



# GENERAL EDUCATION ASSESSMENT

*2014 Report*



This page purposely blank.

## Acknowledgements

We would like to acknowledge the following people who provided information for this report:

The Office of Institutional Research and Assessment, which provided demographic and preparedness data.

Madison Sauls, graduate assistant for General Education Assessment

This page purposely blank.

# TABLE OF CONTENTS

|   |    |
|---|----|
| EXECUTIVE SUMMARY .....   | 1  |
| 1. BACKGROUND, SCOPE, AND METHODOLOGY .....                                       | 3  |
| 2.1 INQUIRY AND ANALYSIS .....  | 8  |
| APPENDIX 2.1.A INQUIRY AND ANALYSIS RUBRIC .....                                  | 17 |
| APPENDIX 2.1.B INQUIRY AND ANALYSIS CORRELATION COEFFICIENTS .....                | 18 |
| 2.2 INQUIRY AND ANALYSIS: DATA ANALYSIS .....                                     | 20 |
| APPENDIX 2.2.A INQUIRY AND ANALYSIS: DATA ANALYSIS RUBRIC .....                   | 25 |
| APPENDIX 2.2.B INQUIRY AND ANALYSIS: DATA ANALYSIS CORRELATION COEFFICIENTS ..... | 26 |
| 3. INFORMATION LITERACY .....   | 27 |
| APPENDIX 3.A INFORMATION LITERACY RUBRIC .....                                    | 37 |
| APPENDIX 3.B INFORMATION LITERACY CORRELATION COEFFICIENTS .....                  | 38 |
| 4. CRITICAL THINKING .....  | 39 |
| APPENDIX 4.A CRITICAL THINKING RUBRIC.....  | 50 |
| APPENDIX 4.B CRITICAL THINKING CORRELATION COEFFICIENTS.....                      | 51 |
| 5. THOUGHTFUL EXPRESSION (ORAL).....  | 53 |
| APPENDIX 5.A THOUGHTFUL EXPRESSION (ORAL) RUBRIC .....                            | 59 |
| APPENDIX 5.B THOUGHTFUL EXPRESSION (ORAL) CORRELATION COEFFICIENTS .....          | 60 |
| 6. THOUGHTFUL EXPRESSION (WRITTEN) .....  | 61 |
| APPENDIX 6.A THOUGHTFUL EXPRESSION (WRITTEN) RUBRIC .....                         | 67 |
| APPENDIX 6.B THOUGHTFUL EXPRESSION (WRITTEN) CORRELATION COEFFICIENTS .....       | 68 |
| 7. GENERAL DISCUSSION AND RECOMMENDATIONS.....                                    | 69 |
| UNCW STUDENT ABILITIES ON LEARNING GOALS .....                                    | 69 |
| SCORER FEEDBACK ON PROCESS.....   | 72 |
| INTERRATER RELIABILITY .....  | 73 |
| FOLLOW UP ON PREVIOUS RECOMMENDATIONS.....  | 75 |
| NEW RECOMMENDATIONS.....  | 75 |
| REFERENCES AND RESOURCES.....   | 76 |
| APPENDIX A UNIVERSITY STUDIES CURRICULUM MAP .....                                | 78 |
| APPENDIX B A NOTE ON INTERRATER RELIABILITY MEASURES .....                        | 80 |

## LIST OF FIGURES

|   |    |
|---|----|
| Figure 2.1.a Distributions of scores for Inquiry and Analysis (lower division), applicable scores only.....                 | 9  |
| Figure 2.1.b Distributions of scores for Inquiry and Analysis (upper division), applicable scores only.....                 | 11 |
| Figure 2.2.a Distributions of scores for Inquiry and Analysis: Data Analysis (lower division), applicable scores only ..... | 21 |
| Figure 3.a Distribution of Scores for Information Literacy (Lower Division), Applicable Scores Only .....                   | 28 |
| Figure 3.b Distribution of Scores for Information Literacy (Lower Division), Applicable Scores Only .....                   | 31 |
| Figure 4.a Distribution of Scores for Critical Thinking (lower division), Applicable Scores Only ..                         | 40 |
| Figure 4.b Distribution of Scores for Critical Thinking (Upper Division), Applicable Scores Only                            | 43 |
| Figure 5.a Thoughtful Expression (Oral) Results by Dimension for Applicable Scores Only.....                                | 54 |
| Figure 6.a Distribution of Scores for Thoughtful Expression (Written), Applicable Scores Only ..                            | 62 |

## LIST OF TABLES

|  |    |
|--|----|
| Table 2.1.a Interrater Reliability for Inquiry.....  | 14 |
| Table 2.1.b Inquiry Percent of Sample Scored at or Above 2 and 3.....                        | 15 |
| Table 2.2.a Interrater Reliability for Data Analysis.....                                    | 23 |
| Table 2.2.b Inquiry: Data Analysis Percent of Sample Scored at or Above 2 and 3 .....        | 24 |
| Table 3.a Interrater Reliability for Information Literacy.....                               | 34 |
| Table 3.b Information Literacy Percent of Sample Scored at or Above 2 and 3 .....            | 36 |
| Table 4.a Interrater Reliability for Critical Thinking .....                                 | 47 |
| Table 4.b Critical Thinking Percent of Sample Scored at Least Two and at Least Three .....   | 48 |
| Table 5.a Interrater Reliability for Thoughtful Expression (Oral) .....                      | 56 |
| Table 5.b Thoughtful Expression (Oral) Percent of Sample Scored at or Above 2 and 3 .....    | 57 |
| Table 6.a Interrater Reliability for Thoughtful Expression (Written).....                    | 65 |
| Table 6.b Thoughtful Expression (Written) Percent of Sample Scored at or Above 2 and 3 ..... | 66 |
| Table 7.a Percent of Work Products Meeting Performance Benchmarks by Grade Level.....        | 70 |
| Table 7.b Scorer Feedback on General Education Assessment Process .....                      | 72 |
| Table 7.c Interrater Reliability .....   | 74 |

## EXECUTIVE SUMMARY

This report provides the results of the General Education Assessment efforts for academic year 2013 – 2014. The UNCW Learning Goals were assessed within University Studies courses using AAC&U VALUE Rubrics and locally created rubrics, using the process recommended by the General Education Assessment Committee’s March 2009 recommendations. Five Learning Goals were assessed using student work products from 14 courses within seven University Studies components.

### FINDINGS

#### *INQUIRY*

In Spring 2014, 113 student work products were sampled from Historical and Philosophical Perspectives (PAR 101), Capstone Courses (EVS 495 and MGT 455), and Quantitative and Logical Reasoning (STT 210 and PSY 225). For lower-division courses, the majority of student work met the benchmark level of 2 for all dimensions scored. For the upper-division courses, scores were higher than in the lower-division courses, with the majority of work receiving a 3 or better on four of the six dimensions. A Data Analysis rubric was piloted for inquiry projects involving data analysis and used to score student work sampled from STT 210 and PSY 225. A large majority (over 93%) of student work, which was from lower-division courses, was scored at least a 2 on all three dimensions of the rubric.

#### *INFORMATION LITERACY*

In Fall 2013, 83 student work products from Composition (ENG 103), First Year Seminar (UNI 101), and Information Literacy (CSC 385) were scored on the UNCW Information Literacy rubric. The majority of lower-division student work achieved the benchmark score of 2 on all dimensions. While scores in the 300-level course were higher, a majority of work was scored at or above 3 on only two dimensions: IL1 Determining the Extent of Information Needed and IL5 Access and Use Information Ethically and Legally.

#### *CRITICAL THINKING*

In Spring 2014, 113 papers from six courses were scored on a rubric based on the VALUE Critical Thinking rubric that contains some minor modifications made based on past scorer feedback. In lower-division courses (HST 106 and PAR 101 from Historical and Philosophical Perspectives and MAT 142 from Quantitative and Logical Reasoning), the majority of work products met the lower-division benchmark on CT1 Explanation of Issues, CT2a Evidence: Selecting and Using, and CT5 Conclusions and Related Outcomes. In upper-division courses (ENG 303 from Quantitative and Logical Reasoning and MGT 455 and SWK 497 from Capstone Courses), the majority of work products were scored at least a three on CT1 and CT2a. Assessment results continue to indicate that critical thinking warrants additional instruction and practice across all course levels.

### *THOUGHTFUL EXPRESSION (ORAL)*

In Spring 2014, 56 oral presentations in COM 400, SWK 497, and ART 476 (all Capstone Courses) were scored using the VALUE Oral Communication rubric. While less than 20% of presentations met the benchmark for graduating seniors (level 4), at least 50% of all the oral presentations achieved a level 3 for all dimensions. Scores were highest on OC1 Organization and OC5 Central Message.

### *THOUGHTFUL EXPRESSION (WRITTEN)*

In Fall 2013, 70 student work products from ENG 103 (Composition) and UNI 101 (First Year Seminar) were scored on the AAC&U VALUE Written Communication rubric. The majority of papers met the lower-division benchmark level 2 for all dimensions of the rubric. Student achievement was highest on WC5 Control of Syntax and Mechanics and lowest on WC4 Sources and Evidence.

## **RECOMMENDATIONS**

The Learning Assessment Council recommends a university-wide concentration of efforts on understanding what critical thinking means, how it is being addressed at UNCW (within the University Studies curriculum, within the majors, and within co-curricular activities), and how to provide students with additional opportunities to develop their critical thinking skills. A committee has been appointed to plan a one-year Critical Thinking campaign that will be designed (1) to engage faculty, staff, students, and community members in exploring what it means to be a critical thinker and what critical thinking skills employers are looking for, and (2) to provide information and professional development on ways to incorporate additional opportunities for students to practice critical thinking skills in and outside the classroom.

# 1. BACKGROUND, SCOPE, AND METHODOLOGY

## BACKGROUND AND SCOPE

The University of North Carolina Wilmington Faculty Senate adopted nine UNCW Learning Goals in March 2009 (modified to [eight learning goals](#) in January 2011). The General Education Assessment process is based on the recommendations contained in the [Report of the General Education Assessment Committee](#) presented to the Provost and the Faculty Senate in March 2009. The Learning Assessment Council provides advice and feedback on the process, and recommendations based on the findings. For a complete background on the development of general education assessment at UNCW, see the *General Education Assessment Spring 2010 Report* (Siefert, 2010).

This report contains information on general education assessment activities for the academic year 2013 – 2014. In Fall 2013 and Spring 2014, the following learning goals were assessed: Information Literacy, Thoughtful Expression (Written), Inquiry, Critical Thinking, and Thoughtful Expression (Oral). This report outlines the methodology of and findings from six separate studies, and provides useful information on the abilities of UNCW students as measured through course-embedded assignments completed during their University Studies courses. This report also provides follow up information on the progress made on recommendations made last year and new recommendations.

## METHODOLOGY

For the purposes of this report, general education assessment activities in academic year 2013 – 2014 are divided into six areas: assessment of student learning in Inquiry, Inquiry Data Analysis, Information Literacy, Critical Thinking, Thoughtful Expression (Oral), and Thoughtful Expression (Written).

The following questions were examined:

- What are the overall abilities of students taking University Studies courses with regard to the UNCW Learning Goals of Inquiry, Information Literacy, Critical Thinking, Thoughtful Expression (Oral), and Thoughtful Expression (Written)?
- What are the relative strengths and weaknesses within the subskills of those goals?
- Are there any differences in performance based on course delivery method or demographic and preparedness variables, such as gender, race or ethnicity, transfer students vs. freshman admits, honors vs. non-honors students, total hours completed, or entrance test scores?

- What are the strengths and weaknesses of the assessment process itself?

UNCW has adopted an approach to assessing its Learning Goals that uses assignments that are a regular part of the course content. One strength of this approach is that the student work products are an authentic part of the curriculum, and hence there is a natural alignment often missing in standardized assessments. Students are motivated to perform at their best because the assignments are part of the course content and course grade. The assessment activities require little additional effort on the part of course faculty because the assignments used for the process are a regular part of the coursework. An additional strength of this method is the faculty collaboration and full participation in both the selection of the assignments and the scoring of the student work products.

The student work products collected for General Education Assessment are scored independently on a common rubric by trained scorers. The results of this scoring provide quantitative estimates of students' performance and qualitative descriptions of what each performance level looks like, which provides valuable information for the process of improvement. The normal disadvantage to this type of approach when compared to standardized tests is that results cannot be compared to other institutions. This disadvantage is mitigated in part by the use of the AAC&U VALUE rubrics for many of the Learning Goals. This concern is also addressed by the regular administration of standardized assessments, in particular, the CLA and the ETS Proficiency Profile, giving the university the opportunity to make national comparisons.

#### *ASSESSMENT TOOLS*

For the UNCW Learning Goals of Inquiry, Information Literacy, Critical Thinking, Thoughtful Expression (Oral), and Thoughtful Expression (Written), the Association of American Colleges and Universities (AAC&U) Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics (Rhodes, 2010) -were used, some with modifications based on scorer feedback in the past. The VALUE rubrics, part of the AAC&U Liberal Education and America's Promise (LEAP) initiative, were developed by over 100 faculty and other university professionals. Each rubric contains the common dimensions and most broadly shared characteristics of quality for each dimension.

A locally created rubric was used for assessing Inquiry Data Analysis. The versions of each of the rubrics that were used in the study are located in the appendices of each chapter.

#### *BENCHMARKS*

The VALUE rubrics and most locally created rubrics are designed on a 0 to 4 scale. According to AAC&U, "the capstone [4] level reflects the demonstration of achievement for the specific criterion for a student who graduates with a baccalaureate degree. Milestones [2 and 3] suggest key characteristics of progressive learning as students move from early in their college

experience to the completion of the baccalaureate degree” (Rhodes, 2010, p.2). Based on the design of these rubrics, UNCW uses the capstone level 4 as the benchmark for attainment of graduating seniors. For first- and second-year students assessed in lower-level general education courses, the milestone level 2 is the benchmark for achievement. The rationale for this is that performance at the milestone level 2 indicates that, given additional opportunities to learn and practice, they are on track for achieving a level 4 by the time of graduation. Most locally-created rubrics were designed to follow these same levels.

### ***SAMPLE SELECTION***

The sampling method used lays the foundation for the generalizability of the results. No one part of the University Studies curriculum, nor for that matter no one part of the university experience, is solely responsible for helping students meet UNCW Learning Goals. These skills are practiced in many courses. Each component of University Studies has its own student learning outcomes, and each of these outcomes is aligned to the Learning Goals. The University Studies Curriculum Map in Appendix A displays this alignment. For General Education Assessment purposes, courses are selected that not only meet the learning goals, but are also among those that are taken by a large number of students, in order to represent as much as possible the work of “typical” UNCW students. Within each course, sections are divided into those taught in the classroom and completely online, taught by full-time and part-time instructors, and taught as honors or regular sections. Within each subgroup, sections are selected randomly in quantities that represent as closely as possible the overall breakdown of sections by these criteria. Within each section, all student work products are collected, and random samples of the work products are selected (sometimes consisting of all papers).

Prior to the start of the semester, the General Education Assessment staff meets with course instructors to familiarize them with the relevant rubric(s). Instructors are asked to review their course content and assignments, and to select one assignment that they feel fits some or all of the dimensions of the rubric(s) being used.

Each student enrolled in the selected course sections fills out a Student Work Product Cover Sheet, which acknowledges the use of their work for the purpose of General Education Assessment. These cover sheets are removed before scoring. The name and student ID information on the cover sheets are matched with student demographic information in university records for the purpose of analysis based on demographic and preparedness variables.

### ***SCORING***

#### **Scorer Recruitment and Selection**

Scorers are recruited from UNCW faculty and, in some cases, teaching assistants. A recruitment email is sent to chairs, sometimes to all university chairs, and sometimes to only chairs in selected departments (based on the Learning Goals and course content being assessed), asking

them to forward the email to all full- and part-time faculty in their department. The desire is to include reviewers from a broad spectrum of departments. The intent is to give all faculty an opportunity to participate, to learn about the process and rubrics, and to see the learning students experience as they begin their programs. However, in some cases, the scoring is best done by discipline experts. It is also important to try to have a least one faculty member from each of the departments from which student work products were being reviewed. For the - 2013-2014 studies, scorers were solicited from all departments for all rubrics except Inquiry Data Analysis. For that rubric, scorers from Statistics and Psychology were asked to pilot the rubric. Scorers were selected from those expressing an interest to make up a broad-based panel consisting of full-time and part-time faculty.

### **Scoring Process**

Metarubrics, such as the VALUE rubrics, are constructed so that they can be used to score a variety of student artifacts across disciplines, across universities, and across preparation levels. Their strength is also a weakness: the generality of the rubric makes it more difficult to use than a rubric that is created for one specific assignment. To address this issue, a process must be created that not only introduces the rubric to the scorers, but also makes its use more manageable. The following describes the process for written work products; the process was similar for oral presentations, with the major difference being in the length of the norming session and the scorers' access to student work (for oral projects, scorers either observed the presentations in real-time or via video recording).

Volunteer scorers initially attended a two to two-and-a-half hour workshop on one rubric. During the workshop, scorers reviewed the rubric in detail and were introduced to the following assumptions adopted for applying the rubrics to basic studies work products.

#### **Initial assumptions**

1. When scoring, we are comparing each separate work product to the characteristics we want the work of UNCW graduates to demonstrate (considered to be Level 4).
2. Goals can be scored independently from each other.
3. Relative strengths and weaknesses within each goal emerge through seeking evidence for each dimension separately.
4. Common practice and the instructor's directions guide the scorer's interpretation of the rubric dimensions in relation to each assignment.
5. Additional assumptions will need to be made when each rubric is applied to individual assignments.

After reviewing the rubric and initial assumptions, the volunteers read and scored two to four student work products. Scoring was followed by a detailed discussion, so that scorers could better see the nuances of the rubric and learn what fellow scorers saw in the work products. From

these discussions, assumptions began to be developed for applying the rubric to each specific assignment.

For all the Learning Goals other than Thoughtful Expression (Oral), the work on common assignment-specific assumptions or guidelines was continued on the day of scoring. Scorers were assigned to groups. Scoring of each assignment began with the group scoring one student work product together and discussing their individual scores. Discussion clarified any implicit assumptions each scorer had used in scoring the first work product. From that discussion, each group created any assignment-specific assumptions that they would use for scoring the rest of the set of assignments. Learning-goal experts were available in each scoring event to answer any questions the faculty scorers had about scoring the student work product against the metarubric.

After completing a packet of work products, each scorer completed a rubric feedback form and turned in the assignment-specific assumptions used by the group. The feedback form asked for information on how well each rubric dimension fit the assignment and student work. It also asked for feedback on the quality criteria for each dimension. Scorers were also asked to complete an end-of-day survey to provide feedback on the entire process.

In order to measure the consistency of the application of the rubric, additional common work products were included in each packet for measuring interrater reliability.

For Thoughtful Expression (Oral), on the first day of scoring, scorers observed two oral presentations, scored them, and immediately after met to discuss scorers.

## 2.1 INQUIRY AND ANALYSIS

The UNCW Inquiry Learning Goal is for students to engage in rigorous, open-minded and imaginative inquiry. For purposes of this Learning Goal, inquiry is the systematic and analytic investigation of an issue or problem with the goal of discovery. Inquiry involves the clear statement of the problem, issue or question to be investigated; examination of relevant existing knowledge; design of an investigation process; analysis of the complexities of the problem, clear rationale supporting conclusions; and identification of limitations of the analysis (UNCW Learning Goals, 2011). The VALUE Inquiry rubric contains six dimensions that are aligned with the UNCW description of Inquiry (see the rubric in Appendix 2.1.A at the end of this chapter). Twelve components of University Studies have at least one student learning outcome that is aligned to Inquiry. For this study, the courses were selected from Historical and Philosophical Perspectives and Capstone Courses.

### LOWER DIVISION

#### *SUMMARY OF SCORES BY DIMENSION*

Three faculty scorers scored twenty-eight work products from one assignment and one course: PAR 101. The assignment was an argumentative essay. Fourteen work products (50 %) were scored by multiple scorers for norming and interrater reliability purposes. Figure 2.1.a provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

## INQUIRY AND ANALYSIS (LOWER DIVISION) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY

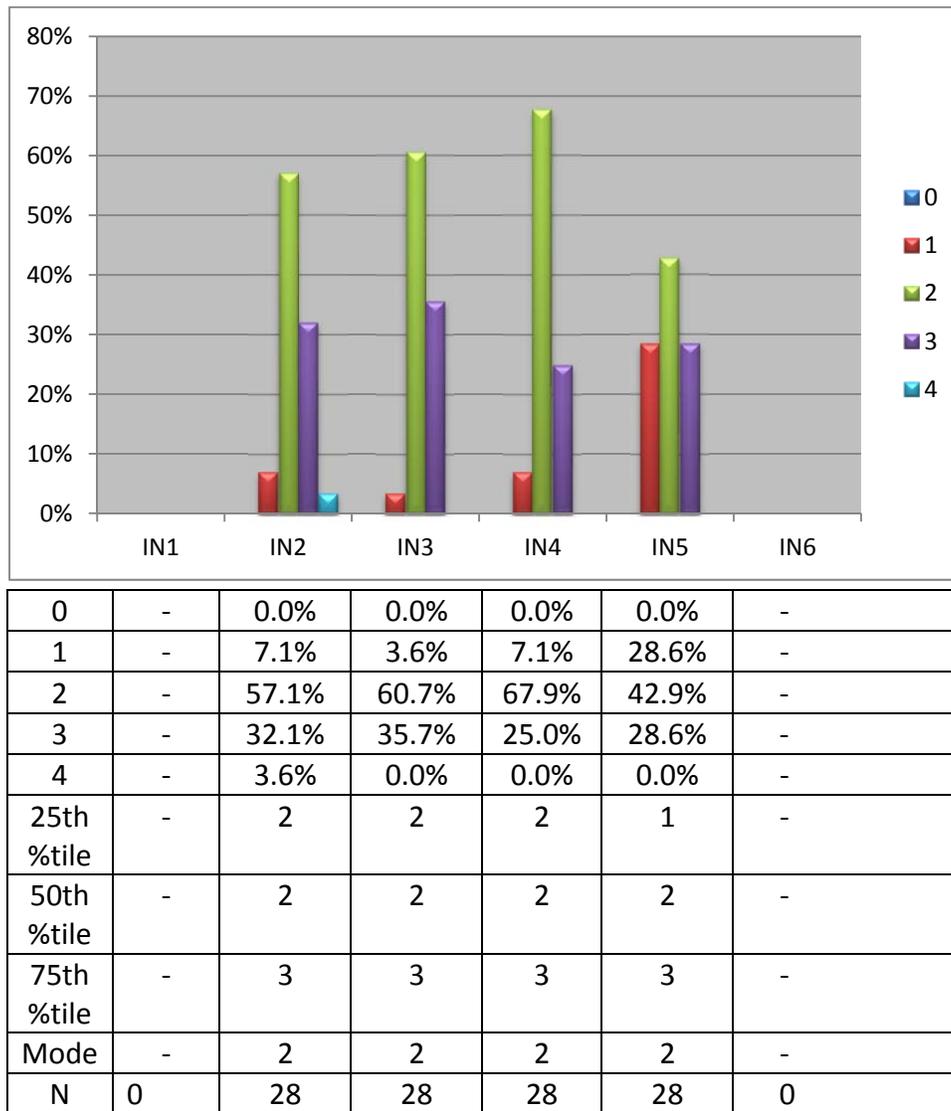


Figure 2.1.a Distributions of scores for Inquiry and Analysis (lower division), applicable scores only

### *RESULTS BY DIMENSION*

#### **IN 1 Topic Selection**

No work products were scored on the IN 1 criteria as the topic was selected for the students.

#### **IN 2 Existing Knowledge, Research and/or views**

No work products scored a zero on this dimension. Two work products presented information from irrelevant sources representing limited points of views/approaches (scores of one). Sixteen work products presented information from relevant sources representing limited points of view/approaches (scores of two). Just over a third of the work products presented in-depth

information from relevant sources representing various points of view/approaches (scores of three). One work product synthesized in-depth information from relevant sources representing various points of view/approaches (score of four).

### **IN 3 Design Process**

In one work product, the inquiry design demonstrated a misunderstanding of the methodology or theoretical framework (score of one). For seventeen work products, the critical elements of the methodology or theoretical framework were missing, incorrectly developed, or unfocused (scores of two). For ten work products, the critical elements of the methodology or theoretical framework were appropriately developed; however, more subtle elements were ignored or unaccounted for (scores of three). None of the work products achieved the capstone score of four.

### **IN 4 Analysis**

Two work products listed evidence, but it was not organized and/or was unrelated to focus (scores of one). Nineteen work products organized evidence, but the organization was not effective in revealing important patterns, differences, or similarities (scores of two). A quarter of the work products organized evidence to reveal important patterns, differences, or similarities related to focus (scores of three). No work products achieved the capstone score of four.

### **IN 5 Conclusions**

Eight work products stated an ambiguous, illogical, or unsupportable conclusion from inquiry findings (score of 1). Twelve work products stated a general conclusion that, because it was so general, also applied beyond the scope of the inquiry findings (scores of two). Eight work products stated a conclusion focused solely on the inquiry findings. The conclusion arose specifically from and responded specifically to the inquiry findings (scores of three). No work products achieved the capstone score of four.

### **IN 6 Limitations and Implications**

No work products were scored on the IN 6 criteria.

#### *CORRELATION BETWEEN DIMENSIONS*

All dimensions were correlated at the .01 level of significance. See Appendix 2.2.B for the complete presentation of dimension correlation coefficients.

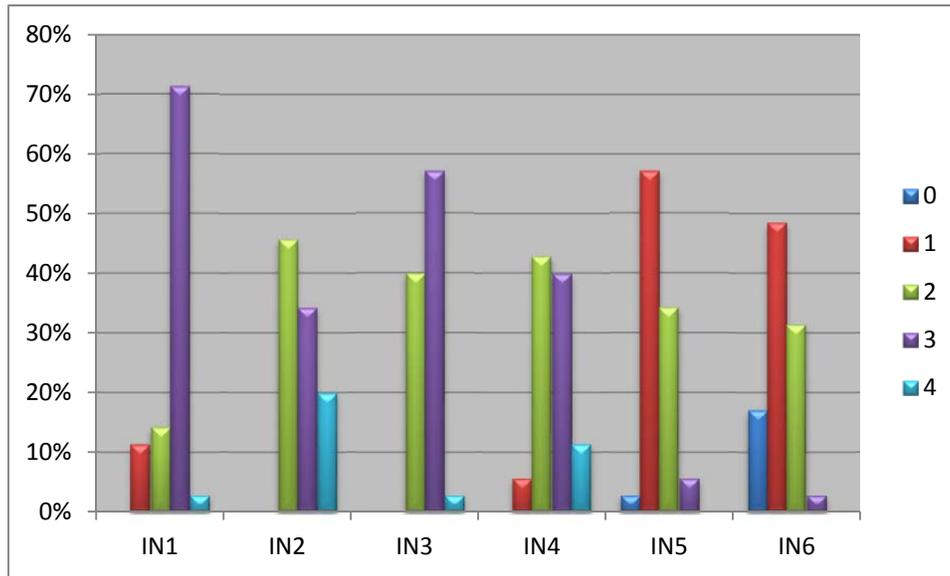
## **UPPER DIVISION**

### *SUMMARY OF SCORES BY DIMENSION*

Seven faculty scorers scored thirty-five work products from two assignments across two upper-division courses: EVS 495 and MGT 455. Both EVS 495 and MGT 455 are from the Capstone component of University Studies. The assignments included projects with case analyses and research papers with presentations. Twenty-one work products (60 %) were scored by multiple

scorers for norming or interrater reliability. Figures 2.1.b provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

**INQUIRY AND ANALYSIS (UPPER DIVISION) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY**



|            |       |       |       |       |       |       |
|------------|-------|-------|-------|-------|-------|-------|
| 0          | 0%    | 0%    | 0%    | 0%    | 2.9%  | 17.1% |
| 1          | 11.4% | 0%    | 0%    | 5.7%  | 57.1% | 48.6% |
| 2          | 14.3% | 45.7% | 40.0% | 42.9% | 34.3% | 31.4% |
| 3          | 71.4% | 34.3% | 57.1% | 40.0% | 5.7%  | 2.9%  |
| 4          | 2.9%  | 20.0% | 2.9%  | 11.4% | 0%    | 0%    |
| 25th %tile | 2     | 2     | 2     | 2     | 2     | 2     |
| 50th %tile | 3     | 3     | 3     | 3     | 2     | 2     |
| 75th %tile | 3     | 3     | 3     | 3     | 3     | 3     |
| Mode       | 3     | 2     | 3     | 2     | 2     | 2     |
| N          | 35    | 35    | 35    | 35    | 35    | 35    |

Figure 2.1.b Distributions of scores for Inquiry and Analysis (upper division), applicable scores only

**RESULTS BY DIMENSION**

**IN 1 Topic Selection**

Four work products identified a topic that was far too general and wide-ranging as to be manageable and doable (score of one). Five work products identified a topic that while manageable/doable, was too narrowly focused and left out relevant aspects of the topic (scores of

two). Twenty-five work products identified a focused and manageable/doable topic that appropriately addressed relevant aspects of the topic (scores of three), and one work product identified a creative, focused, and manageable topic that addressed potentially significant yet previously less-explored aspects of the topic (score of four).

## **IN 2 Existing Knowledge, Research and/or Views**

No work products presented information from irrelevant sources representing limited points of views/approaches (score of one). Sixteen work products presented information from relevant sources representing limited points of view/approaches (scores of two). Just over a third of the work products presented in-depth information from relevant sources representing various points of view/approaches (scores of three). Seven work product synthesized in-depth information from relevant sources representing various points of view/approaches (scores of four).

## **IN 3 Design Process**

None of the work products scored a zero or included an inquiry design that demonstrated a misunderstanding of the methodology or theoretical framework (score of one). For fourteen work products, the critical elements of the methodology or theoretical framework were missing, incorrectly developed, or unfocused (scores of two). For twenty work products, the critical elements of the methodology or theoretical framework were appropriately developed; however, more subtle elements were ignored or unaccounted for (scores of three). In one work product, all elements of the methodology or theoretical framework were skillfully developed. Appropriate methodology or theoretical frameworks could be synthesized from across disciplines or from relevant sub disciplines (score of four).

## **IN 4 Analysis**

Two work products listed evidence, but it was not organized and/or was unrelated to focus (scores of one). Fifteen work products organized evidence, but the organization was not effective in revealing important patterns, differences, or similarities (scores of two). Fourteen work products organized evidence to reveal important patterns, differences, or similarities related to focus (scores of three). Four work products organized and synthesized evidence to reveal insightful patterns, differences, or similarities related to focus (scores of four).

## **IN 5 Conclusions**

One work product did not reach the Benchmark level one of scoring (score of zero). Twenty work products stated an ambiguous, illogical, or unsupported conclusion from inquiry findings (score of one). Just over a third of the work products stated a general conclusion that, because it was so general, also applied beyond the scope of the inquiry findings (scores of two). Two work products stated a conclusion focused solely on the inquiry findings. The conclusion arose specifically from and responded specifically to the inquiry findings (scores of three). No work products achieved the capstone score of four.

## **IN 6 Limitations and Implications**

Six work products did not meet the Benchmark level one of scoring. Slightly less than half of the work products presented limitations and implications, but they were possibly irrelevant and unsupported (scores of one). Eleven work products presented relevant and supported limitations and implications (scores of two). One work product discussed relevant and supported limitations and implications (score of three). No work products reach the Capstone level of scoring (score of four).

### *CORRELATION BETWEEN DIMENSIONS*

All dimensions were correlated at the .01 or .05 level of significance. See Appendix 2.B for the complete presentation of dimension correlation coefficients.

## **ANALYSIS ACROSS COURSE LEVELS**

### *DEMOGRAPHIC AND PREPAREDNESS FINDINGS*

There were no statistically significant differences between the means, medians, and score distributions of male vs. female students, transfer vs. freshman-start students, honors vs. non-honors students, or ethnicity groups. There were no Isaac Bear Early College students in the sample so comparisons could not be made between the scores from Isaac Bear vs. non-Isaac Bear students.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, those having completed 0 – 30 credit hours, 31 – 60 credit hours, 61 – 90, and over 90 credit hours. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed no statistically significant differences between the groups. Looking at Spearman rho correlation coefficients, the number of total hours completed was significantly positively correlated with IN1 (.418\*). There were no other statistically significant correlations found between scores on the rubric dimensions and GPA, SAT scores, and ACT scores.

### *COMPARISONS BETWEEN CRITERIA*

Comparisons were made across a number of criteria. Work was collected from courses taught by both tenure-line faculty and part-time faculty; there was no significant difference in scores on work collected from the two types of faculty. There was, however, a difference in scores between upper- and lower-division courses on all the dimensions scored for both divisions (IN1 and IN6 were only scored on upper-division work). For IN2, IN3, IN4, and IN5, work from upper-division courses scored higher. Because the upper and lower division categories also reflect the University Studies component categories, the same results were seen when the scores from various components were compared: work from Capstone Courses scored higher than did work sampled from the Historical and Philosophical Perspectives component.

### *INTERRATER RELIABILITY*

There were a number of common papers scored between each pair of faculty scorers so that interrater reliability could be assessed (27 out of 63, or 42.9% of the total number of papers, though not all dimensions were necessarily scored for all common papers). Table 2.1 shows the reliability measures for Inquiry.

Table 2.1.a Interrater Reliability for Inquiry

| Dimension                                      | N  | Percent Agreement | Plus Percent Adjacent | Krippendorff's alpha |
|--|----|-------------------|-----------------------|----------------------|
| IN 1 Topic Selection                           | 17 | 76.5%             | 88.2%                 | -.2667               |
| IN 2 Existing Knowledge, Research and/or Views | 27 | 44.4%             | 96.3%                 | .6230                |
| IN 3 Design Process                            | 27 | 88.2%             | 96.3%                 | .4977                |
| IN 4 Analysis                                  | 27 | 59.3%             | 96.3%                 | .6595                |
| IN 5 Conclusions                               | 27 | 40.7%             | 88.9%                 | .4354                |
| IN 6 Limitations and Implications              | 17 | 52.9%             | 94.1%                 | -.2711               |

Interrater reliability is a measure of the degree of agreement between scorers, and provides information about the trustworthiness of the data. It helps answer the question—would a different set of scorers at a different time arrive at the same conclusions? In practice, interrater reliability is enhanced over time through scorer discussion, as well as through improvements to the scoring rubric. Percent Agreement, Percent Agreement Plus Adjacent, and Krippendorff's Alpha measure scorer agreement. The UNCW benchmark is .67 for Krippendorff's Alpha. See Appendix B of this report for a more complete discussion of these statistics and the determination of benchmark levels.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff's Alpha, there are no dimensions of the rubric that meets these standards, though IN 2 Existing Knowledge, Research and/or Views and IN 4 Analysis come close. Also, there is some indication that the small sample size for IN1 and IN6 may invalidate the calculation. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that all dimensions had at least 88% of scores in agreement or within one level of each other.

### *SCORER FEEDBACK ON RUBRIC AND SCORING*

Scorers were asked to provide feedback about the assignment’s alignment to the rubric and, following scoring, the application of the rubric to the assignment. Scorers felt that, in general, the assignments required students to incorporate existing knowledge and/or research about a topic into their work and that the methodology the students were to use for the assignment was clear. Students were not always asked to analyze their own findings within the assignment or to state conclusions and/or implications from their inquiry process. Scorers reported that, in particular, IN 6 Limitations and Implications was difficult to apply to scoring. The majority of scorers’ comments included that dimension IN 6 was actually measuring a number of fuzzy ideas and that perhaps the terms could be defined more explicitly.

### *DISCUSSION*

Table 2.1.b shows the percent of work products scored at or above the benchmark levels.

Table 2.1.b Inquiry Percent of Sample Scored at or Above 2 and 3

| Lower-Division Courses                         |   |   |
|--|---|---|
| Dimension                                      | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| IN 1 Topic Selection                           | --                                      | --  |
| IN 2 Existing Knowledge, Research and/or Views | 92.9%                                   | 35.7%                                     |
| IN 3 Design Process                            | 96.4%                                   | 37.0%                                     |
| IN 4 Analysis                                  | 92.9%                                   | 25.0%                                     |
| IN 5 Conclusions                               | 71.4%                                   | 28.6%                                     |
| IN 6 Limitations and Implications              | --                                      | --  |
| Upper-Division Courses                         |   |   |
| Dimension                                      | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| IN 1 Topic Selection                           | 88.6%                                   | 74.3%                                     |
| IN 2 Existing Knowledge, Research and/or Views | 100%                                    | 54.3%                                     |
| IN 3 Design Process                            | 100%                                    | 60.0%                                     |
| IN 4 Analysis                                  | 94.3%                                   | 51.4%                                     |
| IN 5 Conclusions                               | 97.1%                                   | 40.0%                                     |
| IN 6 Limitations and Implications              | 82.9%                                   | 34.3%                                     |

For the lower-division courses, the benchmark achievement is defined as rubric score level two. A majority of work products from the lower-division courses were scored at this benchmark level or better. A fourth to a third of students scored above the benchmark. For the capstone courses, the benchmark is a score of four. The percent of scores of four is presented in Figure 2.1b. Scores at or above two and three are presented here in order analyze growth between lower-division and upper-division courses. The results show that the percent of students scoring at least a two and at least a three is higher in the capstone courses. However, the findings also indicate that most students are not reaching the graduating senior benchmark. There was a

statistical difference in scores between upper- and lower-division courses on the four dimensions scored for both, with work from upper-division courses scoring higher.

## APPENDIX 2.1.A INQUIRY AND ANALYSIS RUBRIC



### INQUIRY AND ANALYSIS VALUE RUBRIC

*for more information, please contact [value@aacu.org](mailto:value@aacu.org)*

#### Definition

Inquiry is the ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively and responsibly use and share that information for the problem at hand. – The National Forum on Information Literacy

*Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.*

|   | Benchmark<br>1   | Milestones   |  | Capstone<br>4  | Score |
|---|--|--|--|--|-------|
|   |  | 2  | 3  |  |       |
| <b>Topic selection</b>                            | Identifies a topic that is far too general and wide-ranging as to be manageable and doable.  | Identifies a topic that while manageable/doable, is too narrowly focused and leaves out relevant aspects of the topic.   | Identifies a focused and manageable/doable topic that appropriately addresses relevant aspects of the topic.   | Identifies a creative, focused, and manageable topic that addresses potentially significant yet previously less-explored aspects of the topic.   |       |
| <b>Existing Knowledge, Research, and/or Views</b> | Presents information from irrelevant sources representing limited points of view/approaches. | Presents information from relevant sources representing limited points of view/approaches.                               | Presents in-depth information from relevant sources representing various points of view/approaches.  | Synthesizes in-depth information from relevant sources representing various points of view/approaches.   |       |
| <b>Design Process</b>                             | Inquiry design demonstrates a misunderstanding of the methodology or theoretical framework.  | Critical elements of the methodology or theoretical framework are missing, incorrectly developed, or unfocused.          | Critical elements of the methodology or theoretical framework are appropriately developed, however, more subtle elements are ignored or unaccounted for. | All elements of the methodology or theoretical framework are skillfully developed. Appropriate methodology or theoretical frameworks may be synthesized from across disciplines or from relevant subdisciplines. |       |
| <b>Analysis</b>                                   | Lists evidence, but it is not organized and/or is unrelated to focus.                        | Organizes evidence, but the organization is not effective in revealing important patterns, differences, or similarities. | Organizes evidence to reveal important patterns, differences, or similarities related to focus.  | Organizes and synthesizes evidence to reveal insightful patterns, differences, or similarities related to focus.   |       |
| <b>Conclusions</b>                                | States an ambiguous, illogical, or unsupported conclusion from inquiry findings.             | States a general conclusion that, because it is so general, also applies beyond the scope of the inquiry findings.       | States a conclusion focused solely on the inquiry findings. The conclusion arises specifically from and responds specifically to the inquiry findings.   | States a conclusion that is a logical extrapolation from the inquiry findings.   |       |
| <b>Limitations and Implications</b>               | Presents limitations and implications, but they are possibly irrelevant and unsupported.     | Presents relevant and supported limitations and implications.  | Discusses relevant and supported limitations and implications.   | Insightfully discusses in detail relevant and supported limitations and implications.  |       |

## APPENDIX 2.1.B INQUIRY AND ANALYSIS CORRELATION COEFFICIENTS

### *LOWER DIVISION*

|                |     |                         | IN1 | IN2    | IN3    | IN4    | IN5    | IN6 |
|----------------|-----|-------------------------|-----|--------|--------|--------|--------|-----|
| Spearman's rho | IN1 | Correlation Coefficient | .   | .      | .      | .      | .      | .   |
|                |     | Sig. (2-tailed)         | .   | .      | .      | .      | .      | .   |
|                |     | N                       | 0   | 0      | 0      | 0      | 0      | 0   |
|                | IN2 | Correlation Coefficient | .   | 1.000  | .539** | .645** | .612** | .   |
|                |     | Sig. (2-tailed)         | .   | .      | .003   | .000   | .001   | .   |
|                |     | N                       | 0   | 28     | 28     | 28     | 28     | 0   |
|                | IN3 | Correlation Coefficient | .   | .539** | 1.000  | .650** | .523** | .   |
|                |     | Sig. (2-tailed)         | .   | .003   | .      | .000   | .004   | .   |
|                |     | N                       | 0   | 28     | 28     | 28     | 28     | 0   |
|                | IN4 | Correlation Coefficient | .   | .645** | .650** | 1.000  | .539** | .   |
|                |     | Sig. (2-tailed)         | .   | .000   | .000   | .      | .003   | .   |
|                |     | N                       | 0   | 28     | 28     | 28     | 28     | 0   |
|                | IN5 | Correlation Coefficient | .   | .612** | .523** | .539** | 1.000  | .   |
|                |     | Sig. (2-tailed)         | .   | .001   | .004   | .003   | .      | .   |
|                |     | N                       | 0   | 28     | 28     | 28     | 28     | 0   |
|                | IN6 | Correlation Coefficient | .   | .      | .      | .      | .      | .   |
|                |     | Sig. (2-tailed)         | .   | .      | .      | .      | .      | .   |
|                |     | N                       | 0   | 0      | 0      | 0      | 0      | 0   |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

*UPPER DIVISION*

|                |     |                         | IN1    | IN2    | IN3    | IN4    | IN5    | IN6    |
|----------------|-----|-------------------------|--------|--------|--------|--------|--------|--------|
| Spearman's rho | IN1 | Correlation Coefficient | 1.000  | .543** | .481*  | .464** | .352*  | .367*  |
|                |     | Sig. (2-tailed)         | .      | .001   | .003   | .005   | .038   | .030   |
|                |     | N                       | 35     | 35     | 35     | 35     | 35     | 35     |
|                | IN2 | Correlation Coefficient | .543** | 1.000  | .557** | .595** | .508** | .645** |
|                |     | Sig. (2-tailed)         | .001   | .      | .001   | .000   | .002   | .000   |
|                |     | N                       | 35     | 35     | 35     | 35     | 35     | 35     |
|                | IN3 | Correlation Coefficient | .481** | .557** | 1.000  | .737** | .498** | .482** |
|                |     | Sig. (2-tailed)         | .003   | .001   | .      | .000   | .002   | .003   |
|                |     | N                       | 35     | 35     | 35     | 35     | 35     | 35     |
|                | IN4 | Correlation Coefficient | .464** | .595** | .737** | 1.000  | .652** | .750** |
|                |     | Sig. (2-tailed)         | .005   | .000   | .000   | .      | .000   | .000   |
|                |     | N                       | 35     | 35     | 35     | 35     | 35     | 35     |
|                | IN5 | Correlation Coefficient | .352*  | .508** | .498** | .652** | 1.000  | .681** |
|                |     | Sig. (2-tailed)         | .038   | .002   | .002   | .000   | .      | .000   |
|                |     | N                       | 35     | 35     | 35     | 35     | 35     | 35     |
|                | IN6 | Correlation Coefficient | .367*  | .645** | .482** | .750** | .681** | 1.000  |
|                |     | Sig. (2-tailed)         | .030   | .000   | .003   | .000   | .000   | .      |
|                |     | N                       | 35     | 35     | 35     | 35     | 35     | 35     |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

## 2.2 INQUIRY AND ANALYSIS: DATA ANALYSIS

The UNCW Inquiry Learning Goal is for students to engage in rigorous, open-minded and imaginative inquiry. For purposes of this Learning Goal, inquiry is the systematic and analytic investigation of an issue or problem with the goal of discovery. Inquiry involves the clear statement of the problem, issue or question to be investigated; examination of relevant existing knowledge; design of an investigation process; analysis of the complexities of the problem, clear rationale supporting conclusions; and identification of limitations of the analysis (UNCW Learning Goals, 2011). This was the first use of a pilot Data Analysis rubric containing three dimensions that are aligned with the UNCW description of Inquiry (see the rubric in Appendix 2.2.A at the end of this chapter) and that can be used alone or in conjunction with the Inquiry and Analysis rubric, replacing IN 4 and IN 5: Analysis and Conclusions dimensions. Twelve components of University Studies have at least one student learning outcome that is aligned to Inquiry. For this study, the courses were selected from Quantitative and Logical Reasoning.

### *SUMMARY OF SCORES BY DIMENSION*

Two faculty scorers scored forty-eight work products from two assignments and two courses: STT 210 and PSY 225. The assignments consisted of exams. Nine work products (20 %) were scored by multiple scorers for the purposes of norming and determining interrater reliability. Figures 2.2.a provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

## INQUIRY AND ANALYSIS: DATA ANALYSIS RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY

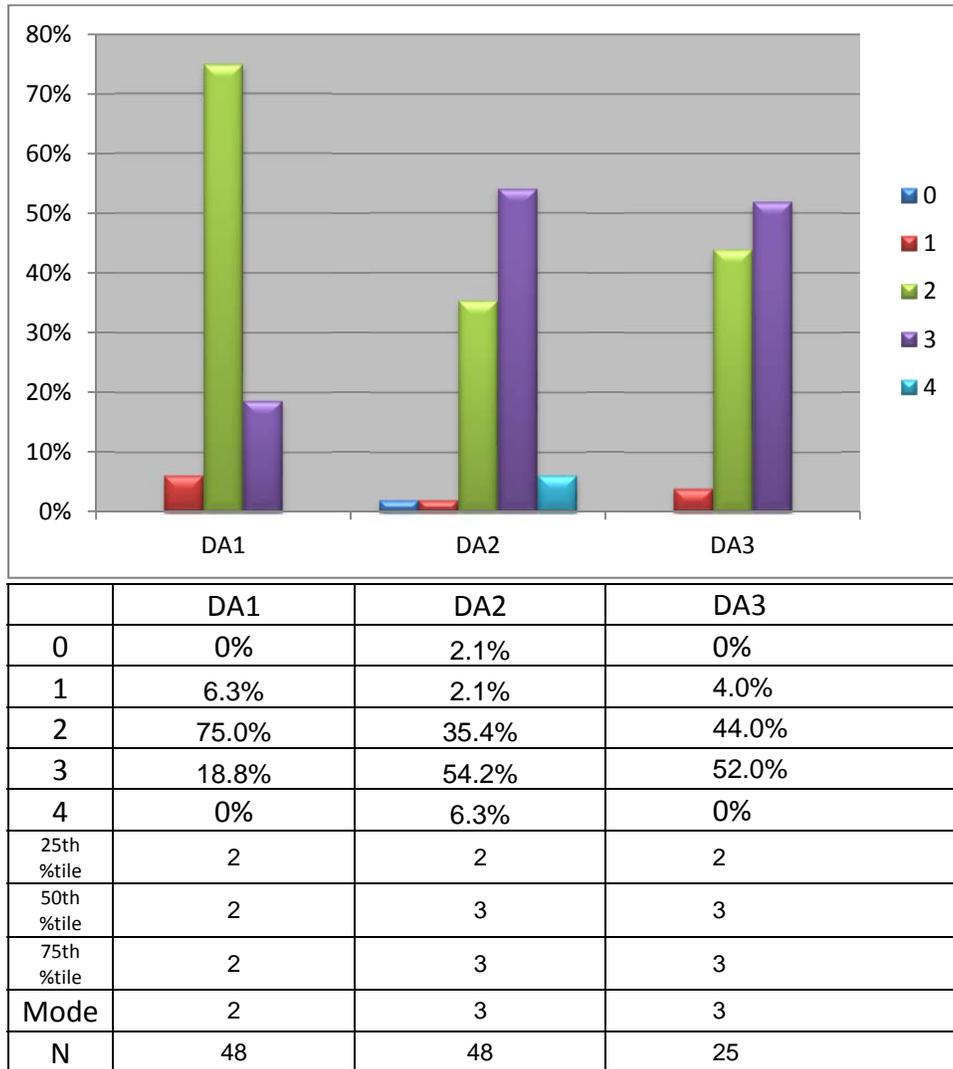


Figure 2.2.a Distributions of scores for Inquiry and Analysis: Data Analysis (lower division), applicable scores only

### *RESULTS BY DIMENSION*

#### **DA 1 Summarizing and Analyzing Data**

No work products scored a zero on this dimension. Three work products converted some data into mathematical portrayal, but with inaccurate or inappropriate results. Three-fourths of the student work converted data into a mathematical portrayal that was partially accurate or appropriate. Nine papers converted the relevant data into a complete mathematical portrayal that was both appropriate and accurate. There were no scores of four on DA 1.

## **DA 2 Explanation of Results**

One paper received a score of zero on this dimension and one paper received a score of one (attempted to explain information presented in mathematical and/or statistical forms, but explanation was inaccurate). Almost one in twenty papers provided partially accurate explanations of information presented in mathematical and/or statistical forms. One-quarter of student work provided mostly accurate explanations of information presented in mathematical and/or statistical forms. Finally, three work products provided accurate explanations of information presented in mathematical and/or statistical forms.

## **DA 3 Inferences**

DA 3 was deemed applicable for twenty three work products. No work products were scored as a level one or a level four. A single work product scored a level one (made inferences that are not supported by the mathematical and/or statistical analysis). Eleven work products received a score of two (made inferences that are partially supported by the mathematical and/or statistical analysis). The majority of work products (thirteen) received a score of three, indicating that the papers made reasonable inferences that were supported by the mathematical and/or statistical analysis.

### *CORRELATION BETWEEN DIMENSIONS*

DA2 was correlated with DA1 and DA 3 at the .01 level of significance. DA1 and DA3 were not significantly correlated with each other. See Appendix 2.2.B for the complete presentation of dimension correlation coefficients.

### *DEMOGRAPHIC AND PREPAREDNESS FINDINGS*

There were no statistically significant differences between the means, medians, and score distributions for different ethnicity categories, transfer vs. freshman-start students, and honors vs. non-honors students. There were no Isaac Bear students in the sample so comparisons could not be made between scores of Isaac Bear students vs. non-Isaac Bear students. There was a statistically significant difference in the scores of males vs. females on DA2, with females scoring higher.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, those having completed 0 – 30 credit hours, 31 – 60 credit hours, 61 – 90, and over 90 credit hours. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed no statistically significant differences between the groups. Looking at Spearman rho correlation coefficients, the number of total hours completed was negatively correlated with DA3 (-.446\*). Other statistically significant correlations were found between GPA and DA1 (.398\*\*), DA2 (.525\*\*), and DA3 (.446\*).

### *COMPARISONS BETWEEN CRITERIA*

The courses from which the student work was sampled were similar in that they were from the same University Studies component and they were both lower-division courses, so comparisons could not be made across these criteria.

### *INTERRATER RELIABILITY*

There were a number of common papers scored between each pair of faculty scorers so that interrater reliability could be assessed (9 out of 48, or 18.8%). Table 2.2.a shows the reliability measures for Data Analysis.

Table 2.2.a Interrater Reliability for Data Analysis

| Dimension                           | N | Percent Agreement | Plus Percent Adjacent | Krippendorff's alpha |
|-------------------------------------|---|-------------------|-----------------------|----------------------|
| DA 1 Summarizing and Analyzing Data | 7 | 57.1%             | 100%                  | -.1818               |
| DA 2 Explanation of Results         | 7 | 57.1%             | 85.7%                 | -.0873               |
| DA 3 Inferences                     | 7 | 28.6%             | 100%                  | .0582                |

Interrater reliability is a measure of the degree of agreement between scorers, and provides information about the trustworthiness of the data. It helps answer the question—would a different set of scorers at a different time arrive at the same conclusions? In practice, interrater reliability is enhanced over time through scorer discussion, as well as through improvements to the scoring rubric. Percent Agreement, Percent Agreement Plus Adjacent, and Krippendorff's Alpha measure scorer agreement. The UNCW benchmark is .67 for Krippendorff's Alpha. See Appendix B of this report for a more complete discussion of these statistics and the determination of benchmark levels.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff's Alpha, there are no dimensions of the rubric that meets these standards. However, there is some indication that the small sample size may invalidate the calculation. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that all dimensions had at least 85% of scores in agreement or within one level of each other.

### *SCORER FEEDBACK ON RUBRIC AND SCORING*

Scorers were asked to provide feedback about the assignment's alignment to the rubric and, following scoring, the application of the rubric to the assignment. Scorers reported that both assignments provided precise instructions for how students should apply methodology and

analysis. Students were also asked to state the conclusions of the findings but not the implications. Scorers felt that for the application of DA 3 to these particular assignments, both score level 3 and score level 4 indicate that students achieved a correct answer, but that students did not have the opportunity to earn up to a level four, given the nature of the assignment.

*DISCUSSION*

Table 2.2.b shows the percent of work products scored at or above the benchmark levels.

Table 2.2.b Inquiry: Data Analysis Percent of Sample Scored at or Above 2 and 3

| Lower-Division Courses              |   |   |
|-------------------------------------|---|---|
| Dimension                           | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| DA 1 Summarizing and Analyzing Data | 93.8%                                   | 18.8%                                     |
| DA 2 Explanation of Results         | 95.8%                                   | 60.4%                                     |
| DA 3 Inferences                     | 96.0%                                   | 52.0%                                     |

For lower-division courses, the benchmark achievement is defined as rubric score level two. A very high percentage of work products scored at this benchmark level (only lower-division work products were sampled).

There were no statistical differences in scores between student class level. There was a significant negative correlation between total number of credit hours completed DA3. This, combined with the scorer feedback that DA3 was somewhat difficult to score in the two higher score levels, may point to the need for a closer look at the dimension criteria for DA3.

## APPENDIX 2.2.A INQUIRY AND ANALYSIS: DATA ANALYSIS RUBRIC

### DATA ANALYSIS RUBRIC

Can be used alone or in conjunction with the AAC&U VALUE Inquiry or Analysis rubric, replacing the Analysis (and Conclusions) dimensions

*Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.*

|                                       | <b>Benchmark<br/>1</b>   | <b>Milestones</b>   |  | <b>Capstone<br/>4</b>   | <b>Score</b> |
|---------------------------------------|--|---|--|---|--------------|
|                                       |  | 2   | 3  |   |              |
| <b>Summarizing and Analyzing Data</b> | Converts some information (data) into a mathematical portrayal, but results are inappropriate or inaccurate.       | Converts some information (data) into a mathematical portrayal that is partially appropriate or accurate.   | Converts some information (data) into a mathematical portrayal that is appropriate or accurate.          | Converts some information (data) into an insightful mathematical portrayal in a way that contributes to deep understanding. |              |
| <b>Explanation of Results</b>         | Attempts to explain information presented in mathematical and/or statistical forms, but explanation is inaccurate. | Provides partially accurate explanations of information presented in mathematical and/or statistical forms. | Provides mostly accurate explanations of information presented in mathematical and/or statistical forms. | Provides accurate explanations of information presented in mathematical and/or statistical form.                            |              |
| <b>Inferences</b>                     | Makes inferences that are not supported by the mathematical and/or statistical analysis.                           | Makes inferences that are partially supported by the mathematical and/or statistical analysis.              | Makes reasonable inferences that are supported by the mathematical and/or statistical analysis.          | Makes carefully considered, appropriate and accurate inferences based on the mathematical and/or statistical analysis.      |              |

NOTES:

## APPENDIX 2.2.B INQUIRY AND ANALYSIS: DATA ANALYSIS CORRELATION COEFFICIENTS

|                |     |                         | DA1    | DA2    | DA3    |
|----------------|-----|-------------------------|--------|--------|--------|
| Spearman's rho | DA1 | Correlation Coefficient | 1.000  | .474** | .354   |
|                |     | Sig. (2-tailed)         | .      | .001   | .083   |
|                |     | N                       | 48     | 48     | 25     |
|                | DA2 | Correlation Coefficient | .474** | 1.000  | .539** |
|                |     | Sig. (2-tailed)         | .001   | .      | .005   |
|                |     | N                       | 48     | 48     | 25     |
|                | DA3 | Correlation Coefficient | .354   | .539** | 1.000  |
|                |     | Sig. (2-tailed)         | .083   | .005   | .      |
|                |     | N                       | 25     | 25     | 25     |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

### 3. INFORMATION LITERACY

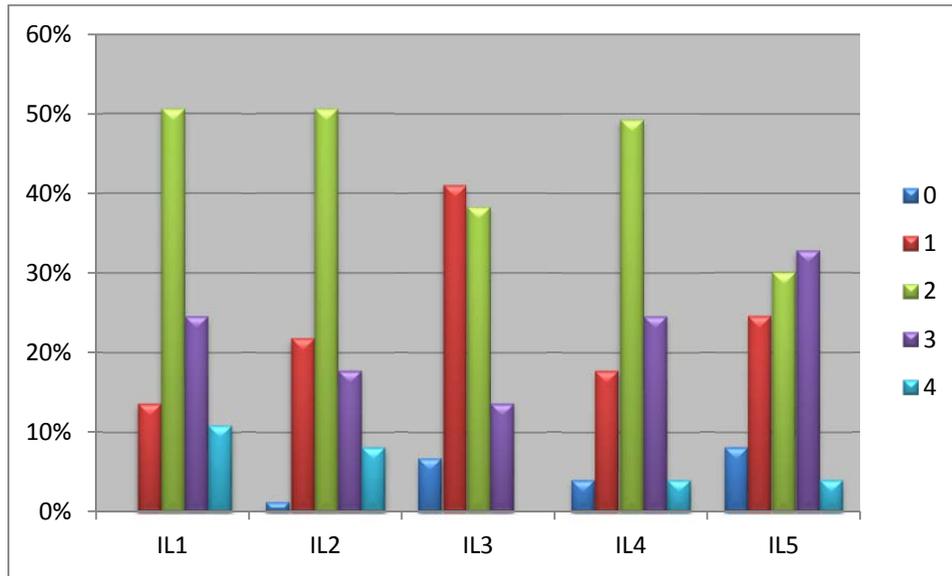
The UNCW Information Literacy learning goal is for students to locate, evaluate, and effectively use information by applying a variety of academic and technological skills. For purposes of this learning goal, information literacy is characterized by the ability to determine the extent of the information needed, accessing the needed information, critically evaluating the information, organizing the information to accomplish a specific purpose, and using the information ethically and legally (UNCW Learning Goals, 2011). The Information Literacy rubric is based largely on the AAC&U VALUE rubric and details five dimensions of knowledge related to information literacy. The Information Literacy rubric can be found in Appendix 3.A at the end of this chapter. In this study, work was sampled from the First Year Seminar, Information Literacy, and Composition components of University Studies.

#### LOWER DIVISION

##### *SUMMARY OF SCORES BY DIMENSION*

Eight faculty scorers scored seventy-three work products from eight assignments across two courses: ENG 103 and UNI 101. ENG 103 is from the Composition and Information Literacy components of University Studies, and UNI 101 is from the First Year Seminar and Information Literacy components of University Studies. The types of assignments consisted of research papers, group projects, and group presentations. Sixteen work products (21.9%) were scored by multiple scorers for the purposes of scorer norming and determining interrater reliability. Figure 3.a provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

**INFORMATION LITERACY (LOWER DIVISION) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY**



|            |       |       |       |       |       |
|------------|-------|-------|-------|-------|-------|
| 0          | 0%    | 1.4%  | 6.8%  | 4.1%  | 8.2%  |
| 1          | 13.7% | 21.9% | 41.1% | 17.8% | 24.7% |
| 2          | 50.7% | 50.7% | 38.4% | 49.3% | 30.1% |
| 3          | 24.7% | 17.8% | 13.7% | 24.7% | 32.9% |
| 4          | 11.0% | 8.2%  | 0%    | 4.1%  | 4.1%  |
| 25th %tile | 2     | 2     | 1     | 2     | 1     |
| 50th %tile | 2     | 2     | 2     | 2     | 2     |
| 75th %tile | 3     | 3     | 2     | 3     | 3     |
| Mode       | 2     | 2     | 1     | 2     | 3     |
| N          | 73    | 73    | 73    | 73    | 73    |

Figure 3.a Distribution of Scores for Information Literacy (Lower Division), Applicable Scores Only

**RESULTS BY DIMENSION**

**IL 1 Determining the Extent of Information Needed**

Ten work products had difficulty defining the scope of the research question or thesis, had difficulty determining key concepts, and the types of information (sources) selected did not relate to concepts or answer research question (scores of one). Half of the work products defined the scope of the research question or thesis incompletely (parts were missing, remained too broad or too narrow, etc.) but could determine key concepts, and the types of information (sources) selected partially related to concepts or answered research question (scores of two). Eighteen

work products defined the scope of the research question or thesis completely, could determine key concepts, and the types of information (sources) selected related to concepts or answered research question (scores of three). Eight work products effectively defined the scope of the research question or thesis, effectively determined key concepts, and the types of information (sources) selected directly related to concepts or answered research question (scores of four).

## **IL 2 Access the Needed Information**

One work product did not reach the benchmark level one of scoring (score of zero). Sixteen work products accessed information randomly and retrieved information that lacked relevance and quality (scores of one). Half of the work products accessed information using simple search strategies and retrieved information from some relevant, though limited and similar, sources (scores of two). Thirteen work products accessed information using a variety of search strategies and from relevant information sources, and demonstrated an ability to refine search (scores of three). Six work products accessed information using effective, well-designed search strategies and from most appropriate information sources (scores of four).

## **IL 3 Evaluate Information and its Sources Critically**

This dimension had the greatest number of papers scoring below a level two. Five work products did not reach the benchmark level one of scoring (scores of zero). Thirty work products took information from source(s) without any interpretation/evaluation of the material, and the viewpoints of authors were taken as fact, without question (scores of one). Twenty-eight work products took information from source(s) with some interpretation/evaluation, but not a coherent analysis of the material, and the authors' viewpoints were taken as mostly fact, with little questioning (scores of two). Ten work products took information from source(s) with enough interpretation/evaluation to develop a coherent analysis of the material, and the authors' viewpoints were subject to questioning (scores of three). No work products reached the Capstone level four of scoring.

## **IL 4 Use Information Effectively to Accomplish a Specific Purpose**

Three work products did not reach the benchmark level one of scoring (scores of zero). Thirteen work products communicated information from sources; however, the information was fragmented and/or used inappropriately (misquoted, taken out of context, or incorrectly paraphrased, etc.), so the intended purpose was not achieved (scores of one). Nearly half of the work products communicated and organized information from sources, though the information was not yet synthesized, so the intended purpose was not fully achieved (scores of two). Eighteen work products communicated, organized and synthesized information from sources, and the intended purpose was achieved (scores of three). Three work products communicated organized and synthesized information from sources to fully achieve a specific purpose, with clarity and depth (scores of four).

## **IL 5 Access and Use Information Ethically and Legally**

This dimension deals with ethical use of information and had the most number of zero scores. Six work products did not reach the benchmark level one of scoring (scores of zero). For any score greater than zero for this dimension, the differences in score level are based on the number of information-use strategies employed. Any score above zero also indicates that a work product demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information. Eighteen work products consistently used only one information use strategy (citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; and distinguishing between common knowledge and ideas requiring attribution) (scores of one). Twenty-two of the work products demonstrated two types of information-use strategies (score of two). About a third of the work products showed evidence of three types of information-use strategies (score of three). Three work products showed evidence of all types of information-use strategies (score of four).

### *CORRELATION BETWEEN DIMENSIONS*

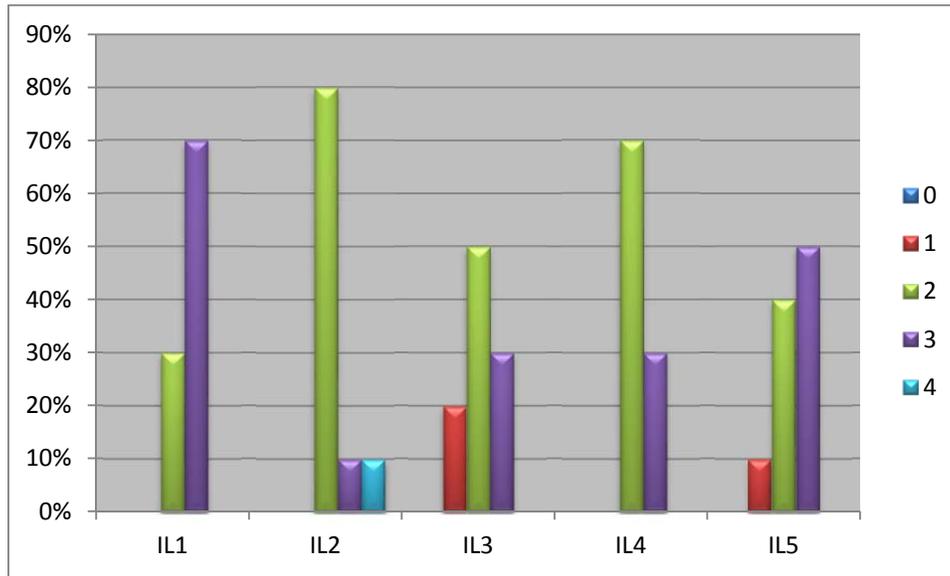
All dimensions were correlated at the .01 level of significance. See Appendix 3.B for the complete presentation of dimension correlation coefficients.

## **UPPER DIVISION**

### *SUMMARY OF SCORES BY DIMENSION*

Two faculty scorers scored ten work products from one assignment from one course: CSC 385. CSC 385 is from the Information Literacy component of University Studies. The assignment was a written essay. Two work products (20 %) were scored by multiple scorers for norming purposes and for determining interrater reliability. Figure 3.b provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

## INFORMATION LITERACY (UPPER DIVISION) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY



|            |       |       |       |       |       |
|------------|-------|-------|-------|-------|-------|
| 0          | 0%    | 0%    | 0%    | 0%    | 0%    |
| 1          | 0%    | 0%    | 20.0% | 0%    | 10.0% |
| 2          | 30.0% | 80.0% | 50.0% | 70.0% | 40.0% |
| 3          | 70.0% | 10.0% | 30.0% | 30.0% | 50.0% |
| 4          | 0%    | 10.0% | 0%    | 0%    | 0%    |
| 25th %tile | 2     | 2     | 2     | 2     | 2     |
| 50th %tile | 3     | 2     | 2     | 2     | 3     |
| 75th %tile | 3     | 2     | 3     | 3     | 3     |
| Mode       | 3     | 2     | 2     | 2     | 3     |
| N          | 10    | 10    | 10    | 10    | 10    |

Figure 3.b Distribution of Scores for Information Literacy (Lower Division), Applicable Scores Only

### *RESULTS BY DIMENSION*

#### **IL 1 Determining the Extent of Information Needed**

No work products scored a zero or had difficulty defining the scope of the research question or thesis, had difficulty determining key concepts, or provided types of information (sources) that did not relate to concepts or answer research question (score of one). Three work products defined the scope of the research question or thesis incompletely (parts were missing, remained too broad or too narrow, etc.) though they papers did determine key concepts, and the types of information (sources) selected partially related to concepts or answered research question (scores of two). Seven work products defined the scope of the research question or thesis completely,

could determine key concepts, and the types of information (sources) selected related to concepts or answered research question (scores of three). No work products effectively defined the scope of the research question or thesis, effectively determined key concepts, and the types of information (sources) selected directly related to concepts or answered research question (scores of four).

## **IL 2 Access the Needed Information**

No work products scored a zero or accessed information randomly and retrieved information that lacked relevance and quality (scores of one). Eight work products accessed information using simple search strategies and retrieved information from some relevant, though limited and similar, sources (scores of two). One work product accessed information using a variety of search strategies and from relevant information sources, and demonstrated an ability to refine search (score of three). One work product accessed information using effective, well-designed search strategies and from most appropriate information sources (score of four).

## **IL 3 Evaluate Information and its Sources Critically**

There were no zero scores on this dimension. Two work products took information from source(s) without any interpretation/evaluation of the material, and the viewpoints of authors were taken as fact, without question (scores of one). Five work products took information from source(s) with some interpretation/evaluation, but not a coherent analysis of the material, and the authors' viewpoints were taken as mostly fact, with little questioning (scores of two). Three work products took information from source(s) with enough interpretation/evaluation to develop a coherent analysis of the material, and the authors' viewpoints were subject to questioning (scores of three). No work products reached the Capstone level four of scoring.

## **IL 4 Use Information Effectively to Accomplish a Specific Purpose**

There were no zero scores nor did any work products communicate or display information from sources that was fragmented and/or used inappropriately (misquoted, taken out of context, or incorrectly paraphrased, etc.), so that the intended purpose was not achieved (scores of one). Seven work products communicated and organized information from sources; however the information was not yet synthesized, so the intended purpose was not fully achieved (scores of two). Three work products communicated, organized and synthesized information from sources, and the intended purpose was achieved (scores of three). No work products communicated, organized, and synthesized information from sources to fully achieve a specific purpose, with clarity and depth (score of four).

## **IL 5 Access and Use Information Ethically and Legally**

This dimension deals with ethical use of information. No work products scored a zero on this dimension. Six work products did not reach the benchmark level one of scoring (scores of zero). For any score greater than zero for this dimension, the differences in score level are based on the number of information-use strategies employed. Any score above zero also indicates that a work product demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information. One work product consistently used only one information use strategy (citations and references; choice of paraphrasing, summary, or quoting; using information in ways that are true to original context; and distinguishing between common knowledge and ideas requiring attribution) (score of one). Four of the work products demonstrated two types of information-use strategies (score of two). Half of the work products showed evidence of three types of information-use strategies (score of three). None of the work products showed evidence of all types of information-use strategies (score of four).

### *CORRELATION BETWEEN DIMENSIONS*

IL1 was correlated with IL3 and IL5 at the .05 level. IL3 was correlated with IL4 and IL5 at the .01 level. See Appendix 3.B for the complete presentation of dimension correlation coefficients.

## **ANALYSIS ACROSS COURSE LEVELS**

### *DEMOGRAPHIC AND PREPAREDNESS FINDINGS*

There were no statistically significant differences between the means, medians, and score distributions of different ethnicity groups of students and Honors vs. non-honors students. Differences did exist between male and female students on IL1, IL2, and IL4, with females scoring higher on all three dimensions. For freshman-start vs. transfer students, scores were statistically similar across the dimensions except for IL3, with Freshman-start students scoring higher. There were no Isaac Bear students in the sample, so comparisons between Isaac Bear vs. non-Isaac Bear students could not be completed.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, those having completed 0 – 30 credit hours, 31 – 60 credit hours, 61 – 90, and over 90 credit hours. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed statistically significant differences between the groups for IL5, with sophomores and juniors scoring higher on the dimension. Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with IL2 (.381\*\*), IL3 (.219\*), IL4 (.251\*), and IL5 (.365\*\*). Other statistically significant correlations were found between ACT scores and IL1 (.368\*\*), IL2 (.432\*\*), and IL4 (.321\*).

### *COMPARISONS BETWEEN CRITERIA*

Scores were compared across a number of criteria. There was no statistical difference between the scores on any dimension for work completed in courses taught by tenure-line faculty vs. part-time faculty. Likewise, there was also no difference in scores on any dimension for work collected from upper division vs. lower division courses.

All work scored was from the Information Literacy component of University Studies. Each course sampled is also in one additional component of University Studies—First Year Experience, Composition, and Writing Intensive. In comparing work collected from courses in these different University Studies components, statistically significant differences were seen in scores on IL2 and IL5. For both dimensions, the highest scores were seen on work collected from the course in the Composition component, followed by work from courses in the Writing Intensive component. Work from the First Year Seminar component scored the lowest on these two dimensions.

### *INTERRATER RELIABILITY*

There were a number of common papers scored between each pair of faculty scorers so that interrater reliability could be assessed (13 out of 83, or 13.3% of the total number of papers). Table 3.a shows the reliability measures for Information Literacy.

Table 3.a Interrater Reliability for Information Literacy

| Dimension   | N  | Percent Agreement | Plus Percent Adjacent | Krippendorff's alpha |
|---|----|-------------------|-----------------------|----------------------|
| IL 1 Determining the Extent of Information Needed                 | 11 | 72.7%             | 90.9%                 | .6781                |
| IL 2 Access the Needed Information                                | 11 | 81.8%             | 90.9%                 | .4851                |
| IL 3 Evaluate Information and its Sources Critically              | 11 | 72.7%             | 90.9%                 | .6321                |
| IL 4 Use Information Effectively to Accomplish a Specific Purpose | 11 | 54.5%             | 100%                  | .6219                |
| IL 5 Access and Use Information Ethically and Legally             | 11 | 63.6%             | 100%                  | .7685                |

Interrater reliability is a measure of the degree of agreement between scorers, and provides information about the trustworthiness of the data. It helps answer the question—would a different set of scorers at a different time arrive at the same conclusions? In practice, interrater reliability is enhanced over time through scorer discussion, as well as through improvements to the scoring rubric. Percent Agreement, Percent Agreement Plus Adjacent, and Krippendorff's Alpha measure scorer agreement. The UNCW benchmark is .67 for Krippendorff's Alpha. See Appendix B of this report for a more complete discussion of these statistics and the determination of benchmark levels.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff's Alpha, there are two dimensions of the rubric that meets these standards, IL1 and IL5. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that all dimensions had at least 90% of scores in agreement or within one level of each other.

#### *SCORER FEEDBACK ON RUBRIC AND SCORING*

Scorer feedback was collected during and following the scoring process. Scorers were asked to comment on the application of the rubric to the assignment. They reported that, often, the type of research asked for in the assignment was not clear, especially considering the types of citations that were used in the assignment. Overall, scorers felt that the assignment instructions did not address "evaluation of information and its sources". Instructions seemed to be presented implicitly, and it was confusing as to whether sources needed to be scholarly or non-scholarly. Some scorers viewed the purpose of the assignments from one course as gaining experience in working as a group rather than to work with information literacy. Scorers mentioned that the assignments could have been better fitted to the Information Literacy rubric.

When asked to comment on specific quality criteria in the rubric, scorers indicated concern with the explanation of what "sources" mean, and whether or not the students were actually aware of the distinction between different types of sources. Suggested improvements to the rubric criteria included that IL1 could be separated into different criteria, rather than being one dimension.

#### *DISCUSSION*

Table 3.b shows the percent of work products scored at or above the benchmark levels.

Table 3.b Information Literacy Percent of Sample Scored at or Above 2 and 3

| Lower-Division Courses                     |   |   |
|--|---|---|
| Dimension                                  | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| IL1 Determine Extent of Information Needed | 86.3%                                   | 35.6%                                     |
| IL2 Access Needed Information              | 76.7%                                   | 26.0%                                     |
| IL3 Evaluate Information and Sources       | 52.1%                                   | 13.7%                                     |
| IL4 Use Information Effectively            | 78.1%                                   | 28.4%                                     |
| IL5 Access and Use Information Ethically   | 67.1%                                   | 37.0%                                     |
| Upper-Division Courses                     |   |   |
| Dimension                                  | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| IL1 Determine Extent of Information Needed | 100%                                    | 70.0%                                     |
| IL2 Access Needed Information              | 100%                                    | 20.0%                                     |
| IL3 Evaluate Information and Sources       | 80.0%                                   | 30.0%                                     |
| IL4 Use Information Effectively            | 100.0%                                  | 30.0%                                     |
| IL5 Access and Use Information Ethically   | 90.0%                                   | 50.0%                                     |

For the lower-division courses, the benchmark achievement is defined as rubric score level two. A majority of work products from the lower-division courses were scored at this benchmark level or better, although for IL3 it was just over 50%. For the upper-division course, which was a 300-level course, a score of three is considered progressing well towards the ultimate benchmark of four for graduating seniors. Only for IL1 were most work products scored at least a three.

There were no statistical differences found in the scores on any dimension between the two levels of courses (upper- vs. lower-division), probably due to the small sample size of upper-division work products. However, papers written by students with sophomore and junior standing were scored statistically significantly higher than papers written by freshman for IL5. Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with IL2 (.381\*\*), IL3 (.219\*), IL4 (.251\*), and IL5 (.365\*\*). These results indicate that students in this study achieve higher scores on information literacy dimensions as the number of credit hours completed increases.

Looking at the percentage of work scored at benchmark level or higher, students achieved highest on IL1, IL2, and IL4. Looking back to past studies of students' information literacy skills, performance on IL1 and IL4 has remained consistently high.

# APPENDIX 3.A INFORMATION LITERACY RUBRIC

## Information Literacy VALUE Rubric

*Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.*

|  | Benchmark<br>1   | Milestones   |  | Capstone<br>4  | Score |
|--|--|--|--|--|-------|
|  |  | 2  | 3  |  |       |
| <b>1. Determine the Extent of Information Needed</b>                   | Has difficulty defining the scope of the research question or thesis. Has difficulty determining key concepts. Types of information (sources) selected do not relate to concepts or answer research question.  | Defines the scope of the research question or thesis incompletely (parts are missing, remains too broad or too narrow, etc.). Can determine key concepts. Types of information (sources) selected partially relate to concepts or answer research question.  | Defines the scope of the research question or thesis completely. Can determine key concepts. Types of information (sources) selected relate to concepts or answer research question.   | Effectively defines the scope of the research question or thesis. Effectively determines key concepts. Types of information (sources) selected directly relate to concepts or answer research question.  |       |
| <b>2. Access the Needed Information</b>                                | Accesses information randomly and retrieves information that lacks relevance and quality.  | Accesses information using simple search strategies and retrieves information from some relevant, though limited and similar, sources.   | Accesses information using variety of search strategies and from relevant information sources. Demonstrates ability to refine search.  | Accesses information using effective, well-designed search strategies and from most appropriate information sources.   |       |
| <b>3. Evaluate Information and its Sources Critically</b>              | Information is taken from source(s) without any interpretation/evaluation of the material; viewpoints of authors are taken as fact, without question.  | Information is taken from source(s) with some interpretation/evaluation, but not a coherent analysis of the material; viewpoints of authors are taken as mostly fact, with little questioning.   | Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis of the material; viewpoints of authors are subject to questioning.  | Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis of the material; viewpoints of authors are questioned thoroughly.  |       |
| <b>4. Use Information Effectively to Accomplish a Specific Purpose</b> | Communicates information from sources. The information is fragmented and/or used inappropriately (misquoted, taken out of context, or incorrectly paraphrased, etc.), so the intended purpose is not achieved.   | Communicates and organizes information from sources. The information is not yet synthesized, so the intended purpose is not fully achieved.  | Communicates, organizes and synthesizes information from sources. Intended purpose is achieved.  | Communicates, organizes and synthesizes information from sources to fully achieve a specific purpose, with clarity and depth   |       |
| <b>5. Access and Use Information Ethically and Legally</b>             | Consistently uses one of the following information use strategies: <ul style="list-style-type: none"> <li>• use of citations and references,</li> <li>• choice of paraphrasing, summary, or quoting,</li> <li>• using information in ways that are true to original context,</li> <li>• distinguishing between common knowledge and ideas requiring attribution;</li> </ul> AND demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information. | Consistently uses two of the following information use strategies: <ul style="list-style-type: none"> <li>• use of citations and references,</li> <li>• choice of paraphrasing, summary, or quoting,</li> <li>• using information in ways that are true to original context,</li> <li>• distinguishing between common knowledge and ideas requiring attribution;</li> </ul> AND demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information. | Consistently uses three of the following information use strategies: <ul style="list-style-type: none"> <li>• use of citations and references,</li> <li>• choice of paraphrasing, summary, or quoting,</li> <li>• using information in ways that are true to original context,</li> <li>• distinguishing between common knowledge and ideas requiring attribution;</li> </ul> AND demonstrate sa full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information. | Consistently uses all of the following information use strategies: <ul style="list-style-type: none"> <li>• use of citations and references,</li> <li>• choice of paraphrasing, summary, or quoting,</li> <li>• using information in ways that are true to original context,</li> <li>• distinguishing between common knowledge and ideas requiring attribution;</li> </ul> AND demonstrates a full understanding of the ethical and legal restrictions on the use of published, confidential, and/or proprietary information. |       |

**NOTES: Modified UNCW.**

## APPENDIX 3.B INFORMATION LITERACY CORRELATION COEFFICIENTS

### *LOWER DIVISION*

|                |     |                         | IL1    | IL2    | IL3    | IL4    | IL5    |
|----------------|-----|-------------------------|--------|--------|--------|--------|--------|
| Spearman's rho | IL1 | Correlation Coefficient | 1.000  | .657** | .596** | .638** | .557** |
|                |     | Sig. (2-tailed)         | .      | .000   | .000   | .000   | .000   |
|                |     | N                       | 73     | 73     | 73     | 73     | 73     |
|                | IL2 | Correlation Coefficient | .657** | 1.000  | .602** | .728** | .701** |
|                |     | Sig. (2-tailed)         | .000   | .      | .000   | .000   | .000   |
|                |     | N                       | 73     | 73     | 73     | 73     | 73     |
|                | IL3 | Correlation Coefficient | .596** | .602** | 1.000  | .682** | .679** |
|                |     | Sig. (2-tailed)         | .000   | .000   | .      | .000   | .000   |
|                |     | N                       | 73     | 73     | 73     | 73     | 73     |
|                | IL4 | Correlation Coefficient | .638** | .728** | .682** | 1.000  | .648** |
|                |     | Sig. (2-tailed)         | .000   | .000   | .000   | .      | .000   |
|                |     | N                       | 73     | 73     | 73     | 73     | 73     |
|                | IL5 | Correlation Coefficient | .557** | .701** | .679** | .648** | 1.000  |
|                |     | Sig. (2-tailed)         | .000   | .000   | .000   | .000   | .      |
|                |     | N                       | 73     | 73     | 73     | 73     | 73     |

### *UPPER DIVISION*

|                |     |                         | IL1   | IL2   | IL3    | IL4    | IL5    |
|----------------|-----|-------------------------|-------|-------|--------|--------|--------|
| Spearman's rho | IL1 | Correlation Coefficient | 1.000 | .325  | .701*  | .429   | .714*  |
|                |     | Sig. (2-tailed)         | .     | .359  | .024   | .217   | .020   |
|                |     | N                       | 10    | 10    | 10     | 10     | 10     |
|                | IL2 | Correlation Coefficient | .325  | 1.000 | .319   | .271   | .478   |
|                |     | Sig. (2-tailed)         | .359  | .     | .369   | .449   | .162   |
|                |     | N                       | 10    | 10    | 10     | 10     | 10     |
|                | IL3 | Correlation Coefficient | .701* | .319  | 1.000  | .866** | .767** |
|                |     | Sig. (2-tailed)         | .024  | .369  | .      | .001   | .010   |
|                |     | N                       | 10    | 10    | 10     | 10     | 10     |
|                | IL4 | Correlation Coefficient | .429  | .271  | .866** | 1.000  | .630   |
|                |     | Sig. (2-tailed)         | .217  | .449  | .001   | .      | .051   |
|                |     | N                       | 10    | 10    | 10     | 10     | 10     |
|                | IL5 | Correlation Coefficient | .714* | .478  | .767** | .630   | 1.000  |
|                |     | Sig. (2-tailed)         | .020  | .162  | .010   | .051   | .      |
|                |     | N                       | 10    | 10    | 10     | 10     | 10     |

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).

## 4. CRITICAL THINKING

The UNCW Critical Thinking learning goal is for students to use multiple methods and perspectives to critically examine complex problems. For purposes of this Learning Goal, “Critical thinking is ‘skilled, active interpretation and evaluation of observations, communications, information and argumentation’ (Fisher and Scriven, 1997). Critical thinking involves a clear explanation of relevant issues, skillful investigation of evidence, purposeful judgments about the influence of context or assumptions, reasoned creation of one’s own perspective, and synthesis of evidence and implications from which conclusions are drawn” (UNCW Learning Goals, 2011). The rubric used to score this learning goal is based largely on the VALUE Critical Thinking rubric; based on feedback from scorers, some minor modifications were made to the rubric. This updated version of the Critical Thinking rubric can be found in Appendix 4.A at the end of this chapter. In this study, work was sampled from the Quantitative and Logical Reasoning, Historical and Philosophical Perspectives, and Capstone components of University Studies.

### **LOWER DIVISION**

#### *SUMMARY OF SCORES BY DIMENSION*

Eight faculty scorers scored 90 work products from five assignments across three courses from the Spring 2014 semester: two sections of HST 106, one section of MAT 142, and two sections of PAR 101. HST 106 and PAR 101 are in the Historical and Philosophical Perspectives component of University Studies, and MAT 142 is in the Quantitative and Logical Reasoning component of University studies. The types of assignments consisted of in-class written exams and out-of-class essays. Forty-four work products were scored by multiple scorers for norming purposes and for determining interrater reliability. Figure 4.a provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

## CRITICAL THINKING (LOWER DIVISION) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY

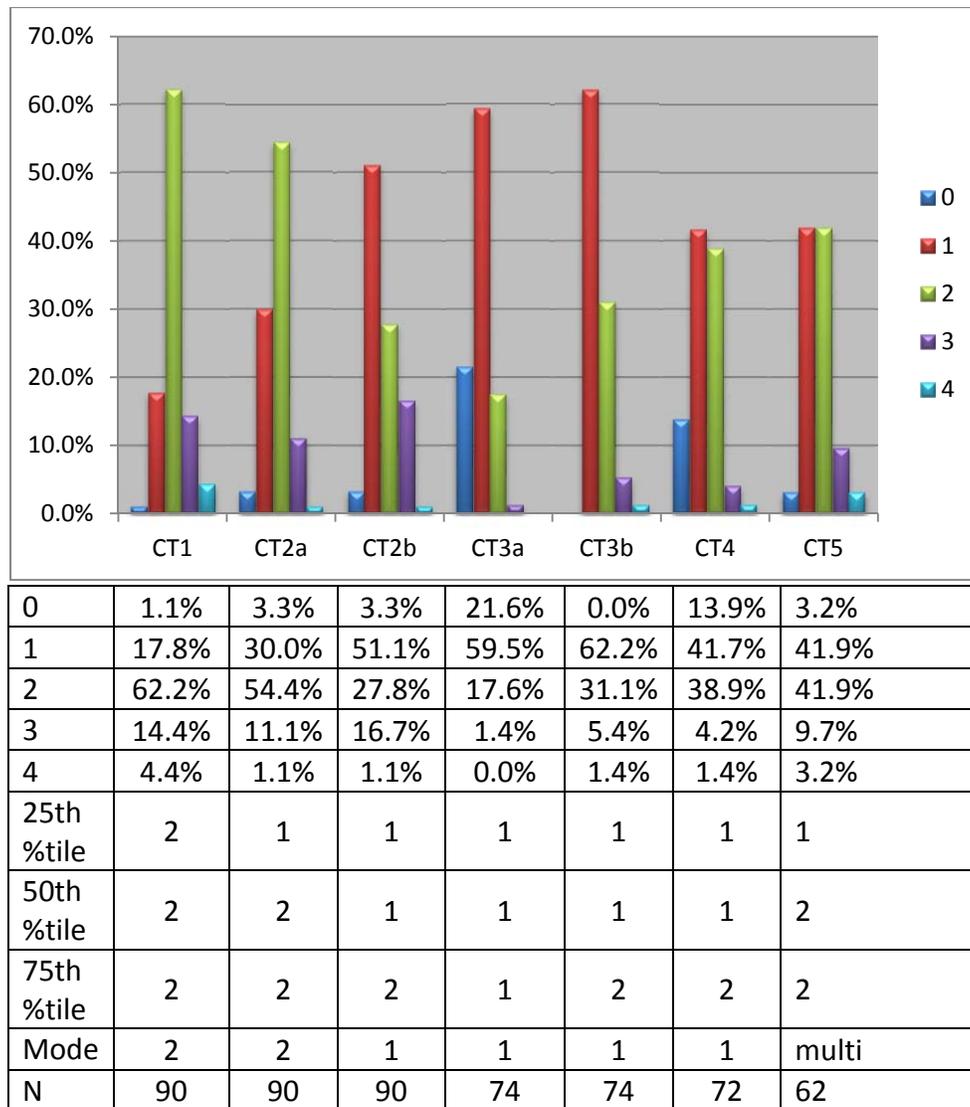


Figure 4.a Distribution of Scores for Critical Thinking (lower division), Applicable Scores Only

### *RESULTS BY DIMENSION*

#### **CT1 Explanation of Issues**

One work product did not reach the benchmark level 1 of scoring (score of zero). Sixteen work products stated the issue/problem without clarification or description (scores of 1). Fifty-six work products stated the issue/problem to be considered critically but the description left some aspects unexplored (scores of 2). Thirteen work products stated, described, and clarified the issue/problem so that understanding was not seriously impeded by omissions (scores of 3), while four work products stated and described the issue/problem comprehensively, delivering all relevant information necessary for full understanding (scores of 4).

**CT2a Evidence: selecting and using information**

Three work products did not reach the benchmark level 1 of scoring (scores of zero). Twenty-seven work products took information from source(s) without any interpretation/evaluation (scores of 1). A little over half of the work products took information from the source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis (scores of 2). Ten work products took information from the source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis (scores of 3). One work product took information from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis (score of 4).

**CT2b Evidence: critically examining evidence for viewpoint**

Three work products failed to reach the benchmark level 1 of scoring (scores of zero). Just over half of the work products took the viewpoints of authors as facts, without question (scores of one). Twenty-five work products took the viewpoints of authors mostly as fact, with little questioning (scores of two). Fifteen work products thought the viewpoints of authors were subject to questioning (scores of three). One work product questioned the viewpoints of the authors thoroughly (score of four).

**CT3a: Influence of context and assumptions: Assumptions**

Seventy-four work products were scored for CT3a. Sixteen work products received scores of zero for CT3a. Forty-four work products showed an emerging awareness of present assumptions (own or others') (scores of one). Thirteen work products questioned some assumptions and may have been more aware of others' own assumptions than their own (or vice versa) (scores of two). One work product identified their and others' assumptions when presenting a position (score of three). No work products reached the capstone level 4 of scoring.

**CT3b: Influence of context and assumptions: Context**

Seventy-four work products were scored on this dimension. None of the scored papers failed to reach at least the level one benchmark. Forty-six work products began to identify some contexts when presenting a position (scores of one). About one quarter of the work products identified several relevant contexts when presenting a position (scores of two). Four work products identified several relevant contexts and discussed at least some aspects of their interconnectedness (scores of four). One work product carefully evaluated the relevance of context when presenting a position (score of four).

**CT4: Student's position (position, perspective, thesis, or hypothesis)**

Seventy-two work products were scored on this dimension. Ten work products did not reach the benchmark level 1 of scoring (scores of zero). One-third of the scored work products stated a position that was simplistic and obvious (scores of one). Twenty-eight work products chose a position that acknowledged different sides of an issue (scores of two). Three work products chose a specific position that took into account the complexities of an issue, and others' points of view were acknowledged within position (scores of three). One work product chose a position that was imaginative, taking into account the complexities of an issue, acknowledging the limits of the chosen position, and synthesizing others' points within position (score of four).

**CT5: Conclusions and related outcomes**

Sixty-two work products were scored for CT5 Conclusions and Related Outcomes. Two work products did not reach the benchmark level 1 of scoring (scores of zero). Twenty-six work products stated a conclusion that was inconsistently tied to some of the information discussed, and the related outcomes were oversimplified (scores of one). Twenty-six work products stated a conclusion that was logically tied to information (because information is chosen to fit the desired conclusion), sometimes identifying related outcomes (scores of two). Six work products stated a conclusion that was logically tied to a range of information, including opposing viewpoints, sometimes clearly identifying related outcomes (scores of three). Two work products stated conclusions and related outcomes that were logical and reflected student's informed evaluation and ability to place evidence and perspectives discussed in priority order (scores of four).

*CORRELATION BETWEEN DIMENSIONS*

All dimensions were correlated at the .01 or .05 level of significance. See Appendix 4.B for the complete presentation of dimension correlation coefficients.

**UPPER DIVISION***SUMMARY OF SCORES BY DIMENSION*

Eight faculty scorers scored 23 work products from three assignments across three courses from the Spring 2014 semester: ENG 303, MGT 455, and SWK 497. ENG 303 is in the Quantitative and Logical Reasoning component of University Studies, while MGT 455 and SWK 497 are part of the Capstone component of University Studies. The types of assignments consisted of written essays and projects and one set of work products consisted of projects completed by teams of students (MGT 455). Eighteen products were scored by multiple scorers for norming and interrater reliability purposes. Figure 4.b provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

**CRITICAL THINKING (UPPER) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY**

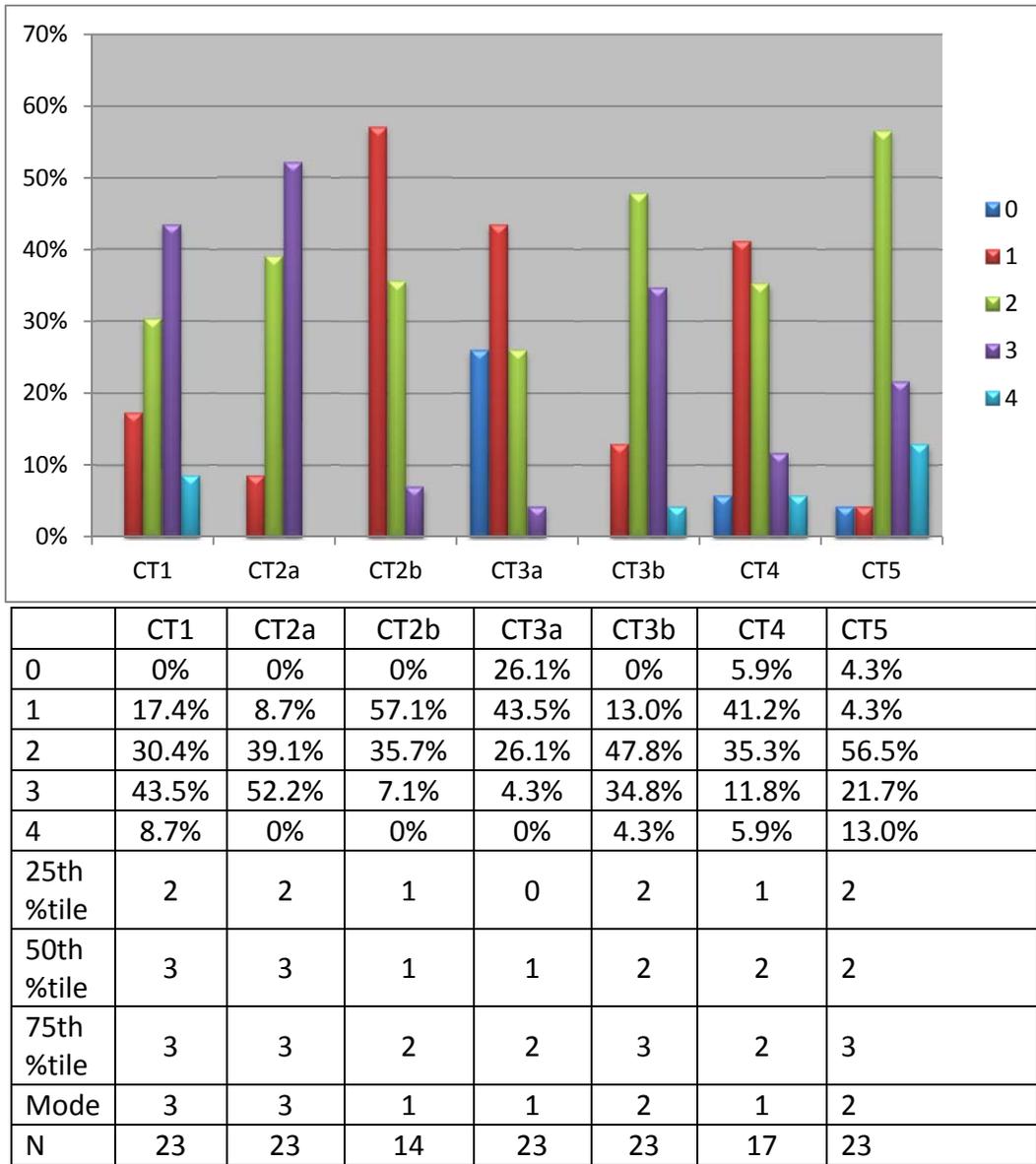


Figure 4.b Distribution of Scores for Critical Thinking (Upper Division), Applicable Scores Only

**RESULTS BY DIMENSION**

**CT1 Explanation of Issues**

Four work products stated the issue/problem without clarification or description (scores of 1). Slightly under a third of the work products stated the issue/problem to be considered critically but the description left some aspects unexplored (scores of 2). Ten work products stated, described, and clarified the issue/problem so that understanding was not seriously impeded by omissions (scores of 3), while two work products stated and described the issue/problem comprehensively, delivering all relevant information necessary for full understanding (scores of 4).

**CT2a Evidence: selecting and using information**

Two work products took information from source(s) without any interpretation/evaluation (scores of 1). Nine work products took information from the source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis (scores of 2). Just over half of the work products took information from the source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis (scores of 3). No work products took information from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis (score of 4).

**CT2b Evidence: critically examining evidence for viewpoint**

Eight work products took the viewpoints of authors as facts, without question (scores of one). Five work products took the viewpoints of authors mostly as fact, with little questioning (scores of two). One work products thought the viewpoints of authors were subject to questioning (scores of three), and no work products questioned the viewpoints of the authors thoroughly (score of four).

**CT3a: Influence of context and assumptions: Assumptions**

Six work products did not reach the benchmark level 1 of scoring (scores of zero). Ten work products showed an emerging awareness of present assumptions (own or others') (scores of one). Six work products questioned some assumptions and may have been more aware of others' own assumptions than their own (or vice versa) (scores of two). One work product identified their and others' assumptions when presenting a position (score of three), and no work products reached the capstone level 4 of scoring.

**CT3b: Influence of context and assumptions: Context**

Three work products began to identify some contexts when presenting a position (scores of one). Nearly half of the work products identified several relevant contexts when presenting a position (scores of two). Eight work products identified several relevant contexts and discussed at least some aspects of their interconnectedness (scores of three). One work product carefully evaluated the relevance of context when presenting a position (score of four).

**CT4: Student's position (position, perspective, thesis, or hypothesis)**

One work product did not reach the benchmark level 1 of scoring (scores of zero). Seven work products stated a position that was simplistic and obvious (scores of one). Six work products chose a position that acknowledged different sides of an issue (scores of two). Two work products chose a specific position that took into account the complexities of an issue, and others' points of view were acknowledged within position (scores of three). One work product chose a position that was imaginative, taking into account the complexities of an issue, acknowledging the limits of the chosen position, and synthesizing others' points within position (score of four).

## **CT5: Conclusions and related outcomes**

One work product did not reach the benchmark level 1 of scoring (scores of zero). One work products stated a conclusion that was inconsistently tied to some of the information discussed, and the related outcomes were oversimplified (scores of one). Thirteen work products stated a conclusion that was logically tied to information (because information is chosen to fit the desired conclusion), sometimes identifying related outcomes (scores of two). Five work products stated a conclusion that was logically tied to a range of information, including opposing viewpoints, sometimes clearly identifying related outcomes (scores of three). Three work products stated conclusions and related outcomes that were logical and reflected student's informed evaluation and ability to place evidence and perspectives discussed in priority order (scores of four).

### *CORRELATION BETWEEN DIMENSIONS*

All dimensions were correlated at the .01 level of significance. See Appendix 4.B for the complete presentation of dimension correlation coefficients.

## **ANALYSIS ACROSS COURSE LEVELS**

### *DEMOGRAPHIC AND PREPAREDNESS FINDINGS*

There were no statistically significant differences between the means, medians, and score distributions of male vs. female students. There was likewise no difference in the scores from the different ethnicity groups except for CT3a Influence of Assumptions, for which the scores of Hispanic students were significantly higher than those for papers from other ethnicity categories. Freshman-start students also scored better on CT3a, CT2b and CT3b than did transfer students. CT2b and CT4 scores were also different between Honors vs. non-honors students (with Honors students scoring higher). There were no differences in scores between Isaac Bear students and non-Isaac Bear students.

To compare scores based on number of credit hours completed, two methods were used. First, students were grouped into four categories, those having completed 0 – 30 credit hours, 31 – 60 credit hours, 61 – 90, and over 90 credit hours. Comparison of means (using ANOVA), medians (using Independent Samples test of medians) and distributions (using the Mann-Whitney U statistic) showed statistically significant differences between the groups for CT1, CT2b, and CT5, with students with 31-60 credit hours scoring the *highest* on all three dimensions. Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with CT5 (.286\*). Other statistically significant correlations were found between SAT Verbal and CT1 (.241\*), and GPA and CT3b (.252\*) and CT4 (.301\*\*).

### *COMPARISONS BETWEEN CRITERIA*

There was statistically significant difference in scores from work that was completed in class vs. out of class for CT2b, CT3b, and CT5. For each of these three dimensions, work that was completed outside of class scored higher than work completed in class.

Scores were similarly compared from courses taught by tenure-line faculty vs. part-time faculty. For all rubric dimensions except for CT2b and CT3a, the scores were the same from course taught by both faculty types. For CT2b, scores were higher from courses taught by tenure-line faculty, and from CT3a, scores were higher from courses taught by part-time faculty. Comparing the lower- and upper-division scores, there was a significant difference in the scores on CT1, CT2a, CT3b, and CT5 (there was no statistically significant difference in scores for the other dimensions). In each case, scores on the work collected from upper-division courses were higher.

Scores were also compared across University Studies components. Scores on CT1, CT3a, and CT4 were statistically similar. Scores on CT2a, CT2b, CT3b, and CT5 were not. For CT2a, CT2b, and CT3b, scores were highest on work collected from the Quantitative and Logical Reasoning component. Scores on CT5 were highest on work collected from the Capstone Course component.

#### *INTERRATER RELIABILITY*

There were a number of common papers scored between each pair of faculty scorers so that interrater reliability could be assessed (62 out of 113, or 55%), though not all dimensions were scored across-the-board. Table 4.a shows the reliability measures for Critical Thinking.

Table 4.a Interrater Reliability for Critical Thinking

| Dimension  | N  | Percent Agreement | Plus Percent Adjacent | Krippendorff's alpha |
|--|----|-------------------|-----------------------|----------------------|
| CT1 Explanation of Issues  | 62 | 50.0%             | 83.9%                 | .3307                |
| CT2a Evidence: selecting and using information                         | 62 | 62.9%             | 90.3%                 | .6774                |
| CT2b Evidence: critically examining evidence for viewpoint             | 56 | 67.9%             | 92.9%                 | -.0119               |
| CT3a: Influence of context and assumptions: Assumptions                | 55 | 45.5%             | 90.9%                 | .2615                |
| CT3b: Influence of context and assumptions: Context                    | 55 | 41.8%             | 85.5%                 | -.0528               |
| CT4: Student's position (position, perspective, thesis, or hypothesis) | 50 | 66.0%             | 90.0%                 | .0032                |
| CT5: Conclusions and related outcomes                                  | 49 | 51.0%             | 81.6%                 | -.0250               |

Interrater reliability is a measure of the degree of agreement between scorers, and provides information about the trustworthiness of the data. It helps answer the question—would a different set of scorers at a different time arrive at the same conclusions? In practice, interrater reliability is enhanced over time through scorer discussion, as well as through improvements to the scoring rubric. Percent Agreement, Percent Agreement Plus Adjacent, and Krippendorff's Alpha measure scorer agreement. The UNCW benchmark is .67 for Krippendorff's Alpha. See Appendix B of this report for a more complete discussion of these statistics and the determination of benchmark levels.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff's Alpha, there is one dimension of the rubric that meets these standards, CT2a. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that all dimensions had at least 81% of scores in agreement or within one level of each other.

### *SCORER FEEDBACK ON RUBRIC AND SCORING*

Scorers provided feedback on the assignment prior to scoring and after scoring. All assignments prompted students to clearly state the issue to be considered. Over half of the scorers reported that students were also prompted to examine the viewpoint of the authors of any texts used to complete the assignment (if applicable). However, fewer scorers (about one quarter) reported that students were asked to consider assumptions (either their own or others) or contexts when completing the assignment. These were also the dimensions (CT3a and CT3b) that were most often cited as being difficult and problematic for the scorers (primarily because the tasks not asked for in the assignment). When asked to comment about problematic quality criteria, all scorer comments mentioned the mismatch between assignment and rubric: that either assumptions were necessary in applying the rubric to the assignment or that whole dimensions were simply not applicable at times.

### *DISCUSSION*

Table 4.b shows the percent of work products scored at a level two or higher and the percent of work products scored at a level three or higher for each dimension.

Table 4.b Critical Thinking Percent of Sample Scored at Least Two and at Least Three

| Lower-Division Courses                            |   |   |
|---|---|---|
| Dimension   | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| CT 1 Explanation of Issues                        | 81.1%                                   | 18.9%                                     |
| CT2a Evidence: Selecting and Using                | 66.7%                                   | 12.2%                                     |
| CT2b Evidence: Critically Examining for Viewpoint | 45.6%                                   | 17.8%                                     |
| CT3a Influence of Assumptions                     | 18.9%                                   | 1.4%                                      |
| CT3b Influence of Context                         | 37.8%                                   | 6.8%                                      |
| CT4 Student's Position                            | 44.4%                                   | 5.6%                                      |
| CT5 Conclusions and Related Outcomes              | 54.8%                                   | 12.9%                                     |
| Upper-Division Courses                            |   |   |
| Dimension   | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| CT 1 Explanation of Issues                        | 82.6%                                   | 52.2%                                     |
| CT2a Evidence: Selecting and Using                | 91.3%                                   | 52.2%                                     |
| CT2b Evidence: Critically Examining for Viewpoint | 42.9%                                   | 7.1%                                      |
| CT3a Influence of Assumptions                     | 30.4%                                   | 4.3%                                      |
| CT3b Influence of Context                         | 87.0%                                   | 39.1%                                     |
| CT4 Student's Position                            | 52.9%                                   | 17.6%                                     |
| CT5 Conclusions and Related Outcomes              | 91.3%                                   | 34.8%                                     |

Comparing scores between lower-division and upper-division courses shows higher scores on work from upper-division courses, except for CT2b Evidence: Critically Examining for

Viewpoint. For the lower-division courses, the benchmark achievement is defined as rubric score level two. Only for CT1, CT2a, and CT5 did a majority of the lower-division work achieve that score level. The benchmark for graduating seniors is a score of four. A score of three for 300-level courses demonstrates adequate progress towards the graduating senior benchmark. A majority of upper-division work products achieved at least a score of three on only two dimensions: CT1 and CT2a.

Comparing the lower- and upper-division scores, there was a significant difference in the scores on CT1, CT2a, CT3b, and CT5 with work from upper-division courses scoring higher. When student work products were grouped according to class level (i.e. freshman, etc.), there were statistically significant differences between the class level groups for CT1, CT2b, and CT5, with students with 31-60 credit hours scoring the highest on all three dimensions. Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with CT5 (.286\*).

These findings that students are stronger on CT1 and CT2a and weaker on the other dimensions align with similar findings from earlier from earlier Critical Thinking studies. And the percent of students scoring at least a two is increasing in lower-division courses.

## APPENDIX 4.A CRITICAL THINKING RUBRIC

### CRITICAL THINKING VALUE RUBRIC (AAC&U)

Modification 2: May 2013 UNCW

*Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.*

|   | <b>Benchmark</b>   | <b>Milestones</b>  |  |  | <b>Capstone</b> | <b>Score</b> |
|---|--|--|--|--|-----------------|--------------|
|   | <b>1</b>   | <b>2</b>   | <b>3</b>   | <b>4</b>   |                 |              |
| <b>1. Explanation of Issues</b>   | Issue/problem to be considered critically is stated without clarification or description.                    | Issue/problem to be considered critically is stated but description leaves some aspects unexplored.  | Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions. | Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.                 |                 |              |
| <b>2. Evidence</b><br><i>a. Selecting and using information</i>                       | Information is taken from source(s) without any interpretation/evaluation.                                   | Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis.             | Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis.                     | Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis.  |                 |              |
| <i>b. Critically examining evidence for viewpoint</i>                                 | Viewpoints of authors are taken as fact, without question.   | Viewpoints of authors are taken as mostly fact, with little questioning.   | Viewpoints of authors are subject to questioning.  | Viewpoints of authors are questioned thoroughly.   |                 |              |
| <b>3. Influence of context and assumptions</b><br><i>a. Assumptions</i>               | Shows an emerging awareness of present assumptions (own or others').   | Questions some assumptions. May be more aware of others' assumptions than one's own (or visa versa).   | Identifies own and others' assumptions when presenting a position.   | Thoroughly (systematically and methodically) analyzes own and others' assumptions when presenting a position.  |                 |              |
| <i>b. Context</i>   | Begins to identify some contexts when presenting a position.   | Identifies several relevant contexts when presenting a position.   | Identifies several relevant contexts and discusses at least some aspects of their interconnectedness.                                      | Carefully evaluates the relevance of context when presenting a position.   |                 |              |
| <b>4. Student's position</b><br><i>(position, perspective, thesis, or hypothesis)</i> | Specific position is stated, but is simplistic and obvious.  | Specific position acknowledges different sides of an issue.  | Specific position takes into account the complexities of an issue. Others' points of view are acknowledged within position.                | Specific position is imaginative, taking into account the complexities of an issue. Limits of position are acknowledged. Others' points of view are synthesized within position. |                 |              |
| <b>5. Conclusions and related outcomes</b><br><i>(implications and consequences)</i>  | Conclusion is inconsistently tied to some of the information discussed; related outcomes are oversimplified. | Conclusion is logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes are identified. | Conclusion is logically tied to a range of information, including opposing viewpoints; related outcomes are identified clearly.            | Conclusions and related outcomes are logical and reflect student's informed evaluation and ability to place evidence and perspectives discussed in priority order.               |                 |              |

## APPENDIX 4.B CRITICAL THINKING CORRELATION COEFFICIENTS

### *LOWER DIVISION*

|                |      |                         | CT1    | CT2a   | CT2b   | CT3a   | CT3b   | CT4    | CT5    |
|----------------|------|-------------------------|--------|--------|--------|--------|--------|--------|--------|
| Spearman's rho | CT1  | Correlation Coefficient | 1.000  | .495** | .270** | .504** | .285*  | .388** | .608** |
|                |      | Sig. (2-tailed)         | .      | .000   | .010   | .000   | .014   | .001   | .000   |
|                |      | N                       | 90     | 90     | 90     | 74     | 74     | 72     | 62     |
|                | CT2a | Correlation Coefficient | .495** | 1.000  | .462** | .587** | .403** | .597** | .584** |
|                |      | Sig. (2-tailed)         | .000   | .      | .000   | .000   | .000   | .000   | .000   |
|                |      | N                       | 90     | 90     | 90     | 74     | 74     | 72     | 62     |
|                | CT2b | Correlation Coefficient | .270** | .462** | 1.000  | .552** | .313** | .615** | .344** |
|                |      | Sig. (2-tailed)         | .010   | .000   | .      | .000   | .007   | .000   | .006   |
|                |      | N                       | 90     | 90     | 90     | 74     | 74     | 72     | 62     |
|                | CT3a | Correlation Coefficient | .504** | .587** | .552** | 1.000  | .293*  | .584** | .483** |
|                |      | Sig. (2-tailed)         | .000   | .000   | .000   | .      | .011   | .000   | .000   |
|                |      | N                       | 74     | 74     | 74     | 74     | 74     | 56     | 60     |
|                | CT3b | Correlation Coefficient | .285*  | .403** | .313** | .293*  | 1.000  | .496** | .427** |
|                |      | Sig. (2-tailed)         | .014   | .000   | .007   | .011   | .      | .000   | .001   |
|                |      | N                       | 74     | 74     | 74     | 74     | 74     | 56     | 60     |
|                | CT4  | Correlation Coefficient | .388** | .597** | .615** | .584** | .496** | 1.000  | .521** |
|                |      | Sig. (2-tailed)         | .001   | .000   | .000   | .000   | .000   | .      | .000   |
|                |      | N                       | 72     | 72     | 72     | 56     | 56     | 72     | 44     |
|                | CT5  | Correlation Coefficient | .608** | .584** | .344** | .483** | .427** | .521** | 1.000  |
|                |      | Sig. (2-tailed)         | .000   | .000   | .006   | .000   | .001   | .000   | .      |
|                |      | N                       | 62     | 62     | 62     | 60     | 60     | 44     | 62     |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

*UPPER DIVISION*

|                |                         |                         | CT1   | CT2a   | CT2b   | CT3a   | CT3b   | CT4    | CT5   |
|----------------|-------------------------|-------------------------|-------|--------|--------|--------|--------|--------|-------|
| Spearman's rho | CT1                     | Correlation Coefficient | 1.000 | .519   | -.224  | .081   | .498   | .478   | .610  |
|                |                         | Sig. (2-tailed)         | .     | .011   | .442   | .712   | .016   | .052   | .002  |
|                |                         | N                       | 23    | 23     | 14     | 23     | 23     | 17     | 23    |
|                | CT2a                    | Correlation Coefficient | .519  | 1.000  | .246   | .074   | .454   | .310   | .475  |
|                |                         | Sig. (2-tailed)         | .011  | .      | .397   | .736   | .030   | .227   | .022  |
|                |                         | N                       | 23    | 23     | 14     | 23     | 23     | 17     | 23    |
|                | CT2b                    | Correlation Coefficient | -.224 | .246   | 1.000  | .806** | -.084  | .524   | -.063 |
|                |                         | Sig. (2-tailed)         | .442  | .397   | .      | .000   | .775   | .183   | .830  |
|                |                         | N                       | 14    | 14     | 14     | 14     | 14     | 8      | 14    |
| CT3a           | Correlation Coefficient | .081                    | .074  | .806** | 1.000  | .058   | .731** | -.071  |       |
|                | Sig. (2-tailed)         | .712                    | .736  | .000   | .      | .794   | .001   | .746   |       |
|                | N                       | 23                      | 23    | 14     | 23     | 23     | 17     | 23     |       |
| CT3b           | Correlation Coefficient | .498*                   | .454* | -.084  | .058   | 1.000  | .712** | .799** |       |
|                | Sig. (2-tailed)         | .016                    | .030  | .775   | .794   | .      | .001   | .000   |       |
|                | N                       | 23                      | 23    | 14     | 23     | 23     | 17     | 23     |       |
| CT4            | Correlation Coefficient | .478                    | .310  | .524   | .731** | .712** | 1.000  | .417   |       |
|                | Sig. (2-tailed)         | .052                    | .227  | .183   | .001   | .001   | .      | .096   |       |
|                | N                       | 17                      | 17    | 8      | 17     | 17     | 17     | 17     |       |
| CT5            | Correlation Coefficient | .610**                  | .475* | -.063  | -.071  | .799** | .417   | 1.000  |       |
|                | Sig. (2-tailed)         | .002                    | .022  | .830   | .746   | .000   | .096   | .      |       |
|                | N                       | 23                      | 23    | 14     | 23     | 23     | 17     | 23     |       |

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\* . Correlation is significant at the 0.01 level (2-tailed).

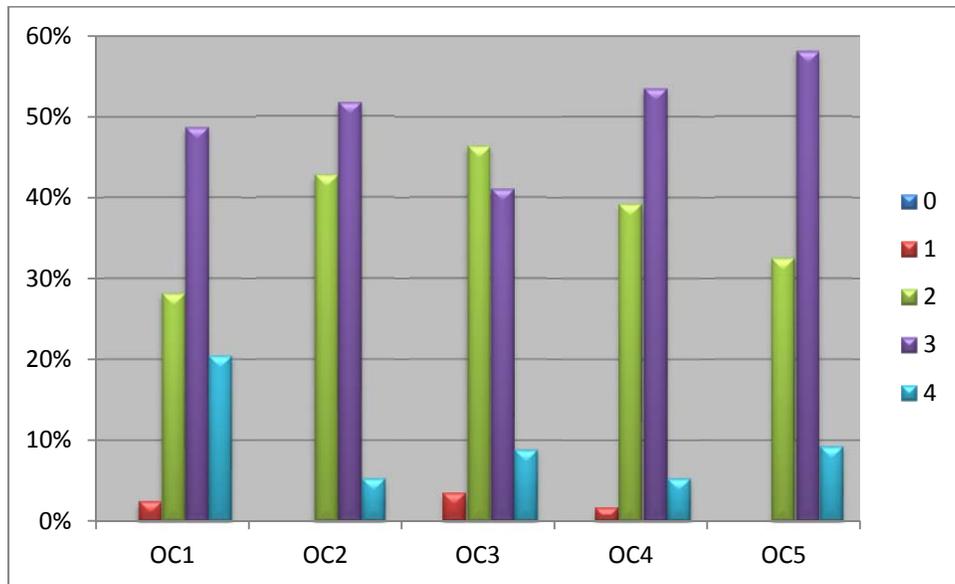
## 5. THOUGHTFUL EXPRESSION (ORAL)

The UNCW Thoughtful Expression (Oral) learning goal is for students to demonstrate an ability to express meaningful ideas orally. For purposes of this Learning Goal, thoughtful expression is “the ability to communicate meaningful ideas in an organized, reasoned and convincing manner. Thoughtful expression involves a purpose responsive to an identified audience, effective organization, insightful reasoning and supporting detail, style appropriate to the relevant discipline, purposeful use of sources and evidence, and error-free syntax and mechanics” (UNCW Learning Goals, 2011). The VALUE Oral Communication rubric contains five dimensions that are aligned with the UNCW description of Thoughtful Expression. The Thoughtful Expression (Oral) rubric can be found in Appendix 5.A at the end of this chapter. In this study, work was sampled from the Capstone component of University Studies.

### *SUMMARY OF SCORES BY DIMENSION*

Six faculty scorers scored 56 oral presentations from four assignments across three courses from the Spring 2014 semester: COM 400, SWK 497 and ART 476, which are all capstone courses. Forty-three of the work products (76.8%) were scored by multiple scorers. The oral presentations were observed by scorers either in real-time or as video recordings. Figure 5.a provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

## THOUGHTFUL EXPRESSION (ORAL) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY



|            | OC1   | OC2   | OC3   | OC4   | OC5   |
|------------|-------|-------|-------|-------|-------|
| 0          | 0%    | 0%    | 0%    | 0%    | 0%    |
| 1          | 2.6%  | 0%    | 3.6%  | 1.8%  | 0%    |
| 2          | 28.2% | 42.9% | 46.4% | 39.3% | 32.6% |
| 3          | 48.7% | 51.8% | 41.1% | 53.6% | 58.1% |
| 4          | 20.5% | 5.4%  | 8.9%  | 5.4%  | 9.3%  |
| 25th %tile | 2     | 2     | 2     | 2     | 2     |
| 50th %tile | 3     | 3     | 2.5   | 3     | 3     |
| 75th %tile | 3     | 3     | 3     | 3     | 3     |
| Mode       | 3     | 3     | 2     | 3     | 3     |
| N          | 39    | 56    | 56    | 56    | 43    |

Figure 5.a Thoughtful Expression (Oral) Results by Dimension for Applicable Scores Only

### *RESULTS BY DIMENSION*

#### **OC1 Organization**

For only one work product was there a lack of organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions)(score of 1). For eleven work products, an organizational pattern was intermittently observable within the presentation (scores of 2). Just under half of the work products contained an organizational pattern that was clearly and consistently observable within the presentation (scores of 3), and eight work products

included an organizational pattern that was clearly and consistently observable, and skillfully made the content of the presentation cohesive (scores of 4).

### **OC2 Language**

Twenty four work products contained language choices that were mundane and commonplace, and only partially supported the effectiveness of the presentation, while still using language appropriate to audience (scores of 2). Just over half of the work products utilized language choices that were thoughtful and generally supported the effectiveness of the presentation, while also being appropriate to the audience (scores of 3). In four work products, the language choices were imaginative, memorable, and compelling, and enhanced the effectiveness of the presentation, while also being appropriate to the audience (scores of 4).

### **OC3 Delivery**

For two work products, delivery techniques detracted from the understandability of the presentation, and speaker appeared uncomfortable (scores of 1). For just under half of the work products, the speaker appeared attentive and the delivery techniques (posture, gesture, eye contact, and vocal expressiveness) made the presentation understandable (scores of 2). About two in five work products included delivery techniques (posture, gesture, eye contact, and vocal expressiveness) that made the presentation interesting, and the speaker appeared comfortable (scores of 3). For just fewer than one in ten work products, the delivery techniques (posture, gesture, eye contact, and vocal expressiveness) made the presentation compelling, and the speaker appeared polished and confident (scores of 4).

### **OC4 Supporting Material**

One work product contained insufficient supporting materials (explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities) and made reference to information or analysis that minimally supported the presentation or established the presenter's credibility/authority on the topic (score of 1). Twenty two work products contained supporting materials that made appropriate reference to information or analysis that partially supported the presentation or established the presenter's credibility/authority on the topic (scores of 2). Just under one half of the work products included supporting materials that made appropriate reference to information or analysis that generally supported the presentation or established the presenter's credibility/authority on the topic (scores of 3). Only three work products utilized a variety of types of supporting materials that made appropriate reference to information or analysis that significantly supported the presentation or established the presenter's credibility/authority on the topic (score of 4).

### **OC5 Central Message**

Fourteen work products contained a central message that was basically understandable but was not often repeated (scores of 2). A little over half of the work products presented a central

message that was clear and consistent with the supporting material (scores of 3), while four of the work products contained a central message that was compelling (precisely stated, appropriately repeated, memorable) (scores of 4).

*CORRELATION BETWEEN DIMENSIONS*

All dimensions were correlated at the .001 or .01 level of significance. See Appendix 5.B for the complete presentation of dimension correlation coefficients.

*DEMOGRAPHIC AND PREPAREDNESS FINDINGS*

There were no statistically significant differences between the means, medians, and score distributions of the different ethnicity categories, gender, or freshman start vs. transfer students. There were no honors students or Isaac Bear students in the sample so comparisons could not be made between honors and non-honors or Isaac Bear students and non-Isaac Bear students. Likewise, all students in the sample were seniors, so there were no different class-level categories to compare. Looking at Spearman rho correlation coefficients, the number of total hours completed was not significantly correlated with any of the rubric dimensions, either. Statistically significant correlations were found between SAT Verbal and OC2 (.446\*), SAT Mathematical and OC1 (.477\*), and GPA and OC2 (.354\*) and OC3 (.462\*).

*COMPARISONS BETWEEN COURSES AND ASSIGNMENT TYPES*

The courses from which work was sampled were similar in terms of the University Studies component and division: all courses were upper-division Capstone Courses. There was no significant difference in scores from courses taught by professors vs. those taught by non-tenure-line faculty.

*INTERRATER RELIABILITY*

About three quarters of the presentations were multiply scored so that interrater reliability could be assessed, though not all dimensions were scored across-the-board. Table 5.a shows the reliability measures for Thoughtful Expression (Oral).

Table 5.a Interrater Reliability for Thoughtful Expression (Oral)

| Dimension               | N  | Percent Agreement | Plus Percent Adjacent | Krippendorff's alpha |
|-------------------------|----|-------------------|-----------------------|----------------------|
| OC1 Organization        | 42 | 23.9%             | 54.8%                 | .3089                |
| OC2 Language            | 42 | 47.6%             | 90.5%                 | .2243                |
| OC3 Delivery            | 42 | 45.2%             | 88.1%                 | .4916                |
| OC4 Supporting Material | 42 | 38.1%             | 83.3%                 | .1433                |
| OC5 Central Message     | 42 | 26.2%             | 66.7%                 | .2502                |

Interrater reliability is a measure of the degree of agreement between scorers, and provides information about the trustworthiness of the data. It helps answer the question—would a different set of scorers at a different time arrive at the same conclusions? In practice, interrater reliability is enhanced over time through scorer discussion, as well as through improvements to the scoring rubric. Percent Agreement, Percent Agreement Plus Adjacent, and Krippendorff’s Alpha measure scorer agreement. The UNCW benchmark is .67 for Krippendorff’s Alpha. See Appendix B of this report for a more complete discussion of these statistics and the determination of benchmark levels.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff’s Alpha, there is no dimension of the rubric that meets these standards. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that three dimensions had at least 83% of scores in agreement or within one level of each other.

***SCORER FEEDBACK ON RUBRIC AND SCORING***

Some scorers provided feedback after scoring the oral presentations. All of the feedback focused on OC2 Delivery. One scorer reported some difficulty in applying OC2 language to students’ work; the comments indicated that level 4 for OC2 might be asking more than could be expected of a graduating senior. Other scorers noted that perhaps the language for OC2 could be changed to include “enthusiasm of presentation”, as it was observed that enthusiasm was often what differentiated an average presentation from an above average one.

***DISCUSSION***

Table 5.b shows the percent of work products scored at or above the benchmark levels.

Table 5.b Thoughtful Expression (Oral) Percent of Sample Scored at or Above 2 and 3

| Upper-Division Courses  |   |   |
|-------------------------|---|---|
| Dimension               | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| OC1 Organization        | 97.4%                                   | 69.2%                                     |
| OC2 Language            | 100%                                    | 57.1%                                     |
| OC3 Delivery            | 96.4%                                   | 50.0%                                     |
| OC4 Supporting Material | 98.2%                                   | 58.9%                                     |
| OC5 Central Message     | 100%                                    | 67.4%                                     |

The benchmark score for graduating seniors is a four. All students in this sample were graduating seniors. The percent of presentations scored at the level four are provided in Figure 5.a, and range from 5% to 20%. At least half of all work products were scored at least a three for all dimensions, with OC1 and OC5 having the greatest percentage of work products scored at this

level. In the initial pilot study using this rubric (Spring 2010), OC3 was also the lowest-scoring dimension.

## APPENDIX 5.A THOUGHTFUL EXPRESSION (ORAL) RUBRIC

### ORAL COMMUNICATION VALUE RUBRIC

*for more information, please contact [value@aacu.org](mailto:value@aacu.org)*

#### Definition

Oral communication is a prepared, purposeful presentation designed to increase knowledge, to foster understanding, or to promote change in the listeners' attitudes, values, beliefs, or behaviors.

*Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.*

|                            | <b>Benchmark<br/>1</b>   | <b>Milestones</b>   |   | <b>Capstone<br/>4</b>   | <b>Score</b> |
|----------------------------|--|---|---|---|--------------|
|                            |  | 2   | 3   |   |              |
| <b>Organization</b>        | Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is not observable within the presentation.  | Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is intermittently observable within the presentation.  | Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is clearly and consistently observable within the presentation.  | Organizational pattern (specific introduction and conclusion, sequenced material within the body, and transitions) is clearly and consistently observable and is skillful and makes the content of the presentation cohesive.   |              |
| <b>Language</b>            | Language choices are unclear and minimally support the effectiveness of the presentation. Language in presentation is not appropriate to audience.   | Language choices are mundane and commonplace and partially support the effectiveness of the presentation. Language in presentation is appropriate to audience.  | Language choices are thoughtful and generally support the effectiveness of the presentation. Language in presentation is appropriate to audience.   | Language choices are imaginative, memorable, and compelling, and enhance the effectiveness of the presentation. Language in presentation is appropriate to audience.  |              |
| <b>Delivery</b>            | Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) detract from the understandability of the presentation, and speaker appears uncomfortable.   | Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation understandable, and speaker appears tentative.  | Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation interesting, and speaker appears comfortable.   | Delivery techniques (posture, gesture, eye contact, and vocal expressiveness) make the presentation compelling, and speaker appears polished and confident.   |              |
| <b>Supporting Material</b> | Insufficient supporting materials (explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities) make reference to information or analysis that minimally supports the presentation or establishes the presenter's credibility/authority on the topic. | Supporting materials (explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities) make appropriate reference to information or analysis that partially supports the presentation or establishes the presenter's credibility/authority on the topic. | Supporting materials (explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities) make appropriate reference to information or analysis that generally supports the presentation or establishes the presenter's credibility/authority on the topic. | A variety of types of supporting materials (explanations, examples, illustrations, statistics, analogies, quotations from relevant authorities) make appropriate reference to information or analysis that significantly supports the presentation or establishes the presenter's credibility/authority on the topic. |              |
| <b>Central Message</b>     | Central message can be deduced, but is not explicitly stated in the presentation.  | Central message is basically understandable but is not often repeated and is not memorable.   | Central message is clear and consistent with the supporting material.   | Central message is compelling (precisely stated, appropriately repeated, memorable, and strongly supported.)  |              |

## APPENDIX 5.B THOUGHTFUL EXPRESSION (ORAL) CORRELATION COEFFICIENTS

|                |      |                         | OC 1   | OC2    | OC3    | OC4    | OC5    |
|----------------|------|-------------------------|--------|--------|--------|--------|--------|
| Spearman's rho | OC 1 | Correlation Coefficient | 1.000  | -.011  | .189   | .457** | .422** |
|                |      | Sig. (2-tailed)         | .      | .947   | .248   | .003   | .007   |
|                |      | N                       | 39     | 39     | 39     | 39     | 39     |
|                | OC2  | Correlation Coefficient | -.011  | 1.000  | .574** | .438** | .393** |
|                |      | Sig. (2-tailed)         | .947   | .      | .000   | .001   | .009   |
|                |      | N                       | 39     | 56     | 56     | 56     | 43     |
|                | OC3  | Correlation Coefficient | .189   | .574** | 1.000  | .402** | .324*  |
|                |      | Sig. (2-tailed)         | .248   | .000   | .      | .002   | .034   |
|                |      | N                       | 39     | 56     | 56     | 56     | 43     |
|                | OC4  | Correlation Coefficient | .457** | .438** | .402** | 1.000  | .402** |
|                |      | Sig. (2-tailed)         | .003   | .001   | .002   | .      | .008   |
|                |      | N                       | 39     | 56     | 56     | 56     | 43     |
|                | OC5  | Correlation Coefficient | .422** | .393** | .324*  | .402** | 1.000  |
|                |      | Sig. (2-tailed)         | .007   | .009   | .034   | .008   | .      |
|                |      | N                       | 39     | 43     | 43     | 43     | 43     |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

## 6. THOUGHTFUL EXPRESSION (WRITTEN)

The UNCW Thoughtful Expression (Written) learning goal is for students to demonstrate an ability to express meaningful ideas in writing. For purposes of this Learning Goal, “thoughtful expression is the ability to communicate meaningful ideas in an organized, reasoned and convincing manner. Thoughtful expression involves a purpose responsive to an identified audience, effective organization, insightful reasoning and supporting detail, style appropriate to the relevant discipline, purposeful use of sources and evidence, and error-free syntax and mechanics” (UNCW Learning Goals, 2011). The VALUE Written Communication rubric contains five dimensions that are aligned with the UNCW description of Thoughtful Expression. The Thoughtful Expression (Written) rubric can be found in Appendix 6.A at the end of this chapter. In this study, work was sampled from the First Year Seminar, Composition, and Information Literacy components of University Studies.

### *SUMMARY OF SCORES BY DIMENSION*

Eight faculty scorers scored 70 work products from seven assignments across two courses from the Fall 2013 semester: ENG 103 and UNI 101. UNI 101 is in the First Year Seminar and Information Literacy components of University Studies. ENG 103 is in the Composition component and the Information Literacy component. The types of assignments consisted of out-of-class research papers. Twenty-nine work products (41.4%) were scored by multiple scorers. Figure 6.a provides the score distributions for each dimension for work products that were scored on that dimension (i.e., work products with blank scores are not included).

## THOUGHTFUL EXPRESSION (WRITTEN) RESULTS BY DIMENSION FOR APPLICABLE SCORES ONLY

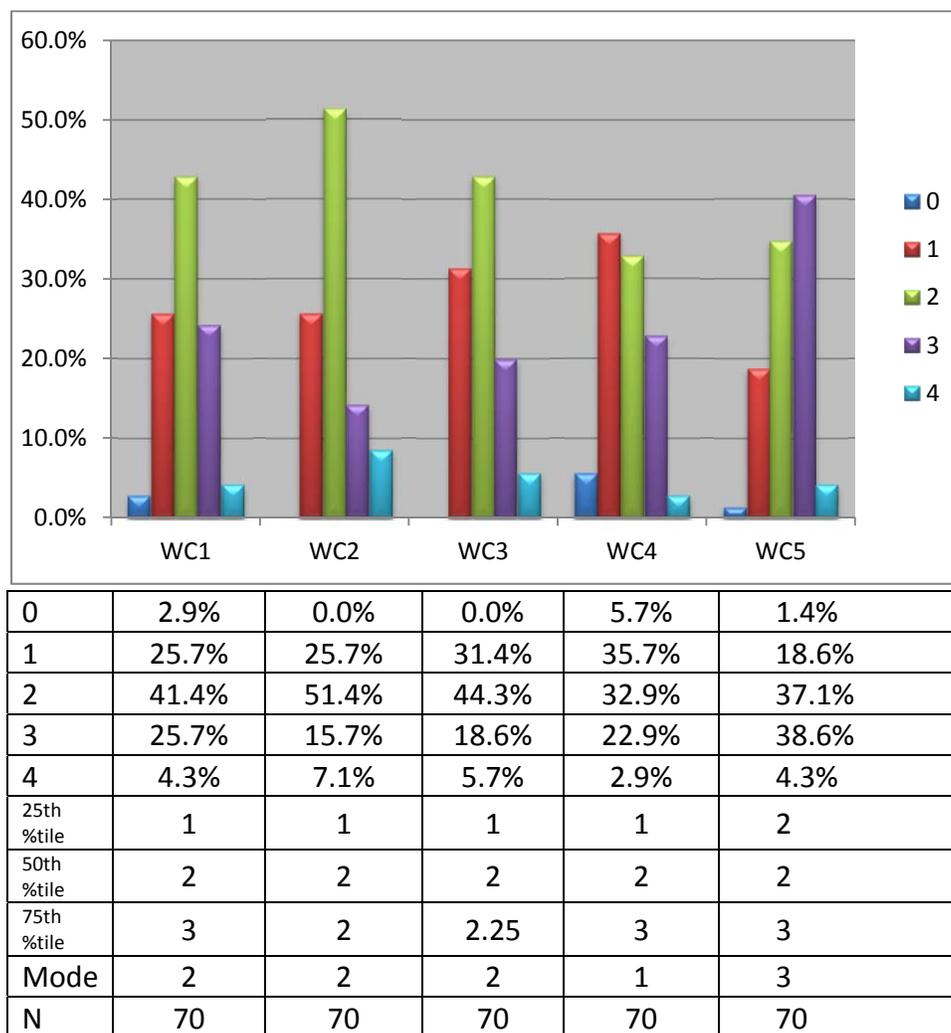


Figure 6.a Distribution of Scores for Thoughtful Expression (Written), Applicable Scores Only

### *RESULTS BY DIMENSION*

#### **WC1 Context of and Purpose for Writing**

Two work products demonstrated complete lack of attention to context, audience, purpose and to the assigned task (score of 0). Just over one in four work products demonstrated minimal attention to context, audience, purpose, and to the assigned task (scores of 1). Over two in 5 work products demonstrated awareness of the context, audience, purpose, and assigned task (scores of 2). Slightly less than one quarter of the work products demonstrated adequate consideration of context, audience, and purpose, and a clear focus on the assigned task (scores of 3). Three work products demonstrated a thorough understanding of context, audience, and purpose that was responsive to the assigned task and focused all elements of the work (scores of 4).

### **WC2 Content Development**

No work products failed to demonstrate any content development (score of 0). Just over one-fourth of work products used appropriate and relevant content to develop simple ideas in some parts of the work (scores of 1). Just over half of the work products used appropriate and relevant content to develop and explore ideas through the work (scores of 2). One out of seven work products used appropriate, relevant and compelling content to explore ideas within the context of the discipline (scores of 3). Finally, six of the 70 work products used appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's understanding, and shaping the whole work (scores of 4).

### **WC3 Genre and Disciplinary Conventions**

No work products failed to show an attempt to use a consistent system for organization and presentation (score of 0). Slightly less than one in three work products demonstrated an attempt to use a consistent system for basic organization and presentation (scores of 1). Thirty of the work products followed expectations appropriate to the specific writing task for basic organization, content, and presentation (scores of 2). One in five work products demonstrated consistent use of important conventions particular to the writing task, including stylistic choices (scores of 3). Four work products received a score of four on this dimension, demonstrating detailed attention to and execution of a wide range of conventions particular to a discipline or writing task.

### **WC4 Sources and Evidence**

This dimension had the highest number of zero-level scores: four work products demonstrated no attempt to use sources to support ideas. Just over one third of the work products demonstrated an attempt to use sources to support ideas (scores of 1). Slightly fewer work products scored a level two, demonstrating an attempt to use credible and/or relevant sources to support ideas that were appropriate to the task. Just under one quarter of the work products demonstrated consistent use of credible, relevant sources to support ideas (scores of 3). Two work products demonstrated skill use of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline (scores of 4).

### **WC5 Control of Syntax and Mechanics**

One work product failed to meet even the level 1 benchmark (score of 0). Slightly more than two in 15 work products used language that sometimes impeded meaning because of errors in usage (score of 1). Over one-third of the work products used language that generally conveyed meaning with clarity, although writing included some errors (score of 2) and four in ten work products used straightforward language that generally conveyed meaning, with few errors (score of 3). Three work products used graceful language that skillfully communicated meaning with clarity and fluency, with virtually no errors (scores of 4).

### *CORRELATION BETWEEN DIMENSIONS*

All dimensions were correlated at the .01 level of significance. See Appendix 6.B for the complete presentation of dimension correlation coefficients.

### *DEMOGRAPHIC AND PREPAREDNESS FINDINGS*

There were no statistically significant differences between the means, medians, and score distributions of male vs. female students, or the different ethnicity categories. There was a difference in the scores between Honors and non-honors students for all dimensions, with Honors students scoring higher across all dimensions. There were no Isaac Bear students in the sample so no comparisons could be made between Isaac Bear and non-Isaac Bear students.

Looking at Spearman rho correlation coefficients, the number of total hours completed was positively correlated with WC2 (.321\*\*). Other statistically significant correlations were found between SAT Verbal and WC1 (.275\*), WC2 (.344\*\*), WC4 (.400\*\*), and WC5 (.346\*\*); SAT Mathematical and WC1 (.292\*), WC2 (.254\*), WC4 (.281\*), and WC5 (.414\*\*); ACT score and WC1 (.301\*), and WC5 (.347\*), and GPA and CT3b (.252\*) and CT4 (.301\*\*). A comparison of scores based on number of credit hours was not completed, as there were only three work products collected from non-freshman students.

### *COMPARISONS BETWEEN CRITERIA*

Scores were compared to determine if difference existed between a number of criteria. There was a difference in scores on work collected from classes taught by professors vs. those taught by non-tenure-line faculty on two dimensions: WC2 and WC4. For both dimensions, work collected from courses taught by non-tenure-line faculty scored higher.

Regarding the University Studies components from which the work was collected, there was a significant difference in the scores from the components on three dimensions: WC1, WC2, and WC4. For all three dimensions, work from courses in Composition scored higher.

### *INTERRATER RELIABILITY*

There were a number of common papers scored between each pair of faculty scorers so that interrater reliability could be assessed (20 out of 70, or 28.6. Table 6.a shows the reliability measures for Thoughtful Expression (Written).

Table 6.a Interrater Reliability for Thoughtful Expression (Written)

| Dimension                              | N  | Percent Agreement | Plus Percent Adjacent | Krippendorff's alpha |
|--|----|-------------------|-----------------------|----------------------|
| WC1 Context of and Purpose for Writing | 20 | 45.0%             | 95.0%                 | .5515                |
| WC2 Content Development                | 20 | 50.0%             | 90.0%                 | .5655                |
| WC3 Genre and Disciplinary Conventions | 20 | 60.0%             | 95.0%                 | .6855                |
| WC4 Sources and Evidence               | 20 | 35.0%             | 80.0%                 | .4433                |
| WC5 Control of Syntax and Mechanics    | 20 | 40.0%             | 95.0%                 | .6186                |

Interrater reliability is a measure of the degree of agreement between scorers, and provides information about the trustworthiness of the data. It helps answer the question—would a different set of scorers at a different time arrive at the same conclusions? In practice, interrater reliability is enhanced over time through scorer discussion, as well as through improvements to the scoring rubric. Percent Agreement, Percent Agreement Plus Adjacent, and Krippendorff's Alpha measure scorer agreement. The UNCW benchmark is .67 for Krippendorff's Alpha. See Appendix B of this report for a more complete discussion of these statistics and the determination of benchmark levels.

Comparing the results of the reliability indices for this study to the benchmark of .67 for Krippendorff's Alpha, there is one dimension of the rubric that meets these standards, WC3. Looking at percent agreement plus adjacent (that is, the scores that were within one level of each other), we find that three dimensions had at least 80% of scores in agreement or within one level of each other with four of the five dimensions having at least 90% of scores within one score level.

***SCORER FEEDBACK ON RUBRIC AND SCORING***

Scorer feedback was collected during and following the scoring process. Scorers were asked to comment on the application of the rubric to the assignment. Regarding the assignments, many scorers reported finding it difficult to apply genre conventions or postulate a purpose for the given assignment when taken out of context of the larger course. Most assignments, according to the scorers, did not supply specific instructions regarding syntax or mechanics. When asked to comment on specific quality criteria that they found to be problem, scorers felt that, in general, the criteria were easy to apply (except for one set of scorers that mentioned the assignment did not provide instructions about citation style). Many scorer comments addressed the need for the assignment instructions to be more detailed. Finally, when asked to provide suggestions about

how the rubric quality criteria could be improved, scorers suggested that the rubric be accompanied by examples of common disciplinary conventions, that score level zero be added as an explicit option, and that the language of the rubric be made “more academic” and with “neutral diction” with fewer “and/or” criteria and compound criteria within a single score level.

*DISCUSSION*

Table 6.b shows the percent of work products scored at or above the benchmark levels.

Table 6.b Thoughtful Expression (Written) Percent of Sample Scored at or Above 2 and 3

| Lower-Division Courses                 |   |   |
|--|---|---|
| Dimension                              | % of Work Products Scored Two or Higher | % of Work Products Scored Three or Higher |
| WC1 Context of and Purpose for Writing | 71.4%                                   | 40.0%                                     |
| WC2 Content Development                | 74.3%                                   | 22.9%                                     |
| WC3 Genre and Disciplinary Conventions | 68.6%                                   | 25.7%                                     |
| WC4 Sources and Evidence               | 58.6%                                   | 25.7%                                     |
| WC5 Control of Syntax and Mechanics    | 79.7%                                   | 44.9%                                     |

For the lower-division courses, the benchmark achievement is defined as rubric score level two. A majority of work products from the lower-division courses were scored at this benchmark level or better. Dimensions WC5, WC2, and WC1 showed the highest percentage of work products meeting the lower-division benchmark.

Looking back to previous Thoughtful Expression (Written) studies, WC1 and WC5 have appeared historically as dimensions on which UNCW students achieve higher scores.

## APPENDIX 6.B THOUGHTFUL EXPRESSION (WRITTEN) RUBRIC

### WRITTEN COMMUNICATION VALUE RUBRIC

*for more information, please contact [value@aacu.org](mailto:value@aacu.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 texts, 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.*

|   | Benchmark<br>1  | Milestones   |   | Capstone<br>4  | Score |
|---|---|--|---|--|-------|
|   |   | 2  | 3   |  |       |
| <b>1. Context of and Purpose for Writing</b><br><i>Includes considerations of audience, purpose, and the circumstances surrounding the writing task(s).</i>           | Demonstrates minimal attention to context, audience, purpose, and to the assigned tasks(s) (e.g., expectation of instructor or self as audience). | Demonstrates awareness of context, audience, purpose, and to the assigned tasks(s) (e.g., begins to show awareness of audience's perceptions and assumptions). | 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 a thorough understanding of context, audience, and purpose that is responsive to the assigned task(s) and focuses all elements of the work.   |       |
| <b>2. Content Development</b>   | Uses appropriate and relevant content to develop simple ideas in some parts of the work.  | Uses appropriate and relevant content to develop and explore ideas through most of the work.   | Uses appropriate, relevant, and compelling content to explore ideas within the context of the discipline and shape the whole work.  | Uses appropriate, relevant, and compelling content to illustrate mastery of the subject, conveying the writer's understanding, and shaping the whole work.   |       |
| <b>3. Genre and Disciplinary Conventions</b><br><i>Formal and informal rules inherent in the expectations for writing in particular forms and/or academic fields.</i> | Attempts to use a consistent system for basic organization and presentation.  | Follows expectations appropriate to a specific discipline and/or writing task(s) for basic organization, content, and presentation                             | Demonstrates consistent use of important conventions particular to a specific discipline and/or writing task(s), including organization, content, presentation, and stylistic choices | 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 |       |
| <b>4. Sources and Evidence</b>  | Demonstrates an attempt to use sources to support ideas in 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 consistent use of credible, relevant sources to support ideas that are situated within the discipline and genre of the writing.  | Demonstrates skillful use of high-quality, credible, relevant sources to develop ideas that are appropriate for the discipline and genre of the writing  |       |
| <b>5. Control of Syntax and Mechanics</b>   | Uses language that sometimes impedes meaning because of errors in usage.  | Uses language that generally conveys meaning to readers with clarity, although writing may include some errors.  | Uses straightforward language that generally conveys meaning to readers. The language in the portfolio has few errors.  | Uses graceful language that skillfully communicates meaning to readers with clarity and fluency, and is virtually error-free.  |       |

## APPENDIX 6.B THOUGHTFUL EXPRESSION (WRITTEN) CORRELATION COEFFICIENTS

|                |     |                         | WC1   | WC2   | WC3   | WC4   | WC5   |
|----------------|-----|-------------------------|-------|-------|-------|-------|-------|
| Spearman's rho | WC1 | Correlation Coefficient | 1.000 | .751  | .691  | .618  | .436  |
|                |     | Sig. (2-tailed)         | .     | .000  | .000  | .000  | .000  |
|                |     | N                       | 70    | 70    | 70    | 70    | 69    |
|                | WC2 | Correlation Coefficient | .751  | 1.000 | .691  | .622  | .544  |
|                |     | Sig. (2-tailed)         | .000  | .     | .000  | .000  | .000  |
|                |     | N                       | 70    | 70    | 70    | 70    | 69    |
|                | WC3 | Correlation Coefficient | .691  | .691  | 1.000 | .618  | .600  |
|                |     | Sig. (2-tailed)         | .000  | .000  | .     | .000  | .000  |
|                |     | N                       | 70    | 70    | 70    | 70    | 69    |
|                | WC4 | Correlation Coefficient | .618  | .622  | .618  | 1.000 | .569  |
|                |     | Sig. (2-tailed)         | .000  | .000  | .000  | .     | .000  |
|                |     | N                       | 70    | 70    | 70    | 70    | 69    |
|                | WC5 | Correlation Coefficient | .436  | .544  | .600  | .569  | 1.000 |
|                |     | Sig. (2-tailed)         | .000  | .000  | .000  | .000  | .     |
|                |     | N                       | 69    | 69    | 69    | 69    | 69    |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

## 7. GENERAL DISCUSSION AND RECOMMENDATIONS

This chapter provides a general discussion across all studies. It includes an overall discussion of findings regarding student abilities, scorer and instructor feedback on the process, an overall discussion of interrater reliability, follow up on prior-year recommendations, and new recommendations.

### *UNCW STUDENT ABILITIES ON LEARNING GOALS*

Five of the eight UNCW Learning Goals were assessed in 2013-2014 using course-embedded assignments. Table 7.a combines the results from all studies, and presents the percentage of work products at each grade level that met or exceeded the performance benchmark, or proficiency level, for each dimension of each rubric.

Table 7.a Percent of Work Products Meeting Performance Benchmarks by Grade Level

| Courses                   | Dimension                                      | % freshman at level 1+ | % sophomores at level 2+ | % juniors at level 3+ | % seniors at level 4 |
|---------------------------|--|------------------------|--------------------------|-----------------------|----------------------|
| PAR 101, EVS 495, MGT 455 | IN 1 Topic Selection                           | --                     | --                       | --                    | 4.3%<br>n=23         |
|                           | IN 2 Existing Knowledge, Research and/or Views | 100%<br>n=5            | 90.9%<br>n=11            | 0%<br>n=1             | 13.3%<br>n=30        |
|                           | IN 3 Design Process                            | 100%<br>n=5            | 100%<br>n=11             | 0%<br>n=1             | 3.3%<br>n=30         |
|                           | IN 4 Analysis                                  | 100%<br>n=5            | 90.9%<br>n=11            | 0%<br>n=1             | 6.7%<br>n=30         |
|                           | IN 5 Conclusions                               | 100%<br>n=5            | 72.7%<br>n=11            | 0%<br>n=1             | 6.7%<br>n=30         |
|                           | IN 6 Limitations and Implications              | --                     | --                       | --                    | 4.3%<br>n=23         |
| STT 210, PSY 225          | DA 1 Summarizing and Analyzing Data            | 100%<br>n=3            | 89.5%<br>n=19            | 21.1%<br>n=19         | 0%<br>n=6            |
|                           | DA 2 Explanation of Results                    | 100%<br>n=3            | 94.7%<br>n=19            | 73.7%<br>n=19         | 0%<br>n=6            |
|                           | DA 3 Inferences                                | --                     | 100%<br>n=5              | 60.0%<br>n=15         | 0%<br>n=4            |
| ENG 103, UNI 101, CSC 385 | IL1 Determine Extent of Information Needed     | 100%<br>n=69           | 100%<br>n=2              | 80.0%<br>n=5          | 0%<br>n=7            |
|                           | IL2 Access Needed Information                  | 98.6%<br>n=69          | 100%<br>n=2              | 60.0%<br>n=5          | 14.3%<br>n=7         |
|                           | IL3 Evaluate Information and Sources           | 92.3%<br>n=69          | 50%<br>n=2               | 20.0%<br>n=5          | 0%<br>n=7            |
|                           | IL4 Use Information Effectively                | 95.7%<br>n=69          | 100%<br>n=2              | 60.0%<br>n=5          | 0%<br>n=7            |
|                           | IL5 Access and Use Information Ethically       | 91.3%<br>n=69          | 100%<br>n=2              | 100%<br>n=5           | 0%<br>n=7            |
|                           | WC2 Content Development                        | 100%<br>n=66           | 100%<br>n=2              | 0%<br>n=1             | --                   |
|                           | WC3 Genre and Disciplinary Conventions         | 100%<br>n=66           | 50.0%<br>n=2             | 0%<br>n=1             | --                   |
|                           | WC4 Sources and Evidence                       | 93.9%<br>n=66          | 100%<br>n=2              | 100%<br>n=1           | --                   |
|                           | WC5 Control of Syntax and Mechanics            | 98.5%<br>n=65          | 100%<br>n=2              | 100%<br>n=1           | --                   |

| Courses   | Dimension   | % freshman at level 1+ | % sophomores at level 2+ | % juniors at level 3+ | % seniors at level 4 |
|---|---|------------------------|--------------------------|-----------------------|----------------------|
| HST 106,<br>MAT 142,<br>PAR 101,<br>ENG 303,<br>MGT 455,<br>SWK 497 | CT 1 Explanation of Issues                        | 100%<br>n=39           | 92.0%<br>n=25            | 0%<br>n=12            | 3.8%<br>n=26         |
|   | CT2a Evidence: Selecting and Using                | 97.4%<br>n=39          | 80.0%<br>n=25            | 0%<br>n=12            | 0%<br>n=26           |
|   | CT2b Evidence: Critically Examining for Viewpoint | 100%<br>n=39           | 64.0%<br>n=25            | 0%<br>n=12            | 0%<br>n=18           |
|   | CT3a Influence of Assumptions                     | 78.8%<br>n=33          | 37.5%<br>n=16            | 0%<br>n=12            | 0%<br>n=26           |
|   | CT3b Influence of Context                         | 100%<br>n=33           | 56.3%<br>n=16            | 0%<br>n=12            | 3.8%<br>n=26         |
|   | CT4 Student's Position                            | 88.8%<br>n=36          | 63.6%<br>n=22            | 0%<br>n=6             | 5.0%<br>n=20         |
|   | CT5 Conclusions and Related Outcomes              | 93.1%<br>n=29          | 75.0%<br>n=16            | 0%<br>n=7             | 8.3%<br>n=24         |
| COM 400,<br>SWK 497,<br>ART 476                                     | OC1 Organization                                  | --                     | --                       | --                    | 19.4%<br>n=36        |
|   | OC2 Language                                      | --                     | --                       | --                    | 3.8%<br>n=52         |
|   | OC3 Delivery                                      | --                     | --                       | --                    | 9.6%<br>n=52         |
|   | OC4 Supporting Material                           | --                     | --                       | --                    | 3.8%<br>n=52         |
|   | OC5 Central Message                               | --                     | --                       | --                    | 7.5%<br>n=40         |
| ENG 103,<br>UNI 101   | WC1 Context of and Purpose for Writing            | 97.0%<br>n=66          | 100%<br>n=2              | 0%<br>n=1             | --                   |

Table 7.a provides information about how students at each class level (0-30 hours, 31-60 hours, 61-90 hours, and 90+ hours) performed on all dimensions across General Education Assessment studies for AY 2013-2-14. It is important to note that, while a level one score indicates the benchmark at which a freshman is expected to perform and a level four score indicates what a graduating senior should demonstrate, levels two and three do not necessarily correspond to sophomore and junior class designation. Rather, score levels two and three indicate developmental milestones that student work is expected to achieve as students move toward satisfying degree requirements.

Looking at the scores of freshman students, all but one dimension across all five studies (there was no freshman work sampled in the Thoughtful Expression (Oral) study) had 80% or more of student work achieving the level one benchmark. CT3a Influence of Assumptions had 78.8% of student work reaching at least a level one. Looking at the scores of senior students across five

studies (no senior work was scored for Written Communication), only three dimensions (IN2 Existing Knowledge, Research, and/or Views, IL2 Access Needed Information , and OC1 Organization) had more than 10% of student work achieving a capstone score of four.

Looking more closely at the percentage of students in sophomore and junior class level categories achieving at the levels two and three milestones, we see that sophomore-level student work more often reached a level two than junior-level work reached a level three. For sophomore-level work, 16 of the 24 dimensions (66.7%) across all studies had more than 80% of scores at a level two or higher. For junior-level work, only four of the 24 dimensions (16.7%) across all studies had more than 80% of scores at a level three or higher.

***SCORER FEEDBACK ON PROCESS***

Table 7.b provides combined results for the survey items for the Information Literacy, Thoughtful Expression (Written), Inquiry, Critical Thinking, and Thoughtful Expression (Oral) scoring processes.

Table 7.b Scorer Feedback on General Education Assessment Process

|  | Strongly Disagree | Disagree    | Neutral     | Agree         | Strongly Agree |
|--|-------------------|-------------|-------------|---------------|----------------|
| The invitation to volunteer accurately described the experience.                                       | 0<br>(0%)         | 0<br>(0%)   | 0<br>(0%)   | 9<br>(30%)    | 21<br>(71%)    |
| The timing of the invitation gave adequate opportunity to arrange for attending workshops and scoring. | 0<br>(0%)         | 0<br>(0%)   | 0<br>(0%)   | 0<br>(0%)     | 31<br>(100%)   |
| The norming session adequately prepared me for what was expected of me during the scoring session.     | 0<br>(0%)         | 0<br>(0%)   | 1<br>(3.3%) | 12<br>(43.3%) | 17<br>(56.7%)  |
| The scoring session was well-organized.  | 0<br>(0%)         | 0<br>(0%)   | 0<br>(0%)   | 5<br>(16.7%)  | 25<br>(83.3%)  |
| The structure of the scoring made it reasonable to work for the full time.                             | 0<br>(0%)         | 2<br>(6.7%) | 0<br>(5.9%) | 6<br>(20.0%)  | 22<br>(73.3%)  |
| When I had questions, one of the leaders was available to answer it.                                   | 0<br>(0%)         | 0<br>(0%)   | 0<br>(0%)   | 4<br>(13.8%)  | 25<br>(86.2%)  |
| When I had questions, the question was answered.   | 0<br>(0%)         | 0<br>(0%)   | 1<br>(3.6%) | 4<br>(14.3%)  | 23<br>(82.1%)  |

|   | Strongly Disagree | Disagree    | Neutral      | Agree         | Strongly Agree |
|---|-------------------|-------------|--------------|---------------|----------------|
| I was comfortable scoring student work products from outside my discipline on the board Learning Goals. | 0<br>(0%)         | 1<br>(3.3%) | 2<br>(6.7%)  | 16<br>(53.3%) | 11<br>(36.7%)  |
| The process is an appropriate way to assess students on the UNCW Learning Goals.                        | 0<br>(0%)         | 1<br>(3.4%) | 4<br>(13.8%) | 14<br>(50.0%) | 10<br>(35.7%)  |
| This process is valuable in improving student learning.   | 0<br>(0%)         | 1<br>(3.4%) | 8<br>(27.6%) | 12<br>(41.4%) | 8<br>(27.6%)   |
| I would participate in this process again.  | 0<br>(0%)         | 0<br>(0%)   | 1<br>(3.3%)  | 7<br>(23.3%)  | 22<br>(73.3%)  |
| I would recommend participating in this process to my colleagues.                                       | 0<br>(0%)         | 0<br>(0%)   | 0<br>(0%)    | 4<br>(13.8%)  | 25<br>(86.2%)  |

There were also two open-ended questions on the survey and a section for comments or suggestions. The results of the Likert-scale, open-ended questions, and comments or suggestions are discussed below.

When asked what parts of the process worked the best there were three major themes. Eleven scorers felt the norming process worked the best, eight more felt that either working with a partner or team collaboration on student artifacts was the best working process, and four other scorers felt that working with colleagues from other disciplines to look at student work was the best part. Other well-noted processes were the organization of the workshop, the amount of time allowed to complete work, discussions on the artifacts, and receiving feedback. The last open-ended question asked in what ways the scoring process could be improved. Two scorers commented that the scoring process could be improved by increasing the amount of norming. Ten other scorers commented on improvement of the rubric and its clarity (including defining specific words and phrases). Other improvements suggested having professors align their assignments with the assignment rubric beforehand, altering the starting time of the workshop, and the room was noisy, making it distracting for some scorers. Overall, participants described the scoring event's invitation, expectations, structure, and applicability positively.

### *INTERRATER RELIABILITY*

Table 7.c combines the interrater reliability findings from all 2013-2014 studies, arranged in descending order by Krippendorff's alpha.

Table 7.c Interrater Reliability

| Dimension   | Krippendorff's Alpha | Percent Agreement | Plus Percent Adjacent |
|---|----------------------|-------------------|-----------------------|
| IL 5 Access and Use Information Ethically and Legally                 | 0.7685               | 63.6%             | 100%                  |
| WC3 Genre and Disciplinary Conventions                                | 0.6855               | 60.0%             | 95.0%                 |
| IL 1 Determining the Extent of Information Needed                     | 0.6781               | 72.7%             | 90.9%                 |
| CT2a Evidence: selecting and using information                        | 0.6774               | 62.90%            | 90.30%                |
| IN 4 Analysis   | 0.6595               | 59.30%            | 96.30%                |
| IL 3 Evaluate Information and its Sources Critically                  | 0.6321               | 72.70%            | 90.90%                |
| IN 2 Existing Knowledge, Research and/or Views                        | 0.6230               | 44.40%            | 96.30%                |
| IL 4 Use Information Effectively to Accomplish a Specific Purpose     | 0.6219               | 54.50%            | 100%                  |
| WC5 Control of Syntax and Mechanics                                   | 0.6186               | 40.00%            | 95.00%                |
| WC2 Content Development   | 0.5655               | 50.00%            | 90.00%                |
| WC1 Context of and Purpose for Writing                                | 0.5515               | 45.00%            | 95.00%                |
| IN 3 Design Process   | 0.4977               | 88.20%            | 96.30%                |
| OC3 Delivery  | 0.4916               | 45.20%            | 88.10%                |
| IL 2 Access the Needed Information                                    | 0.4851               | 81.80%            | 90.90%                |
| WC4 Sources and Evidence  | 0.4433               | 35.00%            | 80.00%                |
| IN 5 Conclusions  | 0.4354               | 40.70%            | 88.90%                |
| CT1 Explanation of Issues   | 0.3307               | 50.00%            | 83.90%                |
| OC1 Organization  | 0.3089               | 23.90%            | 54.80%                |
| CT3a Influence of context and assumptions: Assumptions                | 0.2615               | 45.50%            | 90.90%                |
| OC5 Central Message   | 0.2502               | 26.20%            | 66.70%                |
| OC2 Language  | 0.2243               | 47.60%            | 90.50%                |
| OC4 Supporting Material   | 0.1433               | 38.10%            | 83.30%                |
| DA 3 Inferences   | 0.0582               | 28.60%            | 100%                  |
| CT4 Student's position (position, perspective, thesis, or hypothesis) | 0.0032               | 66.00%            | 90.00%                |
| CT2b Evidence: critically examining evidence for viewpoint            | -0.0119              | 67.90%            | 92.90%                |
| CT5 Conclusions and related outcomes                                  | -0.0250              | 51.00%            | 81.60%                |
| CT3b Influence of context and assumptions: Context                    | -0.0528              | 41.80%            | 85.50%                |
| DA 2 Explanation of Results   | -0.0873              | 57.10%            | 85.70%                |
| DA 1 Summarizing and Analyzing Data                                   | -0.1818              | 57.10%            | 100%                  |
| IN 1 Topic Selection  | -0.2667              | 76.50%            | 88.20%                |
| IN 6 Limitations and Implications                                     | -0.2711              | 52.90%            | 94.10%                |

Two of the Information Literacy dimensions met the Krippendorff's benchmark, along with one of the Thoughtful Expression (Written) dimensions and one of the Critical Thinking dimensions. The higher Krippendorff's alpha scores indicate a greater chance that the scores from different scorers were in agreement not because of chance.

Scores on 20 of the 31 rubric dimensions in this study were in agreement or within one score level was at or above 90%, which indicates good interrater reliability. The IRR for Thoughtful Expression: Oral tended to be lowest, but it bears mention that, for most of the oral presentations,

there were two to four scorers, making IRR more difficult to achieve than for the usual pair of scorers.

#### *FOLLOW UP ON PREVIOUS RECOMMENDATIONS*

The following explains the progress made on the recommendations from last year.

#### **2013 Recommendations**

Following the 2013 report, it was decided that the results and analysis of the assessment process must be disseminated more purposefully and broadly so that faculty members can address these findings in their courses. A team consisting of the Associate Vice Chancellor and Dean of Undergraduate Studies, the Director of General Education Assessment, the Chair of the University Studies Advisory Committee, and the Undergraduate Studies Liaison to University Studies will be presenting Thoughtful Expression (Written) cumulative findings to each department and/or school. These presentations provide a context for the University Studies courses from each area and a jumping off point for meaningful discussion within the unit. The presentations will be completed by the end of Spring 2015.

#### *NEW RECOMMENDATIONS*

Key findings from the 2013-2014 assessment process show the following:

- of all the learning goals, students struggle most with critical thinking;
- within information literacy, students have most difficulty with evaluation information and its sources; and
- within inquiry, students have most difficulty with conclusions, limitations and implications.

It was also noted that scores on critical thinking have not changed over the last five years. Based on this analysis, the Learning Assessment Council recommends a university-wide concentration of efforts on understanding what critical thinking means, how it is being addressed at UNCW (within the University Studies curriculum, within the majors, and within co-curricular activities), and how to provide students with additional opportunities to develop their critical thinking skills. A committee has been appointed to plan a one-year Critical Thinking campaign that will be designed (1) to engage faculty, staff, students, and community members in exploring what it means to be a critical thinker and what critical thinking skills employers are looking for, and (2) to provide information and professional development on ways to incorporate additional opportunities for students to practice critical thinking skills in and outside the classroom.

## REFERENCES AND RESOURCES

- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. (2<sup>nd</sup> edition). Thousand Oaks, CA: Sage Publications.
- Rhodes, T. ed. (2010). *Assessing Outcomes and Improving Achievement Tips and Tools for Using Rubrics*. Washington, DC: Association of American Colleges and Universities. (Copies of the VALUE Rubrics can be found at [http://www.aacu.org/value/rubrics/index\\_p.cfm?CFID=33263360&CFTOKEN=78277616](http://www.aacu.org/value/rubrics/index_p.cfm?CFID=33263360&CFTOKEN=78277616))
- Siefert, L. (2010) *General Education Assessment Spring 2010 Report*. University of North Carolina Wilmington internal document. <http://www.uncw.edu/assessment/Documents/GeneralEducationAssessmentReportSp2010Final.pdf>
- University of North Carolina Wilmington. (2011). *UNCW Learning Goals*. adopted by Faculty Senate March 17, 2009 and modified on January 18, 2011. <http://www.uncw.edu/assessment/uncwLearningGoals.html>
- University of North Carolina Wilmington. (2009). *Report of the General Education Assessment Committee*, March 2009. <http://www.uncw.edu/assessment/Documents/General%20Education/GenEdAssessmentCommitteeReportMarch2009.pdf>



## APPENDIX A UNIVERSITY STUDIES CURRICULUM MAP

### University Studies Component Student Learning Outcomes and UNCW Learning Goals

|                             |  | Creative Inquiry       |                              | Critical Thinking       |                                    | Thoughtful Expression  |                 | Responsible Citizenship |                    |
|-----------------------------|--|------------------------|------------------------------|-------------------------|------------------------------------|------------------------|-----------------|-------------------------|--------------------|
|                             |  | Foundational Knowledge | Inquiry                      | Information Literacy    | Critical Thinking                  | Thoughtful Expression  | Second Language | Diversity               | Global Citizenship |
| Foundations                 | Composition  |                        | CMP2, CMP3                   | CMP3                    | CMP1, CMP2, CMP3, CMP4             | CMP1, CMP2, CMP3, CMP4 |                 |                         |                    |
|                             | Fresh Seminar                                      |                        | FS2                          | FS1                     | FS3                                | FS4                    |                 |                         |                    |
|                             | Foreign Language                                   | SL1, SL2, SL4          | SL4                          |                         | SL1, SL2, SL3, SL4                 |                        | FL1, FL2, FL3   | SL4                     | SL3, SL4           |
|                             | Lifespan Wellness                                  | W1, W2, W3, W4         |                              |                         | W1                                 |                        |                 |                         |                    |
|                             | Mathematics and Statistics                         | MS1, MS2               | MS1, MS2                     | MS2                     | MS1, MS2, MS3                      | MS3                    |                 |                         |                    |
| Approaches and Perspectives | Aesthetic, Interpretive, and Literary Perspectives | AIL1                   | AIL1                         | AIL1                    | AIL1, AIL2, AIL3                   | AIL1                   |                 | AIL2, AIL3              |                    |
|                             | Historical and Philosophical Approaches            | HPA1                   | HPA1, HPA3, HPA4             | HPA2                    | HPA2, HPA4                         |                        |                 | HPA3                    | HPA4               |
|                             | Living in a Global Society                         | GS1, GS2               | GS2                          |                         | GS2                                |                        |                 | GS2                     | GS2, GS3           |
|                             | Living in Our Diverse Nation                       | LDN1, LDN3             | LDN3                         | LDN2, LDN4              | LDN2, LDN4                         |                        |                 | LDN1, LDN3, LDN4        |                    |
|                             | Scientific Approaches to the Natural World         | SAN1, SAN2             | SAN1, SAN2                   | SAN2                    | SAN1, SAN2, SAN3                   | SAN3                   |                 |                         |                    |
|                             | Understanding Human Institutions and Behaviors     | HIB1                   |                              | HIB2                    | HIB2, HIB3, HIB4                   |                        |                 |                         | HIB4               |
| Common Requirements         | Information Literacy                               |                        | IL1, IL3                     | IL1, IL2, IL3, IL4, IL5 | IL1, IL2, IF3, IL4, IL5            | IL4                    |                 |                         |                    |
|                             | Quantitative Logical Reasoning                     | QRE1, QRE2             | QRE1, QRE2, LOG1, LOG2, LOG3 | QRE1, QRE2              | QRE1, QRE2, QRE3, LOG1, LOG2, LOG3 | QRE3, LOG3             |                 |                         |                    |
|                             | Writing Intensive                                  | WI1, WI5               | WI3                          | WI2, WI3, WI5           | WI2, WI4, WI5                      | WI3, WI4, WI5          |                 |                         |                    |
|                             | Capstone   |                        |                              |                         |                                    |                        |                 |                         |                    |

Shaded items are the focus of General Education Assessment activities during present cycle (Fall 2011 to Spring 2014).



## APPENDIX B A NOTE ON INTERRATER RELIABILITY MEASURES

There is much debate about the best means of measuring interrater reliability. There are many measures that are used. Some differences in the measures are due to the types of data (nominal, ordinal, or interval data). Other differences have to do with what is actually being measured. Correlation coefficients describe *consistency* between scorers. For example, if Scorer 1 always scored work products one level higher than Scorer 2, there would be perfect correlation between them. You could always predict one scorer's score by knowing the other's score. It does not, however, yield any information about *agreement*. A value of 0 for a correlation coefficient indicates no association between the scores, and a value of 1 indicates complete association. Spearman rho rank order correlation coefficient is an appropriate correlation coefficient for ordinal data.

Percent agreement measures exactly that—the percentage of scores that are exactly the same. It does not, however, account for chance agreement. Percent adjacent measures the number of times the scores were exactly the same plus the number of times the scores were only one level different. Percent adjacent lets the researcher know how often there is major *disagreement* between the scorers on the quality of the artifact.

Krippendorff's alpha is a measure of agreement that accounts for chance agreement. It can be used with ordinal data, small samples, and with scoring practices where there are multiple scorers. A value of 0 for alpha indicates only chance agreement, and a value of 1 indicates reliable agreement not based on chance. Negative values indicate “systematic disagreement” (Krippendorff, 2004).

Determining acceptable values for interrater reliability measures is not easy. Acceptable levels will depend on the purposes that the results will be used for. These levels must also be chosen in relationship to the type of scoring tool or rubric, and the measure of reliability being used. In this case, the tool is a “metarubric,” a rubric that is designed to be applied across a broad range of artifacts and contexts. This type of instrument requires more scorer interpretation than rubrics designed for specific assignments. For consistency measures, such as correlation coefficients, in a seminal work, Nunnally states that .7 may suffice for some purposes whereas for other purposes “it is frightening to think that any measurement error is permitted” (Nunnally, 1978, pp.245-246). The standard set for Krippendorff's alpha by Krippendorff himself is .8 to ensure that the data are at least similarly interpretable by researchers. However, “where only tentative conclusions are acceptable, alpha greater than or equal to .667 may suffice” (Krippendorff, 2004, p. 241). In the present context, we should aim for values of at least .67, with the recognition that this could be difficult given the broad range of artifacts scored with the metarubrics.