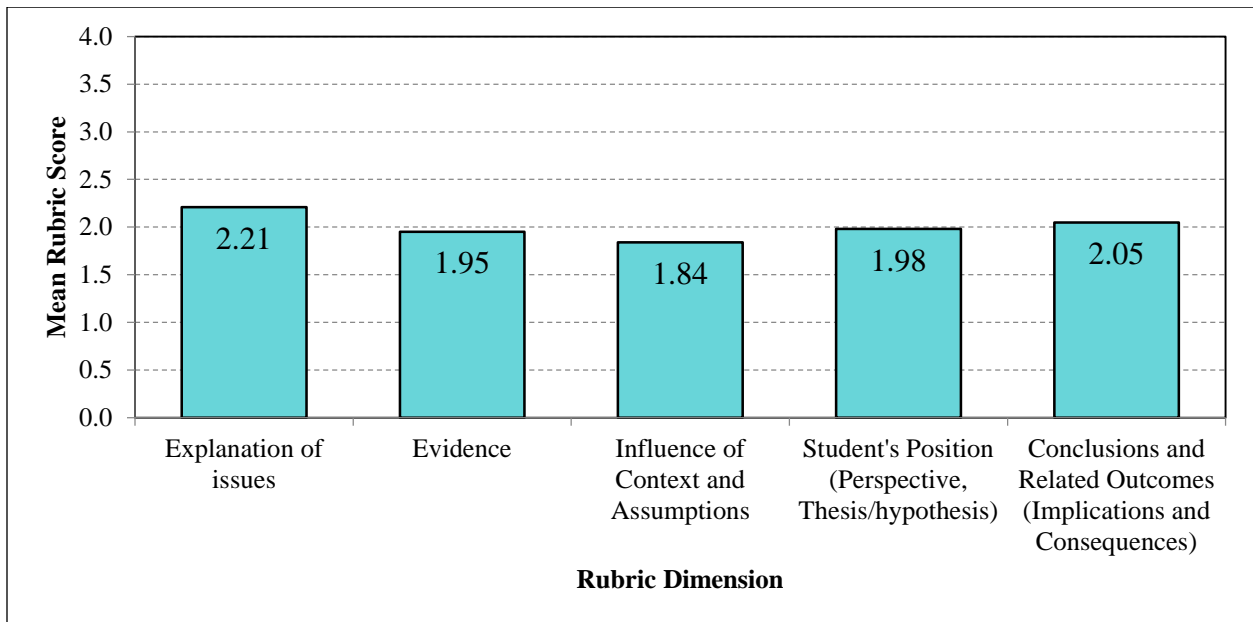

Figure 11. Mean score by rubric dimension for CT for 272 artifacts from 26 volunteered assignments.
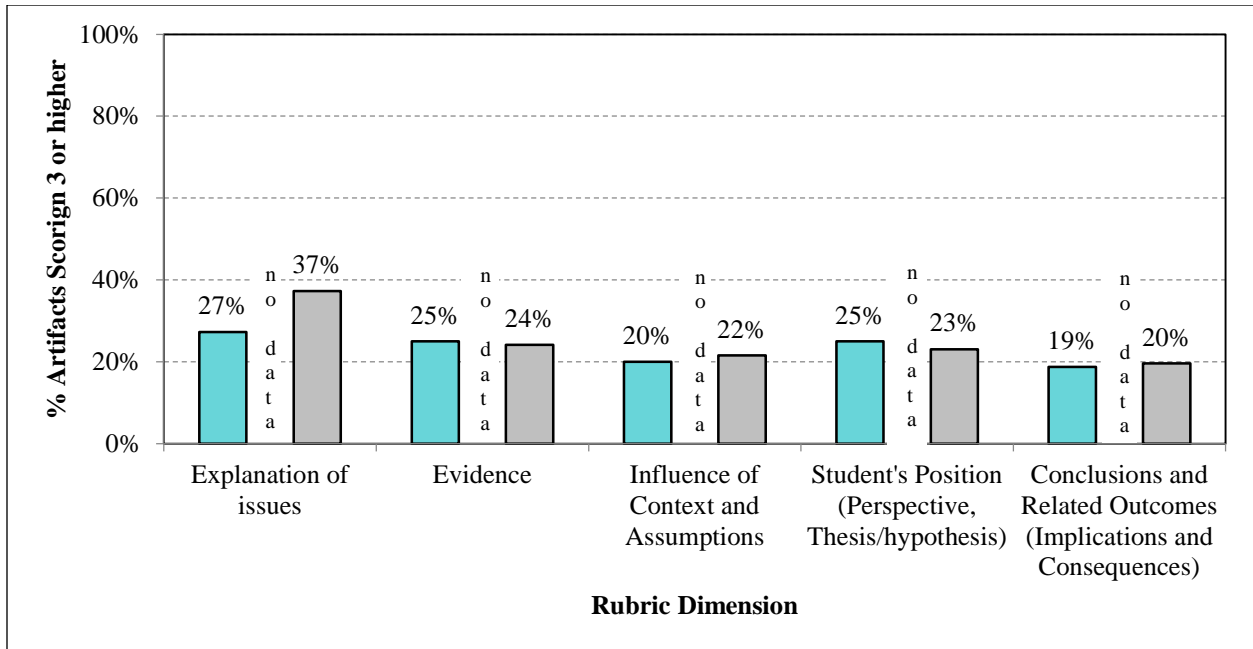
Figure 12. Comparison of CT achievement at 3 or higher across all rubric dimensions for 272 artifacts from 26 volunteered assignments. FSW Online (teal), n=16, Dual Enrollment (purple), n=0, Traditional (gray), n=256.
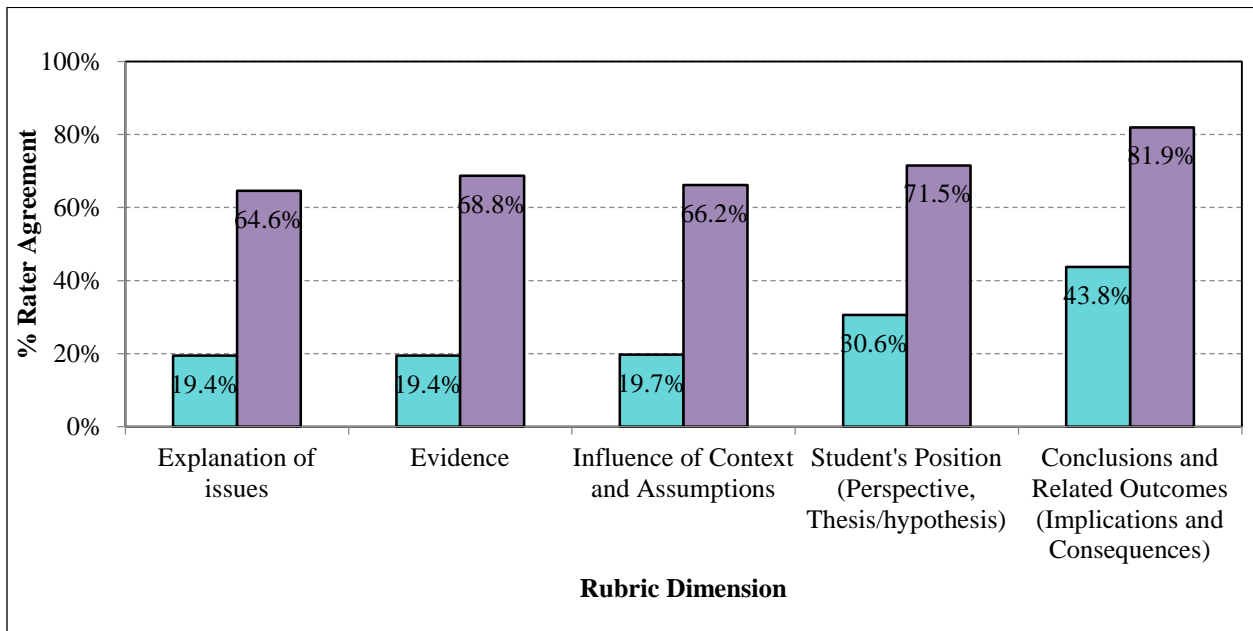

Figure 13. Inter-rater reliability (as %) for CT competency. Each artifact was scored by two scorers. Percentage (%) of agreement (teal) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.
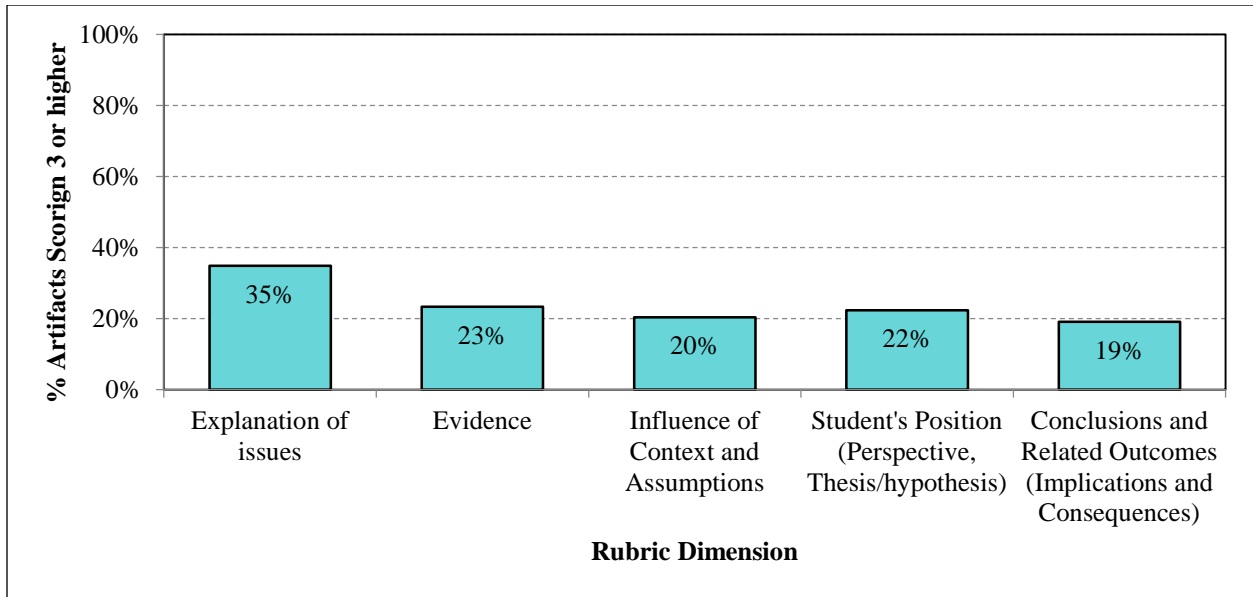
Figure 14. CT Achievement at 3 or higher for AA courses only across all rubric dimensions for 262 artifacts from 23 volunteered assignments.
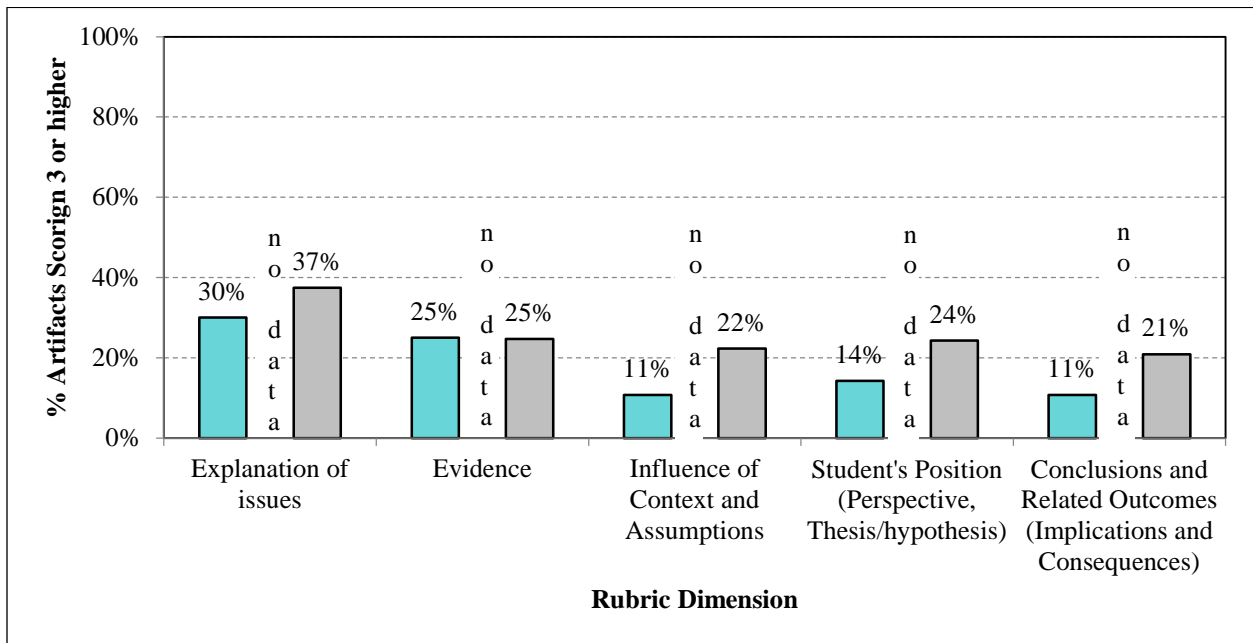

Figure 15. Comparison of CT achievement at 3 or higher for AA courses only across all rubric dimensions for 262 artifacts from 23 volunteered assignments. FSW Online (teal), n=14, Dual Enrollment (purple), n=0, Traditional (gray), n=248.
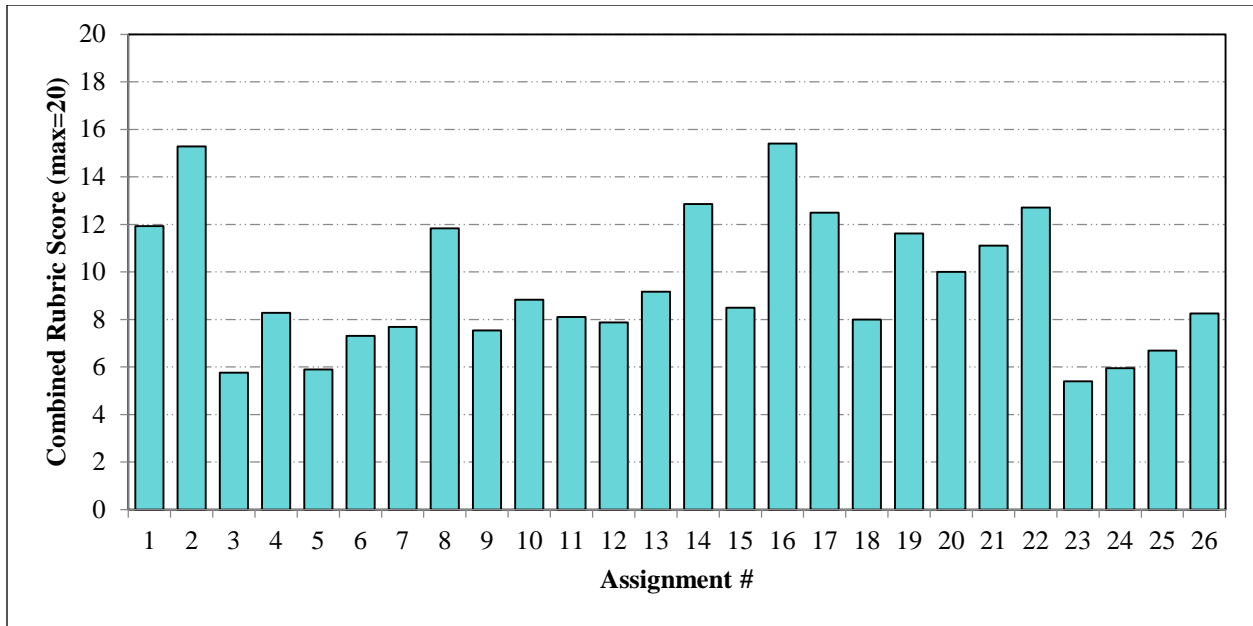
Figure 16. Overall rubric score (cumulative score of all five CT dimensions) by assignment. In cases where one or more dimensions was left unscored an overall score would not be a comparable statistic and is therefore left blank in this representation. Assignments: 1) POS 2041, 2) ENC 1101, 3) SPC 1017, 4) WOH 1030, 5) BSC 1010, 6) DEP 2004, 7) REA 0019, 8) PHY 2048, 9) DEP 2004, 10) AML 2020, 11) CHM 2025, 12) PHI 2010, 13) MAN 4720, 14) ENC 1102, 15) AMH 2010, 16) HUM 2235, 17) RED 4519, 18) AMH 2020, 19) SLS 1515, 20) MAC 1105, 21) PHI 2600, 22) ENC 1102, 23) ECO 2013, 24) WOH 1023, 25) SYG 1000, 26) LIS 2004.
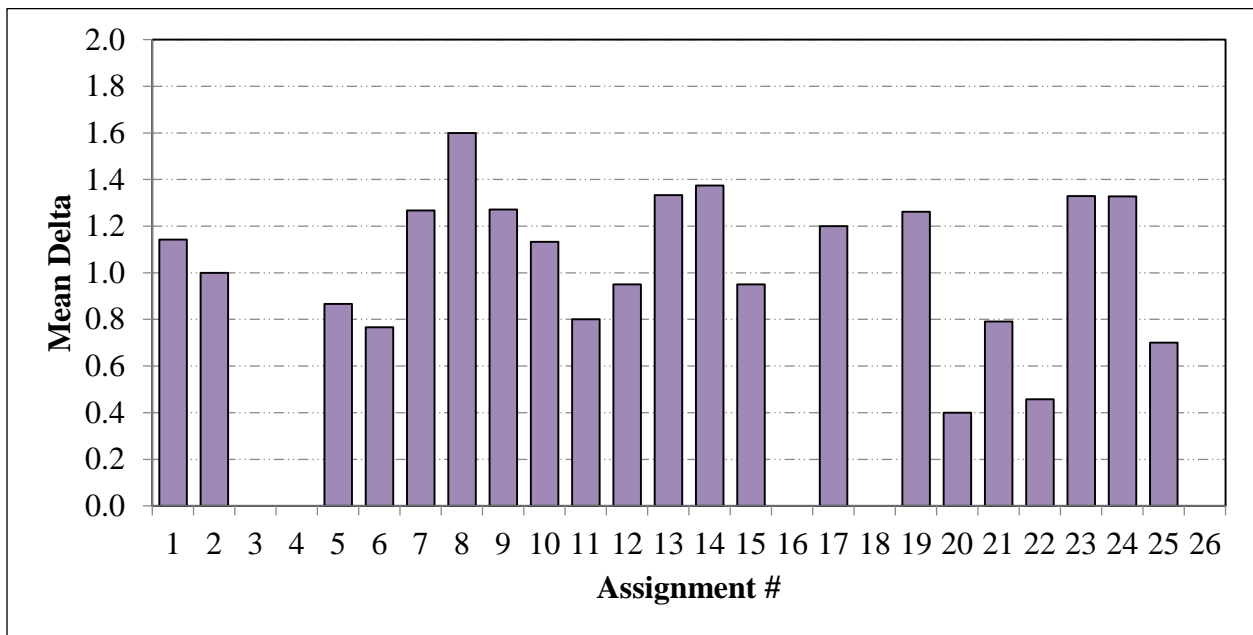

Figure 17. Inter-rater reliability of CT scoring by assignment using mean difference in scoring per dimension by assignment. Example: A mean delta of 1.0 means that, on average, the scoring pair disagrees by 1 achievement level for each rubric dimension scored. In cases where no mean delta is presented, one or more raters did not score the assignment and so comparisons could not be made. Assignments: 1) POS 2041, 2) ENC 1101, 3) SPC 1017, 4) WOH 1030, 5) BSC 1010, 6) DEP 2004, 7) REA 0019, 8) PHY 2048, 9) DEP 2004, 10) AML 2020, 11) CHM 2025, 12) PHI 2010, 13) MAN 4720, 14) ENC 1102, 15) AMH 2010, 16) HUM 2235, 17) RED 4519, 18) AMH 2020, 19) SLS 1515, 20) MAC 1105, 21) PHI 2600, 22) ENC 1102, 23) ECO 2013, 24) WOH 1023, 25) SYG 1000, 26) LIS 2004.

4 TECHNOLOGY/INFORMATION MANAGEMENT (TIM)

The TIM competency goal is by completion of the general education requirements, students will be able to demonstrate the skills and use the technology necessary to collect, verify, document, and organize information from a variety of sources. The FSW Learning Assessment Committee will use the pilot results to establish a baseline for the number of artifacts scored a 3 or higher on relevant dimensions of the rubric. Figures 18 through 23 below depict achievement and inter-rater reliability for the TIM competency in college-wide and Associate of Arts (AA) cohorts. Note that no online artifacts and only two dual enrollment artifacts were collected in the study so no cross-format comparisons could be made.

Feedback from scorers regarding the AAC&U Value rubric for the TIM competency included one common element among scorers regarding rubric suitability. Scorers noted that alignment of assignments with the assessment tool is hindered by the competency as defined at FSW (Technology/Information Management) and that of the AAC&U Value rubric (Information Literacy). In many cases, scorers reported assignments were not capable of being scored on the rubric because of this misalignment.
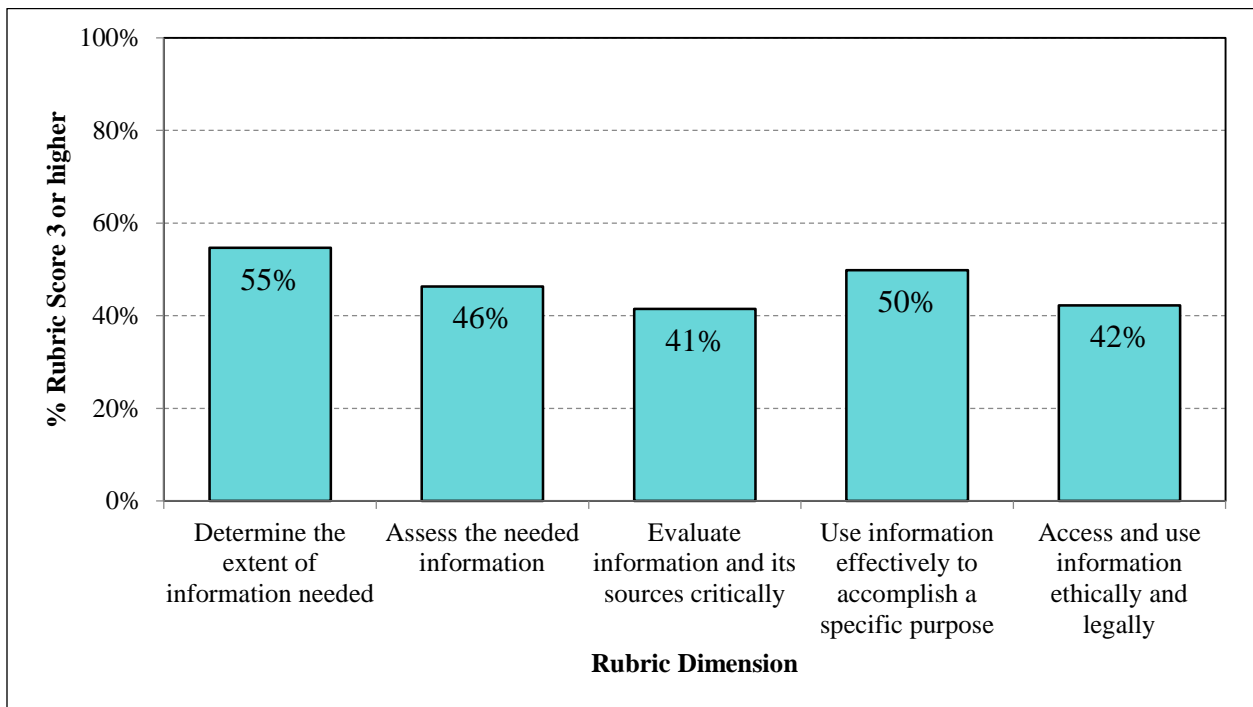

Figure 18. TIM Achievement at 3 or higher across all rubric dimensions for 129 artifacts from 15 volunteered assignments.

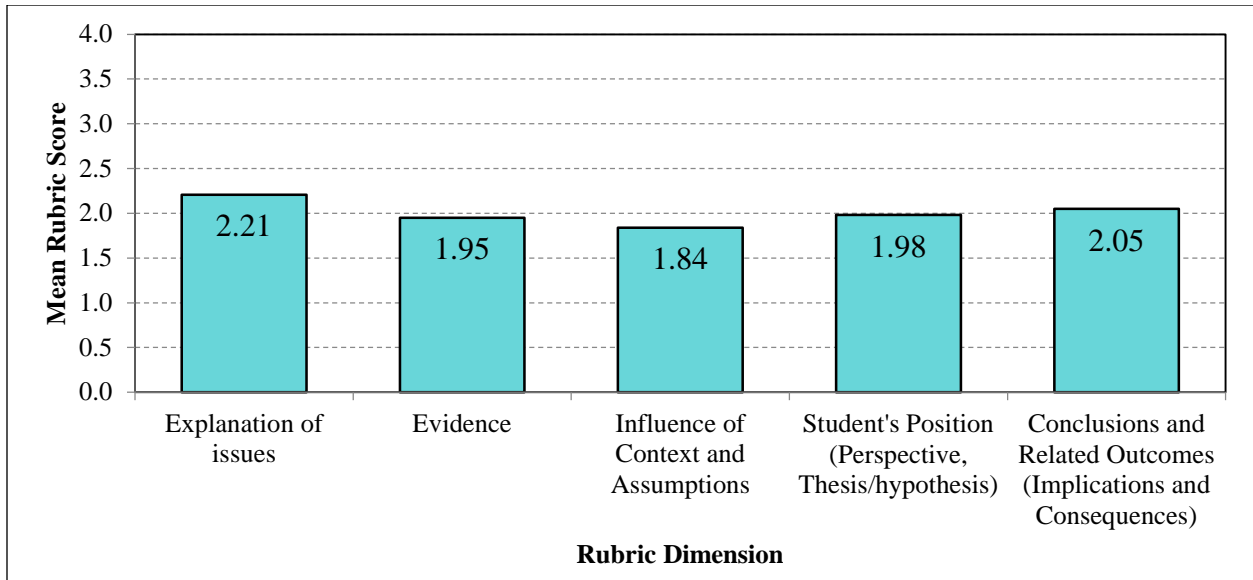*Figure 19. Mean score by rubric dimension for TIM for 129 artifacts from 15 volunteered assignments.*
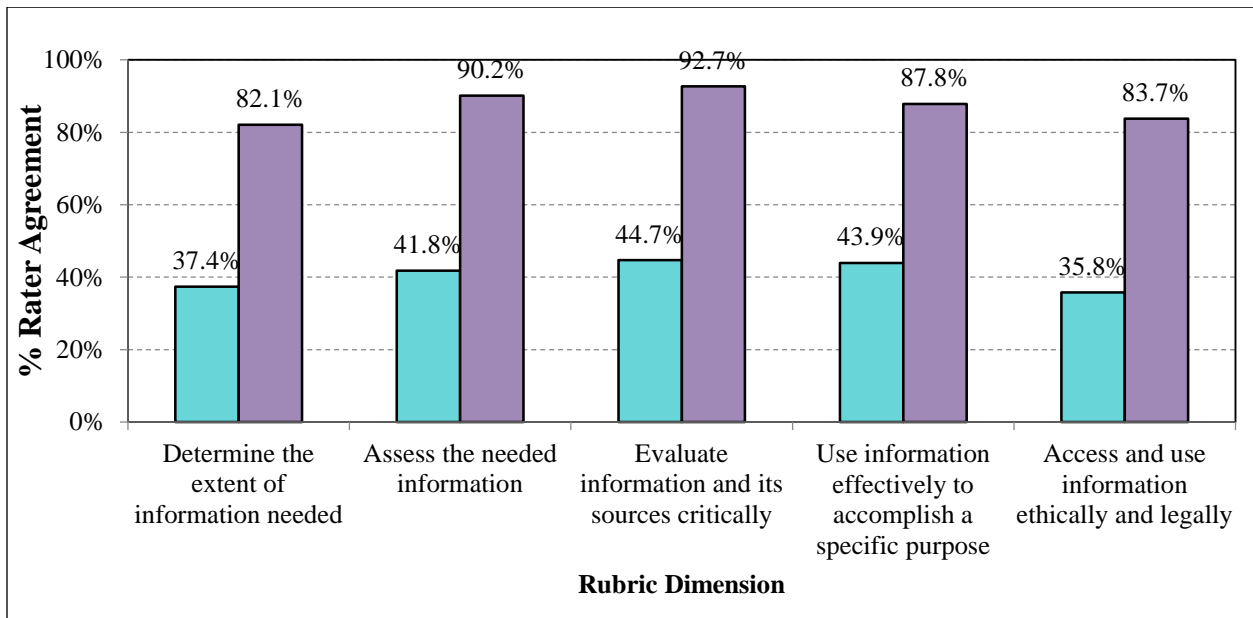

Figure 20. Inter-rater reliability (as %) for TIM competency. Each artifact was scored by two scorers. Percentage (%) of agreement (teal) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.

Figure 21. TIM Achievement at 3 or higher for AA courses only across all rubric dimensions for 119 artifacts from 14 volunteered assignments.
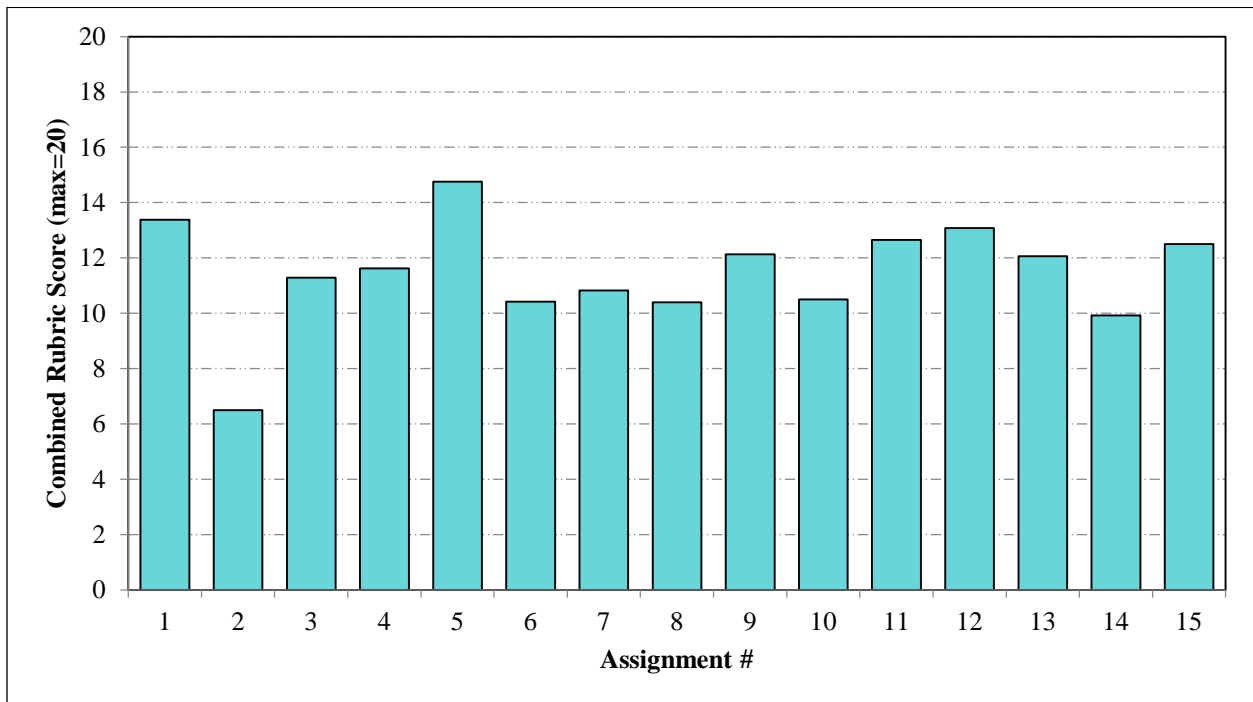
Figure 22. Overall rubric score (cumulative score of all five TIM dimensions) by assignment. In cases where one or more dimensions was left unscored an overall score would not be a comparable statistic and is therefore left blank in this representation. Assignments: 1) AMH 2020, 2) DEP 2004, 3) SPC 1017, 4) WOH 1030, 5. LIS 2004, 6) SPC 1017, 7) WOH 1030, 8) BSC 1010, 9) SPC 1017, 10) ENC 1102, 11) BSC 1011, 12) WOH 1030, 13) ENC 1102, 14) ENC 1102, 15) MCB 2010.
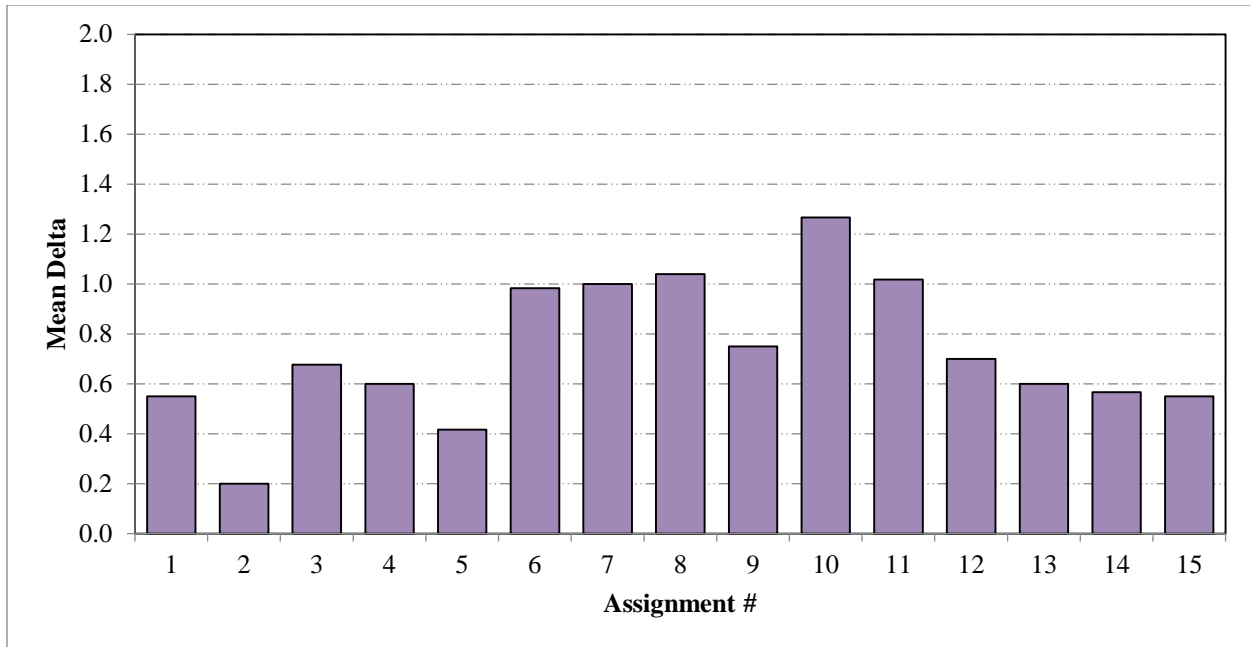
Figure 23. Inter-rater reliability of TIM scoring by assignment using mean difference in scoring per dimension by assignment. Example: A mean delta of 1.0 means that, on average, the scoring pair disagrees by 1 achievement level for each rubric dimension scored. Assignments: 1) AMH 2020, 2) DEP 2004, 3) SPC 1017, 4) WOH 1030, 5. LIS 2004, 6) SPC 1017, 7) WOH 1030, 8) BSC 1010, 9) SPC 1017, 10) ENC 1102, 11) BSC 1011, 12) WOH 1030, 13) ENC 1102, 14) ENC 1102, 15) MCB 2010.

5 GLOBAL SOCIO-CULTURAL RESPONSIBILITY (GSR)

The GSR competency goal is by completion of the general education requirements, students will be able to identify, describe, and apply responsibilities, core civic beliefs, and values present in a diverse society. The FSW Learning Assessment Committee will use the pilot results to establish a baseline for the number of artifacts scored a 3 or higher on relevant dimensions of the rubric. Figures 24 through 31 below depict achievement and inter-rater reliability for the GSR competency in college-wide and Associate of Arts (AA) cohorts.

Feedback from scorers regarding the AAC&U Value rubric for the GSR competency included a number of concerns regarding rubric suitability. Scorers noted that the rubric is often relying on terms that are not clearly defined and therefore interpreted in very different ways by different scorers. One example of this would be the use of the word 'culture' in the rubric. As noted by a scorer with a background in sociological and cultural areas, the word can have both a global sense such as a 'nationalistic culture' and a localized sense such as 'video gamer culture'. As a result, scoring dimensions become rather vague and disconnected from the aim of the rubric possibly affecting scores, again, a result common to other case studies (Rhodes and Finley, 2013).

Additionally, scorers noted the rubric specificity may be most useful only in cases where assignments are tailored to a defined area so that terms like 'culture' can be clearly understood when scoring. In most cases achievement at '2' compared with '3' was not clear enough.
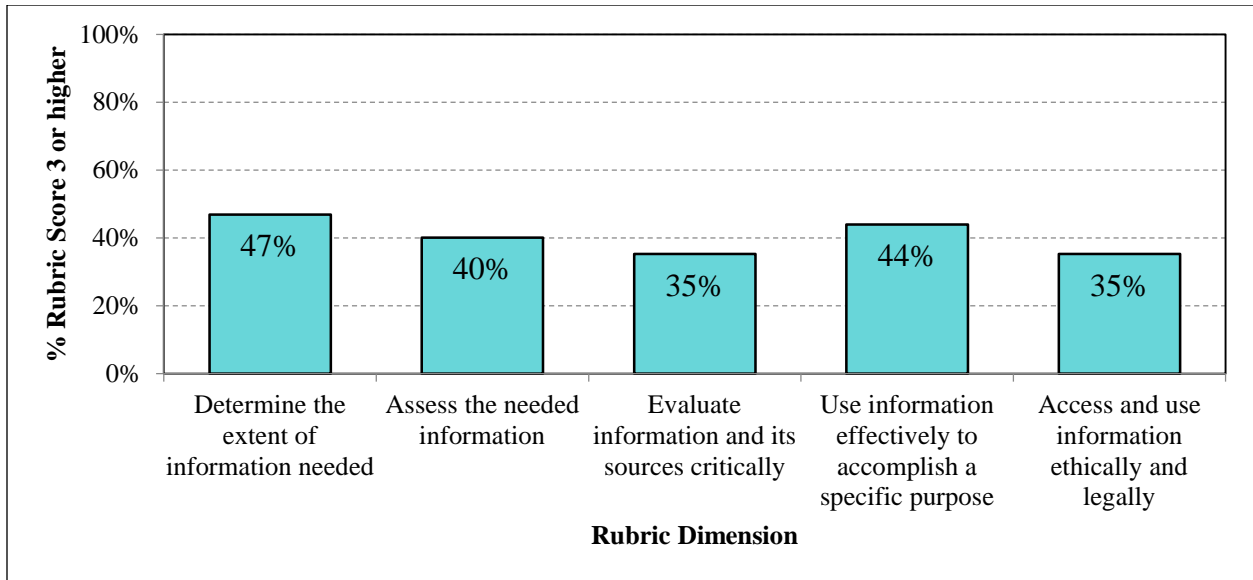
Figure 24. GSR Achievement at 3 or higher across all rubric dimensions for 137 artifacts from 11 volunteered assignments.
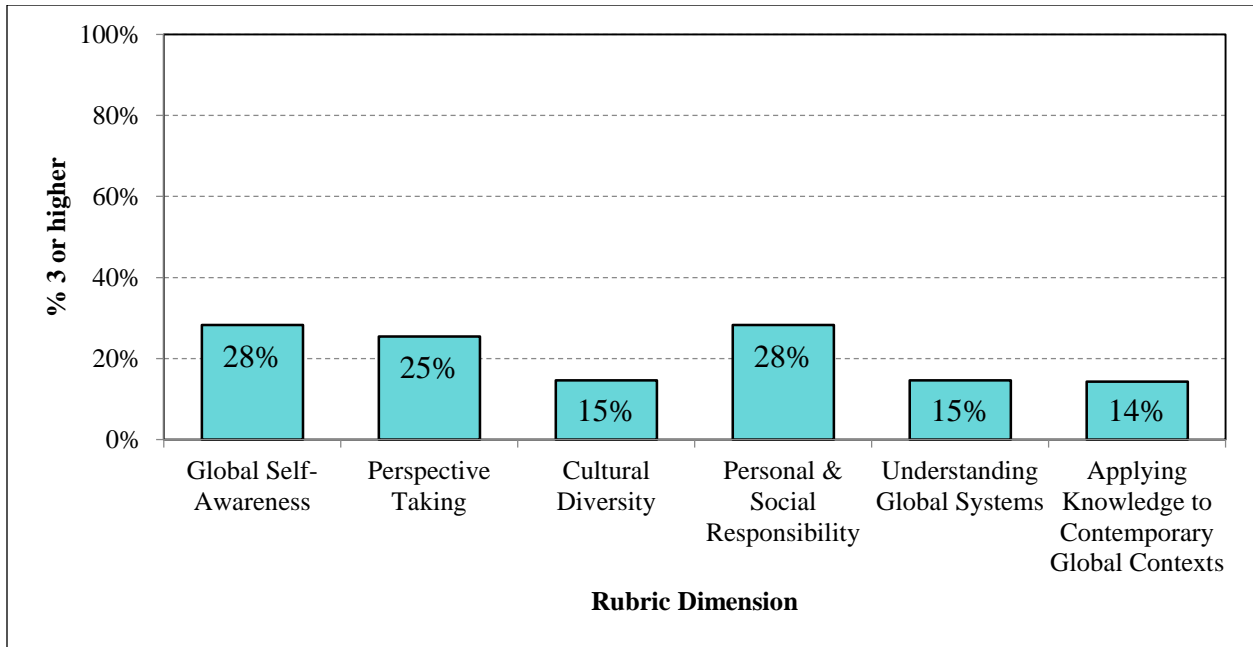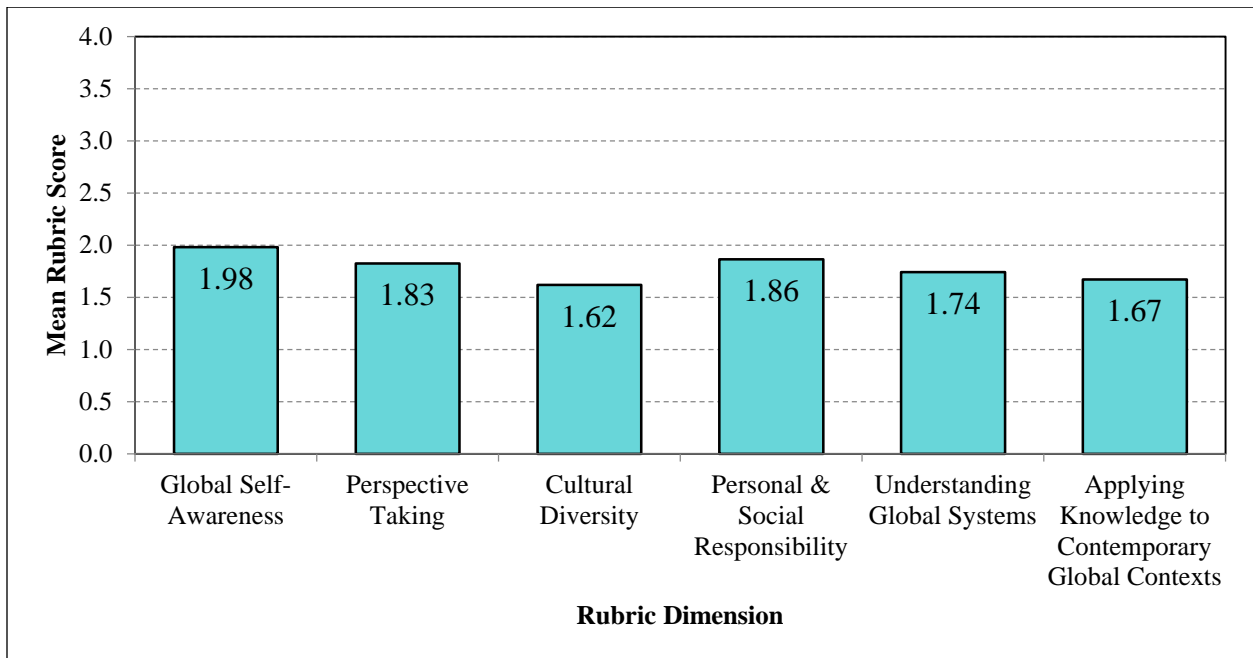

Figure 25. Mean score by rubric dimension for GSR for 137 artifacts from 11 volunteered assignments.
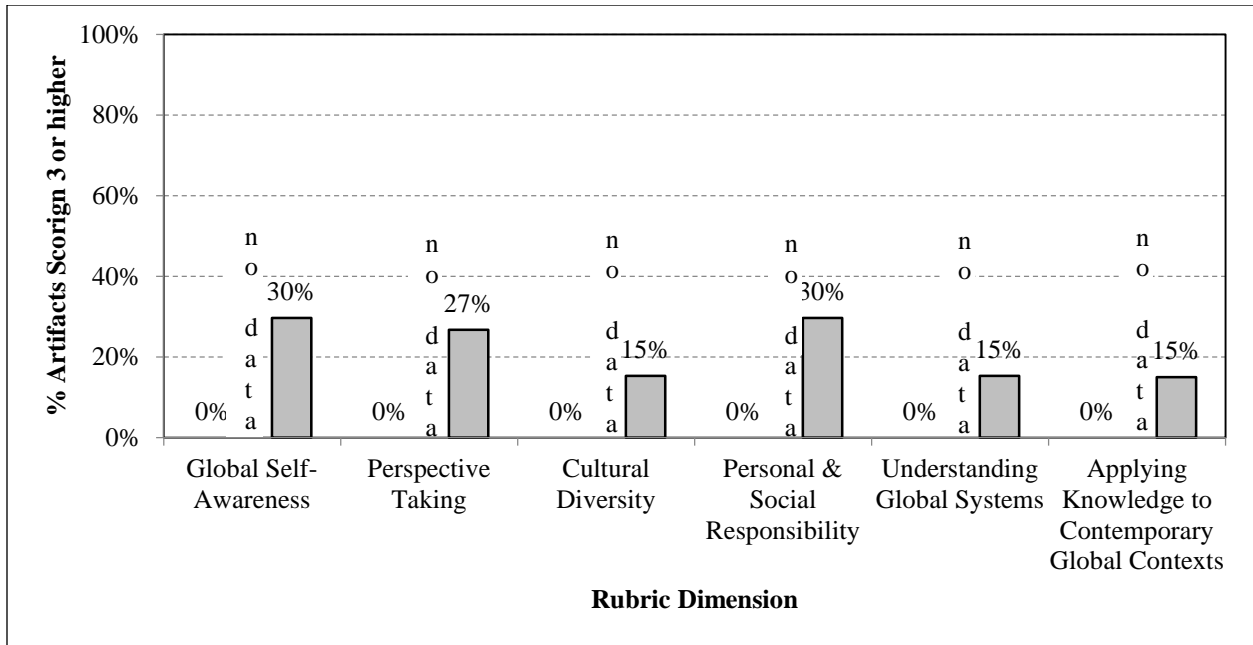
Figure 26. Comparison of GSR achievement at 3 or higher across all rubric dimensions for 137 artifacts from 11 volunteered assignments. FSW Online (teal), n=10, Dual Enrollment (purple), n=0, Traditional (gray), n=127.
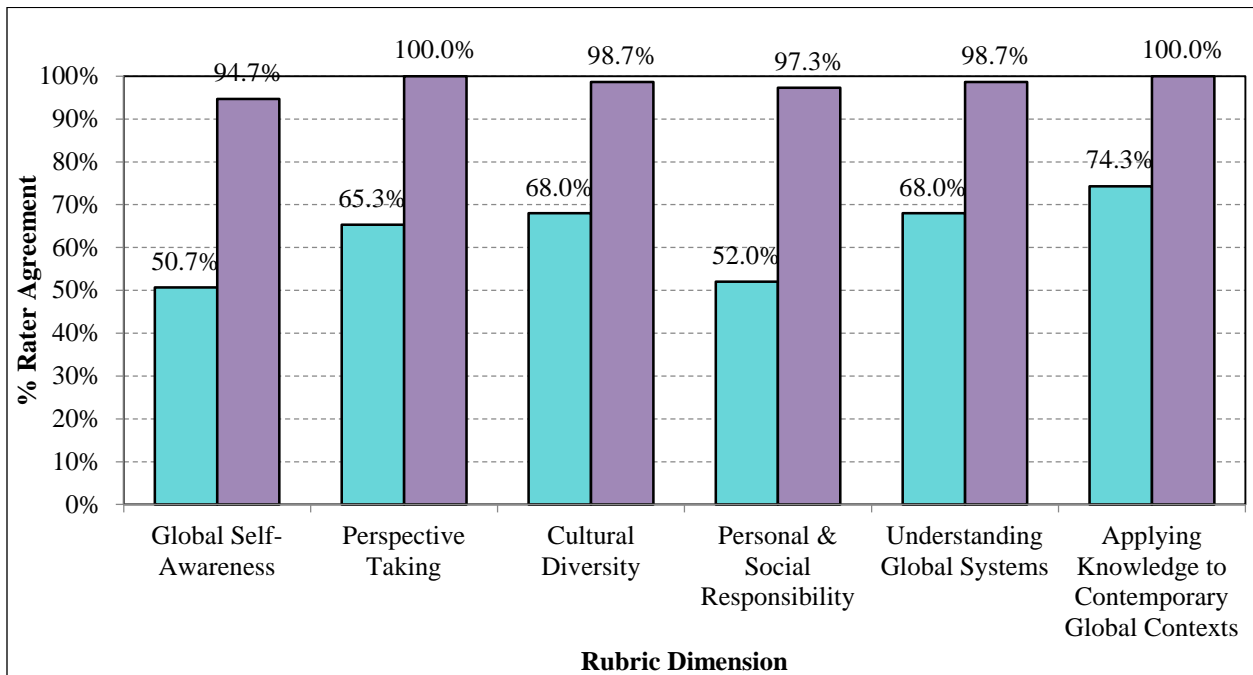
Figure 27. Inter-rater reliability (as %) for GSR competency. Each artifact was scored by two scorers. Percentage (%) of agreement (teal) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.

Figure 28. GSR Achievement at 3 or higher for AA courses only across all rubric dimensions for 98 artifacts from 10 volunteered assignments.

Figure 29. Comparison of GSR achievement at 3 or higher for AA courses only across all rubric dimensions for 88 artifacts from 10 volunteered assignments. FSW Online (teal), n=10, Dual Enrollment (purple), n=0, Traditional (gray), n=88.
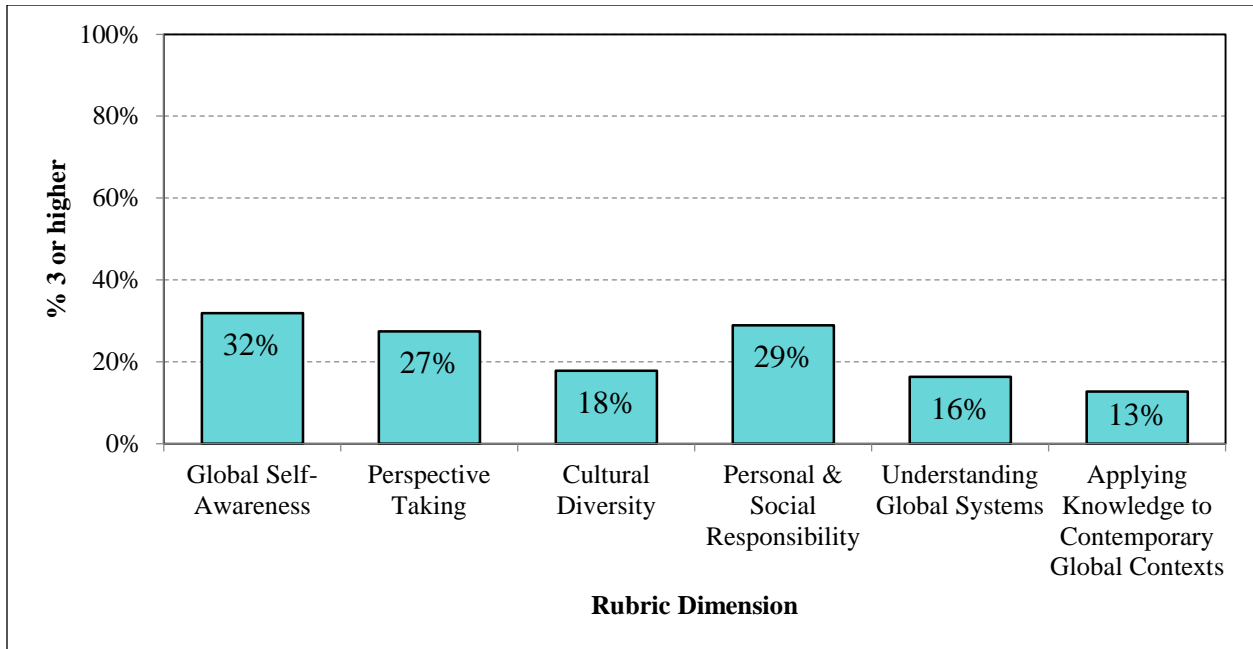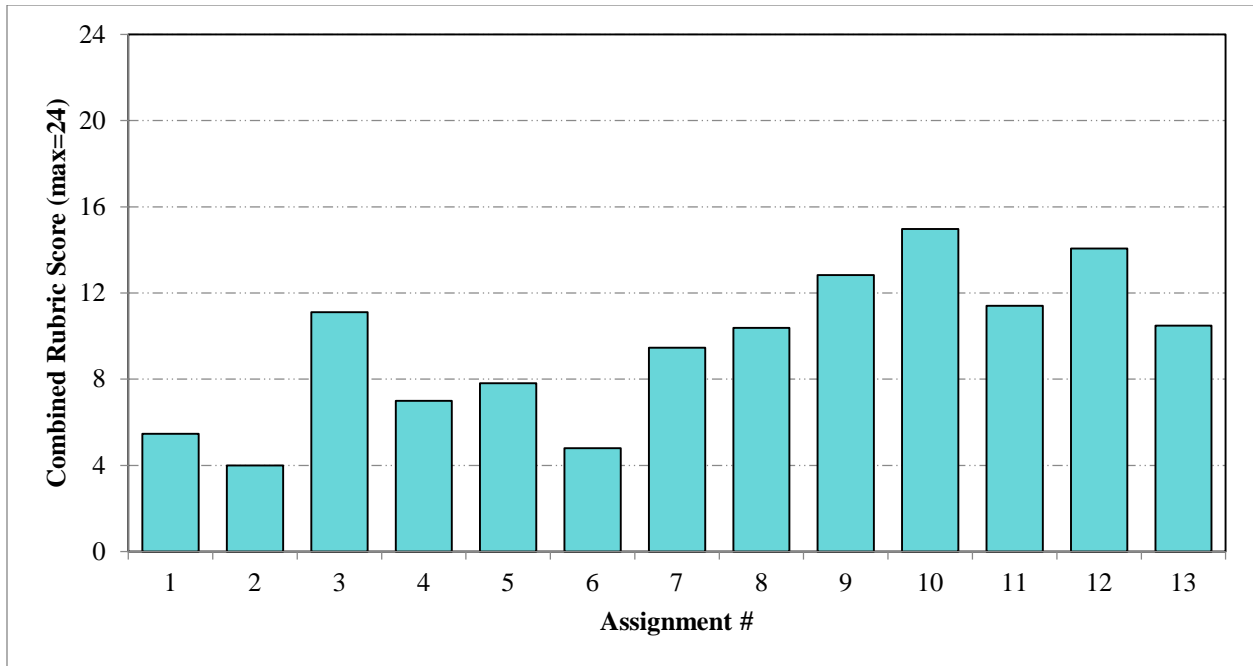
Figure 30. Overall rubric score (cumulative score of all six GSR dimensions) by assignment. In cases where one or more dimensions was left unscored an overall score would not be a comparable statistic and is therefore left blank in this representation. Assignments: 1) AML 2010, 2) PLA 1003, 3) WOH 1030, 4) DEP 2004, 5) PHI 2010, 6) AML 2020, 7) CHM 2025, 8) PHI 2010, 9) MAN 4720, 10) AMH 2010, 11) SPC 1017, 12) PHI 2600, 13) NUR 1022.
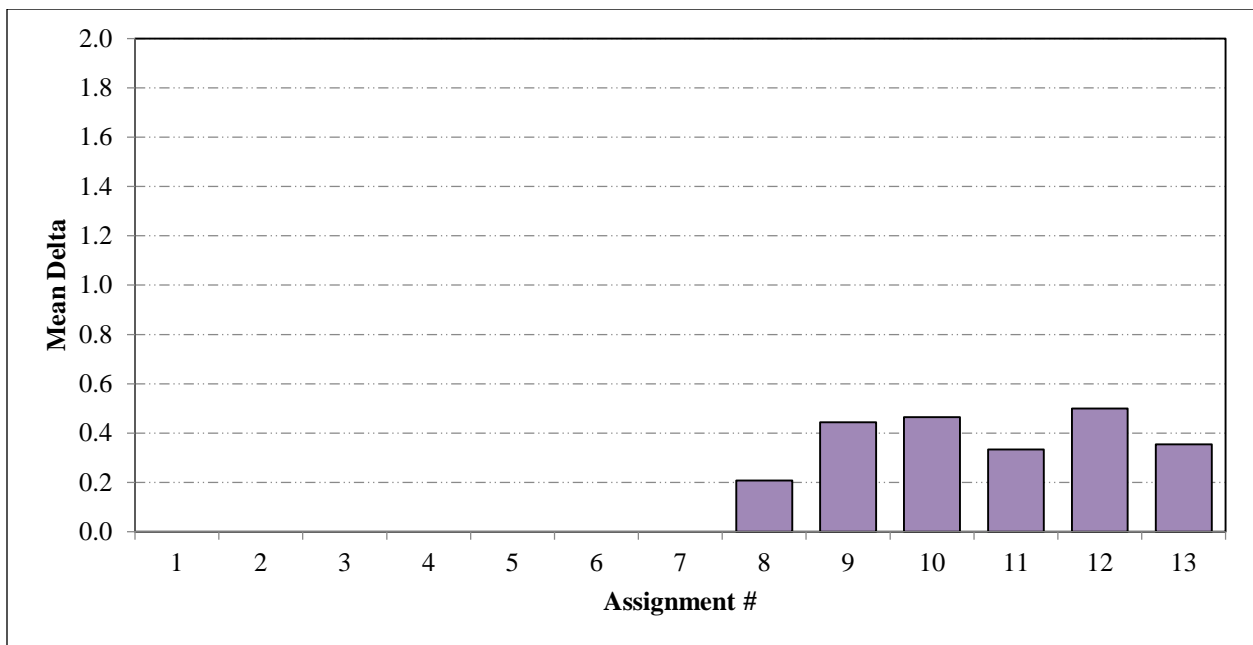
Figure 31. Inter-rater reliability of GSR scoring by assignment using mean difference in scoring per dimension by assignment. Example: A mean delta of 1.0 means that, on average, the scoring pair disagrees by 1 achievement level for each rubric dimension scored. In cases where no mean delta is presented, one or more raters did not score the assignment and so comparisons could not be made. Assignments: 1) AML 2010, 2) PLA 1003, 3) WOH 1030, 4) DEP 2004, 5) PHI 2010, 6) AML 2020, 7) CHM 2025, 8) PHI 2010, 9) MAN 4720, 10) AMH 2010, 11) SPC 1017, 12) PHI 2600, 13) NUR 1022.

6 SCIENTIFIC AND QUANTITATIVE REASONING (QR)

The QR competency goal is 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 use the pilot results to establish a baseline for the number of artifacts scored a 3 or higher on relevant dimensions of the rubric. Figures 32 through 37 below depict achievement and inter-rater reliability for the QR competency in college-wide and Associate of Arts (AA) cohorts. Note that all QR artifacts collected were from the AA cohort, and so no AA-specific figures are represented here.

Although the AAC&U Value rubric for QR was not used, and instead a rubric developed by FSW was implemented, a review of scorer feedback was still tallied to test suitability in General Education assignments. Overall, scorers felt that the rubric was easy to use and aligned well with the QR competency. The only area where scorers addressed concern was perhaps modifying the rubric to distinguish between "analysis/synthesis" and "evaluation".
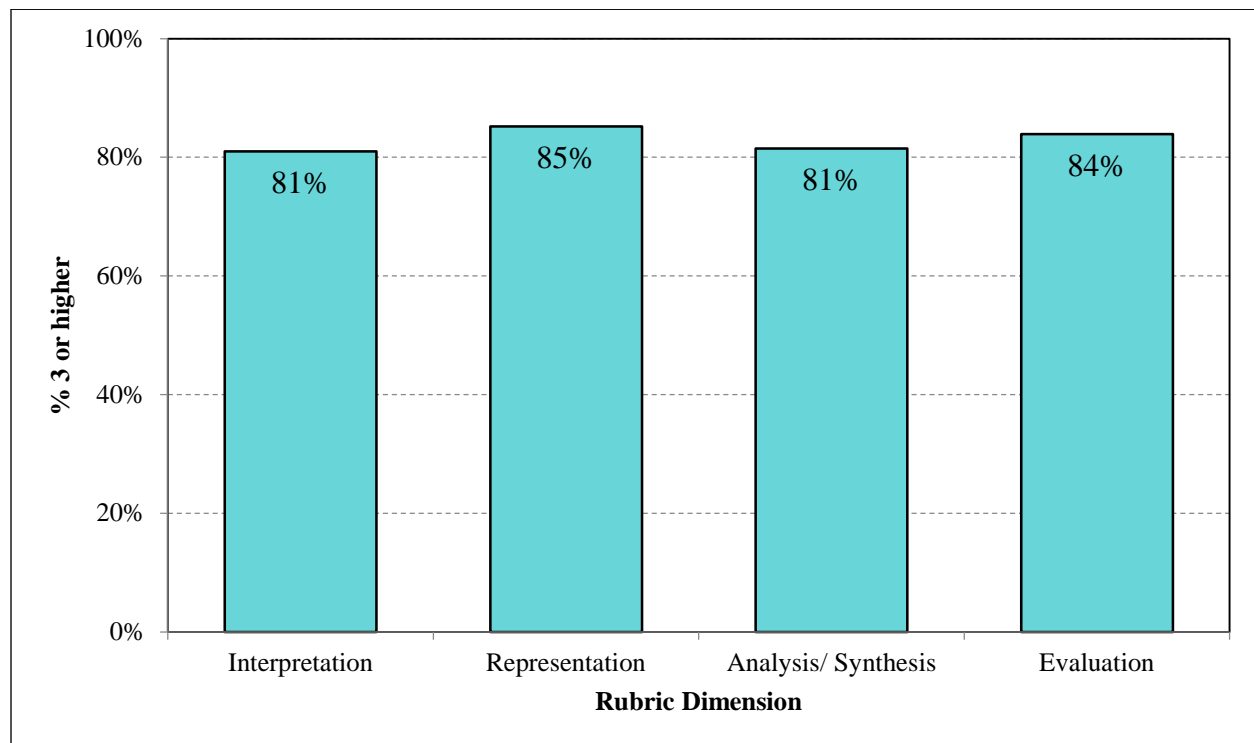
Figure 32. QR Achievement at 3 or higher across all rubric dimensions for 145 artifacts from 8 volunteered assignments.
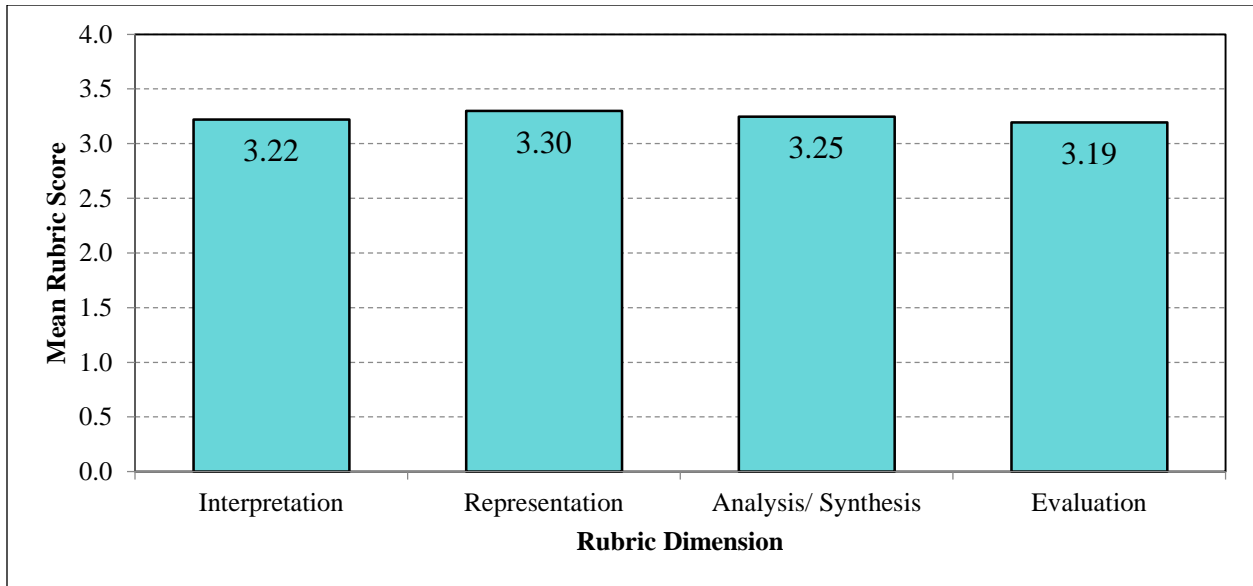
Figure 33. Mean score by rubric dimension for QR for 145 artifacts from 8 volunteered assignments.
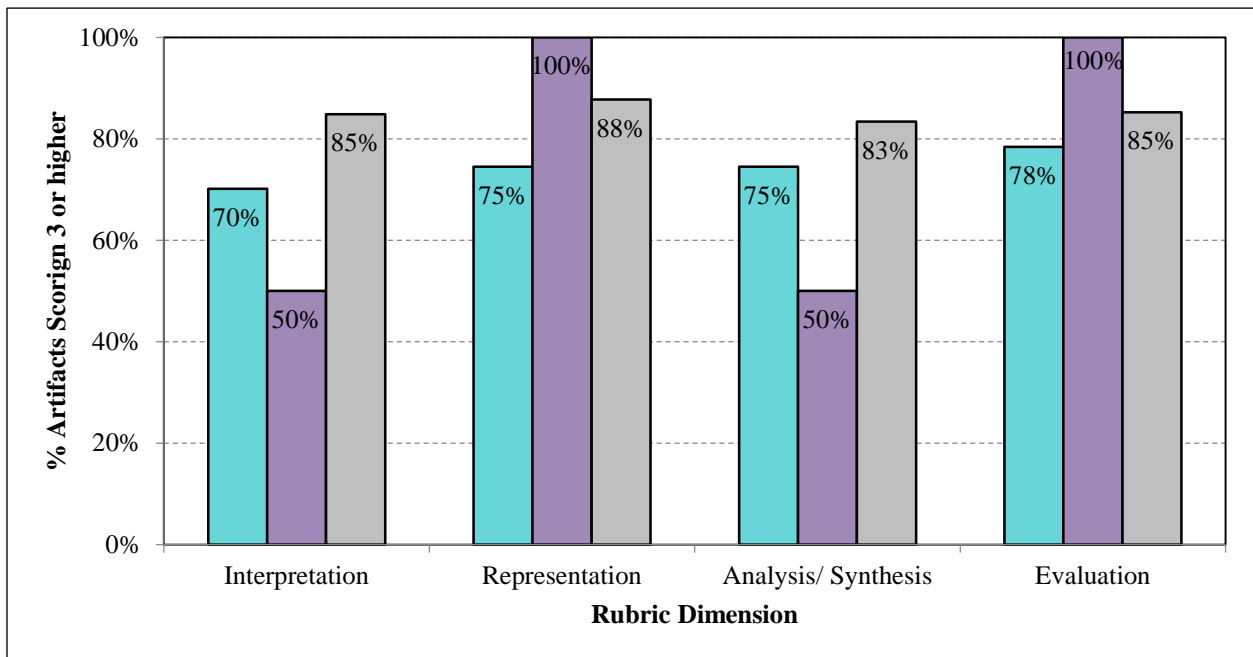
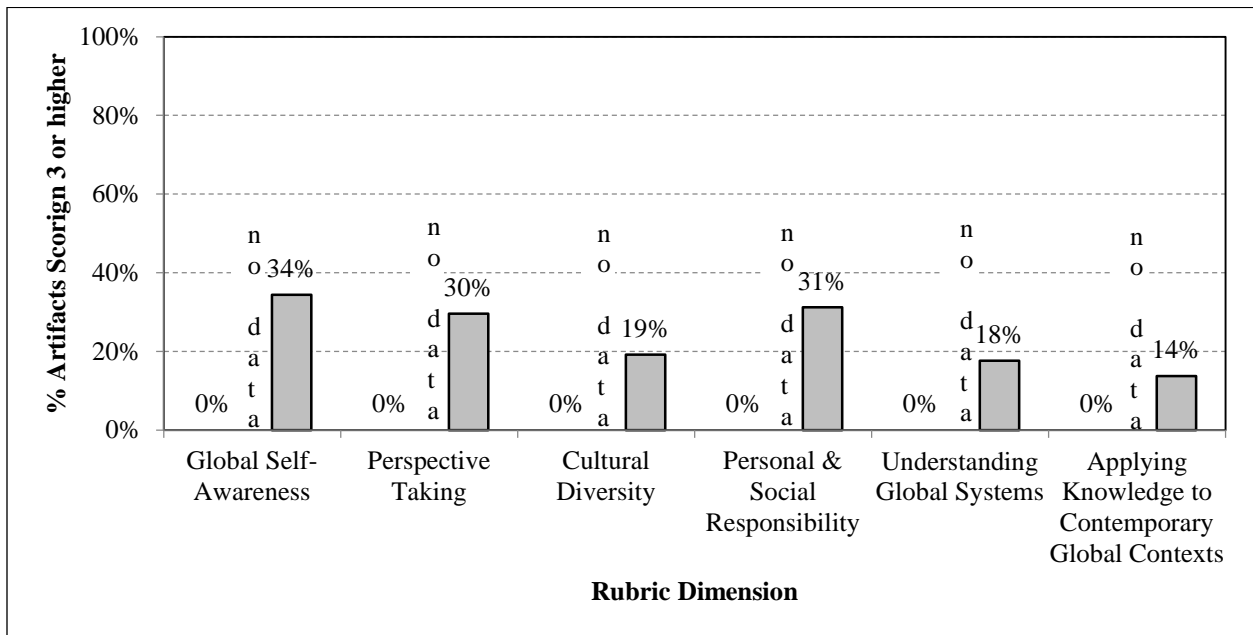*Figure 34. Comparison of QR achievement at 3 or higher across all rubric dimensions for 145 artifacts from 8 volunteered assignments. FSW Online (teal), n=32, Dual Enrollment (purple), n=2, Traditional (gray), n=111.*
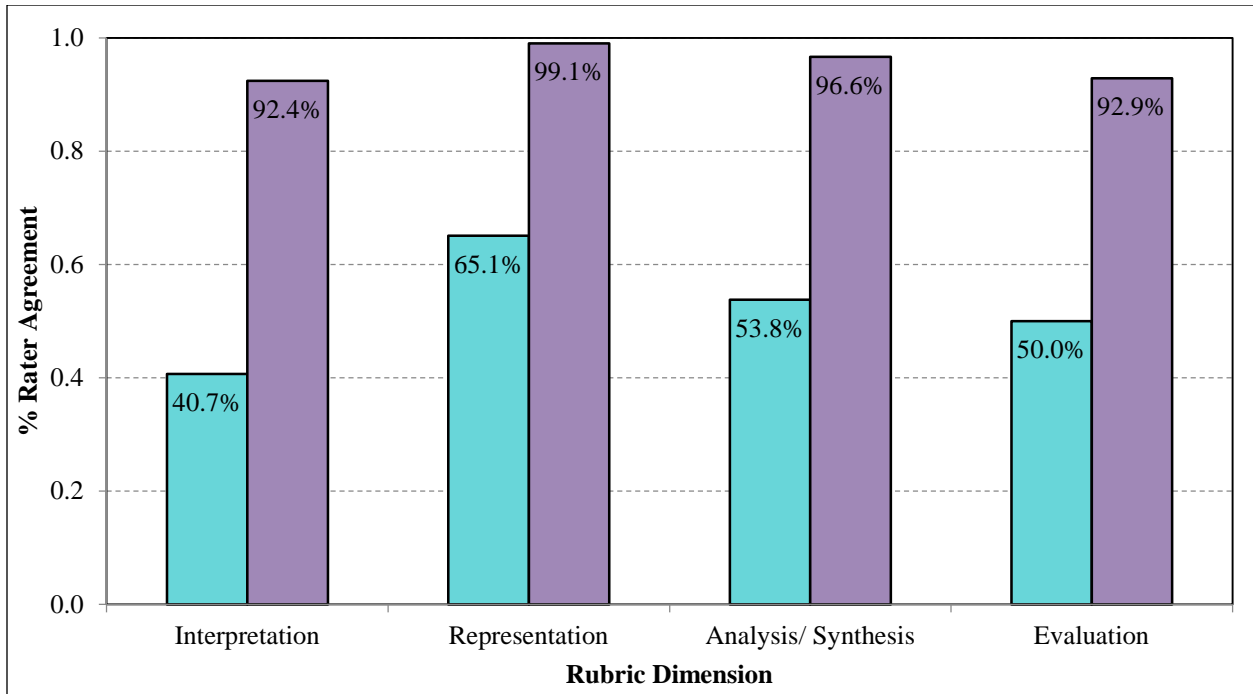
Figure 35. Inter-rater reliability (as %) for QR competency. Each artifact was scored by two scorers. Percentage (%) of agreement (teal) is defined as cases where scores by each scorer were identical. Percentage (%) +/- 1 agreement (purple) is defined as cases where scores by each scorer were within 1 of each other.
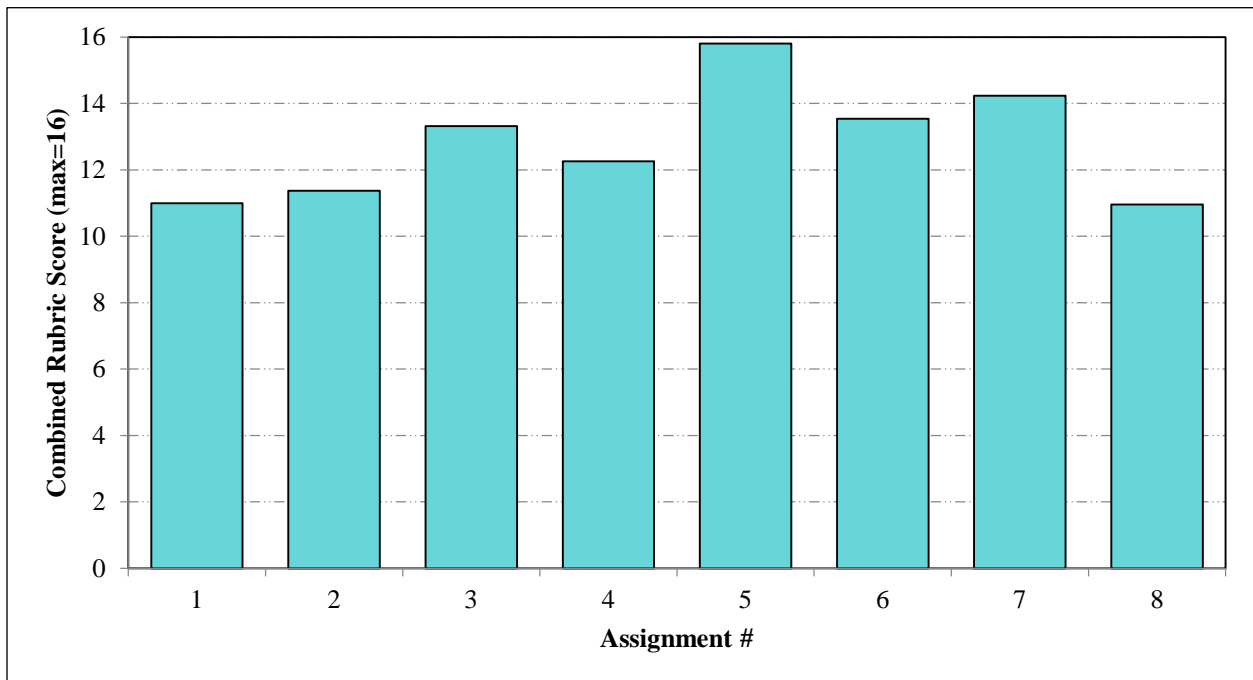
Figure 36. Overall rubric score (cumulative score of all four QR dimensions) by assignment. In cases where one or more dimensions was left unscored an overall score would not be a comparable statistic and is therefore left blank in this representation. Assignments: 1) DEP 2004, 2) PLA 1003, 3) DEP 2004, 4) PHY 2048, 5) MAC 1105, 6) ECO 2013, 7) ECO 2013, 8) CHM 2025.
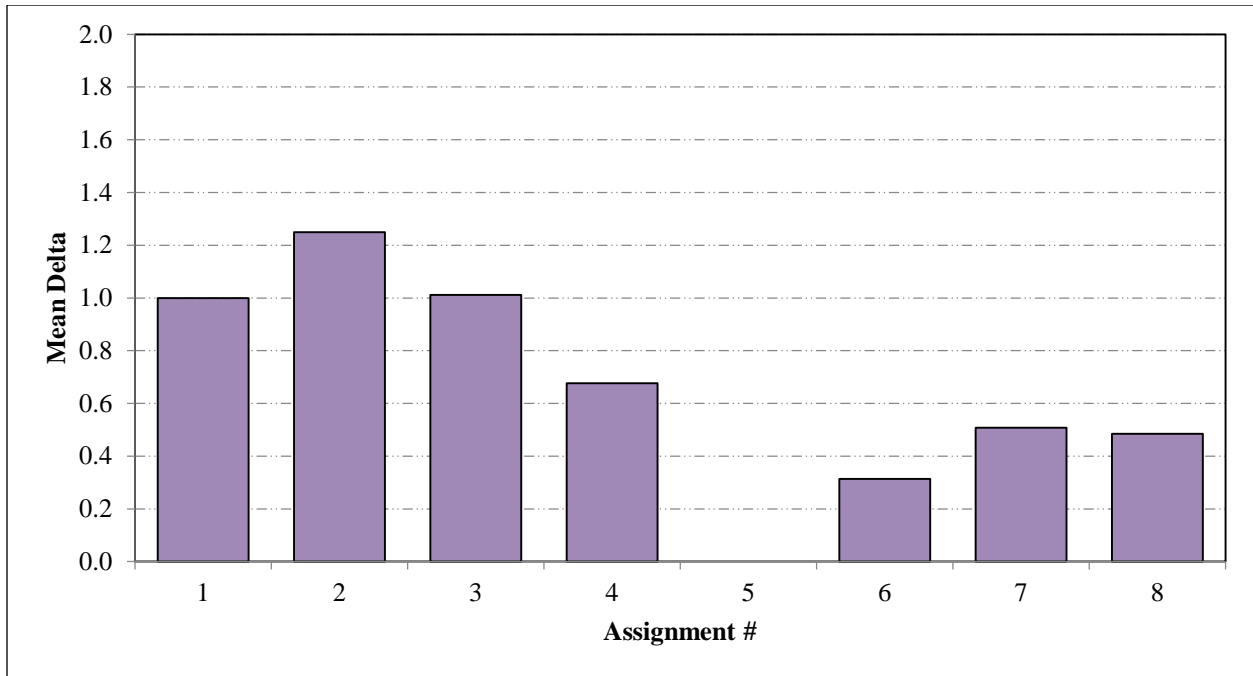
Figure 37. Inter-rater reliability of QR scoring by assignment using mean difference in scoring per dimension by assignment. Example: A mean delta of 1.0 means that, on average, the scoring pair disagrees by 1 achievement level for each rubric dimension scored. In cases where no mean delta is presented, one or more raters did not score the assignment and so comparisons could not be made. Assignments: 1) DEP 2004, 2) PLA 1003, 3) DEP 2004, 4) PHY 2048, 5) MAC 1105, 6) ECO 2013, 7) ECO 2013, 8) CHM 2025.

7 CONCLUSIONS

FSW's General Education Program was assessed through voluntary submission of assignments identified by faculty as fitting the criteria of the five competencies Communication, Critical Thinking, Technology/Information Management, Global Socio-cultural Responsibility, and Scientific and Quantitative Reasoning). The study employed the use of the Association of American Colleges & Universities (AAC&U) rubrics for all but one competency (Scientific and Quantitative Reasoning, QR) which used a rubric developed by faculty from the School of Pure & Applied Sciences. The study established a baseline for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric and a baseline for the number of students receiving scores of 3 or higher on relevant dimensions of the rubric across sites (Online, Dual Enrollment, and Traditional). Results also included these same outcomes with respect to courses included in the AA program.

A drilldown of Communication (COM) results are as follows:
1. One of five rubric dimensions exhibit greater than 60% achievement at level '3' (Context of and Purpose for Writing). The remaining four dimensions exhibit a similar level of achievement ranging from 40% to 48%.
2. Mean achievement levels for each of the five rubric dimensions of COM range from 2.23 to 2.74 on a 4-point scale.
3. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 60% to 90% (n = 10). Dual enrollment artifacts achievement

at level '3' range from 28% to 44% (n = 9). Traditional artifacts achievement at level '3' range from 39% to 60% (n = 201).

4. Inter-rater reliability exhibit rubric scoring agreement ranging from 32% to 44% with a +/- 1 agreement ranging from 81% to 91%.

5. With respect to AA courses, one of 5 rubric dimensions exhibit greater than 55% achievement at level '3' (Context of and Purpose for Writing). The remaining four dimensions exhibit a similar level of achievement ranging from 39% to 46%.

6. In a study comparing AA courses with online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 60% to 90% (n = 10). Dual enrollment artifacts achievement at level '3' range from 28% to 44% (n = 9). Traditional artifacts achievement at level '3' range from 39% to 59% (n = 201).

7. In a study comparing overall rubric score of courses (cumulative score of all five COM dimensions) individual assignment artifacts range from 8.8/20 to 16.6/20 with a wide variety of achievement across the 25 submitted assignments.

8. In a study comparing inter-rater reliability by assignment, mean deltas (difference in scoring by scorers in each rubric dimension) range from 0.3 to 1.1.

9. In a review of scorer feedback, multiple scorers noted that "Sources and Evidence" is not relevant in all COM assignments. Also, scorers noted the use of "some" and "few" at differing achievement levels when interpreting "Control of Syntax and Mechanics" is at times very difficult to distinguish.

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

1. One of five rubric dimensions exhibit greater than 30% achievement at level '3' (Explanation of Issues). The remaining four dimensions exhibit a similar level of achievement ranging from 20% to 24%.

2. Mean achievement levels for each of the five rubric dimensions of CT range from 2.84 to 2.21 on a 4-point scale.

3. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 19% to 27% (n = 16). No data were collected to compare Dual Enrollment artifacts. Traditional artifacts achievement at level '3' range from 20% to 37% (n = 256).

4. Inter-rater reliability exhibit rubric scoring agreement ranging from 19% to 44% with a +/- 1 agreement ranging from 65% to 82%.

5. With respect to AA courses, one of five rubric dimensions exhibit greater than 30% achievement at level '3' (Explanation of Issues). The remaining four dimensions exhibit a similar level of achievement ranging from 19% to 23%.

6. In a study comparing AA courses with online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range from 11% to 30% (n = 14). No data were collected to compare Dual Enrollment artifacts. Traditional artifacts achievement at level '3' range from 21% to 37% (n = 248).

7. In a study comparing overall rubric score of courses (cumulative score of all five CT dimensions) individual assignment artifacts range from 5.4/20 to 15.4/20 with a wide variety of achievement across the 26 submitted assignments.

8. In a study comparing inter-rater reliability by assignment, mean deltas (difference in scoring by scorers in each rubric dimension) range from 0.4 to 1.6.

9. In a review of scorer feedback, scorers noted the rubric seemed to require greater distinction between '2' and '3' and less distinction between '3' and '4'. Additionally, some scorers felt that the rubric was not fluid enough to accommodate the varied types of CT assignments produced in the voluntary submission.

A drilldown of Technology/Information Management (TIM) results are as follows:

1. Two of five rubric dimensions exhibit greater than 50% achievement at level '3'. The remaining three dimensions exhibit a similar level of achievement ranging from 41% to 46%.

2. Mean achievement levels for each of the five rubric dimensions of TIM range from 1.84 to 2.21 on a 4-point scale.

3. No online artifacts and only two dual enrollment artifacts were collected in the study so no cross-format comparisons could be made.

4. Inter-rater reliability exhibit rubric scoring agreement ranging from 37% to 45% with a +/- 1 agreement ranging from 82% to 93%.

5. With respect to AA courses, three of five rubric dimensions exhibit greater than 40% achievement at level '3'. The remaining two dimensions exhibit a similar level of achievement at 35%.

6. No online artifacts and only two dual enrollment artifacts were collected in the study so no cross-format comparisons could be made.

7. In a study comparing overall rubric score of courses (cumulative score of all five TIM dimensions) individual assignment artifacts range from 6.5/20 to 14.8/20 with a wide variety of achievement across the 15 submitted assignments.

8. In a study comparing inter-rater reliability by assignment, mean deltas (difference in scoring by scorers in each rubric dimension) range from 0.2 to 1.3.

9. In a review of scorer feedback, scorers reported assignments were not capable of being scored on the rubric because of misalignment between AAC&U Value rubric and TIM competency as defined at FSW.

A drilldown of Global Socio-cultural Responsibility (GSR) results are as follows:

1. Achievement in all six rubric dimensions exhibit less than 30% achievement at level '3' ranging from 14% to 28%.

2. Mean achievement levels for each of the six rubric dimensions of GSR range from 1.62 to 1.98 on a 4-point scale.

3. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range is 0% for all six dimensions (n = 10). No Dual Enrollment data were collected. Traditional artifacts achievement at level '3' range from 15% to 30% (n = 127).

4. Inter-rater reliability exhibit rubric scoring agreement ranging from 51% to 74% with a +/- 1 agreement ranging from 95% to 100%.

5. With respect to AA courses, one of the six rubric dimensions exhibit greater than 30% achievement at level '3' (Global Self-Awareness). The remaining five dimensions exhibit achievement ranging from 13% to 29%.

6. In a study comparing AA courses with online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' is 0% for all six dimensions (n = 10). No data were collected to compare Dual Enrollment artifacts. Traditional artifacts achievement at level '3' range from 14% to 34% (n = 88).
7. In a study comparing overall rubric score of courses (cumulative score of all six GSR dimensions) individual assignment artifacts range from 4.0/24 to 15.0/24 with a wide variety of achievement across the 13 submitted assignments.
8. In a study comparing inter-rater reliability by assignment, mean deltas (difference in scoring by scorers in each rubric dimension) range from 0.2 to 0.5.
9. In a review of scorer feedback, scorers noted that the rubric is often relying on terms that are not clearly defined and therefore interpreted in very different ways by different scorers. Additionally, in most cases, achievement at '2' compared with '3' was not clear enough.

A drilldown of Scientific and Quantitative Reasoning (QR) results are as follows:
1. Achievement in all four rubric dimensions exhibit greater than 80% achievement at level '3' ranging from 81% to 85%.
2. Mean achievement levels for each of the four rubric dimensions of QR range from 3.19 to 3.30 on a 4-point scale.
3. In a study comparing online, dual enrollment, and traditional artifacts, online artifacts achievement at level '3' range is 70% to 78% for all four dimensions (n = 32). Dual Enrollment data constitutes only two samples. Traditional artifacts achievement at level '3' range from 83% to 88% (n = 111).
4. Inter-rater reliability exhibit rubric scoring agreement ranging from 41% to 65% with a +/- 1 agreement ranging from 92% to 99%.
5. Note that all QR artifacts collected were from the AA cohort, and so no AA-specific figures are represented here.
6. In a study comparing overall rubric score of courses (cumulative score of all four QR dimensions) individual assignment artifacts range from 11.0/16 to 15.8/16 with a wide variety of achievement across the 8 submitted assignments.
7. In a study comparing inter-rater reliability by assignment, mean deltas (difference in scoring by scorers in each rubric dimension) range from 0.3 to 1.3.
8. In a review of scorer feedback, scorers felt that the rubric was easy to use and aligned well with the QR competency noting that the only change might be to better distinguish between "analysis/synthesis" and "evaluation".

8 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.